DISS. ETH NO. ……

INFORMATION-COMPUTATION GAPS IN ROBUST STATISTICS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES

(Dr. sc. ETH Zurich)

presented by

TOMMASO D'ORSI

Doctorate, ETH Zürich

born on 06.04.1993

Prof. Dr. David Steurer
Prof. Dr. Luca Trevisan
Dr. Vincent Cohen-Addad

2023

A Bruno e Liliana.

Information-computation gaps in robust statistics

Tommaso d'Orsi

The success of modern machine learning algorithms in extracting global information from data crucially relies on strong distributional assumptions on the input datasets. However, real-world datasets may contain outliers, fake or malicious data, or measurement errors that are known to substantially degrade the performance of many of these algorithms. The design of *robust* algorithms –that succeed even when the input dataset satisfies the distributional assumptions only approximately– has thus become a fundamental topic across statistics, mathematics and computer science.

In this thesis we resolve several open questions central to this broad research agenda. Our focus is two-fold: on one side we establish statistical and computational tractability of adversarial models; on the other, we introduce novel, efficient and robust algorithms that provide provably optimal guarantees.

With respect to the first goal, we unveil surprising information-computation gaps that show how the computational landscape of semi-random problems may differ from their average-case or worst-case counterparts. For example in the context of sparse principal component analysis or constraint satisfaction problems.

With respect to the second goal, we design new algorithms that achieve provably optimal guarantees in these general semi-random models. When there is a computational price to pay for robustness, our efficient algorithms match the new computational limits we established. When there is no price for robustness – such as for stochastic block models– our algorithms match the guarantees of their fragile counterparts and, in some cases, even improve over them.

By-products of our results are novel techniques to speed-up robust algorithms, and new state-of-the-art algorithms satisfying other important, related, requirements, such as differential privacy.

Acknowledgments

First and foremost I want to acknowledge my advisor David Steurer. David is the person who most influenced my view about math and computer science. I started my PhD with a fervid desire to explore the concept of computation but without any of the experience and mathematical skills required to do so. Armed with patience, generosity and his unique positive pragmatism, David shared with me his knowledge and shaped me into a computer scientist. Among many invaluable lessons, David also taught me what truly working in a team means. I cannot thank him enough for that.

One of the most important thing I did in my PhD was writing an email to Vincent Cohen-Addad. Starting from this email, Vincent shared with me many fascinating questions in computer science (a few we have solved, some we will solve) and has been one of my closest collaborators ever since. He also introduced me to Google and has been my host there for an incredible, eight months long, internship. From our very first interaction, Vincent went above and beyond his role, always seeking the best for me and supporting me towards my goals.

Surprisingly enough, another crucial thing I did in my PhD was reaching out to Luca Trevisan. Early on in my education, it was through his "In Theory" blog that I first learned about many captivating phenomena in mathematics, and science in general. With a background comparable to mine, Luca quickly became one of my scientific heroes. From our first meeting, Luca has been an incredible source of ideas, concerning both computer science and life as a whole. I am beyond grateful for his support and advice.[1]

Next, I want to thank Pravesh Kothari. Pravesh has been a recurring character in my PhD since the very beginning. He has always been available to discuss any idea, doubt or concern I had. Thanks to his contagious optimism and deep understanding of computer science, Pravesh has been a role model to me.

I want also to thank Alessandro Epasto, Warren Schudy, Peilin Zhong and the whole Omega team. I had a fantastic time at Google, which provided me with a completely different perspective and made me a better researcher and individual. Ale, Warren and Peilin spent many hours working with me. Our joint projects have been both fun and instructive. I am very grateful for that.

I have had the great luck to work with an outstanding set of collaborators, without whom this thesis would not exist. Thanks to (excluding people already mentioned): Hongjie Chen, Davin Choo, Jingqiu Ding, Yiding Hua, Jacob Imola, Chih-Hung Liu, Rajai Nasser, Gleb Novikov and Stefan Tiegel. Many of you became close friends, I am very grateful for

[1] We can consider this a counterexample to the saying "Never meet your heroes".

that. Gleb and I have been colleagues for our whole PhD. We have a lot to show for the countless hours spent working together: several joint publications and a deep and sincere friendship. Thanks to Jingqiu Ding and Rajai Nasser for the herculean effort behind our robust algorithm for stochastic block models. That 200 pages proof is the main reason why this thesis is so long. Thanks also to Chih-Hung for hosting me in Taiwan for what has been one of the best journeys of my life.

I am grateful to Afonso Bandeira, for being my second advisor and for keeping his door always open for me. Thanks also to Claudia Günthart and Bernadette Gianesi for being the best administrative team. Things always worked out for me because of you!

During my PhD years I have been lucky to meet some incredible people. Thanks to my friends Andrea, Daniele, Gustavo and Julia. Thanks also to my Rötelvillains: Carl (my Cicero in Zürich), Francesco, Isabel, Kathrin, Matias, Merel, Sarka, Sophie, Tina (adopted) and Ulrike. You are family to me and I cannot be grateful enough for that.

Next, I would like to thank my lifelong friends Big Mike, Federico, Gigi, Irene, Linus[2] and Mattew, for keeping me grounded and always sticking around no matter what. Thanks also to Gianni, Marco, Massimo, Luciana and Vittorina, for their deep affection and support. Via del Ricordo will always be home to me.

Finally, I would also like to thank my family, to whom I owe it all. In particular, thanks to my brother Lorenzo, for being my role model. To my mother Laura, for her unconditional love and for always reminding me what life is about. To my father Fulvio, for instilling in me the values of honesty and hard work.

Last, thanks to Chiara for all that words can say, and more.

[2]and Linus and Linus and Linus and Linus.

Contents

Chapter 1

Introduction

Estimating information from structured data is a central theme in statistics and has found applications in an incredibly wide array of disciplines. In an estimation problem, the starting assumption is the existence of a –known *a priori*– family of probability distributions $\mathcal{P} := \{\mathbb{P}_x \mid x \in \Omega\}$ over some space $\mathcal{Y}$, each indexed by some parameter $x \in \Omega$. One then receives a collection of observations[1] $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)$ drawn from an unknown probability distribution $\mathbb{P}_x \in \mathcal{P}$. The goal is to invert this random process and (approximately) recover the parameter x.

The last two decades have seen tremendous advancements in the understanding of estimation problems, from both a statistical and a computational perspective. These advancements have resulted in novel algorithms (and complementary statistical and computational lower bounds) that can efficiently infer global information from the large, high dimensional datasets used by modern applications (see [Wai19] and references therein). The Achilles heel of this framework is its over-reliance on the model assumptions. For example, consider the basic task of computing the mean of a d-dimensional spherical Gaussian distribution given a sequence of n independent observations. Here the empirical mean estimator fares well (in fact, it achieves the essentially optimal ℓ_2 error convergence $O(\sqrt{d/n})$ with high probability). However, the existence of even a single outlier can lead the algorithm to output a vector completely unrelated to the true mean.

As the assumptions of classical models are too unrealistic to capture the multiple facets of real world datasets, the design of algorithms that are less susceptible to unpredictable, possibly malicious, perturbations have become a pressing challenge. A convenient abstraction to capture these robustness requirements is that of *adversarial perturbations*. In an estimation problem with adversarial perturbations, the collection of observations $\mathbf{Y}$ is secretely replaced by an adversary –assumed to have unbounded computational power– with a modified version Y'. The goal is then to invert this semi-random process and recover the parameter x.

The quantity and quality of changes that an adversary can introduce in Y' result

[1] In this introductory chapter, we use boldface to denote random variables.

in distinct models with different levels of generality and diverse properties. While the high level goal is always that of capturing natural (real world) instances, often multiple reasonable definitions of adversaries exist for the same problem. Hence, finding the right constraint to impose on Y' is often a delicate, problem dependent process. In our example above, a natural adversary would be one that is allowed to replace a small, yet constant, fraction of the observations with arbitrary points.

In recent years, a flurry of works have produced robust algorithms for many estimation problems (see [DK19, LM22] and references therein). Despite these advancements, the statistical and computational landscapes of robust estimation remain largely unknown. Improving our understanding of the *statistical and computation trade-offs* of estimation problems in the presence of adversarial corruptions is the goal of this thesis.

A number of other reasons makes the study of robust algorithms compelling. There is a rich theory about the computational complexity of worst-case problems [AB09] and their (in)approximability [Kho10, FLM20]. A similar understanding is also being developed in the context of average-case complexity [Hop18, Wai19, RSS18], particularly in the settings of high dimensional estimation. To perform well in the worst-case settings, an algorithm needs to be able to perform well on all instances. Conversely, in the average case, the algorithm is only required to perform well on typical instances drawn from the given distribution (estimation problems are average-case problems[2]). The pictures developed in these two settings provide a stark contrast: problems that are computationally hard in the worst-case, can often be approximately solved in the average case in polynomial time. Semi-random adversarial models provide a natural way to *interpolate* between worst-case and average-case and thus, allow one to study how the computational landscape of these problems is affected as one consider larger and larger families of inputs.

Another motivation is that of collecting evidence of a *qualitative separation* between different computational models. Despite this concept being somewhat difficult to formalize, the belief that certain algorithmic techniques are "robust" –in contrast to others which are "fragile"– has driven many of the recent advancements in robust statistics [dKNS20]. To understand how semi-random models can be used in these settings, picture[3] two students A, B which are both able to optimally solve some test T. One way to gather evidence that the former is more proficient in the subject than the latter, consists of providing them with a harder test T' that only student A is able to solve. By observing this same outcome for more and more tests, one begins to wonder whether indeed the first student is more

[2]In this thesis, the distinction between estimation and problems and average-case problems is generally only conceptual. In an average case problem there is not a parameter x to recover but only a function to optimize. The underlying statistical and computational phenomena, as well the algorithmic techniques, are essentially the same. In particular, while for clarity of the exposition we focus on estimation problems, as we will see in Section 1.1.4, our ideas can be naturally applied to average-case problems without the planted structure.

[3]This analogy is inspired by oracle separations in computational complexity theory.

knowledgeable in the subject than the second.

1.1 Themes

1.1.1 The price of robustness

Perhaps the most fascinating phenomenon in high dimensional statistics is the existence of *information-computation* gaps. In many estimation problems, for a given objective and a given set of parameters C, there is a value $\alpha_{\text{stat}}(C)$ such that, when the signal-to-noise ratio[4] satisfies $\text{SNR} > \alpha_{\text{stat}}(C)$, it is information-theoretically possible to solve the problem, and below which the objective is unreachable. That is, there is a statistical *phase transition* at the value $\alpha_{\text{stat}}(C)$.

To make things more concrete, let us introduce our first running example: the *single spiked covariance model* for sparse PCA. Here we are given a collection of d-dimensional vectors $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n)^\top$ drawn from the distribution $N(0, \mathrm{Id}_d + \beta x x^\intercal)$, for an unknown k-sparse unit vector $x \in \mathbb{R}^d$ and a signal strength $\beta \in \mathbb{R}$. The goal is to compute an estimate $\hat{\mathbf{x}}$ for x with correlation[5] bounded away from zero so that $\|\hat{\mathbf{x}}\| = 1$ and $\langle \hat{\mathbf{x}}, x \rangle^2 \geqslant \Omega(1)$.[6] (Here, we square the inner product because x is identifiable only up to sign.) We can achieve optimal statistical guarantees for sparse PCA in the spiked covariance model by the following kind of exhaustive search: among all k-by-k principal submatrices of the empirical covariance matrix of the vectors $\{\mathbf{y}_i\}_{i \in [n]}$, find one with maximum eigenvalue and output a corresponding eigenvector [AW08, BR13]. In particular, this procedure achieves constant correlation with high probability as long as $\min\{\beta, \beta^2\} \cdot n \geqslant \tilde{\Omega}(k)$. However, the running-time is exponential in k.

A similar phenomenon occurs when we restrict ourselves to the set of estimators computable in polynomial time.[7] In this case we indicate the threshold by α_{comp} and the objective is then achievable for $\text{SNR} > \alpha_{\text{comp}}(C)$. When $\alpha_{\text{comp}}(C) > \alpha_{\text{stat}}(C)$ the computational phase transition does not match the statistical phase transition and the problem exhibits an information-computation gap as, for $\alpha_{\text{stat}}(C) < \text{SNR} < \alpha_{\text{comp}}(C)$, solving the objective requires a super-polynomial resources budget.

For a wide range of parameters, sparse principal component analysis indeed exhibits

[4] The signal-to-noise ratio is a (possibly complicated) function of the signal strength and (if relevant) the number of observations.

[5] Instead of asking for the correlation to be bounded away from 0, we could also ask for it to approach 1. Alternatively, we could ask to recover the support of x. At the granularity of our discussion here, these measures of success are equivalent in most regards.

[6] To simplify our discussion, we hide absolute constant multiplicative factors using the standard notations $\lesssim, \gtrsim$, $O(\cdot)$, $\Omega(\cdot)$ and $\Theta(\cdot)$. Similarly, we hide multiplicative factors logarithmic in the parameters at hand using the notation $\tilde{O}(\cdot), \tilde{\Omega}(\cdot)$.

[7] Note that one could also restrict the analysis to subexponential algorithm or, in general, to any specific computational budget.

such a information-computation gap. The model has a sharp transition in the top eigenvalue for $n \cdot \beta^2 \geqslant \Omega(d)$ (called *BPP transition* [BBAP05] in reference to the authors' names). In this regime, called *strong-signal regime*, the following spectral algorithm matches the optimal statistical guarantees of exhaustive search: compute the top right singular vector of $\mathbf{Y}$ (the matrix with rows $\mathbf{y}_1, \ldots, \mathbf{y}_n$) and restrict it to the k largest entries [KNV13]. We refer to this algorithm as *SVD with thresholding*. In other words, there is no information-computation gap in this strong signal regime.

In contrast, whenever $\beta^2 \cdot n \leqslant O(d)$, a principal component analysis of $\{\mathbf{y}_i\}_{i\in[n]}$ cannot be used to recover x. The best known polynomial-time algorithms [JL09, DM14, dKNS20, DKWB23, Cd21] for this *weak-signal* regime succeed with high probability whenever $\beta^2 \cdot n \geqslant \Omega(k^2 \log \frac{d}{k^2})$, almost quadratically worse than exhaustive search! A large and diverse body of work provides formal evidence of this trade-off [AW08, CMW13, BR13, KNV13, HKP+17, DKWB23, PR22, Cd21], in the form of reductions from conjecturally hard problems, such as planted clique [BR13], or concrete lower bounds against restricted classes of algorithms [HKP+17, DKWB23, PR22, Cd21].

For robust estimation problems, the presence of adversarial perturbations adds another dimension to this picture and generates a three-way trade-off between the desired accuracy objective, the signal-to-noise ratio and the strength of the adversarial model. For a given objective, a set of parameters C, and a constrained family of adversarial corruptions $\mathcal{E}$ the problem is information-theoretically solvable for $\text{SNR} > \alpha_{\text{stat}}(C, \mathcal{E})$ and efficiently so for $\text{SNR} > \alpha_{\text{comp}}(C, \mathcal{E})$. Depending on $\mathcal{E}$, these thresholds may differ from their canonical counterparts. When $\alpha_{\text{comp}}(C, \mathcal{E}) > \alpha_{\text{comp}}(C)$, we call the difference $\alpha_{\text{comp}}(C, \mathcal{E}) - \alpha_{\text{comp}}(C)$ the *computational price for robustness* against $\mathcal{E}$. Similarly, we refer to $\alpha_{\text{stat}}(C, \mathcal{E}) - \alpha_{\text{stat}}(C)$ as the *statistical* price for robustness.

In the spiked covariance model, the introduction of adversarial perturbations, where an adversary may change each entry of the input vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ by a small amount, drastically changes the algorithmic landscape. In the strong signal regime $\beta \cdot n \geqslant \Omega(d)$, it is possible to adversarially perturb the vectors $\mathbf{y}_1, \ldots, \mathbf{y}_n$ by at most $\tilde{O}(1/\sqrt{n})$ per entry such that SVD with thresholding achieves only vanishing correlation with the true vector x. An adversarial perturbation with this effect is very natural. It can be viewed as a whitening transformation and corresponds to a natural generative process for $\mathbf{y}_1, \ldots, \mathbf{y}_n$, where the vectors are chosen randomly from an n-dimensional subspace containing an approximately sparse vector (see Section 3.4). The weak-signal regime offers a remarkably different picture. Exhaustive search and *certain* polynomial time algorithms continue to provide essentially the same guarantees, both in terms of error and sample complexity, as in the vanilla single-spike model. In particular, these algorithms can afford entry-wise perturbations bounded by $O(1/n^{1/4})$, substantially larger than the other aforementioned algorithms. This value is significant because with perturbations larger than $\tilde{O}(1/n^{1/4})$ an adversary could completely remove the signal from $\mathbf{Y}$, making the problem information theoretically impossible. In other words, if $\mathcal{E}$ is the family of adversarial matrix perturbations E satisfying $\|E\|_\infty \leqslant O(1/n^{1/4})$,

then $\alpha_{\text{stat}}(C) \approx \alpha_{\text{stat}}(C, \mathcal{E})$ for all configurations C. Among polynomial time algorithms, the above discussion also suggests that in the weak-signal regime $\alpha_{\text{comp}}(C, \mathcal{E}) \approx \alpha_{\text{comp}}(C)$. The picture further shows that no known efficient algorithm optimally solving the strong signal regime in the vanilla settings is robust. Surprisingly, it turns out that this is not a coincidence, but an *inherent* property of the problem: there is a large computational price to pay for robustness.

Theorem 1.1 (Evidence of a computational price of robustness, informal)**.** *Let $t > 0$ be a constant and suppose that*

$$\begin{aligned} d &\leqslant n^{0.99t-1} \\ \beta \cdot n &\leqslant O\Big(k \cdot t \cdot (d/k)^{1/t}\Big) \\ \beta^2 \cdot n^{1.1} &\leqslant k^2 \,. \end{aligned}$$

Then, there exists a distribution μ over $n \times d$ matrices $\mathbf{Y}$ of the form $\mathbf{Y} = \sqrt{\beta}\mathbf{u}\mathbf{x}^T + \mathbf{W} + \mathbf{E}$ where $\|\mathbf{E}\|_\infty \leqslant \tilde{O}\big(1/\sqrt{n}\big)$, with the following properties:

- *μ is indistinguishable from the Gaussian distribution $N(0,1)^{d\times n}$ with respect to all multilinear polynomials of degree at most $n^{0.001}$*
- *the jointly-distributed random variables $\mathbf{W}$, $\mathbf{u}$, $\mathbf{x}$ are independent,*
- *the marginal distribution of $\mathbf{x}$ is supported on unit vectors with entries in $\Big\{-1/\sqrt{k}, 0, 1/\sqrt{k}\Big\}$,*
- *the marginal distribution of $\mathbf{u}$ is uniform over $\{-1,1\}^n$,*
- *the marginal distribution of $\mathbf{W}$ is $N(0,1)^{n\times d}$.*

The third inequality ensures we are in a regime where known polynomial-time algorithms for the weak-signal regime do not work. The other two still admit a wide range of parameters for which exhaustive search would successfully recover the sparse direction. Moreover, for $\mathbf{E} = \mathbf{0}$, SVD with thresholding would also work whenever $\beta \geqslant \Omega(\sqrt{d/n})$ (for example, consider the mildly sparse settings $k = d^{0.6}$, $d^{1/5} = n$, $t = 10$). To fully understand the significance of this result, there a number of underlying questions that needs to be addressed.

First, why are we considering multilinear polynomials? Deep results in recent years have shown that essentially all[8] state-of-the-art polynomial time algorithms in high dimension estimation (without adversarial corruptions) are captured by the computational model of (multilinear) polynomials of logarithmic degree [HKP+17, Hop18, KWB22]. While formalizing this claim into a theorem remains a crucial open question in computer science,

[8]Certain algorithms which heavily rely on the algebraic structure of the input are not captured by this computational model [ZSWB22]. By their very nature these algorithms are extremely brittle and thus can be easily fooled by adversarial corruptions.

one can use low-degree polynomials as a proxy for the class of all efficiently computable algorithms.

Second, what does indistinguishable mean? Here the precise notion involves both classical decision theory [NP33, LCY90] and results on low-degree polynomials [Hop18]. While we thoroughly formalize this in Section 3.4, at the granularity of this discussion we say two distributions are indistinguishable by a family of polynomials if, every such polynomial takes roughly the same values under both distributions, with large probability.

From this perspective, in the strong signal regime, whenever d is significantly smaller than n^t, Theorem 1.1 states that even tiny, but clever, adversarial perturbations can modify the instance so that only algorithms running in time exponential in $n^{\Omega(1)}$ (among those captured by the restricted computational model of polynomials) can distinguish $\mathbf{Y}$ from a typical $n \times d$ Gaussian matrix with no hidden structure!

1.1.2 Certification algorithms

Beyond the striking computational price of robustness just observed, there is a second fascinating phenomenon taking place in the context of sparse PCA. Namely that all known algorithms based on spectral methods [JL09, DM14, KNV13] fail in the presence of the adaptive corruptions above, but the basic SDP[9] [KNV13] provides virtually the same guarantees it achieves in the vanilla settings. In fact, sparse PCA is not just a fortuitous, isolated, example and this happens to be a recurring phenomenon in the context of estimation problems. The underlying question can be phrased as follows:

> *Is there some inherent property that makes an algorithm robust to adversarial perturbations?*

In this thesis we *positively* answer this question for several average-case problems. Continuing with our running example, recall we represent the n-by-d semi-random input matrix as $Y = \sqrt{\beta}\mathbf{u}x^T + \mathbf{W} + E$ where $\mathbf{u}, \mathbf{W}$ are as in Theorem 1.1, E captures adversarial corruptions bounded in $\|\cdot\|_\infty$ and x is the k-sparse structured direction we are seeking. For simplicity let us also assume $\beta \geqslant 1$. The aforementioned exhaustive search algorithm boils down to computing the vector maximizing the k-sparse norm[10] of the empirical covariance $YY^\mathsf{T} = \beta\|\mathbf{u}\|^2 xx^\mathsf{T} + N$, where we used N to capture cross-terms and noise terms in the multiplication. Without adversarial perturbations, whenever $\beta \cdot n \geqslant \tilde{\Omega}(k)$ the algorithm succeeds with high probability because

$$\begin{aligned} \|N\|_{\text{k-sparse}} &\leqslant \tilde{O}(k)\,, \\ \text{and} \quad \beta\|\mathbf{u}\|^2 \cdot \left\|xx^\mathsf{T}\right\|_{\text{k-sparse}} &= O(\beta n)\,. \end{aligned} \tag{1.1.1}$$

[9] Throughout this thesis the term "basic SDP" is used to denote the canonical semidefinite relaxation of the problem. For this reason it can indicate different programs for different problems.

[10] For a matrix $M \in \mathbb{R}^{d\times d}$ the k-sparse norm of M is defined as $\max_{\|v\|=1, v \text{ } k\text{-sparse}} \|Mv\|$.

An insightful way to view this reasoning is through the lens of *sum-of-squares proofs* (we assume knowledge of the sum-of-squares paradigm here and otherwise direct the reader to Chapter 2). There exists a degree $O(d+n)$ sum-of-squares proof (also called a sum-of-squares *certificate*) of Eq. (1.1.1), and consequently a sum-of-squares algorithm that captures exhaustive search (in the sense that the two procedures achieve comparable guarantees). In a similar fashion, the basic SDP can be interpreted as a sum-of-squares algorithm of degree 2. Its worse guarantees then stems from the fact that, unfortunately, we only know degree-2 sum-of-squares certificates of the looser inequality $\|N\|_{\text{k-sparse}} \leqslant \tilde{O}(k^2)$. Now, these algorithms are robust because, as entry-wise perturbations bounded by $O(1/n^{1/4})$ cannot significantly change the value of the k-sparse norm of N, the same certificates can also be obtained in the semi-random case!

More generally, we will see how algorithms that come with certificates of *key statistics* of the problem at hand are intrinsically robust, in the sense that small perturbations – which by virtue of being small cannot significantly change such statistics – cannot be used to fool them. In contrast, fragile algorithms – which do not produce such certificates – may be easily fooled by the same adversarial perturbations.

Certification algorithms for Sparse PCA. In line with the emerging picture above, we are able to completely characterize the computational landscape of sparse PCA with adversarial signs. We summarize these results here in two tables and then discuss them in detail in Chapter 3. Our algorithmic conclusion reads as follows:

> *For the problem of sparse PCA, the sum-of-squares algorithm achieves the best known guarantees among robust polynomial time algorithms. Furthermore, under the restrict computational model of low-degree polynomials, these guarantees are nearly optimal.*

Strong Signal Regime			
Algorithm	*Succeeds if*	*Running Time*	*Robust*
SVD with thresholding	$\beta \gtrsim \sqrt{\frac{d}{n}} + \frac{k \log d}{n}$	$\tilde{O}(nd)$	No
Sum of squares, Theorem 3.2	$\beta \gtrsim \frac{k \cdot t}{n} \left(\frac{d}{k}\right)^{1/t}$ for $d \gtrsim \tilde{O}(tn \log n)^t$	$d^{O(t)}$	Yes

Table 1.1: Algorithmic landscape in the strong signal regime. The requirements of the robust algorithm complements, up to constants, those of the lower bound Theorem 1.1.

Weak Signal Regime			
Algorithm	*Succeeds if*	*Running Time*	*Robust*
Diagonal thresholding	$\beta \gtrsim \frac{k}{\sqrt{n \cdot t}} \sqrt{\log d}$ for $t \leqslant \frac{1}{\ln d} \min\{d, n\}$	$\mathrm{poly}(n) d^{O(t)}$	No
Covariance thresholding / polynomials*[11]	$\beta \gtrsim \frac{k}{\sqrt{n}} \sqrt{\log \frac{d}{k^2}}$ for $k \lesssim \sqrt{d}$ and $k \lesssim \sqrt{n}$	$\mathrm{poly}(n, d)$	No
Sum of squares, Theorem 3.6	$\beta \gtrsim \frac{k}{\sqrt{n \cdot t}} \sqrt{\log d}$ for $t \leqslant \frac{1}{\ln d} \min\{d, n\}$	$\mathrm{poly}(n) d^{O(t)}$	Yes
Sum-of-squares, Theorem 3.5	$\beta \gtrsim \frac{k}{\sqrt{n}} \sqrt{\log \frac{d}{k^2}}$ for $k \lesssim \sqrt{d}$ and $k \lesssim \sqrt{n}$	$\mathrm{poly}(n, d)$	Yes

Table 1.2: Algorithmic landscape in the weak signal regime. Novel robust algorithms tightly match the existing bounds of fragile algorithms. More details can be found in Chapter 3.

1.1.3 Sharp phase transitions in the presence of adversarial corruptions

Among estimation problems, perhaps the most fascinating phase transition phenomenon can be observed in the context of *stochastic block models (SBMs)*. In its most basic form, the stochastic block model describes the following joint distribution $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \gamma)$ between a vector x of n binary labels and an n-vertex graph $\mathbf{G}$:

- draw a vector $\mathbf{x} \in \{\pm 1\}^n$ uniformly at random,
- for every pair of distinct vertices $i, j \in [n]$, independently create an edge $\{i, j\}$ in the graph $\mathbf{G}$ with probability $(1 + \gamma \cdot \mathbf{x}_i \cdot \mathbf{x}_j) \cdot \frac{d}{n}$.

Note that for distinct vertices $i, j \in [n]$, the edge $\{i, j\}$ is present in $\mathbf{G}$ with probability $(1 + \gamma) \cdot \frac{d}{n}$ if the vertices have the same label $\mathbf{x}_i = \mathbf{x}_j$ and with probability $(1 - \gamma) \cdot \frac{d}{n}$ if the vertices have different labels $\mathbf{x}_i \neq \mathbf{x}_j$. Given a graph $\mathbf{G}$ sampled according to this model, the goal is to recover the (unknown) underlying vector of labels as well as possible.

More general versions of the stochastic block model allow for more than two labels, non-uniform probabilities for the labels, and general edge probabilities depending on the label assignment. However, many of the algorithmic phenomena of the general version can in their essence already be observed for the basic version, so we limit our discussion to that.

Due to its simplicity, the stochastic block model has emerged independently in different communities as a way to represent structured graph models, from statistical physics

[11] In [dKNS20], a fragile algorithm (based on low-degree polynomials) extending the guarantees of Covariance Thresholding to a (slightly) larger set of parameters was introduced. At the granularity of our discussion here this is not relevant and hence, we omit it for clarity.

to sociology and computer science (see [Abb17] and references therein). An attractive feature of this model is the existence of a sharp threshold, below which it is information theoretically impossible to recover the community structure, and above which efficient algorithms exists.[12] One downside of studying the spiked covariance model is the fact that the phase transitions are known up to constant factors. This makes it unattainable to provide a fine grain analysis of the trade-off between accuracy, signal-to-noise ratio and robustness. The stochastic block model instead does not present a similar issue and it is a perfect candidate to study the fine-grade trade-offs of adversarial perturbations.

We say that an algorithm achieves *(weak) recovery* for the stochastic block model $\{\mathsf{SBM}_n(d,\gamma)\}_{n\in\mathbb{N}}$ if the correlation of the algorithm's output $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ and the underlying vector $\mathbf{x}$ of labels is bounded away from zero as n grows. [13]

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\gamma)} \left[\tfrac{1}{n}|\langle x, \hat{x}(\mathbf{G})\rangle|\right] \geqslant \Omega_{\gamma,d}(1)\,.$$

(Here, $\Omega_{\gamma,d}(1)$ hides a positive number depending on ε and d but independent of n). A series of seminal works [MNS15b, MNS18, Mas14] showed that weak recovery is possible (also computationally efficiently) if and only if $d > \gamma/\varepsilon^2$, confirming a conjecture [DKMZ11] from statistical physics (this threshold is commonly referred to as the Kesten-Stigum threshold).[14]

Robustness. In an apparently similar fashion to the sparse PCA problem, a fascinating issue arises in the context of the works on weak recovery when one takes into account *robustness*: the algorithms used to show that weak recovery is possible when $d > 1/\gamma^2$ are fragile in the sense that adversarially modifying a vanishing fraction of edges could fool the algorithm into outputting labels completely unrelated to the true labels. The reason is that these algorithms are based on particular kinds of random walks (self-avoiding or non-backtracking) that can be affected disproportionally by adding small cliques or other dense subgraphs.

As expected, other kinds of algorithms (based on semidefinite programs certifying certain matrix norms) have stronger guarantees in robust settings [FK01, GV16, MS16]. However, these algorithms are only known to work for $d > C/\gamma^2$ for an absolute constant $C > 1$, even in the non-robust setting.[15] Further complicating the picture,"monotone adversaries" (a *monotone adversary* is allowed to change an arbitrary number of edges

[12] In the multi-community settings with $k > 4$, there is a gap between the information theoretic threshold and the computational threshold.

[13] We remark that other definitions of weak recovery require that the algorithm achieves constant correlation with probability tending to 1 as n grows. It turns out that in the setting we consider it is always possible to boost the success probability from $\Omega(1)$ to $1 - o(1)$. See Section 4.1.

[14] For k communities a similar computational threshold is known.

[15] For high-degree graphs, the best-known bound for the basic semidefinite programming relaxation to achieve weak-recovery is of the form $\gamma^2 d > 1 + o_d(1)$ for an unspecified function $o_d(1)$ tending to 0 as d grows [MS16].

as long as each change increases the likelihood of the planted labeling) were shown in [MPW16] to: *(i)* change the threshold so that a bound of the form $d > C \cdot 4/\varepsilon^2$ for $C > 1$ is required to ensure that weak recovery remains possible; *(ii)* have the same optimal solutions for the aforementioned semidefinite program as their vanilla counterpart.

Despite this statistical price to pay for robustness against monotone perturbations, for a natural class of adversaries that are allowed to alter any vanishing fraction of edges, it remained open whether the threshold for weak recovery changes, or whether weak recovery robust against this class of adversaries is possible. In the asymptotic setting $d \to \infty$, this kind of robustness was achieved using the basic semidefinite programming relaxation for the likelihood maximization problem [MS16]. However, as previously mentioned, non-rigorous statistical-physics calculations [JMRT16] suggest that the same algorithm cannot achieve the threshold for constant degree parameter d. Some more recent algorithms [ABARS20], based on graph powering, were shown to achieve the Kesten-Stigum (KS) threshold while being robust against certain weak perturbations (e.g. introduction of a few tangles), but turned out to be fragile against the general perturbations discussed here.

Groundbreaking work [BMR21] provided some early insight. Banks, Mohanty and Raghavendra developed a polynomial-time algorithm for the corresponding distinguishing problem: given a graph drawn from $\mathsf{SBM}_n(d, \gamma)$ with $d > 1/\gamma^2$ and an Erdős-Rényi random graph with the same expected number of edges, the algorithm can distinguish between the two graphs and robustly so, i.e., even after altering a small constant fraction of edges. In the vanilla stochastic block models, it is easy to distinguish a graph sampled from $\mathsf{SBM}_{d,\gamma}$ from an Erdős-Rényi graph (counting the number of triangles suffices). Thus it remained unclear whether this result was the consequence of a similar phenomenon. Nevertheless, [BMR21] introduced novel ideas. Similarly to previous algorithms achieving the threshold $d > 1/\gamma^2$, the algorithm takes into account certain kinds of random walks (specifically non-backtracking ones). A crucial difference is that the algorithm considers only walks of constant length (for fixed d and γ) as opposed to walks of logarithmic length like previous algorithms. In order to leverage the more limited information provided by shorter walks, the algorithm in [BMR21] needs to employ heavier convex optimization techniques (specifically sum-of-squares). The upside is that robustness follows almost directly: The altered edges can affect only a small constant fraction of the (constant-length) walks considered by the algorithm. Correspondingly, the effect on the optimal value for the optimization is small.

Surprisingly, inn this we will see that, indeed, there is *no price* to pay for robustness. The underlying message is that by designing more *sophisticated* certification algorithms –which can leverage as much as possible the information underlying the data– it is possible to achieve optimal guarantees even in the robust settings.

Concretely, we say that an algorithm that given a graph $\mathbf{G}$ outputs an estimate $\hat{x}(\mathbf{G})$ for the community labels of $\mathbf{G}$ achieves ρ-robust weak recovery for $\{\mathsf{SBM}_n(d, \gamma)\}_{n\in\mathbb{N}}$ if

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\gamma)} \min_{G^\circ \in N_\rho(\mathbf{G})} \left[\tfrac{1}{n}|\langle \mathbf{x}, \hat{x}(G^\circ)\rangle|\right] \geqslant \Omega_{d,\varepsilon}(1)\,,$$

where $N_\rho(\mathbf{G})$ is the set of graphs G° that can be obtained from $\mathbf{G}$ by changing at most a ρ-fraction of its edges[16] (so that $|E(\mathbf{G}) \triangle E(G^\circ)| \leqslant \rho \cdot (|E(\mathbf{G})| + |E(G^\circ)|)$).

The following theorem, which we prove in Chapter 4, shows that this notion of robustness does not significantly alter the statistical threshold and that robust polynomial-time algorithms exist that work all the way up to this threshold.

Theorem 1.2. *For every γ, d with $d > 1/\gamma^2$, there exists $\rho > 0$ such that ρ-robust weak recovery for $\{\mathsf{SBM}_n(d, \gamma)\}_{n\in\mathbb{N}}$ is possible. Moreover, the underlying algorithm runs in polynomial time.*

The SDP algorithm behind Theorem 1.2 requires several novel careful ideas and significantly departs from previous algorithms for robust recovery (including those discussed in Section 1.1.2). A key challenge is the peculiar optimization landscape underlying our algorithm: the planted partition may be far from optimal in the sense that completely unrelated solutions could achieve the same objective value. This phenomenon is related to the push-out effect at the BBP phase transition for PCA. The algorithm in Theorem 1.2 is the first to achieve robust recovery in the presence of such a push-out effect in a non-asymptotic setting.

Weak recovery in the presence of node corruptions. Is it possible to achieve weak recovery for stochastic block models at the Kesten-Stigum threshold when the number of corruptions is *larger* than the number of edges in the original graph? Without additional constraints the answer is clearly no. Surprisingly, we show that under the *node* corruption model the answer turns to be yes!

Definition 1.3 (Node-corrupted SBM)**.** Given $\mu \in [0, 1)$ and $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \gamma)$, an adversary may choose up to μn vertices in $\mathbf{G}$ and arbitrarily modify edges incident to at least one of them to produce the corrupted graph G°.

That is, upon drawing a graph from a stochastic block model, an adversary may pick a constant fraction of the nodes and arbitrary alter all the edges incident to at least one of them.

There are several reasons why this model is attractive. First, the adversary is allowed to change an arbitrary number of edges for each corrupted vertex and could introduce up to $O(\mu n^2)$ edges. This means that, in contrast to the other models discussed, the *magnitude* of the corruptions can be significantly larger than the signal! While vertices of untypically large degree are algorithmically easy to identify and remove, less naive adversaries may remove all edges of an arbitrary subset of μn vertices, and replace them by roughly d spurious edges of their choosing. Such an adversary would introduce $O(\mu \cdot d \cdot n)$ edges, causing the algorithm previously discussed (as well as the basic SDP [MS16]) to fail when $\mu \geqslant \gamma$.

[16] That is, each G° can be obtained from $\mathbf{G}$ through a sequence of $\rho \cdot |E(\mathbf{G})|$ edits, each consisting of an addition or deletion.

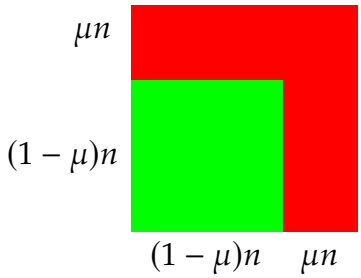


Figure 1.1: For stochastic block models with node corruptions, the entries in the red area are adversarially corrupted, while the entries in the green area are sampled as in stochastic block model.

Second, it is not a priori clear if weak recovery at the KS threshold is possible or if there is some price to pay for robustness. Indeed, it is easy to see that weak recovery becomes information theoretically impossible when more than $\gamma \cdot d \cdot n$ arbitrary edges are corrupted.[17]

Third, this model appears to be "close in spirit" to the canonical μ-Huber contamination model (see [DK19]) studied in the context of clustering mixtures of Gaussians [HL18, KSS18, BDH+20, BDJ+22] (the adjacency matrix of G is expected to look as in Fig. 1.1). Compared to those settings however, it allows for the investigation of sharp phase transitions.

Among previous works, it is important to mention that *none* of the aforementioned algorithms work against such adversarial corruptions. [LM22] showed how to obtain a polynomial-time algorithm –robust to $\mu = o(1)$ node corruptions[18]–that achieves *optimal* recovery rates (i.e. beyond our weak recovery objective) when $\gamma^2 d - 1$ is a sufficiently large constant. Although this algorithm provides weak recovery for large values of $\gamma^2 d$, it falls short of reaching the KS threshold, requiring $\gamma^2 d > C > 1$. [SM19] provided an algorithm achieving the KS threshold in sparse graph, but robust to only $O(n^{0.001})$ vertices corruptions in sparse graphs.

We say that an algorithm that given a graph $\mathbf{G}$ outputs an estimate $\hat{x}(\mathbf{G})$ for the community labels of $\mathbf{G}$ achieves μ-node-robust weak recovery for $\{\mathsf{SBM}_n(d,\gamma)\}_{n\in\mathbb{N}}$ if

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\gamma)} \min_{G^\circ\in V_\mu(\mathbf{G})} \left[\tfrac{1}{n}|\langle \mathbf{x}, \hat{x}(G^\circ)\rangle|\right] \geqslant \Omega_{d,\gamma}(1)\,, \tag{1.1.2}$$

where $V_\mu(\mathbf{G})$ is the set of graphs that can be obtained from $\mathbf{G}$ as in Definition 1.3. Similarly to the context of edge corruptions, the underlying message is that by better leveraging

[17] An adversary may remove each intra-cluster edge with probability $\gamma \cdot d/n$ and add each inter-cluster edge with probability $\gamma \cdot d/n$. With high probability such process alter at most $\gamma \cdot d \cdot n(1 + o(1))$ edges. The graph now is indistinguishable from an Erdős-Rényi graph.

[18] The algorithm in [LM22] is further robust against a different type of adversary corruptions with unbounded monotone changes from semi-random model previously mentioned.

the global information contained in the data, it is possible to achieve node robust weak recovery down to the Kesten-Stigum threshold.

Theorem 1.4. *For every γ, d with $\delta := d\gamma^2 - 1 > 0$, and $\mu \leqslant \Omega_\delta(1)$, μ-node-robust weak recovery is possible.* [19] *Moreover, the underlying algorithm runs in polynomial time.*

This algorithm is the first one that succeeds down to the KS threshold under node corruptions, [LM22] cannot work unless δ is sufficiently large, and the algorithm discussed in Theorem 1.2 cannot tolerate the corruption of $\Omega_\delta(1)$ vertices. The dependence on δ is necessary: if μ is a fixed constant, then recovery is impossible for a small enough constant δ (see Section 5.5 for details). The techniques behind Theorem 1.4 can be applied to related problems such as $\mathbb{Z}_2$-*synchornization*. We show these results in Chapter 5.

1.1.4 Constraint satisfaction problems with adversarial signs

Another natural kind of adversarial perturbations consists of corruptions that *adaptively* break the symmetry of the distribution at hand. To see how powerful such adversaries can be, notice that, in the spiked covariance model previously considered, an adversary that can flip the signs in the observations $\{\mathbf{y}_i\}_{i\in[n]}$ could completely hide the signal in input.

Natural problems in which these corruptions arise are constraint satisfaction problems (henceforth CSPs). CSPs play a major role in computer science. Because of its centrality to the theories of proof complexity [BSB02] and of average-case complexity, and its connection to other questions in cryptography [ABW10], computational complexity[Fei02], and statistical physics [CLP02], the complexity of solving constraint satisfaction problems has been extensively studied since the 1980s.

A canonical example of a CSP that captures the computational phenomena at play, is k-XOR. Indeed, it is possible to essentially reduce arbitrary constraint satisfaction problems to k-XORs [Fei07] (we provide a self-contained simple proof in Section 6.6). A XOR clause over k variables is a constraint of the form $x_{i_1} \cdots x_{i_k} = \sigma$ for $\sigma \in \{\pm 1\}$. Then a k-XOR instance can be represented by a symmetric k-th order tensor T such that $T_{i_1,\dots,i_k} = 0$ if there is no constraint on the k-tuple of variables $x_{i_1}, \dots, x_{i_k}$, and otherwise $T_{i_1,\dots,i_k} \in \{\pm 1\}$ depending on the right-hand-side of the constraint. The value of an assignments $x \in \{\pm 1\}^n$ is given by

$$\sum_{i_1,\dots,i_k} T_{i_1,\dots,i_k} x_{i_1} \cdots x_{i_k} = \langle T, x^{\otimes k} \rangle = \langle S \odot Z, x^{\otimes k} \rangle \,, \tag{1.1.3}$$

where in the last step we decomposed the tensor T as the Hadamard product[20] of the

[19] $\mu \leqslant \Omega_\delta(1)$ here means that μ is bounded by a constant depending on δ.

[20] The Hadamard product of two tensors A, B of same size produces a tensor C of same size where each entry (here indexed by the multi-index β) C_β has value $A_\beta \cdot B_\beta$.

symmetric *sign* tensor S with entries in $\{\pm 1\}$ and the symmetric *indicator* tensor Z such that $Z_{i_1,\ldots,i_k} = 1$ if there is a clause containing $i_1, \ldots, i_k$ and zero otherwise.

There is a vast theory of the complexity of both worst-case satisfiability [BJK05] and approximability [Rag09] of CSPs. The resulting picture is grim: for a large class of CSP instances [Cha16, MR08, FLP15], the Exponential time hypothesis [IKW02] rules out sub-exponential time algorithms that beat the random assignment. In the average case however the computational landscape is more promising. A random k-XOR instance $\boldsymbol{\mathcal{I}}$ over n variables can be generated by drawing m clauses independently and uniformly at random, i.e. by picking a symmetric sign tensor S and a symmetric indicator tensor with km non-zero entries, both uniformly at random (the exact sampling process is inconsequential to our discussion). The computational complexity is captured by the *density* $\alpha = m/n$ of the instance. Rigorous evidence (along with proofs for specific settings[DSS15]) suggests the existence of a critical density α_k, below which a satisfying assignment exists, and above which the instance is unsatisfiable with high probability [COGL07, Ach09, BKS15, RRS17].

Two distinct natural algorithmic tasks are associated with these regimes. For $\alpha > \alpha_k$ the goal is that of *refuting* the instance, by showing a *certificate* of unsatisfiability. A *strong refutation* of a 3-SAT formula is a certificate, verifiable in polynomial time, that every assignment fails to satisfy a constant fraction of the clauses. Conversely, for satisfiable instance, the goal is to find the optimal assignment.

Among several important algorithmic milestones, we mention the idea of using spectral techniques to find refutations and strong refutations (introduced in [FGK05] and then refined in subsequent work) and a reduction from the problem of finding strong refutations for random 3SAT to the problem of finding strong refutations for random 3XOR (introduced in [Fei02] and then refined in subsequent work). We refer the reader to the introduction of [AOW15] for an extended survey of algorithmic ideas and results related to refutations of random constraint satisfaction problems. In particular, the emerging picture shows that refutations and PTAS for solving constraint satisfaction problems [AOW15, AJT19] exist for random instances on n variables and $n^{k/2}(\log n)^{O(1)}$ constraints. (When k is even, $O(n^{k/2})$ constraints suffice.[21]). But below this value the problem becomes computationally intractable [ABW10, DLSS14, Dan16, BKS15].

Semi-random CSPs. There are two natural way one can extend these formulations to semi-random models. One one side, one can consider smoothed instances. In the context of k-XOR these are semirandom instances generated by picking an arbitrary worst-case instance, then randomly and uniformly resampling the sign of each clause.[22] That is, in Eq. (1.1.3) the sign tensor $\mathbf{S}$ is random but the indicator vector Z is worst-case. On the other

[21] In the even settings several of the technical challenges related to the study of tensors disappear, as the instance can be represented as a $n^{k/2} \times n^{k/2}$ matrix. The same ideas do not work for the odd case. An intuition of this difference can be found observing that the spectrum of a rectangular random matrix is characterized by its largest dimension.

[22] In fact, one needs not to resample all signs but only a sufficiently large constant fraction.

side, one can consider instances where the clauses are sampled randomly but the signs are worst-case. This corresponds to picking an arbitrary tensor S and a random tensor $\mathbf{Z}$ in Eq. (1.1.3).

The effects of the two adversaries are different and essentially incomparable. In the former case, the underlying hypergraph spanned by the instance is arbitrary but has hyper-edge weights that are symmetrically distributed. In the latter case, the hypergraph satisfies typical expansion properties of random hypergraphs, but weights are worst case. Surprisingly, [GKM22] showed that there is no price to pay for robustness against *smoothed* instances: efficient strong refutation algorithms exist for $m \geqslant \Omega(n^{k/2})(\log n)^{O(1)}$. We tackle the second family of adversarial corruptions introducing a PTAS that succeeds at the *sharp* threshold $m \gtrsim n^{k/2}$. As often is the case, the underlying insight also immediately leads to sharp strong refutations whenever $m \gtrsim n^{k/2}$ (in fact, both algorithms boil down to producing sharp certificates of injective tensor norms, see Chapter 6). Hence, the result not only extend the state of the art from the average-case to semi-random models, but breaks a long standing barrier that immediately applies to the random settings, matching the computational threshold predicted by lower-bounds (up to constant factors).[23]

Theorem 1.5 (Semi-random k-CSPs, informal)**.** *Let n , k be positive integers, $\varepsilon > 0$, n and $n^{-k/2}/\varepsilon^2 < 1$. Let $P : \{-1,+1\}^k \to \{0,1\}$ be a Boolean k-ary predicate. Let $\mathcal{I}$ be a $CSP(P)$ instance constructed through the following process:*

- *Sample a random $CSP(P)$ instance $\mathcal{I}'$ with at least $n^{k/2}/\varepsilon^2$ constraints.*
- *Given $\mathcal{I}'$, for each clause in $\mathcal{I}'$, replace the sign pattern with an arbitrary (possibly adversarial) sign pattern.*

There exists a polynomial time algorithm that, with probability at least 0.99, returns an assignment $\hat{\mathbf{x}}$ with value $\mathrm{Val}_{\mathcal{I}}(\hat{\mathbf{x}}) \geqslant \mathrm{Opt}_{\mathcal{I}} - O(\varepsilon)$.

1.1.5 From robustness to privacy

Privacy in machine learning and statistical tasks has recently become of critical importance. New regulations, renewed consumer interest as well as privacy leaks, have led the major actors to adopt privacy-preserving solutions for the machine learning [Ela15, app17, 20221]. This new push has resulted in a flurry of activity in algorithm design for private machine learning [MNVT22, KSSU19, CKM+21, TCK+22, KMV22, AL22, DNT15]. Despite this effort, with few exceptions (e.g. [KMV22]) these algorithms provides error guarantees that are significantly worse than their non-private counterparts. Hence, improving on these results remain a pressing open question. Although not immediately obvious, privacy can be seen as a particular extension of robustness.

[23] As we will see in Chapter 6, our result and [GKM22] require solving different technical challenges, therefore it remains unclear whether the gap between random instances and smooth instances is due technical reasons, or whether there is some hidden cost of robustness.

The general picture is the following: a machine learning algorithm receives in input a database, say containing information about individuals, and processes this data to extract meaningful, structured, global information about the whole dataset, say correlation between habits and illnesses. For the algorithm to be private, it should be *impossible*[24] to infer from the output, with reasonable confidence, information about *any* single element in input (or individual in our example). While other notions of privacy exists (e.g. k-anonymity) the de facto privacy standard is the differential-privacy framework of Dwork, McSherry, Nissim, and Smith [DMNS06]. In this framework, the privacy quality is governed by two parameters, ε and δ, which in essence tell us how the probability of seeing a given output changes (both multiplicatively and additively) between two datasets that differ by any individual data element. This notion, in essence, quantifies the amount of information *leaked* by a given algorithm on a single data elements. More precisely:

Definition 1.6 (Differential privacy)**.** An algorithm $\mathcal{M} : \mathcal{Y} \to \mathcal{O}$ is said to be (ε, δ)-differentially private for $\varepsilon, \delta > 0$ if and only if, for every $S \subseteq \mathcal{O}$ and every datasets $Y, Y' \in \mathcal{Y}$ differing in at most one element[25], we have

$$\mathbb{P}[\mathcal{M}(Y) \in S] \leqslant e^{\varepsilon} \cdot \mathbb{P}[\mathcal{M}(Y') \in S] + \delta\,.$$

The goal of the algorithm designer is then to come up with differentially-private algorithms for ε being a small constant and δ being of order $1/n^{\Theta(1)}$.

From robustness to differential privacy. The goal of designing privacy-preserving machine learning models turn out to be closely related to the design of robust algorithms. Indeed on a high level, in both cases the objective is to design algorithms that extract global information without over-relaying on individual data samples.

Concretely, robust parameter estimation tends to morally follow a two-steps process: *(i)* argue that typical inputs are well-behaved, in the sense that they satisfy some property which can be used to accurately infer the desired global information, *(ii)* show that adversarial perturbations cannot significantly alter the quality of well-behaved inputs, so that it is still possible to obtain an accurate estimate. In fact, the estimation algorithms discussed in previous sections are consistent with this paradigm. The analysis of private estimation algorithms can also be conceptually divided in two parts: *utility*, which is concerned with the accuracy of the output, and *privacy*, which ensures there is no leak of sensitive information. In particular, according to Definition 1.6, privacy can be interpreted as the requirement that, for any distinct inputs $Y, Y' \in \mathcal{Y}$, the change in the output is *proportional* to the distance[26] between Y and Y'.

It is easy to see this as a generalization of robustness: while robust algorithm needs this property to hold for typical inputs, private algorithms needs to satisfy it for *any possible*

[24] From an information theoretic perspective.

[25] We say that Y, Y' are neighboring datasets.

[26] The notion of distance is inherently application dependent.

input. Then, stability of the output immediately implies that the introduction of small additive noise to the output (e.g. via standard privatization mechanisms) yields privacy. If the introduced noise is small, then utility is also likely to be preserved. In Chapter 7 we show that natural modifications of known robust (sum-of-squares) algorithms –such as algorithms for learning Gaussian mixtures models [HL18, KSS18, ST21] or community detection in graphs [MS16, GV16]– satisfy the privacy requirement above. We formalize this extended notion through a simple, yet key, insight: if two strongly convex functions over constrainted sets –where both the function and the set may depend on the input– are point-wise close (say in a ℓ_2-sense), their minimizers are also close (in the same sense). The alternative perspective is that projections of points that are close to each other, onto convex sets that are point-wise close, must also be close. The result is a clean, user-friendly, *framework to turn robust estimation algorithms into private algorithms*, while keeping virtually the same guarantees. We apply this paradigm to the spherical Gaussian Mixture Model:

Model 1.7 (Mixtures of spherical Gaussians)**.** Let $D_1, \dots, D_k$ be Gaussian distributions on $\mathbb{R}^d$ with covariance Id and means $\mu_1, \dots, \mu_k$ satisfying $\|\mu_i - \mu_j\| \geqslant \Delta$ for any $i \neq j$. Given a set $\mathbf{Y} = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ of n samples from the uniform mixture over $D_1, \dots, D_k$, estimate $\mu_1, \dots, \mu_k$.

We prove the result below:

Theorem 1.8 (Privately learning mixtures of spherical Gaussians, informal)**.** *Consider an instance of Model 1.7. Let $t > 0$ be such that $\Delta \geqslant O\left(\sqrt{t}k^{1/t}\right)$. For $n \geqslant \Omega\left(k^{O(1)} \cdot d^{O(t)}\right), k \geqslant (\log n)^{1/5}$, there exists an algorithm, running in time $(nd)^{O(t)}$, that outputs vectors $\hat{\boldsymbol{\mu}}_1, \dots, \hat{\boldsymbol{\mu}}_k$ satisfying*

$$\max_{\ell \in [k]} \left\|\hat{\boldsymbol{\mu}}_\ell - \mu_{\pi(\ell)}\right\|_2 \leqslant O(k^{-10}),$$

with high probability, for some permutation $\pi : [k] \to [k]$. Moreover, for $\varepsilon \geqslant k^{-10}$, $\delta \geqslant n^{-10}$, the algorithm is (ε, δ)-differentially private[27] for any input Y.

The Theorem matches the state-of-the-art results in the *non-private* settings [HL18, KSS18, ST21, LL22] (see Chapter 7 for an in-depth discussion). Among differentially private algorithms, prior results could only learn a mixture of k-spherical Gaussian either if: (1) they were given a ball of radius R containing all centers [KSSU19, CKM+21];[28] or (2) the minimum separation between centers needs an additional additive $\Omega(\sqrt{\log n})$ term, i.e. a super-polynomial increase in the minimum required separation whenever $k \leqslant n^{o(1)}$. Theorem 1.8 is the first to get the best of both worlds and, just like its non-private sum-of-squares counterparts, seamlessly works for the significantly more general class of mixtures of Poincaré distributions.

[27] Our notion of adjacent databases here is the obvious one. See Definition 7.51.

[28] In [KSSU19, CKM+21] the sample complexity of the algorithm depends on this radius R.

We also consider stochastic block models, where we focus on *exact recovery*. Here the goal is to actually recover the true partition with high probability. As in the weak-recovery settings, the algorithmic landscape is well-understood [DKMZ11, Mas14, ABH15, MNS15b, MNS18, Abb17]: exact recovery is possible (and can be achieved efficiently) if and only if $\frac{d}{\log n}\left(1-\sqrt{1-\gamma^2}\right) \geqslant 1$.

Theorem 1.9 (Private exact recovery of SBM, informal)**.** *Let* $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \gamma)$*. For any* $\gamma, d, \varepsilon, \delta > 0$ *satisfying*

$$\frac{d}{\log n}\left(1-\sqrt{1-\gamma^2}\right) \geqslant \Omega(1) \quad \textit{and} \quad \frac{\gamma d}{\log n} \geqslant \Omega\left(\frac{1}{\varepsilon^2}\cdot\frac{\log(1/\delta)}{\log n}+\frac{1}{\varepsilon}\right),$$

there exists a polynomial time algorithm that, on input $\mathbf{G}$*, returns* $\hat{x}(\mathbf{G}) \in \{\mathbf{x}, -\mathbf{x}\}$ *with probability* $1 - o(1)$*. Moreover, the algorithm is* (ε, δ)*-differentially private with respect to edge changes.*

For any constant $\varepsilon > 0$, Theorem 1.9 states that (ε, δ)-differentially private exact recovery is possible, in polynomial time, already a constant factor close to the non-private threshold. Previous results [MNVT22] could only achieve comparable guarantees in time $O(n^{O(\log n)})$.

A price for differential privacy? The notion of differential privacy, so tightly connected to that of robustness, also raises a similar question. Namely, whether there is an inherent price an algorithm has to pay to be differentially private and whether this price is in the form of a statistical or computational barrier. That is, analogously to robustness, whether there is a *price for privacy*. In contrast to the results in Section 1.1.3, Theorem 1.9 cannot provide a sharp analysis of the phase transition in the context of differential privacy. Nevertheless, for sparse graphs that admit exact recovery ($d = \Theta(\log n)$), it highlights a natural, *polynomial* trade-off between, the desired accuracy, the bias parameter γ and the privacy parameters ε, δ. Our next result shows that to some extent this trade-off is inherent and the guarantees of our algorithm are almost tight.

Theorem 1.10 (Price for privacy, informal)**.** *Suppose there exists an* ε*-differentially private algorithm such that for any balanced* $x \in \{\pm 1\}^n$*, on input* $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, x)$*,*[29] *outputs* $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ *satisfying*

$$\mathbb{P}(\mathrm{err}(\hat{x}(\mathbf{G}), x) < \zeta) \geqslant 1 - \eta\,.$$

Then,

$$\varepsilon \geqslant \Omega\left(\frac{\log(1/\zeta)}{\gamma d}+\frac{\log(1/\eta)}{\zeta n \gamma d}\right).$$

Notice that this is a lower bound for a large range of error rates (partial to exact recovery). Importantly, while formally incomparable, in the context of exact recovery $\zeta < 1/n$, this lower bound suggests that the guarantees obtained by Theorem 1.9 might

[29] We use this notation to indicate the conditional distribution of $\mathbf{G}$, given $\mathbf{x}$ for stochastic block models.

be close to optimal. In general, setting $\delta = n^{-\Theta(1)}$, implies that Theorem 1.9 achieves $(\varepsilon, n^{-\Theta(1)})$-private exact recover, i.e., whenever[30] $\varepsilon \geqslant \Omega\left(\sqrt{\frac{\log n}{\gamma d}}\right)$. Theorem 1.10 states that in this exact recovery setting $\varepsilon \geqslant \Omega\left(\frac{\log n}{\gamma d}\right)$ is necessary. In particular, for the *gold standard* of differential privacy (i.e. the parameters that are required for real-world applications) $\varepsilon \approx 1, \delta \leqslant O(1/n)$, Theorem 1.9 matches the lower bound up to constant factors. The existence of this separation between our efficient algorithmic result and the statistical lower bound remains an intriguing open question.

1.1.6 Fast and robust algorithms

The robust algorithms discussed in previous sections have one main weakness: their *running time*. On an n-sized input, algorithms based on sum-of-squares runs in time of the order at least $O(n^C)$, for some large constant $C > 1$, even when only requiring low-degree certificates. However, to be useful in practice, modern inference algorithms need to run on high dimensional datasets over billions of observations and thus, have to essentially scale *linearly* in the sample size. In other words, without additional improvements, there is virtually no scenario in which the algorithms seen in the previous chapters can be used on real world datasets.

This phenomenon arises in the context of robustness due to the fact that more easily implementable algorithms, such as spectral methods, heavily rely on extremely fragile distributional assumptions. But for the same reason it is even more central to the study of worst-case complexity, where semidefinite programming has historically had a fundamental role [GW95, KKMO07, KV15, ARV09, Rag08, RST12, ABS15], and where the input is arbitrary.

A very reach and successful theory –often called the matrix multiplicative weights framework (MMW)[31]– has been developed in an effort to design fast algorithms achieving results comparable to their SDP counterparts [AK07, She09, Ste10a]. On a high level, this approach is based on the following steps: *(i)* find an assignment of the program variables that is only approximately feasible, *(ii)* round this infeasible solution into a feasible, integral solution. The key underlying idea is that, for many problems, if the subset of constraints that gets satisfied is chosen carefully, then the rounding algorithm works even though the starting assignment is not feasible! Finally, one obtains a running time improvement if this crucial subset of constraint can be identified and satisfied *quickly*.

Of course, the error guarantees that can be achieved in polynomial time for worst-case instances are significantly worse than what robust algorithms achieve on semi-random inputs. So directly applying results along the outline above does not yield satisfying results.

[30] In addition to the condition independent of ε.

[31] This can be seen as an application of the mirror descent algorithm with the von Neumann negative entropy as the chosen mirror map.

However, these ideas can be adapted to semi-random problems.

Matrix multiplicative weights for semi-random models. Sum-of-squares algorithms for semi-random inputs relies on certifying some crucial property that typical random instances (i.e. without adversarial perturbations) satisfy, *and* that are *still* approximately satisfied even upon introducing adversarial corruptions. Thus, the additional challenge in applying the MMW framework to these settings amounts to showing that, with high probability over the input, the rounding scheme works even for *infeasible* solutions that satisfy this property only approximately.

A convenient example to showcase these ideas is the balance cut problem.

Problem 1.11 (a-balanced cut)**.** Let $a \in [0\,, 1/2]$ and let G be a graph on n vertices. Find the partition (A, B) with maximum cut that also satisfies $\min\{|A|, |B|\} \geqslant an$.

In a celebrated result, [ARV09] Arora, Rao and Vazirani introduced a first $O(\sqrt{\log n})$-approximation algorithm for Problem 1.11, based on a semidefinite relaxation.[32] Subsequent work [AK07, She09] used the MMW framework to design an $O(\sqrt{\log n}/\varepsilon)$-approximation algorithm running in time $\tilde{O}\left(n^{1+\varepsilon}\right)$.[33] A lot of work has been devoted to finding reasonable random and semi-random models that captures the behavior of real world networks, and to solve the balanced cut problem over them [BCLS87, DF86, Bop87, FK01, McS01, JS93, DI98, BL12, MMV12, MMV14, Pen20, CPRT22]. Indeed the stochastic block models of Section 1.1.3 are premier examples. We study here a related but different model, first introduced in [MMV12].

Model 1.12 (Random cut with monotone perturbations)**.** We consider graphs over n vertices generated through the following process. Let $a \in (0, 1/2)$, $\eta(n) \in (0, 1)$:

(i) The adversary partition $[n]$ into sets A, B satisfying $|A|, |B| \geqslant an$.

(ii) Each edge between A and B is drawn randomly and independently with probability η.

(iii) The adversary arbitrarily adds edges within A and within B.

(iv) The adversary arbitrarily removes edges between A and B.

Model 1.12 is significantly more general than stochastic block models. Indeed it captures both the *antiferromagnetic* settings[34], as well as monotone adversarial perturbations [MPW16]. We remark however, that it remains incomparable with the adversarial models discussed in Section 1.1.3.

[32] Which is captured by degree-4 sum-of-squares.

[33] We remark that this running time is achieved combining the work of [AK07, She09] with the max-flow algorithm in [CKL+22].

[34] Where the probability of an intra-cluster edge is smaller than the probability of an inter-cluster edge.

In [MMV12], the authors designed an $O(1)$-approximation algorithm for Model 1.12 (among other problems). This algorithm relies on heavy machinery, among which it requires to solve polylogarithmically many SDPs, each an instance of the canonical semidefinite relaxation for Problem 1.11. It is possible to go beyond the $O(\sqrt{\log n})$-approximation of [ARV09] using this algorithm because it leverages the randomness of the cut in the true partition. Hence, the main challenge faced in designing a fast $O(1)$-approximation algorithm for Model 1.12 consists of finding a subset of constraints that: *(i)* can be satisfied easily, in the sense that one can find a feasible assignment in near-linear time *(ii)* still *captures* enough of the underlying graph structure so that one can still leverage the randomness of the cut in the true partition to obtain a good approximation, from a solution that satisfies those constraints, but not necessarily others.

We present such an algorithm, which provide guarantees analogous (up to constant factors) to those of [MMV12] but runs in almost linear time.

Theorem 1.13 (Fast and robust algorithm for balanced cut)**.** *Let $\rho > 0$. Let G be a graph over n vertices generated through Model 1.12 with parameters $a > 0, \eta \geqslant \Omega(\frac{(\log n)^2 \cdot (\log\log n)^2}{n})$. There exists an algorithm that on input G, with probability $1 - o(1)$, outputs an $\Omega(a)$-balanced cut of value at most $O(n^2 \cdot \eta \cdot \rho)$, namely a cut where each side has size at least $\Omega(a \cdot n)$.*

Moreover, the algorithm runs in time $\tilde{O}\left(|V(G)|^{1+O(1/\rho^2)} + |E(G)|\right)$.

Remark 1.14 (Fast algorithms for other graph partitioning problems). In the worst-case settings, obtaining fast algorithms for max-cut via the MMW framework is significantly easier than for Problem 1.11. Indeed, the natural SDP relaxation Problem 1.11 can be obtained from the classic max-cut semidefinite relaxation introducing additional constraints (ℓ_2^2 triangle inequality costraints). In a similar fashion, for stochastic block models, it is also possible to speed up some of the robust weak-recovery algorithms discussed in Section 1.1.3. The ideas behind Theorem 1.13 can also be combined with known reductions [AK07] to design an $O(1)$-approximation algorithm for sparsest cut, with the same running time. Furthermore, as we will see in Chapter 8, Theorem 1.13 can be further extended to more sophisticated problems such as the *semi-random hierarchical stochastic block model* of [CKMM19].

1.1.6.1 Practical algorithms

Through the matrix multiplicative weights framework it is possible to design, in some cases, fast algorithms for average-case problems that are also robust to adversarial corruptions. However, there is a fundamental drawback in implementing the matrix multiplicative framework in practice, namely that the algorithm is not numerically stable: small approximation errors compound, possibly altering the final outcome, so that the final error is significantly larger than expected.

Is it possible to circumvent this issue and design robust algorithms that run well *in practice*? It turns out that, for certain, specific, weaker adversarial perturbations, this can indeed be achieved.

Practical but fragile spectral algorithms from sum-of-squares. To grasp the required ideas it is necessary to provide additional context and temporarily cast adversarial perturbations aside. There is by now a vast literature on sum-of-squares algorithms for estimation problems, among others we cite [HSS15, BKS14, BKS15, BGG+16, MSS16, PS17, RRS17, dKNS20, dNNS23, dT23, KSS18, HL18]. For a wide range of such problems, it is possible to capture the sum-of-squares algorithm by low-degree spectral methods [HSSS16, HKP+17, SS17, HSS19, dKNS20, DdL+22]. Compared to classical spectral algorithms, these newer algorithms differ as the entries of the matrices considered are particular low-degree polynomials in the instance. The main advantage of these algorithms is that they are often very fast, running in subquadratic time (if not linear) in the input size. However, in stark contrast with their sum-of-squares counterparts, they suffer from the same limitations of classical spectral algorithms in that they are extremely susceptible to adversarial corruptions. That is, even when the starting sum-of-squares program is robust against certain adversarial perturbations, the related spectral algorithm, which does not certify the relevant statistics, is not expected to be.[35]

Practical algorithms robust against adversarial distributions. Despite the picture painted above, there are certain weaker notion of adversaries for which it is possible to design robust and practical spectral algorithms. [CPRT22] designed spectral algorithms that achieve nearly-optimal guarantees for semi-random models in which the adversary can introduce large perturbations, but does not have access to the drawn observations.

Another possibility is to force the adversary to follow some specific distributional assumptions. One may (correctly) argue that, this is equivalent to change the noise distribution in the original planted problem, however the observation becomes interesting if the new distribution *captures* the type of adversarial perturbations that are known to make the problem computationally harder. That is, when these distributional assumptions capture the price of robustness of the problem at hand.

As a concrete example, we consider again the adversarial sparse PCA model discussed in Section 1.1.1. Theorem 1.1 showed that there exists a distribution over matrices, captured by natural adversarial corruptions, for which when $\beta \leqslant O\left(\frac{k}{n} \cdot t \cdot (d/k)^{1/t}\right)$ and $\beta n/k \leqslant n^{0.49}$, no algorithm captured by the computational model of degree $\leqslant n^{0.001}$ polynomials, can approximately recover the hidden vector x. Hence providing formal evidence of the computational price to pay to solve this natural generalization of sparse PCA and of the

[35] We remark there exists specific settings with notable exceptions. In particular, in the context of tensor decomposition, [SS17, HSS19] introduced spectral algorithms that are robust against malicious perturbations of magnitude only logarithmically smaller than their sum-of-squares counterparts.

near optimality of our sum-of-squares algorithm in Section 1.1.2.

We prove that there exists a near-linear time spectral algorithm achieving comparable guarantees to sum-of-squares of degree $\leqslant 6$ *against* the family of adversaries used to prove Theorem 1.1.

Theorem 1.15 (Practical spectral algorithm for the strong signal regime, informal)**.** *Given an n-by-d matrix $\mathbf{Y}$ of the form,*

$$\mathbf{Y} = \sqrt{\beta}\mathbf{u}x^{\mathsf{T}} + \mathbf{W} + \mathbf{E}\,,$$

for $\beta > 0$, a unit k-sparse vector $x \in \mathbb{R}^d$, a Gaussian matrix $\mathbf{W} \sim N(0,1)^{n\times d}$, a Gaussian vector $\mathbf{u} \sim N(0, \mathrm{Id}_n)$ such that $\mathbf{u}$, $\mathbf{W}$ are distributionally independent, and $\mathbf{E}$ is a n-by$-d$ matrix as in Theorem 1.1 for $t \leqslant 3$.[36] Suppose that $d \gtrsim n^3 \log d \log n$, $k \gtrsim n \log n$ and

$$\beta \gtrsim \frac{k}{\sqrt{n}}\left(\frac{d}{k}\right)^{1/3}.$$

Then there exits an algorithm that computes in time $O(nd \log n)$ a unit vector $\hat{\mathbf{x}} \in \mathbb{R}^d$ such that

$$1 - \langle x, \hat{\mathbf{x}}\rangle \leqslant 0.01$$

with probability at least 0.99.

The algorithm behind the Theorem (which we call SVD-t, where t is the corresponding sum-of-squares degree) captures the behavior of degree $\leqslant 6$ sum-of-squares in Theorem 3.2, but runs in time *nearly linear* in the input size. Unsurprisingly, the algorithm cannot certify upper norm bounds, and so it is not expected to be robust in general. However, it solves the problem when the noise distribution is as in Theorem 1.1 (or when there are no adversarial perturbations). Such adversarial settings are especially interesting as the problem has a nice geometric description, in which the objective is to recover an approximately sparse vector planted in a random subspace. (see Section 3.4.2.1 and Theorem 9.5). Indeed, the famous fast algorithm of [HSSS16] for recovering sparse vectors, corresponds to SVD-4. In other words, this algorithm can be seen as a generalization of [HSSS16]! Similarly, for $t = 2$ the algorithm corresponds to the SVD+thresholding algorithm outlined in Section 1.1.1.

It is important to remark that we do not have a mechanical way to study spectral procedures capturing higher-order sum-of-squares algorithm. Hence, extending Theorem 1.15 to degree $t > 6$ remains a fascinating open question.

We remark that, as the algorithm behind Theorem 1.15 indeed runs well in practice. We provide experiments showing how it improves other previous algorithms in Section 9.2.

[36] More precisely, recall that in Theorem 1.1 we consider a specific distribution over matrices $\mathbf{E}$ (this distribution depends on x , $\mathbf{u}$ and $\mathbf{W}$), and here we mean that $\mathbf{E}$ is sampled from this distribution.

1.2 Main contributions and road-map of the thesis

The central part of the thesis *"The price of robustness"* is devoted to the general study of semi-random models with adversarial perturbations. Specifically in the context of sparse principal component analysis (Chapter 3), stochastic block models (Chapter 4 and Chapter 5) and constraint satisfaction problems (Chapter 6). The rest of the thesis investigates related topics. Respectively privacy (Chapter 7) and algorithms that are both fast and robust (Chapter 8 and Chapter 9). With the exception of Chapter 9 which builds on Chapter 3, each chapter is essentially self-contained and relies only on the common introduction (Chapter 1) and preliminaries (Chapter 2). In particular, the notational convention in each chapter is optimized for the context and hence, while it is consistent with the common preliminaries, it *may* not be consistent with other chapters. Finally, the appendices contain additional discussion and deferred proofs of the various chapters.

Part I: The price of robustness

- Chapter 3 provides algorithms and lower bounds for sparse PCA with adversarial perturbations. It is based on the FOCS'20 paper [dKNS20].
- Chapter 4 provides an algorithm for stochastic block model robust against a constant fraction of adversarial edge perturbations. It is based on the FOCS'22 paper [DdNS22].
- Chapter 5 provides an algorithm for stochastic block models robust against a constant fraction of adversarial node perturbations. It is based on the COLT'23 paper [DdH23].
- Chapter 6 provides sharp strong refutations and robust algorithms for random CSPs. It is based on the CCC'23 paper [dT23].

Part II: Privacy from robustness

- Chapter 7 provides algorithms and lower bound for stochastic block models and Gaussian mixture models that achieve guarantees comparable to their non-private counterparts. It is based on the (in submission) paper [CKM+21] .

Part III: Speeding up robust algorithms

- Chapter 8 provides fast and robust algorithms for balanced cut and hierarchical stochastic block models. It is based on the (in submission) paper [CdM23].
- Chapter 9 provides a fast algorithm for a weaker sparse PCA adversarial model. It is based on the FOCS'20 paper [dKNS20].

Remark 1.16. We remark that the following published papers do not appear in this thesis:

- [dNS21] ICML'21.
- [dLN+21] NeurIPS'21.
- [Cd21] NeurIPS'21.
- [Cd22] COLT '22.
- [DdL+22] COLT'22.
- [dNNS23] SODA'23.

Chapter 2

Preliminaries

This chapter introduces common notation used throughout the rest of the thesis and contains some preliminary results required for the subsequent chapters. Other specific notions and notation will be directly introduced in the chapter themselves.

2.1 General definitions and notation

Often times, to simplify our discussion, we hide multiplicative factors logarithmic in the parameters at hand using the notation $\tilde{O}(\cdot)$. When specified, we may use the same notation to hide *poly-logarithmic* factors. Similarly, we hide absolute constant multiplicative factors using the standard notations $\lesssim$, $O(\cdot)$, $\Omega(\cdot)$ and $\Theta(\cdot)$. Often we use the letter C do denote universal constants independent of the parameters at play. We write $o_n(1), \omega_n(1)$ for functions tending to zero (resp. infinity) as n grows. We drop the subscript when the context is clear. We say that an event happens with high probability if this probability is at least $1 - o(1)$. Throughout this thesis, when we say "an algorithm runs in time $O(q)$" we mean that the number of basic arithmetic operations involved is $O(q)$. That is, we ignore bit complexity issues. For sets S, S' we denote by $S \times S'$ their Cartesian product.

Matrix and vector notations. We say that a unit vector $v \in \mathbb{R}^d$ is *flat* if its entries are in $\left\{\pm\frac{1}{\sqrt{t}}, 0\right\}$ for some t. We use $\mathbf{1}$ to denote the all 1's vector and J to denote the all 1's matrix, i.e. $J = \mathbf{1}\mathbf{1}^\top$. For a vector u, we use u_i to denote its i-th entry. We use Id_n to denote the n-by-n dimensional matrix and $\mathbf{0}$ to denote the zero matrix. For a matrix M, we use M_{ij} to denote the (i, j)-th entry of M, $M \succeq \mathbf{0}$ to denote that M is positive semidefinite, $\mathrm{Tr}(M)$ to denote the trace of M. Similarly, for matrices $A, B \in \mathbb{R}^{n\times n}$ we write $A \succeq B$ if $A - B$ is positive semidefinite. For a matrix $M \in \mathbb{R}^{n\times d}$ we denote with $\|M\|$ or $\|M\|_{\mathrm{op}}$ its spectral norm and with $\|M\|_{\mathrm{F}}$ its Frobenius norm. For a matrix $M \in \mathbb{R}^{n\times n}$, we denote by $\lambda_1(M) \geqslant \ldots \geqslant \lambda_n(M)$ its eigenvalues. Then $\rho(M) := \max_i |\lambda_i(M)|$ is the spectral radius of M. W When the context is clear we simply write $\lambda_1, \ldots, \lambda_n$. The spectral radius of a matrix satisfies the following

inequality.

Fact 2.1 (Gelfand's Formula)**.** *Let $M \in \mathbb{R}^{n\times n}$ and let $\|\cdot\|_*$ be a norm. Then for any positive integer z*

$$\rho(M) \leqslant \|M^z\|_*^{1/z}.$$

For two matrices X and Y of the same size, we use $\odot$ to denote the Hadamard product and we define their inner product by $\langle X, Y\rangle = \sum_{i,j=0}^{n} X_{ij}Y_{ij} = \mathrm{Tr}(X^\top Y)$. We introduce here several additional matrix norms.

Definition 2.2 (L_1 norm)**.** For a matrix $M \in \mathbb{R}^{n\times d}$ we denote with $\|M\|_1$ its L_1 norm:

$$\|M\|_1 = \sum_{i\in[n],j\in[d]} \left|M_{ij}\right|.$$

Definition 2.3 (Infinity norm)**.** For a matrix $M \in \mathbb{R}^{n\times d}$ we denote with $\|M\|_\infty$ its infinity norm:

$$\|M\|_\infty = \max_{i\in[n],j\in[d]} \left|M_{ij}\right|.$$

We also use the notation $\|M\|_{\max}$.

Definition 2.4 (Nuclear norm)**.** For a matrix $M \in \mathbb{R}^{n\times n}$ we denote with $\|M\|_{\mathrm{nuc}}$ its nuclear norm:

$$\|M\|_{\mathrm{nuc}} = \sum_{i\in[n]} |\lambda_i(M)|.$$

where $\lambda_i(M)$ is the i-th eigenvalue of matrix M.

Definition 2.5 (Schatten norm)**.** For a matrix $M \in \mathbb{R}^{n\times n}$ and $t \geqslant 1$ we denote by $\|M\|_t$ its t-Schatten norm:

$$\|M\|_t = \left(\mathrm{Tr}\left(M^t\right)\right)^{1/t}.$$

Notice that for $t = \infty$ we recover the spectral norm of M.

Graphs. We denote graphs with the notation $G(V, E)$. We use $V(G)$ to denote the set of vertices in G and similarly $E(G)$ to denote its set of edges. When the context is clear we simply write V and E. We denote by K_n be the complete graph on n vertices.

Definition 2.6 (Walk)**.** A walk W in a graph G is a sequence of vertices $(v_1, \ldots, v_{z+1})$. We say W is a *self-avoiding walk* if no vertex is visited twice.

For a graph G with $V(G) \subseteq [n]$ and vertices $i, j \in V(G)$, we let $\mathrm{SAW}_{ij}^s(G)$ be the set of self-avoiding walks between i and j in G of length s. For $i, j \in [n]$, $s \geqslant 1$, we use SAW_{ij}^s to denote the set of self-avoiding walks between i and j in K_n. We let $\mathrm{SAW}_i^s(G) := \bigcup_{j\in[n]} \mathrm{SAW}_{ij}^s(G)$.

2.2 Sum-of-squares

We introduce here of pseudo-distributions and sum-of-squares proofs (see the lecture notes [BS16] for more details and the appendix in [MSS16] for proofs of some the propositions appearing here).

Let $x = (x_1, x_2, \ldots, x_n)$ be a tuple of n indeterminates and let $\mathbb{R}[x]$ be the set of polynomials with real coefficients and indeterminates $x_1, \ldots, x_n$. We say that a polynomial $p \in \mathbb{R}[x]$ is a *sum-of-squares (sos)* if there are polynomials $q_1, \ldots, q_r$ such that $p = q_1^2 + \cdots + q_r^2$.

2.2.1 Pseudo-distributions

Pseudo-distributions are generalizations of probability distributions. We can represent a discrete (i.e., finitely supported) probability distribution over $\mathbb{R}^n$ by its probability mass function $D \colon \mathbb{R}^n \to \mathbb{R}$ such that $D \geqslant 0$ and $\sum_{x \in \mathrm{supp}(D)} D(x) = 1$. Similarly, we can describe a pseudo-distribution by its mass function. Here, we relax the constraint $D \geqslant 0$ and only require that D passes certain low-degree non-negativity tests.

Concretely, a *level-ℓ pseudo-distribution* is a finitely-supported function $D : \mathbb{R}^n \to \mathbb{R}$ such that $\sum_x D(x) = 1$ and $\sum_x D(x) f(x)^2 \geqslant 0$ for every polynomial f of degree at most $\ell/2$. (Here, the summations are over the support of D.) A straightforward polynomial-interpolation argument shows that every level-∞-pseudo distribution satisfies $D \geqslant 0$ and is thus an actual probability distribution. We define the *pseudo-expectation* of a function f on $\mathbb{R}^d$ with respect to a pseudo-distribution D, denoted $\tilde{\mathbb{E}}_{D(x)} f(x)$, as

$$\tilde{\mathbb{E}}_{D(x)} f(x) = \sum_x D(x) f(x) \ . \tag{2.2.1}$$

The degree-ℓ moment tensor of a pseudo-distribution D is the tensor $\mathbb{E}_{D(x)}(1, x_1, x_2, \ldots, x_n)^{\otimes \ell}$. In particular, the moment tensor has an entry corresponding to the pseudo-expectation of all monomials of degree at most ℓ in x. The set of all degree-ℓ moment tensors of probability distribution is a convex set. Similarly, the set of all degree-ℓ moment tensors of degree d pseudo-distributions is also convex. Key to the algorithmic utility of pseudo-distributions is the fact that while there can be no efficient separation oracle for the convex set of all degree-ℓ moment tensors of an actual probability distribution, there's a separation oracle running in time $n^{O(\ell)}$ for the convex set of the degree-ℓ moment tensors of all level-ℓ pseudodistributions.

Fact 2.7 ([Sho87, Par00, Nes00, Las01]). *For any $n, \ell \in \mathbb{N}$, the following set has a $n^{O(\ell)}$-time weak separation oracle (in the sense of [GLS81][1]):*

$$\left\{ \tilde{\mathbb{E}}_{D(x)}(1, x_1, x_2, \ldots, x_n)^{\otimes d} \mid \textit{degree-d pseudo-distribution D over } \mathbb{R}^n \right\} . \tag{2.2.2}$$

[1] Note that in general there may be bit complexity issues for running sum-of-squares algorithms, see [O'D17].

This fact, together with the equivalence of weak separation and optimization [GLS81] allows us to efficiently optimize over pseudo-distributions (approximately)—this algorithm is referred to as the sum-of-squares algorithm.

The *level-ℓ sum-of-squares algorithm* optimizes over the space of all level-ℓ pseudo-distributions that satisfy a given set of polynomial constraints—we formally define this next.

Definition 2.8 (Constrained pseudo-distributions)**.** Let D be a level-ℓ pseudo-distribution over $\mathbb{R}^n$. Let $\mathcal{A} = \{f_1 \geqslant 0, f_2 \geqslant 0, \ldots, f_m \geqslant 0\}$ be a system of m polynomial inequality constraints. We say that D *satisfies the system of constraints $\mathcal{A}$ at degree r*, denoted $D \sststile{r}{} \mathcal{A}$, if for every $S \subseteq [m]$ and every sum-of-squares polynomial h with $\deg h + \sum_{i \in S} \max\{\deg f_i, r\} \leqslant \ell$,

$$\tilde{\mathbb{E}}_D h \cdot \prod_{i \in S} f_i \geqslant 0 \,.$$

We write $D \sststile{}{} \mathcal{A}$ (without specifying the degree) if $D \sststile{0}{} \mathcal{A}$ holds. Furthermore, we say that $D \sststile{r}{} \mathcal{A}$ holds *approximately* if the above inequalities are satisfied up to an error of $2^{-n^\ell} \cdot \|h\| \cdot \prod_{i \in S} \|f_i\|$, where $\|\cdot\|$ denotes the Euclidean norm[2] of the cofficients of a polynomial in the monomial basis.

We remark that if D is an actual (discrete) probability distribution, then we have $D \sststile{}{} \mathcal{A}$ if and only if D is supported on solutions to the constraints $\mathcal{A}$.

We say that a system $\mathcal{A}$ of polynomial constraints is *explicitly bounded* if it contains a constraint of the form $\{\|x\|^2 \leqslant M\}$. The following fact is a consequence of Fact 2.7 and [GLS81],

Fact 2.9 (Efficient Optimization over Pseudo-distributions)**.** *There exists an $(n+m)^{O(\ell)}$-time algorithm that, given any explicitly bounded and satisfiable system[3] $\mathcal{A}$ of m polynomial constraints in n variables, outputs a level-ℓ pseudo-distribution that satisfies $\mathcal{A}$ approximately.*

2.2.2 Sum-of-squares proofs

Let $f_1, f_2, \ldots, f_r$ and g be multivariate polynomials in x. A *sum-of-squares proof* that the constraints $\{f_1 \geqslant 0, \ldots, f_m \geqslant 0\}$ imply the constraint $\{g \geqslant 0\}$ consists of sum-of-squares polynomials $(p_S)_{S \subseteq [m]}$ such that

$$g = \sum_{S \subseteq [m]} p_S \cdot \Pi_{i \in S} f_i \,. \tag{2.2.3}$$

[2]The choice of norm is not important here because the factor 2^{-n^ℓ} swamps the effects of choosing another norm.

[3]Here, we assume that the bitcomplexity of the constraints in $\mathcal{A}$ is $(n+m)^{O(1)}$.

We say that this proof has *degree* ℓ if for every set $S \subseteq [m]$, the polynomial $p_S \Pi_{i \in S} f_i$ has degree at most ℓ. If there is a degree ℓ SoS proof that $\{f_i \geqslant 0 \mid i \leqslant r\}$ implies $\{g \geqslant 0\}$, we write:

$$\{f_i \geqslant 0 \mid i \leqslant r\} \left|\frac{}{\ell}\right. \{g \geqslant 0\}\,. \tag{2.2.4}$$

Sum-of-squares proofs satisfy the following inference rules. For all polynomials $f, g \colon \mathbb{R}^n \to \mathbb{R}$ and for all functions $F \colon \mathbb{R}^n \to \mathbb{R}^m$, $G \colon \mathbb{R}^n \to \mathbb{R}^k$, $H \colon \mathbb{R}^p \to \mathbb{R}^n$ such that each of the coordinates of the outputs are polynomials of the inputs, we have:

$$\frac{\mathcal{A} \left|\frac{}{\ell}\right. \{f \geqslant 0, g \geqslant 0\}}{\mathcal{A} \left|\frac{}{\ell}\right. \{f + g \geqslant 0\}}, \frac{\mathcal{A} \left|\frac{}{\ell}\right. \{f \geqslant 0\}, \mathcal{A} \left|\frac{}{\ell'}\right. \{g \geqslant 0\}}{\mathcal{A} \left|\frac{}{\ell+\ell'}\right. \{f \cdot g \geqslant 0\}} \qquad \text{(addition and multiplication)}$$

$$\frac{\mathcal{A} \left|\frac{}{\ell}\right. \mathcal{B}, \mathcal{B} \left|\frac{}{\ell'}\right. C}{\mathcal{A} \left|\frac{}{\ell \cdot \ell'}\right. C} \qquad \text{(transitivity)}$$

$$\frac{\{F \geqslant 0\} \left|\frac{}{\ell}\right. \{G \geqslant 0\}}{\{F(H) \geqslant 0\} \left|\frac{}{\ell \cdot \deg(H)}\right. \{G(H) \geqslant 0\}}\,. \qquad \text{(substitution)}$$

Low-degree sum-of-squares proofs are sound and complete if we take low-level pseudo-distributions as models.

Concretely, sum-of-squares proofs allow us to deduce properties of pseudo-distributions that satisfy some constraints.

Fact 2.10 (Soundness)**.** *If $D \left|\frac{}{r}\right. \mathcal{A}$ for a level-ℓ pseudo-distribution D and there exists a sum-of-squares proof $\mathcal{A} \left|\frac{}{r'}\right. \mathcal{B}$, then $D \left|\frac{}{r \cdot r' + r'}\right. \mathcal{B}$.*

If the pseudo-distribution D satisfies $\mathcal{A}$ only approximately, soundness continues to hold if we require an upper bound on the bit-complexity of the sum-of-squares $\mathcal{A} \left|\frac{}{r'}\right. B$ (number of bits required to write down the proof).

In our applications, the bit complexity of all sum of squares proofs will be $n^{O(\ell)}$ (assuming that all numbers in the input have bit complexity $n^{O(1)}$). This bound suffices in order to argue about pseudo-distributions that satisfy polynomial constraints approximately.

The following fact shows that every property of low-level pseudo-distributions can be derived by low-degree sum-of-squares proofs.

Fact 2.11 (Completeness)**.** *Suppose $d \geqslant r' \geqslant r$ and $\mathcal{A}$ is a collection of polynomial constraints with degree at most r, and $\mathcal{A} \vdash \{\sum_{i=1}^n x_i^2 \leqslant B\}$ for some finite B.*

Let $\{g \geqslant 0\}$ be a polynomial constraint. If every degree-d pseudo-distribution that satisfies $D \left|\frac{}{r}\right. \mathcal{A}$ also satisfies $D \left|\frac{}{r'}\right. \{g \geqslant 0\}$, then for every $\varepsilon > 0$, there is a sum-of-squares proof $\mathcal{A} \left|\frac{}{d}\right. \{g \geqslant -\varepsilon\}$.

2.2.3 Sum-of-squares toolkit

We introduce here several statements involving sum of squares that will be used throughout the Thesis. We start with a Cauchy-Schwarz inequality for pseudo-distributions.

Fact 2.12 (Cauchy-Schwarz for pseudo-distributions [BBH+12])**.** *Let f, g be vector polynomials of degree at most d in indeterminate $x \in \mathbb{R}^n$. Then, for any degree $2d$ pseudo-distribution D,*

$$\tilde{\mathbb{E}}_D\big[\langle f, g\rangle\big] \leqslant \sqrt{\tilde{\mathbb{E}}_D[\|f\|^2]} \cdot \sqrt{\tilde{\mathbb{E}}_D[\|g\|^2]}\,.$$

We will also repeatedly use the following SoS version of Cauchy-Schwarz inequality and its generalization, Hölder's inequality:

Fact 2.13 (Sum-of-Squares Cauchy-Schwarz)**.** *Let $x, y \in \mathbb{R}^d$ be indeterminites. Then,*

$$\left|\frac{x,y}{4}\right\} \left\{\left(\sum_i x_i y_i\right)^2 \leqslant \left(\sum_i x_i^2\right)\left(\sum_i y_i^2\right)\right\}$$

We will use the following fact that shows how spectral certificates are captured within the SoS proof system.

Fact 2.14 (Spectral Certificates)**.** *For any $m \times m$ matrix A,*

$$\left|\frac{u}{2}\right\} \left\{\langle u, Au\rangle \leqslant \|A\|\|u\|_2^2\right\}.$$

The next fact establishes a certificate on the infinity-to-one norm of a matrix.

Fact 2.15 ([AN04])**.** *There exists an absolute constant K_G (Grothendieck's constant) and a polynomial time algorithm (based on sum-of-squares) that, for every $A \in \mathbb{R}^{n\times m}$, certifies an upper bound to $\|A\|_{\infty\to 1}$ tight up to a factor K_G. More specifically, for any degree-2 pseudo-distribution $D : \{\pm 1\}^n \times \{\pm 1\}^m \to \mathbb{R}$*

$$\tilde{\mathbb{E}}_{D(x,y)}\langle A, xy^{\mathsf{T}}\rangle \leqslant K_G \cdot \|A\|_{\infty\to 1}\,.$$

We will use the notions of pseudo-covariance and conditional pseudo-distributions.

Definition 2.16 (Pseudo-covariance)**.** Let α, β be multi-indices over $[n]$. Let $d/2 \geqslant |\alpha| + |\beta|$. Let D be a degree-d pseudo-distribution in indeterminates $x_1, \ldots, x_n$ . . Then we write

$$\mathrm{Cov}_D\big(x^\alpha, x^\beta\big) = \tilde{\mathbb{E}}_D[x^\alpha x^\beta] - \tilde{\mathbb{E}}_D[x^\alpha]\tilde{\mathbb{E}}_D[x^\beta]\,.$$

Similarly, we define $\mathbb{V}_D(x^\alpha) = \mathrm{Cov}_D(x^\alpha, x^\alpha)$.

Definition 2.17 (Conditional pseudo-distribution)**.** Let D be a degree-d pseudo-distribution in indeterminates $x_1, \ldots, x_n$. Let $t \geqslant 0$. Suppose D satisfies $\left\{x_i^2 = 1 \,, \forall i \in [n]\right\}$. Then for any $\alpha \in [n]^t$ such that $\tilde{\mathbb{E}}[\frac{1+x^\alpha}{2}] > 0$ we may define the conditional pseudo-distribution of degree $d - t$ as:

$$\tilde{\mathbb{E}}_D[p(x) \mid x^\alpha = 1] = \frac{\tilde{\mathbb{E}}_D[p(x)\frac{1+x^\alpha}{2}]}{\tilde{\mathbb{E}}_D[\frac{1+x^\alpha}{2}]} \,.$$

Similarly, if $\tilde{\mathbb{E}}[\frac{1+x^\alpha}{2}] < 1$, we may define the conditional pseudo-distribution of degree $d - t$ as:

$$\tilde{\mathbb{E}}_D[p(x) \mid x^\alpha = -1] = \frac{\tilde{\mathbb{E}}_D[p(x)\frac{1-x^\alpha}{2}]}{\tilde{\mathbb{E}}_D[\frac{1-x^\alpha}{2}]} \,.$$

It is straightforward to see that, after conditioning, the result is a valid pseudo-distribution of degree $d - t$. Notice also that, when D is an actual distribution, then we simply recover the corresponding conditional distribution.

Last, we introduce the following crucial observation about pseudo-distributions.

Lemma 2.18 (E.g. see [Sch22])**.** *Let D be a degree d pseudo-distribution over indeterminates $x_1, \ldots, x_n$ satisfying $\left\{x^2 = 1 \,, \forall i \in [n]\right\}$. Then, for any $S \subseteq [n]$ with $|S| \leqslant d$, there exists a distribution D' over $\{\pm 1\}^n$ such that, for all multi-indices α over S,*

$$\tilde{E}_D[x^\alpha] = \tilde{E}_{D'}[x^\alpha]$$

In other words, Lemma 2.18 states that, for any degree-d pseudo-distribution D over the hypercube and any subset S of d indeterminates, there exists an actual distribution D' over the hypercube matching its first d moments on S. Notably, combining this results with Definition 2.17, one gets that these low-degree moments of D and D' match even after conditioning.

Part I

The price of robustness

Chapter 3

Sparse PCA with adversarial perturbations

In this opening technical chapter we prove the results outlined in Section 1.1.1 and Section 1.1.2 in the context of *sparse principal component analysis (sparse PCA)*. The content of the chapter is mostly based on [dKNS20]. We consider the following model:

Problem 3.1 (Robust sparse PCA)**.** Given a matrix of the form

$$Y = W + \sqrt{\beta} u_0 v_0^{\mathsf{T}} + E, \text{ where} \tag{3.0.1}$$

- $v_0 \in \mathbb{R}^d$ is a unit k-sparse vector,
- $u_0 \sim N(0, \mathrm{Id}_n)$ is a standard Gaussian vector,
- $W \sim N(0,1)^{n \times d}$ is a Gaussian matrix and W, u_0, v_0 are distributionally independent,
- $E \in \mathbb{R}^{n \times d}$ is an arbitrary perturbation matrix satisfying

$$\|E\|_\infty \lesssim \sqrt{\beta/k} \cdot \min\{\sqrt{\beta}, 1\}\,. \tag{3.0.2}$$

Return a unit vector $\hat{v}$ having non-vanishing correlation with v_0.

Non-robust settings can be recovered enforcing the constraint $\|E\|_\infty = 0$. To get an intuition why bound Eq. (3.0.2) is canonical, observe that for $\beta \geqslant \Omega(1)$ adversarial perturbations of magnitude $\tilde{O}(\sqrt{\beta/k})$ could remove all information about v_0 (see Section 3.1). In other words, the definition of Problem 3.1 allows us to properly define robust algorithms (in the context of sparse PCA) rather than base our notion of robustness on a relative comparison between different algorithms.

We say that an algorithm is robust for a specific signal-to-noise ratio and sample complexity, if in such settings it achieves correlation with v_0 bounded away from zero, *as long as v_0 remains the principal sparse component*. Specifically, we say that an algorithm is

$(n, d, k, \beta, \delta, p)$–*robust* if, for parameters (n, d, k, β), with probability at least p it outputs a unit vector $\hat{v}$ such that $1 - \langle \hat{v}, v_0 \rangle^2 \leqslant \delta$.

To better keep track of the multiple results presented in this chapter, we provide three tables summarizing the main results.

Strong Signal Regime			
Algorithm	*Succeeds if*	*Running Time*	*Robust*
SVD with thresholding	$\beta \gtrsim \sqrt{\frac{d}{n}} + \frac{k \log d}{n}$	$O(nd \log n)$	No
Sum of squares, Theorem 3.2	$\beta \gtrsim \frac{k \cdot t}{n} \left(\frac{d}{k}\right)^{1/t}$ for $d \gtrsim (nt \log^{1+1/t} n)^t$	$d^{O(t)}$	Yes
Spectral algorithm, Theorem 9.2	$\beta \gtrsim \frac{k}{n} \left(\frac{d}{k}\right)^{1/3}$ for $d \gtrsim n^3 \log d \log n$	$O\left(nd \log n\right)$	*[1]

Table 3.1: Algorithmic landscape in the strong signal regime. The spectral algorithm is provably resilient to the adversary used to fool SVD with thresholding but we do not expect it to be robust to arbitrary adversaries. See Chapter 9.

Weak Signal Regime			
Algorithm	*Succeeds if*	*Running Time*	*Robust*
(Generalized) diagonal thresholding	$\beta \gtrsim \frac{k}{\sqrt{n \cdot t}} \sqrt{\log d}$ for $t \leqslant \frac{1}{\ln d} \min\{d, n\}$	$n^{O(1)} d^{O(t)}$	No
Covariance thresholding	$\beta \gtrsim \frac{k}{\sqrt{n}} \sqrt{\log \frac{d}{k^2}}$ for $k \lesssim \sqrt{d}$ and $k \lesssim \sqrt{n}$	$n^{O(1)} d^{O(1)}$	No
Low-degree polynomials, [dKNS20]	$\beta \gtrsim \frac{k}{\sqrt{n}} \sqrt{\log \frac{d}{k^2} + \frac{\log d}{\log n}}$ for $d^{1-o(1)} \lesssim k^2 \lesssim d$ and $n \gtrsim \log^5 d$	$n^{O(1)} d^{O(1)}$	No
Sum of squares, Theorem 3.6	$\beta \gtrsim \frac{k}{\sqrt{n \cdot t}} \sqrt{\log d}$ for $t \leqslant \frac{1}{\ln d} \min\{d, n\}$	$n^{O(1)} d^{O(t)}$	Yes
Basic SDP, Theorem 3.5	$\beta \gtrsim \frac{k}{\sqrt{n}} \sqrt{\log \frac{d}{k^2}}$ for $k \lesssim \sqrt{d}$ and $k \lesssim \sqrt{n}$	$n^{O(1)} d^{O(1)}$	Yes

Table 3.2: Algorithmic landscape in the weak signal regime.

[1] Robust to the distribution of Theorem 3.42.

Computational Lower Bounds for Polynomials			
Settings	*Work*	*Polynomials of degree D cannot distinguish if*	*Up to degree*
Fragile	[dKNS20]	$\beta \lesssim \left\{ \sqrt{\frac{d}{n}}, \frac{k \log\left(2+\frac{Dd}{k^2}\right)}{\sqrt{Dn}} \right\}$	$D \leqslant \frac{n}{\log^2 n}$
Robust	Theorem 3.3	$\beta \leqslant O\left(\frac{k}{\sqrt{n}}\left(\frac{d}{k}\right)^{1/t}\right)$ for $\beta n/k \leqslant n^{0.49}$ and $d \leqslant n^{0.99t-1}$	$D \leqslant n^{0.001}$

Table 3.3: Computational landscape for low-degree polynomials.

Robust algorithms in the strong signal regime. Without adversarial perturbations, in the strong-signal regime $n \gtrsim \frac{d}{\beta^2}$, the following spectral algorithm (*SVD with thresholding*) matches the optimal statistical guarantees of exhaustive search (introduced in Section 1.1.1): compute the top right singular vector of Y and restrict it to the k largest entries [BBAP05, KNV13]. Since adversarial perturbations of the order $\tilde{O}(1/\sqrt{n})$ can change the top eigenvalue of the covariance matrix, PCA arguments cannot be used to obtain resilient algorithms. Different kinds of certificates are needed.

We provide a Sum-of-Squares algorithm that recovers in time $d^{O(t)}$ the sparse vector whenever $n \gtrsim \frac{k}{\beta} \cdot t\left(\frac{d}{k}\right)^{1/t}$ and $d^{1/t} \geqslant \tilde{\Omega}(n)$. The key contribution is indeed an efficient algorithm to certify upper bounds on random quadratic forms. For subgaussian[2] low-rank quadratic forms, these upper bounds approach information-theoretically optimal bounds.

Concretely, for an n-by-d matrix W with i.i.d. Gaussian entries, with high probability the degree-t sum-of-squares algorithm (with running time $d^{O(t)}$) certifies an upper bound of $O(k \cdot (k/d)^{-1/t} \cdot t)$ on the quadratic form $Q(x) = \|Wx\|^2$ over all k-sparse unit vectors x if $d^{1/t} \geqslant \tilde{\Omega}(n)$. With these certificates, a robust algorithm for sparse PCA follows then as a specific corollary.

It is also important to notice how this result for sparse PCA is interesting regardless of its robustness properties. As t approaches $\log(d/k)$, the algorithm approaches the information theoretic optimal bound $O(\frac{k}{\beta} \cdot \log(d/k))$. For example, consider the case $n = 2^{\Theta\left(\sqrt{\log d}\right)}$. If also $\frac{d}{k} = 2^{\Theta\left(\sqrt{\log d}\right)}$, the Sum of Squares algorithm works in time $d^{O\left(\sqrt{\log d}\right)} = n^{O(\log^2 n)}$ with information theoretically optimal guarantees, while exhaustive search takes time exponential in n.

The specific algorithmic result is shown in the following theorem.

Theorem 3.2 (Robust algorithm in the strong signal regime)**.** *Given an n-by-d matrix Y of the form,*

$$Y = \sqrt{\beta} \cdot u_0 v_0^{\mathsf{T}} + W + E \,,$$

[2]Formally we require a stronger property, we need matrices to be *certifiably subgaussian*.

for $\beta > 0$, a unit k-sparse vector $v_0 \in \mathbb{R}^d$, a Gaussian matrix $W \sim N(0,1)^{n\times d}$, a vector $u_0 \in \mathbb{R}^n$ independent of W with $\|u_0\|^2 = \Theta(n)$, and a matrix $E \in \mathbb{R}^{n\times d}$ satisfying $\|E\|_\infty \lesssim \sqrt{\beta/k}\cdot\min\{\sqrt{\beta},1\}$.

For $t \in \mathbb{N}$ suppose that $d \gtrsim n^t \log^{t+1}(n) t^t$ and

$$\beta \gtrsim \frac{k}{n}\cdot t\cdot \left(\tfrac{d}{k}\right)^{1/t}.$$

Then, there exists an algorithm that computes in time $d^{O(t)}$ a unit vector $\hat{v} \in \mathbb{R}^d$ such that

$$1 - \langle \hat{v}, v_0\rangle^2 \leqslant 0.01$$

with probability at least 0.99.

In any case, the fundamental limitation of the above algorithm is the requirement $d^{1/t} \geqslant \tilde{\Omega}(n)$. This constraint makes it impossible to match the guarantees of SVD+ thresholding in most regimes. Theorem 1.1, which we restate here, shows evidence that this limitation is inherent by providing a lower bound against the restricted computational model of low-degree polynomials [BHK+19, HS17, HKP+17, Hop18] This hardness results suggests the aforementioned *fundamental separation between fragile and robust algorithms*: an inherent cost to pay in exchange for robustness.

Theorem 3.3 (Restatement of Theorem 1.1)**.** *Let t be a constant and let $d \leqslant n^{0.99t-1}$. Suppose that*

$$\beta \leqslant O\left(\frac{k}{n}\cdot t\cdot (d/k)^{1/t}\right).$$

and[3] $\beta n/k \leqslant n^{0.49}$. Then, there exists a distribution μ over $n\times d$ matrices Y of the form $Y = \sqrt{\beta}u_0 v_0^T + W + E$ where $\|E\|_\infty \leqslant \tilde{O}\left(1/\sqrt{n}\right)$, with the following properties:

- *μ is indistinguishable from the Gaussian distribution $N(0,1)^{d\times n}$ with respect to all multilinear polynomials of degree at most $n^{0.001}$ in the sense described in Section 3.4.1.2,*
- *the jointly-distributed random variables W, u_0, v_0 are independent,*
- *the marginal distribution of v_0 is supported on unit vectors with entries in $\left\{-1/\sqrt{k}, 0, 1/\sqrt{k}\right\}$,*
- *the marginal distribution of u_0 is uniform over $\{-1,1\}^n$,*
- *the marginal distribution of W is $N(0,1)^{n\times d}$.*

[3] This constraint is used to ensure that inequalities of the form $\beta \gtrsim \frac{k}{\sqrt{n\cdot D}}$ for any $D \leqslant n^{0.001}$ are never satisfied. Informally speaking, we restrict our statement to the settings where algorithms with guarantees similar to diagonal thresholding do not work.

Remark 3.4 (Sparse PCA and Gaussian mixtures). Sparse principal component analysis is intimately related to the problem of learning Gaussian mixtures. Indeed, for a vector v_0 with entries in $\{\pm 1/\sqrt{k}, 0\}$, sparse PCA can be rephrased as the problem of learning a non-uniform mixture M of three subgaussian distributions, one centered at zero, one centered at $\sqrt{\beta/k} \cdot u_0$ and the last at $-\sqrt{\beta/k} \cdot u_0$. As we will see, this is true even for the distribution μ used in Theorem 3.3 Thus, from this perspective the result also provides interesting insight on the complexity of this problem. The theorem suggests that to distinguish between M and a standard Gaussian $W \sim N(0,1)^{n \times d}$, an algorithm would either need $d \gtrsim n^t$ samples or should not be computable by polynomials of degree at most $n^{0.001}$.

Robust algorithms in the weak signal regime. Without adversarial corruptions, the algorithmic landscape in the weak-signal regime is more nuanced. One of the best known polynomial-time algorithms is *diagonal thresholding* [JL09]: restrict the empirical covariance matrix to the principal submatrix that contains the k largest diagonal entries and output the top eigenvector of this submatrix. This algorithm succeeds with high probability whenever $n \gtrsim \frac{k^2}{\beta^2} \log \frac{d}{k}$ — almost quadratically worse than exhaustive search.

Similar guarantees were shown to be achievable in polynomial time through a semidefinite relaxation [dGJL04, AW08] (which we refer to as the *basic SDP*, see Section 3.2 for a precise formulation). Remarkably, more refined algorithms, such as covariance thresholding [DM14] and the low-degree polynomial estimator in [dKNS20], where shown to achieve a logarithmic improvement for $d^{1-o(1)} \leqslant k^2 \leqslant o(d)$, recovering the sparse vector whenever $\beta \gtrsim \frac{k}{\sqrt{n}} \sqrt{\log \frac{d}{k^2}}$.

All aforementioned algorithms *except* the basic SDP can be fooled into outputting vectors uncorrelated with v_0 upon introducing entry-wise adaptive corruptions of magnitude $O(1/\sqrt{n})$ (see Appendix A.1).[4] As the basic SDP was not known to match the guaranteed of Covariance Thresholding and the low-degree polynomial algorithm in [dKNS20], this picture left a logarithmic gap between fragile and robust algorithms. Our first contribution to this regime consists of showing that the basic SDP can actually certify tighter bounds. Hence, it can match the guarantees of these apparently more refined, but fragile, algorithms.

Theorem 3.5 (Robust algorithm in the weak signal regime)**.** *Given an n-by-d matrix Y of the form,*

$$Y = \sqrt{\beta} \cdot u_0 v_0^{\mathsf{T}} + W + E \,,$$

for $\beta > 0$, a unit k-sparse vector $v_0 \in \mathbb{R}^d$, a Gaussian matrix $W \sim N(0,1)^{n \times d}$, a vector $u_0 \in \mathbb{R}^n$ independent of W with $\|u_0\|^2 = \Theta(n)$, and a matrix $E \in \mathbb{R}^{n \times d}$ satisfying $\|E\|_\infty \lesssim \sqrt{\beta/k} \cdot \min\{\sqrt{\beta}, 1\}$.

[4] We remark that a certain informal notion of robustness to entry-wise perturbations of the basic SDP program was already argued in [dGJL04]. Additionally, in [BR13] the authors observed that the algorithm is robust to small perturbations of the empirical covariance matrix. We allow here substantially more general perturbations.

Suppose that

$$\beta \gtrsim \min\left\{\frac{k}{\sqrt{n}}\sqrt{\log\left(2+\frac{d}{k^2}+\frac{d}{n}\right)}, \frac{d}{n}+\sqrt{\frac{d}{n}}\right\}.$$

Then, there exists an algorithm that uses the basic SDP program for sparse PCA, and computes in polynomial time a unit vector $\hat{v} \in \mathbb{R}^d$ such that

$$1-\langle \hat{v}, v_0\rangle^2 \leqslant 0.01$$

with probability at least 0.99.

High degree certificates in the weak signal regime. [DKWB23] and [Cd21] generalized the idea behind Diagonal Threhsolding, providing algorithms that interpolates between Diagonal Thresholding and brute force search. Concretely, given any natural number $t \leqslant n/\log d$, these algorithms recover the sparse vector in time $d^{O(t)}$ if $\beta \gtrsim \frac{k}{\sqrt{tn}}\sqrt{\log d}$. Hence offering a smooth trade-off between sample complexity and running time.[5] Naturally, these algorithms are also fragile. Our second contribution to the weak signal regime is a family of robust Sum-of-Squares algorithms which match the guarantees in [DKWB23, Cd21] in terms of error convergence, sample complexity and running time. The key insight is a novel degree-$O(t)$ Sum-of-Squares certificates of the bound $\|Wx\|^2 \leqslant n + k\sqrt{(n/t)\log d}$.

Theorem 3.6 (Robust algorithm via limited exhaustive search)**.** *Given an n-by-d matrix Y of the form,*

$$Y = \sqrt{\beta}\cdot u_0 v_0^{\mathsf{T}} + W + E\,,$$

for $\beta > 0$, a unit k-sparse vector $v_0 \in \mathbb{R}^d$, a Gaussian matrix $W \sim N(0,1)^{n\times d}$, a vector $u_0 \in \mathbb{R}^n$ independent of W with $\|u_0\|^2 = \Theta(n)$ and a matrix $E \in \mathbb{R}^{n\times d}$ satisfying $\|E\|_\infty \lesssim \sqrt{\beta/k}\cdot\min\{\sqrt{\beta}, 1\}$.

Suppose that for some positive integer $t \leqslant \frac{1}{\ln d}\min\{d, n\}$,

$$\beta \gtrsim \frac{k}{\sqrt{nt}}\sqrt{\log d}\,.$$

Then, there exists an algorithm that computes in time $n^{O(1)}d^{O(t)}$ a unit vector $\hat{v} \in \mathbb{R}^d$ such that

$$1-\langle \hat{v}, v_0\rangle^2 \leqslant 0.01$$

with probability 0.99.

Whenever $k^2 \leqslant d^{1-\Omega(1)}$, Theorem 3.6 provides better guarantees than Theorem 3.5 (with worse running time).

It is also interesting to compare this result with the bound of Theorem 3.2. For some t, we can determine the parameter regimes when one theorem provides better guarantees then the other for running time $d^{O(t)}$. Assume that $(t+1)^{t+1}n^{t+1}(\log n)^{t+2} \gtrsim d \gtrsim t^t n^t (\log n)^{t+1}$. Then there exist constants $0 < C < C'$ such that:

[5] This approach can be generalized to *sparse tensors*, see [Cd21].

- If $k^2 \leqslant d \cdot (Ct)^t$, we get $t \cdot \left(\frac{d}{k}\right)^{1/t} > \sqrt{\frac{n}{t}\log d}$, so in this case the guarantees in Theorem 3.6 are better.
- If $k^2 \geqslant d \cdot \left(n \log^2 n\right)^2 \cdot (C't)^t$, we get $t \cdot \left(\frac{d}{k}\right)^{1/t} < \sqrt{\frac{n}{t}\log d}$, so in this case the guarantees in Theorem 3.2 are better.

Informally speaking, these conditions show that the guarantees in Theorem 3.2 are better when the vector is only mildly sparse: $k^2 \gg d$, and the number of samples n is very small.

Outline and notation

We conclude the section with an outline of the structure of the chapter and some notation.

In Section 3.1 we give an overview of the techniques and the ideas required to obtain the results. We use the prelminary notions introduced in Section 2.2. Section 3.2 and Section 3.3 contains the results for the basic SDP and the Sum-of-Squares algorithms. In Section 3.4 we show our lower bounds on polynomials. Additionally, Appendix A.1 contains formal proofs that thresholding algorithms are not robust.

Notation. We use the notation in Chapter 2. For a matrix $M \in \mathbb{R}^{n\times d}$, we will denote its entry ij with M_{ij}. Depending on the context we may refer to the i-th row or the i-th column of M with M_i or m_i, we will specify it each time to avoid ambiguity. We call $\|M\|_1 = \sum_{i,j\in[d]} \left|M_{ij}\right|$ the "absolute norm" of M. For a Gaussian matrix $W \sim N(0,1)^{n\times d}$, we denote with $w_1, \dots, w_d$ its columns. We hide absolute constant multiplicative factors using the standard notations $\lesssim$, $O(\cdot)$, $\Omega(\cdot)$ and $\Theta(\cdot)$, we hide multiplicative factors logarithmic in d using the notation $\tilde{O}(\cdot)$. For a set $S \subseteq [d]\times[d]$, and a matrix $M \in \mathbb{R}^{d\times d}$, we denote by $M[S]$ the matrix with entries $M[S]_{ij} = M_{ij}$ if $(i,j) \in S$, and $M[S]_{ij} = 0$ otherwise. For a matrix $M \in \mathbb{R}^{d\times d}$ and $\tau \in \mathbb{R}$, we define $\eta_\tau(M) \in \mathbb{R}^{d\times d}$ to be the matrix with entries

$$\eta_\tau(M)_{ij} = \begin{cases} M_{ij} & \text{if } \left|M_{ij}\right| \geqslant \tau \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, we define $\zeta_\tau(M) \in \mathbb{R}^{d\times d}$ to be the matrix with entries

$$\zeta_\tau(M)_{ij} = \begin{cases} M_{ij} - \operatorname{sign}(M_{ij}) \cdot \tau & \text{if } \left|M_{ij}\right| \geqslant \tau \\ 0 & \text{otherwise.} \end{cases}$$

Additional notation will be introduced when needed.

Remark 3.7 (Strong and weak signal regimes in robust settings). The attentive reader may have noticed how the notions of strong and weak signal regime should differ in the robust settings. Indeed there is no easy algorithm that looks at the spectrum of Y and begins to

work as β approaches $\sqrt{\frac{d}{n}}$. In this sense, in the presence of an adversary the bound $\beta \lesssim \sqrt{\frac{d}{n}}$ looses significance. However we will continue using these terms to orientate ourselves and implicitly describe which are the desirable guarantees an algorithm should possess in a given regime. For this reason, when talking about weak-signal regime, our discussion will implicitly revolve around settings in which $\beta \gtrsim \frac{k}{\sqrt{n}}\sqrt{\log \frac{d}{k^2}}$.

3.1 Techniques

Here we outline the main ideas used to design the algorithm and lower bounds behind the results in the chapter. We assume the reader to be familiar with the knowledge in Section 2.2.

3.1.1 Robustness from sparse eigenvalue certificates

How robust should an algorithm be? In light of our discussion, we would like efficient algorithms to be as resilient[6] as exhaustive search. In order for such brute-force algorithm to recover the sparse vector v_0, there must be no other sparse vector x far from v_0 such that $\|Yx\| \approx \|Yv\|$. This also means that the adversary should not be able to plant a k-sparse vector z far from v_0 such that $\|Yx\| \gtrsim \|Yv_0\|$. To see what bound to enforce on the adversarial matrix, first observe that if E were the zero matrix then

$$\|Yv_0\| = \left\|Wv_0 + \sqrt{\beta}u_0\right\| \gtrsim \sqrt{n+\beta n}.$$

Now consider the following adversarial matrix, let x be a k-sparse unit vector with entries in $\left\{0, \pm 1/\sqrt{k}\right\}$ and such that the intersection between $\mathrm{supp}\{x\}$ and $\mathrm{supp}\{v_0\}$ is the empty set. With high probability $\|Wx\| \approx \sqrt{n}$. So let $z = \frac{1}{\|Wx\|}Wx$ and define E as the matrix with entries $E_{ij} = b \cdot z_i \cdot \mathrm{sign}(x_j)$, where $b > 0$ is some parameter that we will choose later. Then

$$\|Yx\| = \|(W+E)x\| = \left\|\left(\|Wx\| + b\sqrt{k}\right)z\right\| \approx \sqrt{n} + b\sqrt{k}.$$

Consequently, $\|Yx\| \geqslant \|Yv_0\|$ whenever $\sqrt{n} + b\sqrt{k} \geqslant \sqrt{n+\beta n}$. The inequality is true for $b \gtrsim \sqrt{\frac{\beta n}{k}} \cdot \min\left\{\sqrt{\beta}, 1\right\}$. In other words, the perturbation matrix must satisfy the bound:

$$\|E\|_\infty \leqslant \tilde{\Omega}\left(\sqrt{\frac{\beta}{k}} \cdot \min\left\{\sqrt{\beta}, 1\right\}\right). \tag{Bound-1}$$

For a set of parameters d, n, k, β, we call an algorithm *perturbation resilient* if it can successfully recover the sparse vector for any adversarial perturbation satisfying bound Bound-1.

[6] We interchangeably use the terms robust and resilient.

Remark 3.8. In the proofs presented here, we will measure adversarial corruptions with the norm $\|E\|_{1\to 2}$, which denotes the largest norm of a column of E. Clearly this choice allows for a larger class of adversaries. There are a two main reasons behind our choice. The first one being that the adversarial matrices we consider are more naturally described using such norm. Furthermore, this norm has a direct correspondence with the infinity norm of the adversarial perturbation in the covariance matrix. Indeed, $\left\|E^\top E\right\|_\infty = \|E\|_{1\to 2}^2$. This also allows one to draw a better comparison between the Wishart and the Wigner model. We remark that the reasoning above can be used as well to show the bound:

$$\|E\|_{1\to 2} \lesssim \min\left\{\sqrt{\frac{\beta n}{k}} \cdot \min\left\{\sqrt{\beta}, 1\right\}\right\}. \tag{Bound-2}$$

Algorithms that certify sparse eigenvalues. For simplicity of the discussion we illustrate the idea of sparse eigenvaluex certificates for the Wigner model: $Y = \gamma v_0 v_0^\top + W + E$, where $\gamma > 0$, $v_0 \in \mathbb{R}^d$ is a k-sparse unit vector, $W \sim N(0,1)^{d\times d}$ and E is some matrix with small entries. Denote the set of k-sparse unit vectors by S_k. The starting idea is to turn the following intuition into an identifiability proof and then a Sum of Squares program: if $\hat{v}$ is a k-sparse unit vector which maximizes $v^\top Y v$ over S_k and γ is large enough, then with high probability $\langle \hat{v}, v_0\rangle^2 \geqslant 0.99$.

Concretely, observe that

$$\begin{aligned} &\text{on one side} && v_0^\top Y v_0 = \gamma + v_0^\top W v_0 + v_0^\top E v_0, \\ &\text{on the other} && \hat{v}^\top Y \hat{v} = \gamma \langle \hat{v}, v_0\rangle^2 + \hat{v}^\top W \hat{v} + \hat{v}^\top E \hat{v}. \end{aligned}$$

Combining the two and rearranging we obtain the inequality

$$\langle \hat{v}, v_0\rangle^2 \geqslant 1 - \frac{1}{\gamma} O\left(\max_{v\in S_k} v^\top W v + \max_{v\in S_k} v^\top E v\right).$$

Now, this is where certified upper bounds come in to the picture. There is an easy certificate (capture by SoS and the basic SDP) of the fact that for any matrix M, $\max_{v\in S_k} v^\top M v \leqslant \|M\|_\infty k$ Using such bound we get

$$\langle \hat{v}, v_0\rangle^2 \geqslant 1 - \frac{1}{\gamma} O\left(\max_{v\in S_k} v^\top W v + k\|E\|_\infty\right). \tag{3.1.1}$$

Eq. (3.1.1) already shows how an algorithm that can certify sparse eigenvalues is perturbation resilient (in the sense of the previous paragraph). Indeed for $\|E\|_\infty = \varepsilon \cdot \gamma/k$, the inequality becomes

$$\langle \hat{v}, v_0\rangle^2 \geqslant 1 - O(\varepsilon) - \frac{1}{\gamma} O\left(\max_{v\in S_k} v^\top W v\right). \tag{3.1.2}$$

At this point, the guarantees of the algorithm depend only on the specific certified upper bound on $\max_{v\in S_k} v^\top W v$ it can obtain.

For the Wishart Model $Y = \sqrt{\beta} u_0 v_0^\mathsf{T} + W + E$, the reasoning is essentially the same. However we need to work with $Y^\mathsf{T} Y - n\mathrm{Id}$ and carefully bound the cross terms. Similar to the Wigner model, the guarantees of the algorithm depend only on the certified upper bound on $\max_{v \in S_k} v^\mathsf{T}\left(W^\mathsf{T} W - n\mathrm{Id}\right) v$ it can obtain. For the rest of our preliminary discussion we go back to the Wishart model.

Refined certificates via basic SDP. For a matrix $M \in \mathbb{R}^{d \times d}$, the basic SDP program[7]

$$\operatorname{argmax}\left\{\langle Y^\mathsf{T} Y, X\rangle \;\middle|\; X \succeq 0, \operatorname{Tr} X = 1, \|X\|_1 \leqslant k\right\} \tag{3.1.3}$$

can certify two types of upper bound:

$$\langle M, X\rangle \leqslant \|M\|_\infty \cdot k \tag{3.1.4}$$

$$\langle M, X\rangle \leqslant \|M\| . \tag{3.1.5}$$

The first follows using $\|X\|_1 \leqslant k$ and the second applying $X \succeq 0, \operatorname{Tr} X = 1$. These are enough to capture standard principal component analysis as well as diagonal and covariance thresholding.

Specifically, Eq. (3.1.5) can be used to certify the upper bound $\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle \leqslant O\left(d + \sqrt{dn}\right)$ – obtaining the guarantees of PCA – and Eq. (3.1.4) the bound $\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle \leqslant O\left(k \cdot \sqrt{n \log d}\right)$, as in diagonal thresholding[8]. Now these results were already known, but surprisingly a combination of the two bounds can also be used to show $\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle \leqslant k \cdot \sqrt{n \log(d/k^2)}$. Thus allowing us to match the guarantees of covariance thresholding.

Concretely, using the notation from the introduction,

$$\begin{aligned}\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle = \langle \eta_\tau\left(W^\mathsf{T} W - n\mathrm{Id}\right), X\rangle + \\ \langle W^\mathsf{T} W - \eta_t\left(W^\mathsf{T} W\right), X\rangle .\end{aligned}$$

Here $W^\mathsf{T} W - \eta_t\left(W^\mathsf{T} W\right)$ is a matrix with entries bounded (in absolute value) by τ for which we can plug in Eq. (3.1.4) and get

$$\langle W^\mathsf{T} W - \eta_t\left(W^\mathsf{T} W\right), X\rangle \leqslant \tau \cdot k$$

The same argument cannot be used for $\eta_\tau\left(W^\mathsf{T} W\right)$, but note that this matrix is suspiciously close (up to an addition of $n \cdot \mathrm{Id}$) to the thresholded covariance matrix obtained in covariance thresholding. Hence, taking $\tau = \sqrt{n \log(d/k^2)}$ and using Eq. (3.1.5), we get

$$\langle \eta_\tau\left(W^\mathsf{T} W - n\mathrm{Id}\right), X\rangle \leqslant O\left(k\sqrt{n \log \frac{d}{k^2}}\right),$$

where we get the spectral bound (almost) for free by the analysis in [DM14].

[7] Recall $\|X\|_1 = \sum_{i,j \in [d]} \left|X_{ij}\right|$ is the "absolute norm".

[8] A more careful analysis can get $k \cdot \sqrt{n \log(d/k)}$, but we ignore it here.

3.1.1.1 Refined certificates via higher-level Sum-of-Squares

Refined certificates via Certifiable Subgaussianity. The Sum-of-Squares algorithm can certify more refined bounds on sparse eigenvalues of $W \sim N(0,1)^{n\times d}$. In particular we can exploit Gaussian moments bound $\mathbb{E}\langle W_i, u\rangle^{2t} \leqslant t^t \cdot \|u\|^{2t}$ for all $t \in \mathbb{N}$, $u \in \mathbb{R}^d$.

Concretely let's see how to use such property to obtain an identifiability proof of a bound on the k-sparse norm of W. To this end let v be a k-sparse vector and let $s \in \{0,1\}^d$ be the indicator vector of its support (here we drop the subscript v_0 to ease the notation). Using Cauchy-Schwarz,

$$\|Wv\|^4 = \left(\sum_{i\leqslant d} v_i\langle W_i, Wv\rangle\right)^2 \leqslant \left(\sum_{i\leqslant d} v_i^2\right)\left(\sum_{i\leqslant d} s_i^2\langle W_i, Wv\rangle^2\right) \leqslant \left(\sum_{i\leqslant d} s_i^2\langle W_i, Wv\rangle^2\right).$$

Then applying Holder's inequality with $1/p + 1/t = 1$, and using the fact that s is binary with norm k,

$$\left(\sum_{i\leqslant d} s_i^2\langle W_i, Wv\rangle^2\right) \leqslant \left(\sum_{i\leqslant d} s_i^{2p}\right)^{1/p}\left(\sum_{i\leqslant d}\langle W_i, Wv\rangle^{2t}\right)^{1/t} \leqslant \|Wv\|^2 \cdot k^{1-1/t}\left(\sum_{i\leqslant d}\langle W_i, \frac{1}{\|Wv\|}Wv\rangle^{2t}\right)^{1/t}.$$

This gets us to,

$$\|Wv\|^2 \leqslant k^{1-1/t} \cdot \left(\sum_{i\leqslant d}\langle W_i, \frac{1}{\|Wv\|}Wv\rangle^{2t}\right)^{1/t}. \tag{3.1.6}$$

Now, whenever $d \gtrsim n^t t^t \log^t n$, the t-moment of the column vectors $W_1 \ldots, W_d$ converges with high probability. That is, for any unit vector u,

$$\frac{1}{d}\sum_{i\leqslant d}\langle W_i, u\rangle^{2t} \leqslant O(t^t). \tag{3.1.7}$$

Thus, combining Eq. (3.1.6) and Eq. (3.1.7) we can conclude

$$\|Wv\|^2 \lesssim k^{1-1/t} \cdot d^{1/t} \cdot t.$$

The catch is that all the steps taken can be written as polynomial inequalities of degree at most $O(t)$. So we can certify the same bound through the Sum-of-Squares proof system.

Certificates via limited brute force. Whenever the sparse vector v_0 is almost flat, that is when for all $i \in \mathrm{supp}\{v_0\}$ we have $|v_{0i}| \in \left[\frac{1}{C\sqrt{k}}, \frac{C}{\sqrt{k}}\right]$, the guarantees of diagonal thresholding can be improved at the cost of increasing its running time (see [DKWB23, Cd21]).

Diagonal thresholding can be viewed as selecting the k vectors of the standard basis $e_1, \ldots, e_d$ maximizing $\|Ye_i\|^2$, and then returning a top eigenvector of the covariance

matrix projected onto the span of such vectors. Indeed this formulation has an intuitive generalization, namely instead of looking at 1-sparse vectors, the algorithm could look into t-sparse vectors u with entries in $\{\pm 1/\sqrt{t}, 0\}$, pick the top $\binom{k}{t}$ and use them to recover v_0.

This idea can be translated into a certified upper bound for the sparse eigenvalues of $W \sim N(0,1)^{n\times d}$. Although we will be able to recover general sparse vectors, for the sake of this discussion we assume v_0 is flat.[9] Let's denote the set of t-sparse flat vectors by $\mathcal{N}_t$. Let $v_0 \in \mathbb{R}^d$ be a k-sparse vector and denote with D the uniform distribution over the vectors in $\mathcal{N}_t$ such that $\langle u, v_0\rangle = \sqrt{t/k}$. That is, the set of vectors u such that $\mathrm{supp}\{u\} \subseteq \mathrm{supp}\{v_0\}$ and with sign pattern matching the sign pattern of v restricted to $\mathrm{supp}\{u\}$.

Note that for any matrix $M \in \mathbb{R}^{d\times d}$,

$$v_0{}^{\mathsf{T}} M v_0 = \frac{k}{t} \mathop{\mathbb{E}}_{u\sim D} \mathop{\mathbb{E}}_{u'\sim D} u^{\mathsf{T}} M u' .$$

This equality *per se* is not interesting, but for a Gaussian matrix $W \sim N(0,1)^{n\times d}$, with high probability,

$$\max_{u,u'\in\mathcal{N}_t} \left|u^{\mathsf{T}}\left(W^{\mathsf{T}}W - n\mathrm{Id}\right)u'\right| \leqslant O\left(\sqrt{nt\log d}\right).$$

Thus, combining the two we get

$$\begin{aligned} v_0{}^{\mathsf{T}}\left(W^{\mathsf{T}}W - n\mathrm{Id}\right)v_0 =& \frac{k}{t}\mathop{\mathbb{E}}_{D} u^{\mathsf{T}}\left(W^{\mathsf{T}}W - n\mathrm{Id}\right)u' \\ \leqslant& \frac{k}{t}\max_{u,u'\in\mathcal{N}_t}\left|u^{\mathsf{T}}\left(W^{\mathsf{T}}W - n\mathrm{Id}\right)u'\right| \\ \leqslant& \frac{k}{\sqrt{t}}\sqrt{n\log d}\,, \end{aligned}$$

which allows us to conclude that $\|Wv_0\|^2 \leqslant n + \frac{k}{\sqrt{t}}\sqrt{n\log d}$. This certificates can be proved using Sum-of-Squares, hence allowing us to improve over the basic SDP by a factor t in the settings $k^2 \leqslant d^{1-\Omega(1)}$.

3.1.2 Concrete lower bounds for robust algorithms

Sparse principal component analysis is what we often call a *planted problem*. These are problems that ask to recover some signal hidden by random or adversarial noise. The easiest way one could formulate a planted problem is its *distinguishing* version: where given two distributions, a *null* distribution without structure and a *planted* distribution containing the hidden signal, the objective is to determine with high probability whether a given instance was sampled from one distribution or the other.

A common strategy to provide evidence for information-computation gap in a certain planted problem is to prove that powerful classes of efficient algorithms are unable to

[9] So the Sum-of-Squares algorithm works in more general settings than the algorithm from [DKWB23].

solve it in the (conjecturally) hard regime. Indeed our goal here will be that of constructing two distributions under which low-degree polynomials take roughly the same values and hence cannot distinguish (in the sense of Section 3.4.1) from which distribution the instance Y was sampled. Since low-degree polynomials cannot tell if Y has indeed the form $W + \sqrt{\beta} u_0 v_0^\top + E$ (and therefore cannot solve the problem), this would mean they cannot be used to improve over the guarantees of Theorem 3.2.

Our null distribution ν will be the standard Gaussian $N(0,1)^{n \times d}$. However, the main question is how to design the planted distribution μ. Recall Y takes the form $W + \sqrt{\beta} u_0 v_0^\top + E$. If we set $E = 0$, then our planted distribution corresponds to the single spike covariance model. We could get a lower bound for such problem (see [dKNS20]) but this would not help us in showing that the guarantees of Theorem 3.2 are tight. On the other hand, if for example we choose E with the goal of planting a large eigenvalue, then the problem of distinguishing between ν and μ may become even easier than without adversarial perturbations.

This suggests that we should choose E very carefully, in particular we should design E so that $Y = W + \sqrt{\beta} u_0 v_0^\top + E$ appears – to the eyes of a low-degree polynomial estimator – as a Gaussian distribution. Our approach will be that of constructing E so that the first few moments of μ will be Gaussian. This will lead us to Theorem 3.3 through two basic observations: first, given two distributions with same first $2t$ moments, computing those first $2t$ moments won't help distinguishing between the two distributions. Second, for a Gaussian distribution $N(0, \mathrm{Id}_n)$, at least n^t samples are required in order for the $2t$-th moment of the empirical distribution to converge to $\mathbb{E}\left[w^{\otimes 2t}\right]$.

Concretely, we consider the following model: we choose iid gaussian vectors $z_1, \dots, z_{n-1} \sim N(0,1)^d$, and a random vector $z_0 \in \mathbb{R}^d$ with iid symmetric (about zero) coordinates that satisfies the following properties:

1. z_0 has approximately k large coordinates (larger than $\lambda \approx \sqrt{\beta n / k}$ by absolute value).
2. For any coordinate of z_0 its first $2t - 2$ moments coincide with moments of $N(0,1)$, and its higher r-moments (for even r) are close to $\frac{k}{d} \lambda^r$.

Then we obtain the matrix $Y \in \mathbb{R}^{n \times d}$ applying a random rotation $R \in \mathbb{R}^{n \times n}$ to the $n \times d$ matrix with rows $z_0^\top, z_1^\top, \dots, z_{n-1}^\top$. It is not difficult to see that indeed such Y can be written as $Y = W + \sqrt{\beta} u_0 v_0^\top + E$, as in the model of Problem 3.1.

Now, assume for simplicity that t is constant and denote the distribution of Y described above by μ and the standard Gaussian distribution $N(0,1)^{n \times d}$ by ν. An immediate consequence of our construction is that for any polynomial p of degree at most $2t - 2$, $\mathbb{E}_{Y \sim \mu}\left[p(Y)\right] = \mathbb{E}_{Y \sim \nu}\left[p(Y)\right]$. Furthermore, in order to reliably tell the difference between $\mathbb{E}_{\mu}\left[p'(Y)\right]$ and $\mathbb{E}_{\nu}\left[p'(W)\right]$ for a polynomial of even degree $r \geqslant 2t$ (say up to $r = n^{0.001}$), we will need a precise estimate of such r-th moments and *hence* at least $n^{r/2 \geqslant t}$ samples. This effect is then shown by proving that for multilinear polynomials $p(Y)$ of degree $D \leqslant n^{0.001}$, if $d \leqslant n^{0.99t - 1}$ and $\beta n / k \leqslant n^{0.49}$, then the low-degree analogue of χ^2-divergence

$\max_{p(Y) \text{ of degree } \leqslant D} \frac{\left(\mathbb{E}_\nu p(Y) - \mathbb{E}_\mu p(Y)\right)^2}{\mathbb{V}_\nu P(Y)}$ is close to zero. Note that for technical reasons our analysis is restricted to the multilinear polynomials. As shown in [BHK+19, HS17, Hop18], this restricted model of computation captures the best known algorithms for many planted problems.

3.2 Robustness of the basic SDP and certified upper bounds

In this section we show the guarantees of the basic SDP algorithm [dGJL04, AW08], thus proving Theorem 3.5.

We will first prove that for any matrix $M \in \mathbb{R}^{d \times d}$ the basic SDP can certify an upper bound $\|Mx\|^2 \leqslant k \cdot \|M\|_\infty^2$ on k-sparse quadratic forms over M. Furthermore we will show that for random Gaussian matrices $W \sim N(0,1)^{n \times d}$ this bound can be significantly improved in various ways, depending on the regime. Most notably, we will show that the basic SDP can certify a bound $\|Wx\|^2 \leqslant n + k\sqrt{n \log(d/\min\{k^2, n\})}$, thus matching the guarantees of Covariance Thresholding. As a corollary, we also get that for $\beta < 1$ the algorithm achieves the best known guarantees among polynomial time algorithms in *both* the fragile and the robust settings.

Formally the Sparse PCA problem can be defined as follows.

Problem 3.9. Given an instance Y of 3.1 let $\hat{\Sigma} = Y^\intercal Y$. Then the Sparse PCA problem is defined by

$$\operatorname{argmax}\left\{v^\intercal \hat{\Sigma} v \;\middle|\; \|v\|^2 = 1, \|v\|_0 \leqslant k\right\}$$

where $\|v\|_0$ is the number of non-zero entries in v.

Solving Problem 3.9 is NP-hard in general [MWA06, Nat95, KNV13], however the following concrete SDP relaxation [dGJL04] can be efficiently solved

$$\operatorname{argmax}\left\{\langle \hat{\Sigma}, X\rangle \;\middle|\; X \succeq 0, \operatorname{Tr} X = 1, \|X\|_1 \leqslant k\right\} \tag{SDP-1}$$

where $\|X\|_1 = \sum_{i,j \in [d]} \left|X_{ij}\right|$ is the "absolute norm". We will show how to recover v_0 using such program.

We start by restating some of the notation from the introduction. For a set $S \subseteq [d] \times [d]$, and a matrix $M \in \mathbb{R}^{d \times d}$, we denote by $M[S]$ the matrix with entries $M[S]_{ij} = M_{ij}$ if $(i,j) \in S$, and $M[S]_{ij} = 0$ otherwise. For a matrix $M \in \mathbb{R}^{d \times d}$ and $\tau \in \mathbb{R}$, we define $\eta_\tau(M) \in \mathbb{R}^{d \times d}$ to be the matrix with entries

$$\eta_\tau(M) = \begin{cases} M_{ij} & \text{if } \left|M_{ij}\right| \geqslant \tau \\ 0 & \text{otherwise.} \end{cases}$$

Furthermore, we define $\zeta_\tau(M) \in \mathbb{R}^{d\times d}$ to be the matrix with entries

$$\zeta_\tau(M) = \begin{cases} M_{ij} - \operatorname{sign}(M_{ij}) \cdot \tau & \text{if } |M_{ij}| \geqslant \tau \\ 0 & \text{otherwise.} \end{cases}$$

3.2.1 Basic certificates for sparse quadratic forms

We show here what certificates over sparse quadratic forms SDP-1 can provide. These certificates are already enough to match the best known guarantees in the weak signal regime. The first observation is that it is straightforward to bound the product between X and matrices with small infinity norm. By construction of X this is indeed a certificate of an upper bound over k-sparse quadratic forms.

Lemma 3.10. *For $k \in \mathbb{N}$, let $X \in \mathbb{R}^{d\times d}$ such that $\|X\|_1 \leqslant k$. Then for any matrix $M \in \mathbb{R}^{d\times d}$*

$$|\langle M, X\rangle| \leqslant k \cdot \|M\|_\infty.$$

Proof. The Lemma follows immediately by choice of X,

$$|\langle X, M\rangle| = \left|\sum_{i,j\in[d]} M_{ij}X_{ij}\right| \leqslant \sum_{i,j\in[d]} |M_{ij}X_{ij}| \leqslant \|M\|_\infty \sum_{i,j\in[d]} |X_{ij}| \leqslant k \cdot \|M\|_\infty.$$

□

Now we improve this bound for random matrices. In particular we look into the Hilbert-Schmidt inner product $\langle \eta_\tau(W^\top W - n\mathrm{Id}), X\rangle$.

Lemma 3.11. *Let $X \in \mathbb{R}^{d\times d}$ be a positive semidefinite matrix such that $\operatorname{Tr} X = 1$ and $\|X\|_1 \leqslant k$. Let $W \sim N(0,1)^{n\times d}$, then with probability $1 - o(1)$*

$$\left|\langle W^\top W - n\mathrm{Id}, X\rangle\right| \leqslant O\left(\min\left\{k\sqrt{n\log\left(1 + \frac{d}{k^2} + \frac{d}{n + k\sqrt{n}}\right)},\ d + \sqrt{dn}\right\}\right).$$

Proof. By Theorem A.20, $\|W^\top W - n\mathrm{Id}\| \leqslant O\left(d + \sqrt{dn}\right)$ with probability $1 - d^{-10}$, so by Lemma A.31

$$\left|\langle W^\top W - n\mathrm{Id}, X\rangle\right| \leqslant O\left(d + \sqrt{dn}\right).$$

Let $D \subseteq [d]\times[d]$ be the set of diagonal entries of $(W^\top W - n\mathrm{Id})$ and $\bar{D}$ its complement. For any $\tau \geqslant 0$ we can rewrite the matrix $(W^\top W - n\mathrm{Id})$ as

$$W^\top W - n\mathrm{Id} = (W^\top W - n\mathrm{Id})[D] + \eta_\tau(W^\top W - n\mathrm{Id})\left[\bar{D}\right] + (W^\top W - n\mathrm{Id} - \eta_\tau(W^\top W - n\mathrm{Id}))\left[\bar{D}\right].$$

Now, by Fact A.18 with probability $1 - o(1)$, $\left\|(W^\top W - n\mathrm{Id})[D]\right\| \leqslant 10\sqrt{n\log d}$. Furthermore,

$$\eta_\tau(W^\top W - n\mathrm{Id})[\bar{D}] = \zeta_\tau(W^\top W - n\mathrm{Id})[\bar{D}] + M\,,$$

where $M \in \mathbb{R}^{d\times d}$ is a matrix with $\|M\|_\infty \leqslant \tau$ and by Theorem A.26 there is a constant $C \geqslant 1$ such that $\left\|\zeta_\tau(W^\mathsf{T}W - n\mathrm{Id})[\bar{D}]\right\| \leqslant C\left(d + \sqrt{dn}\right)\exp\left[-\frac{\tau^2}{Cn}\right]$ with probability $1 - o(1)$. Let $\tau = 10C \cdot \sqrt{n\log\left(1 + \frac{d}{k^2} + \frac{d}{n+k\sqrt{n}}\right)}$. If $d \leqslant n$,

$$\left\|\zeta_\tau(W^\mathsf{T}W - n\mathrm{Id})[\bar{D}]\right\| \leqslant 3Ck\sqrt{n + k\sqrt{n}} \cdot \left(\frac{2\sqrt{dn}}{k\sqrt{n+k} + \sqrt{d(n+k)} + k\sqrt{d}}\right) \leqslant 10Ck\sqrt{n}\,.$$

If $k^2 \leqslant n \leqslant d$,

$$\left\|\zeta_\tau(W^\mathsf{T}W - n\mathrm{Id})[\bar{D}]\right\| \leqslant Ck^2 \cdot \left(\frac{2d}{k^2 + d}\right) \leqslant 2Ck\sqrt{n}\,.$$

And if $n \leqslant \max\{k^2, d\}$,

$$\left\|\zeta_\tau(W^\mathsf{T}W - n\mathrm{Id})[\bar{D}]\right\| \leqslant C(n + k\sqrt{n}) \cdot \left(\frac{2d}{n + k\sqrt{n} + d}\right) \leqslant 4Ck\sqrt{n}\,.$$

So, applying Lemma A.31, we get

$$\left|\langle\eta_\tau(W^\mathsf{T}W - n\mathrm{Id})\left[\bar{D}\right], X\rangle\right| \leqslant \left\|\zeta_\tau(W^\mathsf{T}W - n\mathrm{Id})\left[\bar{D}\right]\right\| + |\langle M, X\rangle| \leqslant 2k\tau.$$

Since X is k-bounded,

$$\left|\langle(W^\mathsf{T}W - n\mathrm{Id} - \eta_\tau(W^\mathsf{T}W - n\mathrm{Id}))\left[\bar{D}\right], X\rangle\right| \leqslant k\tau.$$

Hence with probability $1 - o(1)$

$$\begin{aligned}\left|\langle W^\mathsf{T}W - n\mathrm{Id}, X\rangle\right| &\leqslant 30Ck\sqrt{n\log\left(2 + \frac{d}{k^2} + \frac{d}{n + k\sqrt{n}}\right)} + 10\sqrt{n\log d}\\ &\leqslant 100Ck\sqrt{n\log\left(2 + \frac{d}{k^2} + \frac{d}{n + k\sqrt{n}}\right)},\end{aligned}$$

since if $k \leqslant \log d$, $\log\left(2 + \frac{d}{k^2}\right) \geqslant \frac{1}{2}\log d$. □

3.2.2 The basic SDP algorithm

Having providing certificates on sparse quadratic form, we can now use Eq. (SDP-1) to obtain a robust algorithm for Sparse PCA.

Algorithm 3.12 (SDP-based Algorithm)**.**

Input: Sample matrix $Y = \sqrt{\beta}\cdot u_0 v_0^T + W + E \in \mathbb{R}^{n\times d}$ from 3.1.

Estimate: The sparse vector v_0.

Operation:

1. Compute matrix $X \in \mathbb{R}^{d\times d}$ solving program SDP-1.
2. Output top eigenvector $\hat{v}$ of X.

Indeed we will show that Algorithm 3.12 is perturbation resilient (in the sense of Appendix A.1) and its guarantees matches those of the state-of-the-art *fragile* algorithms such as SVD, Diagonal Thresholding and Covariance Thresholding. The following theorem formalize this result.

Theorem 3.13. *Let Y be a n-by-d matrix of the form,*

$$Y = \sqrt{\beta} \cdot u_0 v_0^\mathsf{T} + W + E\,,$$

for a unit k-sparse vector $v_0 \in \mathbb{R}^d$, a standard Gaussian vector $u_0 \sim N(0, \mathrm{Id}_n)$, an arbitrary matrix $E \in \mathbb{R}^{n\times d}$ and a Gaussian matrix $W \sim N(0,1)^{n\times d}$ such that W, u_0, are distributionally independent. Then algorithm 3.12 outputs a unit vector $\hat{v} \in \mathbb{R}^d$ such that with probability $1 - o(1)$,

$$1 - \langle v_0, \hat{v}\rangle^2 \lesssim \frac{k}{\beta n} \cdot q + \sqrt{\frac{k}{\beta n}}\left(\sqrt{\log\frac{d}{k}} + \|E\|_{1\to 2}\right) \cdot \left(1 + \frac{1}{\sqrt{\beta}}\right).$$

where $q := \min\left\{\sqrt{n\log\left(2 + \frac{d}{k^2} + \frac{d}{n+k\sqrt{n}}\right)}, \frac{d+\sqrt{dn}}{k}\right\}$ and $\|E\|_{1\to 2}$ denotes the largest norm of a column of E. Furthermore, the same kind of guarantees hold if u_0 is a vector with $\|u_0\|^2 = \Theta(n)$ independent of W.

We prove Theorem 3.13 through the result below, which will be useful in the Sum-of-Squares proofs as well.

Theorem 3.14 (Meta-theorem)**.** *Let Y be a n-by-d matrix of the form,*

$$Y = \sqrt{\beta} \cdot u_0 v_0^\mathsf{T} + W + E\,,$$

for a unit k-sparse vector $v_0 \in \mathbb{R}^d$, a standard Gaussian vector $u_0 \sim N(0, \mathrm{Id}_n)$, an arbitrary matrix $E \in \mathbb{R}^{n\times d}$ and a Gaussian matrix $W \sim N(0,1)^{n\times d}$ such that W, u_0, v_0 are distributionally independent. Let X be a feasible solution of SDP-1 satisfying $\langle \hat{\Sigma}, X\rangle \geqslant \langle \hat{\Sigma}, v_0 v_0^\mathsf{T}\rangle$. Then with probability $1 - o(1)$,

$$1 - \langle v_0 v_0^\mathsf{T}, X\rangle \lesssim \frac{1}{\beta n} \cdot \left|\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle\right| + \sqrt{\frac{k}{\beta n}}\left(\sqrt{\log\frac{d}{k}} + \|E\|_{1\to 2}\right) \cdot \left(1 + \frac{1}{\sqrt{\beta}}\right),$$

where $\|E\|_{1\to 2}$ denotes the largest norm of a column of E. Furthermore, the same kind of guarantees hold if u_0 is a vector with $\|u_0\|^2 = \Theta(n)$ independent of W.

Indeed Theorem 3.14 immediately implies Theorem 3.13.

Proof of Theorem 3.13. Assume Theorem 3.14 is true. By definition X satisfies its premises. By Lemma 3.11

$$\left|\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle\right| \leqslant O\left(\min\left\{k\sqrt{n\log\left(1+\frac{d}{k^2}+\frac{d}{n+k\sqrt{n}}\right)},\ d+\sqrt{dn}\right\}\right).$$

Applying Lemma A.29 the result follows. □

Now let's prove Theorem 3.14. First we look into cross-terms containing the signal.

Lemma 3.15. *Let Y be as in Theorem 3.14 and suppose $E \in \mathbb{R}^{n\times d}$ is a matrix with maximal column norm $\|E\|_{1\to 2} \leqslant b$. Let X be a feasible solution to SDP-1. Then with probability $1-o(1)$,*

$$\left|\langle W^\mathsf{T}\sqrt{\beta}u_0 v_0^\mathsf{T}, X\rangle\right| \leqslant O\left(\sqrt{\beta n k\log\frac{d}{k}}\right).$$

Proof. With probability $1-o(1)$, $\|u_0\| \leqslant O(\sqrt{n})$. Let $g = \frac{1}{\|u_0\|}W^\mathsf{T}u_0$. Since u_0 and W are independent, $g \sim N(0,1)$. Let S be the set of k largest coordinates in g, and let $g' = g[S]$. Then $g = g' + g''$, where vector g'' has entries bounded by $O\left(\sqrt{\log\frac{d}{k}}\right)$ and $\|g'\| \leqslant O\left(\sqrt{k\log\frac{d}{k}}\right)$ with probability $1-o(1)$ (by Lemma A.24). Hence by Lemma A.31,

$$\begin{aligned}\left|\langle W^\mathsf{T}\sqrt{\beta}u_0 v_0^\mathsf{T}, X\rangle\right| &\leqslant O\left(\sqrt{n\beta}\left|\langle g' v_0^\mathsf{T}, X\rangle\right| + \sqrt{n\beta}\left|\langle g'' v_0^\mathsf{T}, X\rangle\right|\right)\\ &\leqslant O\left(\sqrt{\beta n k\log\frac{d}{k}}\right) + O\left(\sqrt{n\beta}\left|\langle g'' v_0^\mathsf{T}, X\rangle\right|\right).\end{aligned}$$

By Lemma A.32 and Lemma A.31,

$$\left|\langle g'' v_0^\mathsf{T}, X\rangle\right| \leqslant \sqrt{\langle g''(g'')^\mathsf{T}, X\rangle\cdot\langle v_0 v_0^\mathsf{T}, X\rangle} \leqslant \sqrt{\langle g''(g'')^\mathsf{T}, X\rangle}.$$

The desired bound follows from Lemma 3.10, since the entries of $g''(g'')^\mathsf{T}$ are bounded by $O(\log\frac{d}{k})$ with probability $1-o(1)$. □

Lemma 3.16. *Let Y be as in Theorem 3.14 and suppose $E \in \mathbb{R}^{n\times d}$ is a matrix with maximal column norm $\|E\|_{1\to 2} \leqslant b$. Let X be a feasible solution to SDP-1. Then with probability $1-o(1)$,*

$$\left|\langle E^\mathsf{T}\sqrt{\beta}u_0 v_0^\mathsf{T}, X\rangle\right| \leqslant O\left(b\sqrt{\beta n k}\right).$$

Proof. With probability $1-o(1)$, $\|u_0\| \leqslant O(\sqrt{n})$. Let $z = E^\mathsf{T}u_0$. With probability $1-o(1)$ the entries of z are bounded by $O\left(b\sqrt{n}\right)$. By Lemma A.32 and Lemma A.31,

$$\left|\langle z v_0^\mathsf{T}, X\rangle\right| \leqslant \sqrt{\langle zz^\mathsf{T}, X\rangle\cdot\langle v_0 v_0^\mathsf{T}, X\rangle} \leqslant \sqrt{\langle zz^\mathsf{T}, X\rangle} \leqslant O\left(b\sqrt{nk}\right).$$

□

The following lemma shows how to bound the remaining cross-terms.

Lemma 3.17. *Let Y be as in Theorem 3.14 and suppose $E \in \mathbb{R}^{n\times d}$ is a matrix with maximal column norm $\|E\|_{1\to 2} \leqslant b$. Let X be a feasible solution to SDP-1. Then*

$$\left|\langle E^\mathsf{T} W + W^\mathsf{T} E, X\rangle\right| \leqslant 2b\sqrt{kn} + b^2 k + \left|\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle\right|.$$

Proof. Applying Fact A.30 with setting $A = (W - c\cdot E)^\mathsf{T}(W - c\cdot E)$ for some $c > 0$ and $B = X$ we immediately get

$$c\left|\langle E^\mathsf{T} W + W^\mathsf{T} E, X\rangle\right| \leqslant \langle W^\mathsf{T} W, X\rangle + c^2\langle E^\mathsf{T} E, X\rangle = n + \langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle + c^2\langle E^\mathsf{T} E, X\rangle.$$

By Lemma 3.10

$$\left|\langle E^\mathsf{T} W + W^\mathsf{T} E, X\rangle\right| \leqslant \frac{1}{c}\left(n + \left|\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle\right|\right) + c\cdot b^2 k.$$

Minimizing over c, we get

$$\begin{aligned}\left|\langle E^\mathsf{T} W + W^\mathsf{T} E, X\rangle\right| &\leqslant 2b\sqrt{kn + k\cdot|\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle|}\\ &\leqslant 2b\sqrt{kn} + 2b\sqrt{k\cdot|\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle|}\\ &\leqslant 2b\sqrt{kn} + b^2 k + \left|\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle\right|.\end{aligned}$$

□

We are now ready to prove Theorem 3.14.

Proof of Theorem 3.14. Opening up the product,

$$\begin{aligned}\langle\hat{\Sigma}, X\rangle =&\langle\hat{\Sigma} - n\mathrm{Id} + n\mathrm{Id}, X\rangle\\ =&\beta\|u_0\|^2\langle v_0 v_0{}^\mathsf{T}, X\rangle + n\\ &+ \langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle\\ &+ \langle E^\mathsf{T} E, X\rangle\\ &+ \langle E^\mathsf{T} W + W^\mathsf{T} E, X\rangle\\ &+ \sqrt{\beta}\langle v_0 u_0{}^\mathsf{T} W + W^\mathsf{T} u_0 v_0{}^\mathsf{T} + v_0 u_0{}^\mathsf{T} E + E^\mathsf{T} u_0 v_0{}^\mathsf{T}, X\rangle.\end{aligned}$$

Applying Lemmata 3.10, 3.11, 3.17, 3.16, 3.15 and we get

$$\langle\hat{\Sigma}, X\rangle \leqslant \beta\|u_0\|^2\langle v_0 v_0{}^\mathsf{T}, X\rangle + n + 2\left|\langle W^\mathsf{T} W - n\mathrm{Id}, X\rangle\right| + 2b^2 k + 2b\sqrt{kn} + O\left(\left(\sqrt{\log\frac{d}{k}} + b\right)\sqrt{\beta nk}\right).$$

Furthermore, by choice of X,

$$\langle\hat{\Sigma}, X\rangle \geqslant \langle\hat{\Sigma}, v_0 v_0{}^\mathsf{T}\rangle$$

$$
\begin{aligned}
&= \langle \hat{\Sigma} + n\mathrm{Id} - n\mathrm{Id}, v_0 v_0^\mathsf{T} \rangle \\
&\geqslant \beta \|u_0\|^2 + n \\
&\quad - \left| \langle W^\mathsf{T} W - n\mathrm{Id}, v_0 v_0^\mathsf{T} \rangle \right| \\
&\quad - \left| \langle E^\mathsf{T} E, v_0 v_0^\mathsf{T} \rangle \right| \\
&\quad - \left| \langle E^\mathsf{T} W + W^\mathsf{T} E, v_0 v_0^\mathsf{T} \rangle \right| \\
&\quad - \left| \sqrt{\beta} \langle v_0 u_0^\mathsf{T} W + W^\mathsf{T} u_0 v_0^\mathsf{T} + v_0 u_0^\mathsf{T} E + E^\mathsf{T} u_0 v_0^\mathsf{T}, v_0 v_0^\mathsf{T} \rangle \right| \\
&\geqslant \beta \|u_0\|^2 + n - 2\left| \langle W^\mathsf{T} W - n\mathrm{Id}, v_0 v_0^\mathsf{T} \rangle \right| - 2b^2 k - 2b\sqrt{kn} - O\left(\left(\sqrt{\log \frac{d}{k}} + b \right) \sqrt{\beta n k} \right) \\
&\geqslant \beta \|u_0\|^2 + n - O\left(\left| \langle W^\mathsf{T} W - n\mathrm{Id}, v_0 v_0^\mathsf{T} \rangle \right| - b^2 k - b\sqrt{kn} - \left(\sqrt{\log \frac{d}{k}} + b \right) \sqrt{\beta n k} \right).
\end{aligned}
$$

Now by Theorem A.23 $\left| \langle W^\mathsf{T} W - n\mathrm{Id}, v_0 v_0^\mathsf{T} \rangle \right| \leqslant 10k \log(d/k) + 20\sqrt{nk \log(d/k)}$ with probability $1 - o(1)$. Let $m = \left| \langle W^\mathsf{T} W - n\mathrm{Id}, X \rangle \right| + \left| \langle W^\mathsf{T} W - n\mathrm{Id}, v_0 v_0^\mathsf{T} \rangle \right|$. Combining the two inequalities and rearranging, we get

$$
\beta \|u_0\|^2 \cdot \left(1 - \langle v_0 v_0^\mathsf{T}, X \rangle\right) \leqslant O\left(m + b^2 k + b\sqrt{kn} + \left(\sqrt{\log \frac{d}{k}} + b \right) \sqrt{\beta n k} \right).
$$

With probability $1 - o(1)$, $\|u_0\|^2 \geqslant n/2$. Recall that $b \leqslant \sqrt{\frac{\beta n}{k}}$. Hence

$$
1 - \langle v_0 v_0^\mathsf{T}, X \rangle \leqslant \frac{1}{\beta n} O\left(m + \sqrt{\beta n k \log \frac{d}{k}} + (1 + \beta) b \sqrt{kn} \right).
$$

The result follows rearranging and observing that with probability $1 - o(1)$, $\left| \langle W^\mathsf{T} W - n\mathrm{Id}, v_0 v_0^\mathsf{T} \rangle \right| \lesssim k \log(d/k) + \sqrt{kn \log(d/k)}$ by Lemma A.23. □

3.3 Robustness of SoS and stronger certified upper bounds

In this section we prove Theorem 3.2 and Theorem 3.6. We will show that the Sum-of-Squares algorithm can certify various upper bounds on sparse eigenvalues. In Section 3.3.1 we will prove increasingly stronger certified upper bounds on sparse eigenvalues of subgaussian matrices. These certified upper bounds will require increasingly stronger assumptions on d and n, but for degree $\log(d/k)$ will approach information theoretic guarantees. In Section 3.3.2 we will prove alternative certified upper bounds fo sparse eigenvalues of Gaussian matrices. These bounds will not require any additional assumption on d and n. We will then use these bounds in Section 3.3.3 to obtain maximally robust algorithms for Sparse PCA.

3.3.1 SoS certificates for sparse eigenvalues via certifiable subgaussianity

Let $\mathcal{A}_{s,v}$ be the following system of quadratic constraints. Observe for any (s, v) satisfying $\mathcal{A}_{s,v}$, v is a k-sparse unit vector supported on coordinates i such that $s_i = 1$.

$$\mathcal{A}_{s,v} : \begin{Bmatrix} & \sum_{i=1}^{d} s_i = k \\ \forall i \in [d]. & s_i^2 = s_i \\ \forall i \in [d]. & s_i \cdot v_i = v_i \\ & \sum_{i=1}^{d} v_i^2 = 1 \end{Bmatrix} \tag{3.3.1}$$

We prove a certified upper bound for sparse eigenvalues of random rectangular matrices $W \in \mathbb{R}^{n \times d}$ with independent subgaussian entries. This upper bound differs considerably from the one obtained using SDP-1. Let us recall the definition of subgaussian random variables before proceeding.

Definition 3.18 (C-Subgaussian Random Variables)**.** A $\mathbb{R}$-valued random variable x is said to be C-subgaussian if for every t, $\mathbb{E}|x|^t \leqslant C^{t/2} t^{t/2}$.

Let $W_1, W_2, \ldots, W_d$ be the columns of W. We will use the following lemma:

Lemma 3.19. *Let $W_1, W_2, \ldots, W_d \in \mathbb{R}^n$ be independently drawn from a product distribution with each 1-subgaussian coordinates with mean 0 and variance 1. Then, with probability at least 0.99 over the draw of $W_1, W_2, \ldots, W_d$,*

$$\left| \frac{u}{2t} \right\} \frac{1}{d} \sum_{i \leqslant d} \langle W_i, u \rangle^{2t} \leqslant \|u\|_2^{2t} \left(t^t + \frac{n^{t/2} \log^{(t+1)/2}(n) (C't)^t}{\sqrt{d}} \right) \right\}.$$

for some absolute constant $C' > 0$.

We will prove the lemma whenever the columns of W are *certifiably subgaussian*. Informally, certifiably subgaussianity means that a random variable has its moments upper-bounded as in the the definition above and that this bound has a SoS proof. Formally, we have:

Definition 3.20 (Certifiable Subgaussianity)**.** A $\mathbb{R}^n$-valued random variable Y is said to be t-certifiably C-subgaussian if for all $t' \leqslant t$, $\left| \frac{u}{2t} \right\} \left\{ \mathbb{E}\langle Y, u \rangle^{2t} \leqslant C^t t^t \left(\mathbb{E}\langle Y, u \rangle^2 \right)^t \right\}$. A matrix $W \in \mathbb{R}^{n \times d}$ is said to be t-certifiably C-subgaussian if the uniform distribution on the columns of W is t-certifiably C-subgaussian.

Certifiable subgaussianity has, by now, appeared in several works [KSS18, KS17, HL18, KKM18] that employ the sum-of-squares method for statistical estimation problems.

Given the above lemma, to prove Lemma 3.19, we need to show certified subgaussianity of W when W is a random matrix in $\mathbb{R}^{n \times d}$. To show this, we will use the following fact:

Fact 3.21 (Certifiable Subgaussianity of Product Subgaussians, Lemma 5.9, Page 25 of [KSS18])**.** *Let Y be a $\mathbb{R}^d$-valued random variable with independent, C-subgaussian coordinates of mean 0 and variance 1. Then, Y is t-certifiably C-subgaussian for every t.*

We are now ready to prove Lemma 3.19.

Proof of Lemma 3.19. We have:

$$\left|\frac{u}{2t}\right. \left\{\frac{1}{d}\left(\sum_{i\leqslant d}\langle W_i,u\rangle^{2t} - \mathbb{E}\langle W_i,u\rangle^{2t}\right) = \left\langle u^{\otimes t}, \left(\frac{1}{d}\sum_{i\leqslant d}\left(W_i^{\otimes t}\right)\left(W_i^{\otimes t}\right)^\top - \mathbb{E}\left(W_i^{\otimes t}\right)\left(W_i^{\otimes t}\right)^\top\right)u^{\otimes t}\right\rangle\right\}.$$

Using Fact 2.14 and $\left\|u^{\otimes t}\right\|_2^2 = \|u\|_2^{2t}$, we have:

$$\left|\frac{u}{2t}\right. \left\{\frac{1}{d}\sum_{i\leqslant d}\langle W_i,u\rangle^{2t} - \mathbb{E}\langle W_i,u\rangle^{2t} \leqslant \|u\|_2^{2t}\cdot\left\|\left(\frac{1}{d}\left(W_i^{\otimes t}\right)\left(W_i^{\otimes t}\right)^\top - \mathbb{E}\left(W_i^{\otimes t}\right)\left(W_i^{\otimes t}\right)^\top\right)\right\|\right\}. \tag{3.3.2}$$

From Lemma A.17, we know that with probability at least 0.99 over the draw of $W_1, W_2, \ldots, W_d$, it holds that:

$$\left\|\left(\frac{1}{d}\left(W_i^{\otimes t}\right)\left(W_i^{\otimes t}\right)^\top - \mathbb{E}\left(W_i^{\otimes t}\right)\left(W_i^{\otimes t}\right)^\top\right)\right\| \leqslant \frac{n^{t/2}\log^{(t+1)/2}(n)(C't)^t}{\sqrt{d}}. \tag{3.3.3}$$

Using Fact 3.21,

$$\left|\frac{u}{2t}\right. \left\{\mathbb{E}\langle W_i,u\rangle^{2t} \leqslant t^t\|u\|_2^{2t}\right\}. \tag{3.3.4}$$

Combining (3.3.2), (3.3.3) and (3.3.4), we have:

$$\left|\frac{u}{2t}\right. \left\{\frac{1}{d}\sum_{i\leqslant d}\langle W_i,u\rangle^{2t} \leqslant \|u\|_2^{2t}\left(t^t + \frac{n^{t/2}\log^{(t+1)/2}(n)(C't)^t}{\sqrt{d}}\right)\right\}.$$

□

Lemma 3.19 implies the following lemma:

Lemma 3.22. *Let W satisfy the assumptions of Lemma 3.19. Suppose that $d \geqslant t^t n^t \log^{(t+1)}(n)$. Then with probability at least 0.99,*

$$\mathcal{A}_{s,v}\left|\frac{s,v}{2t}\right. \left\{\|Wv\|_2^{4t} \leqslant dk^{t-1}(C't)^t\|Wv\|_2^{2t}\right\}.$$

for some absolute constant $C' > 0$.

Proof. For $u = Wv$, using $\mathcal{A}_{s,v}\left|\frac{s}{2t}\right. \{s_iv_i = v_i \mid \forall i\}$ and Cauchy-Schwarz inequality, we have:

$$\mathcal{A}_{s,v}\left|\frac{s,v,u}{2t}\right. \left\{\left(\sum_{i\leqslant d}s_iv_i\langle W_i,u\rangle\right)^{2t} \leqslant \left(\sum_{i\leqslant d}v_i^2\right)^t\left(\sum_{i\leqslant d}s_i^2\langle W_i,u\rangle^2\right)^t\right\}$$

Using $\mathcal{A}_{s,v} \left|\frac{s}{2t}\right. \{s_i^{t-1} = s_i^2 \mid \forall i\}$, we have:

$$\mathcal{A}_{s,v} \left|\frac{s,v,u}{t}\right. \left\{\left(\sum_{i\leqslant d} s_i v_i \langle W_i, u\rangle\right)^{2t} \leqslant \left(\sum_{i\leqslant d} v_i^2\right)^t \left(\sum_{i\leqslant d} s_i^2 \langle W_i, u\rangle^2\right)^t\right\}$$

Now, using $\mathcal{A}_{s,v} \left|\frac{s,v}{2}\right. \{\sum_i s_i = k\}$ and Lemma 3.19, we have:

$$\begin{aligned}\mathcal{A}_{s,v} \left|\frac{s,v,u}{t}\right. \left\{\left(\sum_{i\leqslant d} s_i \langle W_i, u\rangle^2\right)^t = \left(\sum_{i\leqslant d} s_i^{t-1} \langle W_i, u\rangle^2\right)^t \leqslant \left(\sum_{i\leqslant d} s_i^t\right)^{t-1} \left(\sum_{i\leqslant d} \langle W_i, u\rangle^{2t}\right) \right. \\ \left. \leqslant k^{t-1} d \|u\|_2^{2t} \left(t^t + \frac{n^{t/2} \log^{(t+1)/2}(n) (C't)^t}{\sqrt{d}}\right)\right\}\end{aligned} \tag{3.3.5}$$

Plugging back $u = Wv$, we get the desired bound. □

Now we are ready to derive the certified upper bound on $\|Wv\|_2^2$.

Lemma 3.23. *Suppose that $d \geqslant C^* t^t n^t \log^t(n)$ for large enough absolute constant C^*. Let D be a pesudo-distribution satisfying $\mathcal{A}_{s,v}$. Let $W \in \mathbb{R}^{n\times d}$ with i.i.d. 1-subgaussian entries with mean 0 and variance 1. Then, with probability at least 0.99 over the draw of $W_1, W_2, \ldots, W_d$,*

$$\tilde{\mathbb{E}}_D \|Wv\|_2^2 \leqslant C' \cdot d^{1/t} k^{1-\frac{1}{t}} t\,,$$

for some absolute constant $C' > 0$.

Proof. Using Lemma 3.22 and taking pseudo-expectations with respect to D that satisfies $\mathcal{A}_{s,v}$, we have:

$$\tilde{\mathbb{E}}_D \|Wv\|_2^{4t} \leqslant d k^{t-1} (C't)^t \tilde{\mathbb{E}}_D \|Wv\|_2^{2t}\,.$$

By Cauchy-Schwarz inequality for pseudo-distributions, $\tilde{\mathbb{E}}_D \|Wv\|_2^{2t} \leqslant \left(\tilde{\mathbb{E}}_D \|Wv\|_2^{4t}\right)^{1/2}$, and by Hölder's indequality $\left(\tilde{\mathbb{E}}_D \|Wv\|_2^2\right)^{2t} \leqslant \tilde{\mathbb{E}}_D \|Wv\|_2^{4t}$. Thus, we have:

$$\left(\tilde{\mathbb{E}}_D \|Wv\|_2^2\right)^t \leqslant d k^{t-1} (C't)^t\,.$$

Taking t-th roots gives: $\tilde{\mathbb{E}}_D \|Wv\|_2^2 \leqslant C' \cdot d^{1/t} k^{1-\frac{1}{t}} t$. □

3.3.2 SoS certificates for sparse eigenvalues via limited brute force

We show here that, using additional constraints over the system $\mathcal{A}_{s,v}$, we can provide different certified upper bounds on the sparse eigenvalues of Gaussian matrices W.

Let $\mathcal{S}_t$ be a set of all vectors with values in $\{0, 1\}$ that have exactly t nonzero coordinates. We start with a definition.

Definition 3.24. For any $u \in \mathcal{S}_t$ we define a polynomial in variables $s_1, \dots, s_d =: s$

$$p_u(s) = \binom{k}{t}^{-1} \cdot \prod_{i \in \mathrm{supp}\{u\}} s_i \,.$$

Note that if v denotes a k-sparse vector and s is the indicator of its support, then for any $u \in \mathcal{S}_t$,

$$p_u(s) = \begin{cases} \binom{k}{t}^{-1} & \text{if } \mathrm{supp}\{u\} \subseteq \mathrm{supp}\{v\} \\ 0 & \text{otherwise} \end{cases}$$

Now consider the following system $\mathcal{B}_{s,v}$ of polynomial constraints.

$$\mathcal{B}_{s,v} : \left\{ \begin{array}{ll} \forall i \in [d], & s_i^2 = s_i \\ & \sum_{i \in [d]} s_i = k \\ \forall i \in [d], & s_i \cdot v_i = v_i \\ & \sum_{i \in [d]} v_i^2 = 1 \\ & \sum_{u \in \mathcal{S}_t} p_u(s) = 1 \\ \forall i \in [d], & \sum_{u \in \mathcal{S}_t} u_i p_u(s) = \frac{t}{k} \cdot s_i \end{array} \right\} \tag{3.3.6}$$

We will use the following preliminary fact.

Fact 3.25. *Let $W \sim N(0,1)^{n \times d}$, let $n \geqslant \log d$ and let $t \leqslant \frac{1}{\log d} \min\{d, n\}$. Then with probability $1 - o(1)$ all principle submatrices of $W^\mathsf{T} W - n\mathrm{Id}$ of size $t \times t$ have spectral norm bounded by $O\left(\sqrt{nt \log d}\right)$.*

Proof. Fix a $t \times t$ principal submatrix N. By Theorem A.20 there exists a constant $C > 0$, such that $\|N\| \leqslant C\sqrt{nt \log d}$ with probability at least $1 - d^{10t}$. The fact follows taking a union bound over all possible $\binom{d}{t}$ submatrices. □

We are now ready to show the upper bound on quadratic forms of sparse vectors.

Theorem 3.26. *Let $W \sim N(0,1)^{n \times d}$. Then there exists a constant $C > 0$ such that with probability at least $1 - o(1)$,*

$$\mathcal{B}_{s,v} \left| \frac{s,v}{4t} \right. \left\{ -C \cdot \frac{k}{\sqrt{t}} \sqrt{n \log d} \leqslant v^\mathsf{T} (W^\mathsf{T} W - n\mathrm{Id}) v \leqslant C \cdot \frac{k}{\sqrt{t}} \sqrt{n \log d} \right\}.$$

Proof. Note that

$$\mathcal{B}_{s,v} \left|\frac{s,v}{2t}\right. \left\{ ss^\mathsf{T} = \frac{k^2}{t^2} \sum_{u,u' \in \mathcal{S}_t} u'u^\mathsf{T} p_{u'}(s) p_u(s) \right\}.$$

For vectors $x, y \in \mathbb{R}^d$ we denote the vector with entries $x_i \cdot y_i$ by $\big(xy\big)$. It follows that

$$\begin{aligned}
&\mathcal{B}_{s,v} \left|\frac{s,v}{4t}\right. \left\{ vv^\mathsf{T} = (vs)(vs)^\mathsf{T} \right\} \\
&\mathcal{B}_{s,v} \left|\frac{s,v}{4t}\right. \left\{ vv^\mathsf{T} = \frac{k^2}{t^2} \sum_{u,u' \in \mathcal{S}_t} (vu')(vu)^\mathsf{T} p_{u'}(s) p_u(s) \right\}.
\end{aligned}$$

Let $M = \big(W^\mathsf{T} W - n\mathrm{Id}\big)$. Then

$$\begin{aligned}
\mathcal{B}_{s,v} &\left|\frac{s,v}{4t}\right. \left\{ v^\mathsf{T} M v = \langle M, vv^\mathsf{T} \rangle \right\} \\
&\left|\frac{s,v}{4t}\right. \left\{ v^\mathsf{T} M v = \langle M, \frac{k^2}{t^2} \sum_{u,u' \in \mathcal{S}_t} (vu')(vu)^\mathsf{T} p_{u'}(s) p_u(s) \rangle \right\} \\
&\left|\frac{s,v}{4t}\right. \left\{ v^\mathsf{T} M v = \frac{k^2}{t^2} \sum_{u,u' \in \mathcal{S}_t} (vu)^\mathsf{T} M (vu') p_{u'}(s) p_u(s) \right\}
\end{aligned}$$

Now for any $u, u' \in \mathcal{S}_t$,

$$\begin{aligned}
\mathcal{B}_{s,v} &\left|\frac{s,v}{2}\right. \left\{ (vu)^\mathsf{T} M (vu') = (vu)^\mathsf{T} W^\mathsf{T} W (vu') - n(vu)^\mathsf{T} (vu') \right\} \\
&\left|\frac{s,v}{2}\right. \left\{ 2(vu)^\mathsf{T} M (vu') \leqslant (vu)^\mathsf{T} W^\mathsf{T} W (vu) + (vu')^\mathsf{T} W^\mathsf{T} W (vu') - 2n(vu)^\mathsf{T} (vu') \right\} \\
&\left|\frac{s,v}{2}\right. \left\{ 2(vu)^\mathsf{T} M (vu') \leqslant (vu)^\mathsf{T} M (vu) + (vu')^\mathsf{T} M (vu') + n \cdot \big(\|(vu)\|^2 + \|(vu')\|^2 - 2(vu)^\mathsf{T} (vu') \big) \right\} \\
&\left|\frac{s,v}{2}\right. \left\{ 2(vu)^\mathsf{T} M (vu') \leqslant (vu)^\mathsf{T} M (vu) + (vu')^\mathsf{T} M (vu') \right\}.
\end{aligned}$$

where the first equality follows by definition, the second using the fact that for any $N \succeq 0$, and $a, b \in \mathbb{R}^d$, $\left|\frac{s,v}{2}\right. \big\{ \langle (a-b)(a-b)^\mathsf{T}, N \rangle \geqslant 0 \big\}$. The last follows from the fact that $\left|\frac{s,v}{2}\right. \Big\{ \|a - b\|^2 \geqslant 0 \Big\}$. Similar derivation shows that $\mathcal{B}_{s,v} \left|\frac{s,v}{4t}\right.$ $\big\{ -2(vu)^\mathsf{T} M (vu') \leqslant -(vu)^\mathsf{T} M (vu) - (vu')^\mathsf{T} M (vu') \big\}$.

Now let q be the maximal norm of any $t \times t$ principal submatrices of M. Note that for any $u \in \mathcal{S}_t$, $|\mathrm{supp}\{(vu)\}| \leqslant t$. Since $\mathcal{B}_{s,v} \left|\frac{s,v}{4t}\right. \{ s_i \geqslant 0 \}$, $\mathcal{B}_{s,v} \left|\frac{s,v}{4t}\right. \big\{ p_u(s) \geqslant 0 \big\}$,

$$\begin{aligned}
\mathcal{B}_{s,v} &\left|\frac{s,v}{4t}\right. \left\{ v^\mathsf{T} M v \leqslant \frac{k^2}{2t^2} \sum_{u,u' \in \mathcal{S}_t} \big((vu)^\mathsf{T} M (vu) + (vu')^\mathsf{T} M (vu') \big) p_{u'}(s) p_u(s) \right\} \\
&\left|\frac{s,v}{4t}\right. \left\{ v^\mathsf{T} M v \leqslant \frac{k^2}{2t^2} \sum_{u,u' \in \mathcal{S}_t} \big(q \cdot \|(vu)\|^2 + q \cdot \|(vu')\|^2 \big) p_{u'}(s) p_u(s) \right\} \\
&\left|\frac{s,v}{4t}\right. \left\{ v^\mathsf{T} M v \leqslant q \frac{k^2}{2t^2} \left(\sum_{u \in \mathcal{S}_t} \|(vu)\|^2 p_u(s) \left(\sum_{u' \in \mathcal{S}_t} p_{u'}(s) \right) + \sum_{u' \in \mathcal{S}_t} \|(vu')\|^2 p_{u'}(s) \left(\sum_{u \in \mathcal{S}_t} p_u(s) \right) \right) \right\}
\end{aligned}$$

$$\left|\frac{s,v}{4t}\right. \left\{ v^\top M v \leqslant q\frac{k^2}{2t^2}\left(\sum_{u\in\mathcal{S}_t}\|(vu)\|^2 p_u(s) + \sum_{u'\in\mathcal{S}_t}\|(vu')\|^2 p_{u'}(s)\right)\right\}$$
$$\left|\frac{s,v}{4t}\right. \left\{ v^\top M v \leqslant q\frac{k^2}{t^2}\sum_{u\in\mathcal{S}_t}\|(vu)\|^2 p_u(s).\right\}$$

Here the second inequality follows from choice of q, the third uses the fact that $\mathcal{B}_{s,v}\left|\frac{s,v}{4t}\right. \left\{\left(\sum_{u\in\mathcal{S}_t} p_u(s)\right) = 1\right\}$. Finally observe that

$$\mathcal{B}_{s,v}\left|\frac{s,v}{4t}\right. \left\{\sum_{u\in\mathcal{S}_t}\|(vu)\|^2 p_u(s) = \sum_{u\in\mathcal{S}_t}\sum_{i=1}^{d} v_i^2 u_i^2 \cdot p_u(s)\right\}$$
$$\left|\frac{s,v}{4t}\right. \left\{\sum_{u\in\mathcal{S}_t}\|(vu)\|^2 p_u(s) = \sum_{i=1}^{d} v_i^2 \sum_{u\in\mathcal{S}_t} u_i \cdot p_u(s)\right\}$$
$$\left|\frac{s,v}{4t}\right. \left\{\sum_{u\in\mathcal{S}_t}\|(vu)\|^2 p_u(s) = \frac{t}{k}\sum_{i=1}^{d} v_i^2 s_i\right\}$$
$$\left|\frac{s,v}{4t}\right. \left\{\sum_{u\in\mathcal{S}_t}\|(vu)\|^2 p_u(s) = \frac{t}{k}\right\},$$

where we used the facts $cB_{s,v}\left|\frac{s,v}{2}\right. \{v_i = v_i \cdot s_i\}$ and $cB_{s,v}\left|\frac{s,v}{4t}\right. \left\{\sum_{u\in\mathcal{S}_t} u_i p_u(s) = \frac{t}{k}\cdot s_i\right\}$. By Fact 3.25, there exists an absolute constant $C > 0$ such that with probability $1 - o(1)$, $q \leqslant C\sqrt{nt\log t}$. Hence with probability $1 - o(1)$,

$$\mathcal{B}_{s,v}\left|\frac{s,v}{4t}\right. \left\{v^\top M v \leqslant C\frac{k}{\sqrt{t}}\sqrt{n\log d}\right\}.$$

Similar derivation shows that

$$\mathcal{B}_{s,v}\left|\frac{s,v}{4t}\right. \left\{-v^\top M v \leqslant C\frac{k}{\sqrt{t}}\sqrt{n\log d}\right\}.$$

□

3.3.3 SoS algorithms

We now use the certified upper bounds from the previous sections to obtain efficient algorithms for Sparse PCA with adversarial errors, thus proving Theorem 3.2 and Theorem 3.6 which we formally restate.

Theorem 3.27. *Suppose $d \gtrsim n^t \log^t(n) t^t$ for $t \in \mathbb{N}$. Let Y be an n-by-d matrix of the form,*

$$Y = \sqrt{\beta} \cdot u_0 v_0^\mathsf{T} + W + E\,,$$

for a unit k-sparse vector $v_0 \in \mathbb{R}^d$, a standard Gaussian vector $u_0 \sim N(0, \mathrm{Id}_n)$, an arbitrary matrix $E \in \mathbb{R}^{n\times d}$ and a Gaussian matrix $W \sim N(0,1)^{n\times d}$ such that W, u_0 are distributionally independent. Then we can compute in time $d^{O(t)}$ a unit vector $\hat{v} \in \mathbb{R}^d$ such that with probability at least 0.99,

$$1 - \langle \hat{v}, v_0\rangle^2 \lesssim \frac{k}{\beta n} \cdot t \cdot \left(\tfrac{d}{k}\right)^{1/t} + \frac{1}{\beta} + \sqrt{\frac{k}{\beta n}}\left(\sqrt{\log \frac{d}{k}} + \|E\|_{1\to 2}\right) \cdot \left(1 + \frac{1}{\sqrt{\beta}}\right),$$

where $\|E\|_{1\to 2}$ denotes the largest norm of a column of E. Furthermore, the same kind of guarantees hold if u_0 is a vector with $\|u_0\|^2 = \Theta(n)$ independent of W.

Theorem 3.28. *Suppose $n \gtrsim \log d$ and $t \leqslant k$. Let Y be an n-by-d matrix of the form,*

$$Y = \sqrt{\beta} \cdot u_0 v_0^\mathsf{T} + W + E\,,$$

for a unit k-sparse vector $v_0 \in \mathbb{R}^d$, a standard Gaussian vector $u_0 \sim N(0, \mathrm{Id}_n)$, an arbitrary matrix $E \in \mathbb{R}^{n\times d}$ and a Gaussian matrix $W \sim N(0,1)^{n\times d}$ such that W, u_0 are distributionally independent. Then we can compute in time $n^{O(1)} d^{O(t)}$ a unit vector $\hat{v} \in \mathbb{R}^d$ such that with probability $1 - o(1)$,

$$1 - \langle \hat{v}, v_0\rangle^2 \lesssim \frac{k}{\beta} \cdot \sqrt{\frac{\log d}{nt}} + \sqrt{\frac{k}{\beta n}}\left(\sqrt{\log \frac{d}{k}} + \|E\|_{1\to 2}\right) \cdot \left(1 + \frac{1}{\sqrt{\beta}}\right),$$

where $\|E\|_{1\to 2}$ denotes the largest norm of a column of E. Furthermore, the same kind of guarantees hold if u_0 is a vector with $\|u_0\|^2 = \Theta(n)$ independent of W.

We will prove Theorem 3.27 and Theorem 3.28 using Algorithm 3.29.

Algorithm 3.29 (Algorithm for Sparse PCA with Adversarial Corruptions)**.**

Given: Sample matrix $Y = \sqrt{\beta} \cdot u_0 v_0^T + W + E \in \mathbb{R}^{n\times d}$ from model 3.1, system $\mathcal{C}_{sv} \in \{\mathcal{A}_{s,v}, \mathcal{B}_{s,v}\}$

Estimate: The sparse vector v_0.

Operation:

1. find a level-$4t$ pseudo-distribution D that satisfies $\mathcal{C}_{s,v}$ and maximizes $\tilde{\mathbb{E}}\|Yv\|_2^2$.
2. Output a top eigenvector $\hat{v}$ of $\tilde{\mathbb{E}} vv^\mathsf{T}$.

Let us analyze the algorithm. The first observation is that any pseudo-distribution satisfying $\mathcal{B}_{s,v}$ also satisfies $\mathcal{A}_{s,v}$. Next we show that any pseudo-distribution satisfying $\mathcal{A}_{s,v}$ is a feasible solution to SDP-1. This will allows us to use Theorem 3.14 and conclude the proofs of Theorem 3.27 and Theorem 3.28.

Lemma 3.30. *Let D be any pseudo-distribution of degree $\geqslant 4$ satisfying $\mathcal{A}_{s,v}$. Then $\tilde{\mathbb{E}}_D vv^\top$ is a feasible solution to SDP-1.*

Proof. Since D satisfies $\mathcal{A}_{s,v}$, $\operatorname{Tr} \tilde{\mathbb{E}}_D vv^\top = \tilde{\mathbb{E}}_D \sum_{i\leqslant d} v_i^2 = 1$. Now, there exists a vector $x \in \mathbb{R}^d$ with entries in $\{-1,+1\}$ such that $\left\|\tilde{\mathbb{E}}_D vv^\top\right\|_1 = \langle xx^\top, \tilde{\mathbb{E}}_D vv^\top\rangle$. By Cauchy-Schwarz inequality for pseudo-distributions,

$$\begin{aligned}
\langle xx^\top, \tilde{\mathbb{E}}_D vv^\top\rangle =& \tilde{\mathbb{E}}_D \sum_{i,j\leqslant d} x_i s_i v_i x_j s_j v_j \\
\leqslant& \sqrt{\tilde{\mathbb{E}}_D \sum_{i,j\leqslant d} x_i^2 x_j^2 s_i^2 s_j^2} \cdot \sqrt{\tilde{\mathbb{E}}_D \sum_{i,j\leqslant d} v_i^2 v_j^2} \\
=& \tilde{\mathbb{E}}_D \sum_{i\leqslant d} s_i^2 \\
=& k\,.
\end{aligned}$$

The result follows as $\tilde{\mathbb{E}} vv^\top \succeq 0$. □

We can now finish the analyses using the certified upper bounds from the previous sections.

Proof of Theorem 3.27. Let D be the pseudo-distribution in Algorithm 3.29. By Lemma 3.23, with probability at least 0.99, $\tilde{\mathbb{E}}_D \|Wv\|_2^2 \leqslant O\left(d^{1/t} k^{1-\frac{1}{t}} t\right)$. Note that since the pseudo-distribution that outputs v_0 satisfies $\mathcal{A}_{s,v}$, by Lemma 3.30, $\tilde{\mathbb{E}}_D vv^\top$ satisfies the premises of Theorem 3.14. Then we immediately get,

$$1 - \tilde{\mathbb{E}}_D \langle v, v_0\rangle^2 \lesssim \frac{k}{\beta n} \cdot t \cdot \left(\tfrac{d}{k}\right)^{1/t} + \frac{1}{\beta} + \sqrt{\frac{k}{\beta n}} \left(\sqrt{\log \frac{d}{k}} + \|E\|_{1\to 2}\right) \cdot \left(1 + \frac{1}{\sqrt{\beta}}\right).$$

The result follows applying Lemma A.29. □

Similarly,

Proof of Theorem 3.28. Let D be the pseudo-distribution in Algorithm 3.29. By Theorem 3.26, with probability $1 - o(1)$, $\left|\tilde{\mathbb{E}}_D v^\top (W^\top W - n\mathrm{Id}) v\right| \leqslant O\left(\frac{k}{\sqrt{t}} \sqrt{n \log d}\right)$. Note that since the pseudo-distribution that outputs v_0 with probability 1 satisfies $\mathcal{B}_{s,v}$, by Lemma 3.30, $\tilde{\mathbb{E}}_D vv^\top$ satisfies the premises of Theorem 3.14. Then we immediately get,

$$1 - \tilde{\mathbb{E}}_D \langle v, v_0\rangle^2 \lesssim \frac{k}{\beta\sqrt{nt}} \log d + \sqrt{\frac{k}{\beta n}} \left(\sqrt{\log \frac{d}{k}} + \|E\|_{1\to 2}\right) \cdot \left(1 + \frac{1}{\sqrt{\beta}}\right).$$

The result follows applying Lemma A.29. □

3.4 Unconditional lower bound in the presence of adversarial perturbations

The goal of this section is to formalize Theorem 3.3. We do so first providing an overview of the low-degree likelihood ratio and low-degree polynomials and their role in our proof. Then we obtain the Theorem.

3.4.1 Low-degree likelihood ratio

The low-degree likelihood ratio is a proxy to model efficiently computable functions. It is closely related to the pseudo-calibration technique and it has been developed in a recent line of work on the Sum-of-Squares hierarchy [BHK+19, HS17, HKP+17, Hop18]. Our description is also based on [BKW20].

The objects of study are distinguishing versions of planted problems, in which given two distributions and an instance, the goal is to decide from which distribution the instance was sampled. For example, in the context of Sparse PCA, the distinguishing formulation takes the form of deciding whether the matrix Y was sampled according to the (planted) distribution as described in 3.1, or if it was sampled from the (null) Gaussian distribution $N(0,1)^{n\times d}$. In general, we denote with ν the null distribution and with μ the planted distribution with the hidden structure.

3.4.1.1 Background on classical decision theory

From the point of view of classical Decision Theory, the optimal algorithm to distinguish between two distribution is well-understood. Given distributions ν and μ on a measurable space $\mathcal{S}$, the likelihood ratio $L(Y) := d\mathbb{P}_\mu(Y)/d\mathbb{P}_\nu(Y)$[10] is the optimal function to distinguish whether $Y \sim \nu$ or $Y \sim \mu$ in the following sense.

Proposition 3.31. *[NP33] If μ is absolutely continuous with respect to ν, then the unique solution of the optimization problem*

$$\max \mathop{\mathbb{E}}_{\mu}\left[f(Y)\right] \qquad \textit{subject to } \mathop{\mathbb{E}}_{\nu}\left[f(Y)^2\right] = 1$$

is the normalized likelihood ratio $L(Y)/\mathbb{E}_\nu\left[L(Y)^2\right]$ and the value of the optimization problem is $\mathbb{E}_\nu\left[L(Y)^2\right]$.

Similarly, arguments about statistical distinguishability are known as well. Unsurprisingly, the likelihood ratio plays a major role here as well. The key concept is the Le Cam's contiguity.

[10] The Radon-Nikodym derivative

Definition 3.32. [LCY90] Let $\underline{\mu} = \left(\mu_n\right)_{n\in\mathbb{N}}$ and $\underline{\nu} = (\nu_n)_{n\in\mathbb{N}}$ be sequences of probability measures on a common probability space $\mathcal{S}_n$. Then $\underline{\mu}$ and $\underline{\nu}$ are *contiguous*, written $\underline{\mu} \triangleleft \underline{\nu}$, if as $n \to \infty$, whenever for $A_n \in \mathcal{S}_n$, $\mathbb{P}_{\underline{\mu}}(A_n) \to 0$ then $\mathbb{P}_{\underline{\nu}}(A_n) \to 0$.

Contiguity allows us to capture the idea of indistinguishability of probability measures. Indeed two contiguous sequences $\underline{\mu}, \underline{\nu}$ of probability measures are indistinguishable in the sense than there is no function $f : \mathcal{S}_n \to \{0, 1\}$ such that $f(Y) = 1$ with high probability whenever $Y \sim \underline{\mu}$ and $f(Y) = 0$ with high probability whenever $Y \sim \underline{\nu}$. The key tool now is the so called *Second Moment Method*, which allows us to establish contiguity through the likelihood ratio.

Proposition 3.33. *If $\mathbb{E}_{\nu}\left[L_n(Y)^2\right]$ remains bounded as $n \to \infty$, then $\underline{\mu} \triangleleft \underline{\nu}$.*

This discussion allows us to argue whether a given function can be used to distinguish between our planted and null distributions.

3.4.1.2 Background on the low-degree method

The main problem with the likelihood ratio is that it is in general hard to compute, thus we need to restrict these classical analysis to the space of efficiently computable functions. Concretely, we use low-degree multivariate polynomials in the entries of the observation Y as a proxy for efficiently computable functions. Denoting with $\mathbb{R}_{\leqslant D}[Y]$ the space of polynomials in Y of degree at most D we can establish a low-degree version of the Neyman-Pearson lemma.

Proposition 3.34 (e.g. [Hop18])**.** *The unique solution of the optimization problem*

$$\max_{f\in\mathbb{R}_{\leqslant D}[Y]} \mathop{\mathbb{E}}_{\mu}\left[f(Y)\right] \qquad \text{subject to } \mathop{\mathbb{E}}_{\nu}\left[f(Y)^2\right] = 1$$

is the normalized orthogonal projection $L^{\leqslant D}(Y)/\mathbb{E}_{\nu}\left[L^{\leqslant D}(Y)^2\right]$ of the likelihood ratio $L(Y)$ onto $\mathbb{R}_{\leqslant D}[Y]$ and the value of the optimization problem is $\mathbb{E}_{\nu}\left[L^{\leqslant D}(Y)^2\right]$.

It is important to remark that at the heart of our discussion, there is the belief that in the study of planted problems, low-degree polynomials capture the computational power of efficiently computable functions. This can be phrased as the following conjecture.

Conjecture 3.35 (Informal)**.** *[BHK+19, HS17, HKP+17, Hop18] For "nice" sequences of probability measures $\underline{\mu}$ and $\underline{\nu}$, if there exists $D = D(d) \geqslant \omega(\log d)$ for which $\mathbb{E}_{\nu}\left[L^{\leqslant D}(Y)^2\right]$ remains bounded as $d \to \infty$, then there is no polynomial-time algorithm that distinguishes in the sense described in 3.4.1.1.*[11]

A large body of work provide support for this conjecture (see any of the citations above), mostly in the form of evidence of an intimate relation between polynomials and Sum of Squares algorithms and lower bounds. For a more in detail discussion we point the interested reader to [HKP+17, Hop18].

[11] We do not explain what "nice" means and direct the reader to [Hop18].

3.4.1.3 Low-degree polynomials

In light of the discussions in Section 3.1.2 and Section 3.4.1 we study a distinguishing problems between two distributions over matrices: the null distribution ν, which in our case is a standard Gaussian, and the planted distribution μ that contains some sparse signal hidden in random (and adversarial) noise. That is, given an instance Y sampled either from the null or from the planted distribution, the goal is to determine whether Y contains a planted signal. We will show that a large class of polynomial time algorithms (capturing the best known algorithms) cannot distinguish between the null and the planted case even when information-theoretically possible. Specifically, we will show that low degree polynomial estimators cannot solve these problem. Similarly to [HKP+17, HS17, DKWB23], we study the low degree analogue of the χ^2-divergence between probability measures.

Definition 3.36. Let μ and ν be probability distributions over $\mathbb{R}^{n\times d}$, and denote by F the set of all functions $f : \mathbb{R}^{n\times d} \to \mathbb{R}$ such that $\left|\mathbb{E}_\mu f\right| < \infty$ and $0 < \mathbb{V}_\nu f < \infty$. The *$\chi^2$-divergence* of μ with respect to ν is defined as

$$\chi^2(\mu \parallel \nu) = \sup_{f\in F} \frac{\left(\mathbb{E}_\mu f - \mathbb{E}_\nu f\right)^2}{\mathbb{V}_\nu f}\,.$$

Note that this value is related to the likelihood ratio L described in Section 3.4.1: the fraction in the right hand side is maximized for $f = L$, and $\chi^2(\mu \parallel \nu) = \mathbb{E}_\nu L^2 - 1$.

Recall that, if $\chi^2(\mu \parallel \nu)$ is bounded, then μ and ν are information-theoretically indistinguishable in the sense of Section 3.4.1.1. The low-degree analogue of χ^2-divergence is defined similarly. Denote by $\mathbb{R}[Y]_{\leqslant D}$ the set of polynomials of degree at most D in $\mathbb{R}[Y]$ (where $\mathbb{R}[Y]$ is the space of polynomials of $n \cdot d$ variables corresponding to the entries of Y).

Definition 3.37. Let $D > 0$ and let μ and ν be probability distributions over $\mathbb{R}^{n\times d}$ such that ν is absolutely continious and for all $p \in \mathbb{R}[Y]_{\leqslant D}$, $\left|\mathbb{E}_\mu p\right| < \infty$ and $\mathbb{V}_\nu p < \infty$. The *degree-$D$ χ^2-divergence* of μ with respect to ν is defined as

$$\chi^2_{\leqslant D}(\mu \parallel \nu) = \sup_{p\in\mathbb{R}[Y]_{\leqslant D}} \frac{\left(\mathbb{E}_\mu p - \mathbb{E}_\nu p\right)^2}{\mathbb{V}_\nu p}\,,$$

where we assume that $0/0 = 0$.

Note that since ν is absolutely continuous, the denominator $\mathbb{V}_\nu p$ is zero if and only if p is constant (and in this case the numerator is also zero).

3.4.1.4 Chi-squared-divergence and orthogonal polynomials

Recall that given a hypothesis testing problem with null distribution ν and planted distribution μ, we say a polynomial $p(Y) \in \mathbb{R}[Y]_{\leqslant D}$ cannot distinguish between μ and ν if

$$\frac{\left|\mathbb{E}_\mu p(Y) - \mathbb{E}_\nu p(Y)\right|}{\sqrt{\mathbb{V}_\nu p(Y)}} \leqslant o(1)\,. \tag{3.4.1}$$

So, if for some distinguishing problem this ratio is small for all $p \in \mathbb{R}[Y]_{\leqslant D}$, then polynomial estimators of degree at most D cannot solve this distinguishing problem. The key observation used to prove bounds for low degree polynomials is the fact that the polynomial which maximizes the ratio (3.4.1) has a convenient characterization in terms of orthogonal polynomials with respect to the null distribution.

Formally, for any linear subspace of polynomials $\mathcal{S}_{\leqslant D} \subseteq \mathbb{R}[Y]_{\leqslant D}$ and any absolutely continuous probability distribution ν such that all polynomials of degree at most $2D$ are ν-integrable, one can define an inner product in the space $\mathcal{S}_{\leqslant D}$ as follows

$$\forall p, q \in \mathcal{S}_{\leqslant D} \quad \left\langle p, q \right\rangle = \mathop{\mathbb{E}}_{Y\sim\nu} p(Y)q(Y)\,.$$

Hence we can talk about orthonormal basis in $\mathcal{S}_{\leqslant D}$ with respect to this inner product.

Proposition 3.38. *Let $\mathcal{S}_{\leqslant D} \subseteq \mathbb{R}[Y]_{\leqslant D}$ be a linear subspace of polynomials of dimension N. Suppose that ν and μ are probability distributions over $Y \in \mathbb{R}^{n\times d}$ such that any polynomial of degree at most D is μ-integrable and any polynomial of degree at most $2D$ is ν-integrable. Suppose also that ν is absolutely continuous. Let $\{\psi_i(Y)\}_{i=1}^N$ be an orthonormal basis in $\mathcal{S}_{\leqslant D}[Y]$ with respect to ν. Then*

$$\max_{p\in\mathcal{S}_{\leqslant D}} \frac{\left(\mathbb{E}_\mu p(Y)\right)^2}{\mathbb{E}_\nu p^2(Y)} = \sum_{i=1}^N \left(\mathop{\mathbb{E}}_{\mu} \psi_i\right)^2.$$

Proof. For any $p \in \mathcal{S}_{\leqslant D}$

$$\mathop{\mathbb{E}}_{\mu} p(Y) = \mathop{\mathbb{E}}_{\mu} \sum_{i=1}^N p_i\psi_i(Y) = \sum_{i=1}^N p_i \mathop{\mathbb{E}}_{\mu} \psi_i(Y) \leqslant \left(\sum_{i=1}^N p_i^2\right)^{1/2} \left(\sum_{i=1}^N \left(\mathop{\mathbb{E}}_{\mu} \psi_i(Y)\right)^2\right)^{1/2}.$$

Since the system $\{\psi_i(Y)\}_{i=1}^N$ is orthonormal with respect to ν,

$$\mathop{\mathbb{E}}_{\nu} p^2(Y) = \sum_{i=1}^N p_i^2\,.$$

Hence we get

$$\frac{\mathbb{E}_\mu p(Y)}{\left(\mathbb{E}_\nu p^2(Y)\right)^{\frac{1}{2}}} \leqslant \left[\sum_{i=1}^N \left(\mathop{\mathbb{E}}_{\mu} \psi_i\right)^2\right]^{1/2}.$$

Note that the polynomial $\sum_{i=1}^N \mathbb{E}_{Y'\sim\mu}\left[\psi_i(Y')\right]\psi_i(Y)$ maximizes the ratio. □

From now on we assume that the distribution ν is Gaussian. In this case a useful orthonormal basis in $\mathbb{R}[Y]_{\leqslant D}$ is the system of Hermite polynomials.

To work with Hermite polynomials we introduce some useful notation. For a multi-index α over $[n] \times [d]$, let $I_\alpha := \{i \in [n] : (i,j) \in \alpha \text{ for some } j \in [d]\}$ and similarly $J_\alpha := \{j \in [d] : (i,j) \in \alpha \text{ for some } i \in [n]\}$. For $j \in [d]$, let $I_{\alpha,j} := \{i \in [n] : (i,j) \in \alpha\}$, and similarly let $J_{\alpha,i} := \{j \in [d] : (i,j) \in \alpha\}$. We will use the notation $\alpha! := \prod_{(i,j)\in\alpha} \alpha_{ij}!$ and for a matrix $X \in \mathbb{R}^{n\times d}$, $X^\alpha := \prod_{(i,j)\in\alpha} X_{ij}^{\alpha_{ij}}$. Note that every multi-index α over $[n] \times [d]$ can be represented as a bipartite multigraph $G_\alpha = (I_\alpha \bigcup J_\alpha, E_\alpha)$ such that each edge $\{i,j\}$ has multiplicity α_{ij}. In this representation the set $J_{\alpha,i}$ corresponds to the neighborhood of the vertex i and the set $I_{\alpha,j}$ corresponds to the neighborhood of j. If α is multilinear, G_α is just a graph (i.e. multiplicity of each edge is 1).

For a multi-index α over $[n] \times [d]$ the corresponding Hermite polynomial is

$$H_\alpha(Y) = \prod_{j\in J_\alpha} \prod_{i\in I_{\alpha,j}} H_{\alpha_{ij}}(Y_{ij}) \,,$$

where H_l for $l \in \mathbb{Z}$ is a degree l one variable Hermite polynomial, defined as follows

$$H_l(x) = \sum_{\substack{0\leqslant r\leqslant l \\ l-r \text{ is even}}} \left(-\frac{1}{2}\right)^{\frac{l-r}{2}} \frac{1}{r!\left(\frac{l-r}{2}\right)!}\, x^r \,.$$

Note that $H_\emptyset(Y) = 1$. Hence by applying Proposition 3.38 to the subspace of polynomials such that $\mathbb{E}_\nu\, p(Y) = 0$, we get

Corollary 3.39. *Let ν be Gaussian. Suppose that the distribution μ is so that any polynomial of degree at most D is μ-integrable. Then*

$$\max_{p\in\mathbb{R}[Y]_{\leqslant D}} \frac{\left(\mathbb{E}_\mu\, p(Y) - \mathbb{E}_\nu\, p(Y)\right)^2}{\mathbb{V}_\nu\, p(Y)} = \sum_{0<|\alpha|\leqslant D} \left(\mathop{\mathbb{E}}_{\mu} H_\alpha(Y)\right)^2 .$$

Denote by $\mathcal{M}_{\leqslant D}$ the space of multilinear polynomials of degree at most D (we do not include constant polynomials in $\mathcal{M}_{\leqslant D}$). Note that multilinear Hermite polynomials H_α (which correspond to multilinear multiindices α) are exactly

$$H_\alpha(Y) = \prod_{j\in J_\alpha} \prod_{i\in I_{\alpha,j}} y_{ij} \,.$$

They form a basis in the space $\mathcal{M}_{\leqslant D}$ (for $0 < |\alpha| \leqslant D$). Let's denote $\mathcal{HM}_{\leqslant D} := \mathcal{H}_{\leqslant D} \bigcap \mathcal{M}_{\leqslant D}$. Applying Proposition 3.38 to the space $\mathcal{M}_{\leqslant D}$ we get

Corollary 3.40. *Let ν be Gaussian. Suppose that the distribution μ is so that any polynomial of degree at most D is μ-integrable. Then*

$$\max_{p\in\mathcal{M}_{\leqslant D}} \frac{\left(\mathbb{E}_\mu\, p(Y) - \mathbb{E}_\nu\, p(Y)\right)^2}{\mathbb{V}_\nu\, p(Y)} = \max_{p\in\mathcal{M}_{\leqslant D}} \frac{\left(\mathbb{E}_\mu\, p(Y)\right)^2}{\mathbb{E}_\nu\, p^2(Y)} = \sum_{H_\alpha(Y)\in\mathcal{HM}_{\leqslant D}} \left(\mathop{\mathbb{E}}_{\mu} H_\alpha(Y)\right)^2 .$$

Hence the key part of proving lower bounds for low degree polynomial estimators is bounding $\mathbb{E}_\mu H_\alpha(Y)$.

3.4.2 Almost Gaussian vector in random subspace

We can now tackle Theorem 3.3. To prove the Theorem, we will show that in the presence of adversarial corruptions, whenever $t \cdot \left(\frac{d}{k}\right)^{1/t} \gtrsim n^{0.499}$ and $d \geqslant \tilde{\Omega}\left(n^t t^t\right)$,, so that the degree t SoS Algorithm 3.29 outperforms other known algorithms, no multilinear polynomial of degree $\lesssim n^{0.001}$ can obtain similar guarantees unless $d \gtrsim \tilde{\Omega}\left(\frac{n}{\ln^2 t}\right)^t$.

As outlined above, we design a specific distinguishing problem. In order to prove a lower bound in the presence of adversarial corruptions, we need to carefully chose the adversarial matrix.

Problem 3.41. (Almost-Gaussian vector in a random subspace) Given a matrix Y in $\mathbb{R}^{n\times d}$, decide whether:

H_0: $Y = W$ where $W \sim N(0,1)^{n\times d}$ is a standard Gaussian matrix.

H_1: $Y = \lambda u\tilde{v}^\mathsf{T} + W + E$, where $W \sim N(0,1)^{n\times d}$ is a standard Gaussian matrix, $u \in \mathbb{R}^n$ is a unit vector with i.i.d. coordinates that take values $\pm 1/\sqrt{n}$ with probability $1/2$ each, and $\tilde{v} \in \mathbb{R}^d$ is a vector with i.i.d. coordinates that take values

$$v_i = \begin{cases} -1 & \text{with probability } \delta/2\,, \\ 1 & \text{with probability } \delta/2\,, \\ 0 & \text{otherwise,} \end{cases}$$

for some $\delta \in [0,1]$. Furthermore $E = u\left(v' - W^\mathsf{T}u\right)^\mathsf{T}$, where v' is sampled according to the following distribution. Let $s \geqslant 0$ be the largest even number such that $\delta\lambda^s \leqslant 2^{-10s}$. For all $j \in [d]$,

- if $\tilde{v}_j \neq 0$, then $v'_j = 0$;
- otherwise v'_j is sampled from the distribution η that has finite support $\operatorname{supp}(\eta) \subseteq [-10s, 10s]$ and moments:

$$\mathop{\mathbb{E}}_{x\sim\eta} x^r = \begin{cases} \frac{(r-1)!!-\lambda^r\delta}{1-\delta} & \text{if } 0 \leqslant r \leqslant s \text{ and even} \\ \leqslant (10s)^r & \text{if } r \geqslant s+2 \text{ and even} \\ 0 & \text{if } r \text{ is odd.} \end{cases}$$

Proposition A.14 shows that if $\delta\lambda^s \leqslant 2^{-10s}$, then such η exists. Note that for $s = 0$ the condition $\delta\lambda^s \leqslant 2^{-10s}$ is always satisfied, so s in the problem description is well-defined. Also note that if $s = 0$, v' is just a zero vector.

If $\delta d \to \infty$, then with high probability $\tilde{v}$ is $\delta d(1-o(1))$-sparse. So $\lambda u \tilde{v}^{\mathsf{T}} = \sqrt{\beta} u_0 v_0^{\mathsf{T}}$ for k-sparse unit vector $v_0 = \frac{1}{\|\tilde{v}\|}\tilde{v}$ (where $k := \delta d(1-o(1))$), $\beta = \frac{k\lambda^2}{n}(1+o(1))$ and $u_0 = \sqrt{n}u$.

We will use the notation $v = \lambda\tilde{v} + v'$. Note that the coordinates of v are independent, have Gaussian moments up to s, and with high probability v has at least $\delta d(1-o(1))$ coordinates $v_j \in \{\pm\lambda\}$.

Geometric description. The planted distribution can be also described in geometric terms, where the problem becomes that of distinguishing between a subspace spanned by independent Gaussian vectors or a subspace spanned by independent Gaussian vectors and the planted vector v.

The construction is the following: at first we sample a signal vector $v \in \mathbb{R}^d$ that has at least $\delta d(1-o(1))$ coordinates with absolute values at least λ (using the construction described above). Then we sample $n-1$ i.i.d. standard Gaussian vectors $\tilde{w}_1, \dots, \tilde{w}_{n-1} \in \mathbb{R}^d$, and perform a random rotation $U \in \mathbb{R}^{n\times n}$ with first column vector u (such that U is independent of $v, \tilde{w}_1, \dots, \tilde{w}_{n-1}$) on $v, \tilde{w}_1, \dots, \tilde{w}_{n-1}$. That is,

$$Y = U \cdot \begin{pmatrix} v^T \\ \tilde{w}_1^T \\ \vdots \\ \tilde{w}_{n-1}^T \end{pmatrix}.$$

This formulation is equivalent to the one described above. Indeed,

$$Y = uv^T + \sum_{i=1}^{n-1} u_i \tilde{w}_i^T,$$

where $u, u_1, \dots, u_{n-1}$ are the columns of U. Note that $\sum_{i=1}^{n-1} u_i \tilde{w}_i^T$ is distributed as a standard (singular) Gaussian supported in the hyperplane orthogonal to u, and $(\mathrm{Id} - uu^T)W$ is also a standard Gaussian supported in the same hyperplane.

The theorem below provides a lower bound for the Problem 3.41. The proof is in section 3.4.2.1.

Theorem 3.42. *Suppose that $n \leqslant d$, $1 \leqslant D \leqslant n^{0.33}$, $0 < \delta < 1$, $k = \delta d$, $\lambda \geqslant 2$, and let $s \geqslant 2$ be the maximal even number such that $\delta\lambda^s \leqslant 2^{-10s}$, and $t = s/2 + 1$. Let ν and μ denote respectively the null and planted distribution (with parameters δ, λ, s) of Problem 3.41. Suppose that $\lambda \geqslant 1000\sqrt{t \ln t}$ and that $\lambda^4 D t^2 \ln^2 t = o\left(n \cdot \left(\log^2\left(\frac{d}{k^2}\right) + 1\right)\right)$ as $n \to \infty$. If*

$$d = o\left(\frac{1}{\lambda^4} \cdot \left(\frac{n}{C \cdot \ln^2 t \cdot D}\right)^t\right)$$

as $n \to \infty$ (for some constant C that does not depend on n, d, δ, λ, s and D), then for any non-constant multilinear polynomial $p : \mathbb{R}^{n\times d} \to \mathbb{R}$ of degree at most D,

$$\frac{\left(\mathbb{E}_\mu p(Y) - \mathbb{E}_\nu p(Y)\right)^2}{\mathbb{V}_\nu p(Y)} \to 0\,,$$

as $n \to \infty$.

Let's try to illustrate the meaning of Theorem 3.42. If $\lambda \geqslant B\sqrt{\log d}$ for sufficiently large B, then δ is so that $\delta\lambda^s = 2^{-10s}$ for an even constant s and $\delta d \to \infty$. Here Algorithm 3.29 can distinguish between the null and the planted distribution if $d \gtrsim n^t \log^t(n) t^t$ in polynomial time. Indeed, in this case with probability at least 0.99 the algorithm 3.29 outputs $\hat{v}$ such that

$$1 - \langle \hat{v}, v_0\rangle^2 \lesssim \frac{1}{\lambda^2}\left(t \cdot \left(\frac{1}{\delta}\right)^{1/t} + \|E\|_{1\to 2}^2\right) \leqslant t \cdot \left(\frac{1}{\lambda^{s+1}\delta}\right)^{1/t} + \left(\frac{\|E\|_{1\to 2}}{\lambda}\right)^2 = \frac{2^{20}t}{\lambda^{1/t}} + \left(\frac{\|E\|_{1\to 2}}{\lambda}\right)^2.$$

The first term tends to 0 and $\|E\|_{1\to 2}$ can be bounded as follows:

$$\|E\|_{1\to 2} \leqslant \|u(v')^\mathsf{T}\|_{1\to 2} + \|uu^\mathsf{T}W\|_{1\to 2} \leqslant \max_{1\leqslant i\leqslant d} v'_i + \max_{1\leqslant i\leqslant d} \left(u^\mathsf{T}W\right)_i \lesssim s + \sqrt{\log d}\,,$$

since $u^\mathsf{T}W$ is a standard Gaussian vector. Hence for sufficiently large B, $\langle \hat{v}, v_0\rangle \geqslant 0.99$.

If in addition $D \leqslant n^{0.001}$ and $\lambda \leqslant n^{0.24}$, then the conditions of Theorem 3.42 are satisfied. Hence in this case for $d \leqslant n^{0.999t-1}$ no multilinear polynomial of degree at most $n^{0.001}$ can distinguish between the planted and the null distribution as $n \to \infty$. Furthermore note that if $\lambda^4 \gtrsim n \log d$, then Diagonal thresholding can distinguish between the planted and the null distribution in polynomial time (even if $d \ll n^{0.999t-1}$). Finally, it easy to see that exhaustive search works as long as $\lambda \gtrsim \sqrt{\log d/k}$.

3.4.2.1 Proof of indistinguishability

The proof of Theorem 3.42 relies on key lemmata which we provide below. The proof itself is then presented at the end of the section.

Lemma 3.43. *Let α be a multiindex over $[n]\times[d]$ such that $H_\alpha(Y) \in \mathcal{HM}_{\leqslant D}$. Then*

$$\mathop{\mathbb{E}}_{\mu} H_\alpha(Y) = \mathbb{E}\left(\prod_{i\in I_\alpha} \sigma_i^{|J_{\alpha,i}|}\right) \prod_{j\in J_\alpha} \mathbb{E}\left[\prod_{i\in I_{\alpha,j}} \left(z_{ij} + \tfrac{1}{\sqrt{n}}(v_j - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_{lj})\right)\right],$$

where $j \in [d]$, $i \in [n]$, $\sigma_i := \sqrt{n}u_i$ and $z_{ij} := \sigma_i w_{ij}$.

Proof. We drop the subscript α for the exposition of the proof.

$$\mathop{\mathbb{E}}_{\mu} H_\alpha(Y) = \mathop{\mathbb{E}}_{\mu} \prod_{j\in J}\prod_{i\in I_j} y_{ij}$$

$$
\begin{aligned}
&= \mathbb{E} \prod_{j\in J}\prod_{i\in I}\left[w_{ij} + u_i\left(v_j - \langle u, w_j\rangle\right)\right] \\
&= \mathbb{E} \prod_{j\in J}\prod_{i\in I_j}\left[w_{ij} + \sigma_i\left(\frac{1}{\sqrt{n}}v_j - \frac{1}{n}\sum_{l\in[n]} z_{lj}\right)\right] \\
&= \mathbb{E} \prod_{j\in J}\prod_{i\in I_j}\sigma_i\left(z_{ij} + \frac{1}{\sqrt{n}}v_j - \frac{1}{n}\sum_{l\in[n]} z_{lj}\right) && (\text{as } \sigma_i w_{ij} = \tfrac{w_{ij}}{\sigma_i}) \\
&= \mathbb{E}\left[\left(\prod_{j\in J}\prod_{i\in I_j}\sigma_i\right)\prod_{j\in J}\prod_{i\in I_j}\left(z_{ij} + \frac{1}{\sqrt{n}}v_j - \frac{1}{n}\sum_{l\in[n]} z_{lj}\right)\right] \\
&= \mathbb{E}\left(\prod_{j\in J}\prod_{i\in I_j}\sigma_i\right)\mathbb{E}\left[\prod_{j\in J}\prod_{i\in I_j}\left(z_{ij} + \frac{1}{\sqrt{n}}(v_j - \frac{1}{\sqrt{n}}\sum_{l\in[n]} z_{lj})\right)\right] \\
&= \mathbb{E}\left(\prod_{i\in I}\sigma_i^{|J_i|}\right)\prod_{j\in J}\mathbb{E}\left[\prod_{i\in I_j}\left(z_{ij} + \frac{1}{\sqrt{n}}(v_j - \frac{1}{\sqrt{n}}\sum_{l\in[n]} z_{lj})\right)\right].
\end{aligned}
$$

□

An immediate consequence of Lemma 3.43 is the following statement:

Corollary 3.44. *Let α be a multiindex over $[n]\times[d]$ such that $H_\alpha(Y) \in \mathcal{H}_{\leqslant D}$. If there exists $j \in J_\alpha$ (or $i \in I_\alpha$) such that $|I_{\alpha,j}|$ (respectively, $|J_{\alpha,i}|$) is odd, then $\mathbb{E}_\mu H_\alpha(Y) = 0$.*

In the following lemma we use the fact that first s moments of coordinates of v coincide with Gaussian moments.

Lemma 3.45. *Let s be the parameter of the planted distribution, let α be a multiindex over $[n]\times[d]$. Suppose that there exists $j_0 \in J_\alpha$ such that $|I_{j_0}| \leqslant s$. Then $\mathbb{E}_\mu H_\alpha(Y) = 0$.*

Proof. For simplicity we will the subscript α. If $\mathbb{E}\left(\prod_{i\in I}\sigma_i^{|J_i|}\right) = 0$, the statement is obviously true. Assume that $\mathbb{E}\left(\prod_{i\in I}\sigma_i^{|J_i|}\right) = 1$ (notice that this expectation can be only 0 or 1). Thus

$$
\begin{aligned}
\mathop{\mathbb{E}}_{\mu} H_\alpha(Y) &= \prod_{j\in J}\mathbb{E}\prod_{i\in I_j}\left(z_{ij} + \frac{1}{\sqrt{n}}(v_j - \frac{1}{\sqrt{n}}\sum_{l\in[n]} z_{lj})\right) \\
&= \mathbb{E}\prod_{i\in I_{j_0}}\left(z_{ij_0} + \frac{1}{\sqrt{n}}(v_{j_0} - \frac{1}{\sqrt{n}}\sum_{l\in[n]} z_{lj_0})\right)\cdot\prod_{j\in J\setminus\{j_0\}}\mathbb{E}\prod_{i\in I_j}\left(z_{ij} + \frac{1}{\sqrt{n}}(v_j - \frac{1}{\sqrt{n}}\sum_{l\in[n]} z_{lj})\right).
\end{aligned}
$$

Since first s moments of v_{j_0} coincide with Gaussian moments,

$$\mathbb{E}\prod_{i\in I_{j_0}}\left(z_{ij_0}+\tfrac{1}{\sqrt{n}}(v_{j_0}-\tfrac{1}{\sqrt{n}}\sum_{l\in[n]}z_{lj_0})\right)=\mathbb{E}\prod_{i\in I_{j_0}}\left(z_{ij_0}+\tfrac{1}{\sqrt{n}}(\zeta-\tfrac{1}{\sqrt{n}}\sum_{l\in[n]}z_{lj_0})\right),$$

where ζ is a standard Gaussian variable that is independent from all z_{ij_0}. Let $\xi_i = z_{ij_0}+\frac{1}{\sqrt{n}}(\zeta-\frac{1}{\sqrt{n}}\sum_{l\in[n]}z_{lj_0})$. Let's show that $\xi\sim N(0,\mathrm{Id}_n)$. $\xi\in\mathbb{R}^n$ is a linear transformation of $\zeta, z_{1j_0},\dots,z_{nj_0}$:

$$\xi = A\begin{pmatrix}\zeta\\ z_{1j_0}\\ \vdots\\ z_{nj_0}\end{pmatrix},$$

where A is an $n\times(n+1)$ matrix with rows $A_i^{\mathsf{T}}=(\frac{1}{\sqrt{n}},\frac{1}{n},\dots,\frac{1}{n},\underbrace{(1-\frac{1}{n})}_{i+1},\frac{1}{n},\dots,\frac{1}{n})$. The rows of A are orthonormal: for all $i\in[n]$

$$(AA^{\mathsf{T}})_{ii}=\frac{1}{n}+(1-\frac{1}{n})^2+\frac{n-1}{n^2}=1-\frac{2}{n}+\frac{1}{n^2}+\frac{1}{n}+\frac{1}{n}-\frac{1}{n^2}=1\,,$$

and for all different $i,l\in[n]$

$$(AA^{\mathsf{T}})_{il}=\frac{1}{n}-\frac{2}{n}(1-\frac{1}{n})+\frac{n-2}{n^2}=\frac{1}{n}-\frac{2}{n}+\frac{2}{n^2}+\frac{1}{n}-\frac{2}{n^2}=0\,.$$

Hence $AA^{\mathsf{T}}=\mathrm{Id}_n$ and $\xi\sim N(0,\mathrm{Id}_n)$. Therefore,

$$\mathbb{E}\prod_{i\in I_{j_0}}\left(z_{ij_0}+\tfrac{1}{\sqrt{n}}(v_{j_0}-\tfrac{1}{\sqrt{n}}\sum_{l\in[n]}z_{lj_0})\right)=\mathbb{E}\prod_{i\in I_{j_0}}\xi_i=0\,.$$

□

Lemma 3.46. *Let s,δ,λ be the same as in the statement of Theorem 3.42. Let $j\in[d], I_j\subseteq[n]$ with even cardinality $|I_j|>s$. Then, if $|I_j|\leqslant\frac{\lambda^2}{100}$,*

$$\mathbb{E}\left[\prod_{i\in I_j}\left(z_{ij}+\tfrac{1}{\sqrt{n}}(v_j-\tfrac{1}{\sqrt{n}}\sum_{l\in[n]}z_l)\right)\right]\leqslant\delta\left(\frac{2^{20}\cdot s\cdot\lambda}{\sqrt{n}}\right)^{|I_j|},$$

and if $|I_j|>\frac{\lambda^2}{100}$,

$$\mathbb{E}\left[\prod_{i\in I_j}\left(z_{ij}+\tfrac{1}{\sqrt{n}}(v_j-\tfrac{1}{\sqrt{n}}\sum_{l\in[n]}z_l)\right)\right]\leqslant\left(\frac{100\sqrt{|I_j|}}{\sqrt{n}}\right)^{|I_j|}.$$

Proof. We drop the subscript j to simplify the notation (in particular, in this proof we denote v_j by v). By symmetry of the Gaussian distribution, opening up the product we see that in order for a monomial to have non-zero expectation, for any left end term z_i there must be a corresponding right term $\frac{1}{\sqrt{n}}(v - \frac{1}{\sqrt{n}} \sum_{l\in[n]} z_l)$. Hence:

$$\mathbb{E}\left[\prod_{i\in I}\left(z_i + \tfrac{1}{\sqrt{n}}(v - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_l)\right)\right] = \sum_{r=0}^{|I|/2}\binom{|I|}{2r}\binom{2r}{r}\mathbb{E}\left[\left(\prod_{i\in[r]} z_i\right)\frac{1}{n^{|I|/2-r/2}}\left(v - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_l\right)^{|I|-r}\right]$$
$$= \frac{1}{n^{|I|/2}}\sum_{r=0}^{|I|/2}\binom{|I|}{2r}\binom{2r}{r}\mathbb{E}\left[\left(\prod_{i\in[r]} z_i^2\right)\left(v - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_l\right)^{|I|-2r}\right].$$

Since v is symmetric:

$$\mathbb{E}\left[\left(\prod_{i\in[r]} z_i^2\right)\left(v - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_l\right)^{|I|-2r}\right] = \sum_{m=0}^{|I|/2-r}\mathbb{E}[v^{|I|-2r-2m}]\,\mathbb{E}\left[\left(\prod_{i\in[r]} z_i^2\right)\cdot\left(\frac{1}{\sqrt{n}}\sum_{l\in[n]} z_l\right)^{2m}\right].$$

By Cauchy–Schwarz:

$$\mathbb{E}\left[\left(\prod_{i\in[r]} z_i^2\right)\left(\frac{1}{\sqrt{n}}\sum_{l\in[n]} z_l\right)^{2m}\right] \leqslant \left(\mathbb{E}\prod_{i\in[r]} z_i^4\right)^{1/2}\left(\mathbb{E}\left(\frac{1}{\sqrt{n}}\sum_{l\in[n]} z_l\right)^{4m}\right)^{1/2} \leqslant 3^{r/2}\cdot(2m)^m.$$

Hence,

$$\mathbb{E}\left[\left(\prod_{i\in[r]} z_i^2\right)\left(v - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_l\right)^{|I|-2r}\right] \leqslant \sum_{m=0}^{|I|/2-r}\mathbb{E}[v^{|I|-2r-2k}]\cdot 3^{r/2}\cdot(2m)^m$$
$$\leqslant 3^{r/2}\sum_{m=0}^{|I|/2-r}\left(\delta\lambda^{|I|-2r-2m} + (10\sqrt{s\ln s})^{|I|-2r-2m}\right)\cdot(2m)^m$$
$$\leqslant 3^{|I|/4}\sum_{m=0}^{|I|/2}\left(\delta\lambda^{|I|-2m} + (10\sqrt{s\ln s})^{|I|-2m}\right)\cdot(2m)^m.$$

Let $M = \max\{\delta\lambda^{|I|}, |I|^{|I|/2}, (10\sqrt{s\ln s})^{|I|}\}$. Thus $2M \geqslant \left(\delta\lambda^{|I|-2m} + (10\sqrt{s\ln s})^{|I|-2m}\right)\cdot(2m)^m$. We get:

$$\mathbb{E}\left[\prod_{i\in I}\left(z_i + \tfrac{1}{\sqrt{n}}(v - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_l)\right)\right] \leqslant \frac{1}{n^{|I|/2}}\sum_{r=0}^{|I|/2}\binom{|I|}{2r}\binom{2r}{r}3^{|I|/4}\cdot|I|\cdot 2M$$

$$\leqslant \frac{1}{n^{|I|/2}} \cdot 2^{|I|} \cdot 2^{|I|} \cdot 3^{|I|/4} \cdot 2^{|I|/2} \cdot M$$
$$\leqslant \left(\frac{10}{\sqrt{n}}\right)^{|I|} \cdot M$$

Consider the case $|I| > \frac{\lambda^2}{100}$. In this case, $M \leqslant 10^{|I|} \cdot |I|^{|I|/2}$. Hence

$$\mathbb{E}\left[\prod_{i\in I}\left(z_i + \tfrac{1}{\sqrt{n}}(v - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_l)\right)\right] \leqslant \left(\frac{100\sqrt{|I|}}{\sqrt{n}}\right)^{|I|}.$$

Now consider the case $|I| \leqslant \frac{\lambda^2}{100}$. If $|I| \geqslant 10s$, then $\delta\lambda^{|I|} \geqslant (\lambda/2)^{|I|-s} \geqslant |I|^{|I|/2}$. Indeed, the inequality holds if $|I| \leqslant (\lambda/2)^{1.8}$, and if $|I| > (\lambda/2)^{1.8}$, then $(\lambda/2)^{s/|I|}\sqrt{|I|}$ is monotone in I, so

$$(\lambda/2)^{s/|I|}\sqrt{|I|} \leqslant 0.1 \cdot \lambda \cdot (\lambda/2)^{100s/\lambda^2} \leqslant 0.1 \cdot \lambda \cdot (\lambda/2)^{1/\ln\lambda} \leqslant \frac{1}{2}\lambda\,,$$

since $\lambda^2 \geqslant 10000s\ln s$. If $|I| < 10s$, then $|I|^{|I|/2} < (10\sqrt{s\ln s})^{|I|}$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[\prod_{i\in I}\left(z_i + \tfrac{1}{\sqrt{n}}(v - \tfrac{1}{\sqrt{n}}\sum_{l\in[n]} z_l)\right)\right] &\leqslant \left(\frac{10}{\sqrt{n}}\right)^{|I|} \cdot \max\{\delta\lambda^{|I|}, (10\sqrt{s\ln s})^{|I|}\} \\
&\leqslant \left(\frac{10}{\sqrt{n}}\right)^{|I|} \delta\lambda^{s+2} \max\{\lambda^{|I|-(s+2)}, \frac{1}{\delta\lambda^{s+2}}(10\sqrt{s\ln s})^{|I|}\} \\
&\leqslant \left(\frac{10}{\sqrt{n}}\right)^{|I|} \delta\lambda^{s+2} \max\{\lambda^{|I|-(s+2)}, 2^{10s}(10\sqrt{s\ln s})^{|I|}\} \\
&\leqslant \delta\left(\frac{2^{20}\cdot\sqrt{s\ln s}\cdot\lambda}{\sqrt{n}}\right)^{|I|}.
\end{aligned}$$

□

We are now ready to prove Theorem 3.42.

Proof of Theorem 3.42. For all positive integers A, B, B', E and E' consider the set $\mathcal{G}_s(A,B,B',E,E')$ of bipartite graphs G_α such that $|I_\alpha| = A$, $|J_\alpha| = B$ and $|\alpha| = E$, $B' = \left|\{j \in J_\alpha \mid |I_j| \leqslant \frac{\lambda^2}{100}\}\right|$, E' is a number of edges adjacent to $\{j \in J_\alpha \mid |I_j| \leqslant \frac{\lambda^2}{100}\}$, and all vertices of G_α have even degree strictly greater than s. Let $B'' = B - B'$ and $E'' = E - E'$.

By lemma 3.46,

$$\sum_{H_\alpha(Y)\in\mathcal{H}\mathcal{M}_{\leqslant D}} \left(\mathop{\mathbb{E}}_{\mu} H_\alpha(Y)\right)^2 \leqslant \sum_{2(s+2)\leqslant E\leqslant D} \sum_{\substack{A,B,B',E' \\ \mathcal{G}_s(A,B,B',E,E')\neq\emptyset}} \binom{n}{A}\binom{d}{B}\frac{(AB)^E}{E!} \cdot \delta^{2B'}\left(\frac{2^{20}\cdot\sqrt{s\ln s}\cdot\lambda}{\sqrt{n}}\right)^{2E'}\left(\frac{100\sqrt{E}}{\sqrt{n}}\right)^{2E''}$$

$$\leqslant \sum_{2(s+2)\leqslant E\leqslant D} \sum_{\substack{A,B,B',E' \\ \mathcal{G}_s(A,B,B',E,E')\neq\emptyset}} \left(\frac{en}{A}\right)^A \left(\frac{eAB}{E}\right)^E \cdot d^B \left(\delta^2\right)^{B'} \left(\frac{2^{20}\cdot\sqrt{s\ln s}\cdot\lambda}{\sqrt{n}}\right)^{2E'} \left(\frac{100\sqrt{E}}{\sqrt{n}}\right)^{2E''}.$$

Since $A \leqslant D/2 = o(n)$, $\left(\frac{en}{A}\right)^A$ is monotone in A. Also notice that if $\mathcal{G}_s(A,B,B',E,E') \neq \emptyset$, $B' \leqslant E'/(s+2)$ and $B'' \leqslant 100\cdot E''/\lambda^2 \leqslant E''/(s+2)$. Let $\phi(B',B'',E',E'')$ be an zero-one indicator that is one if and only if there exists A such that $\mathcal{G}_s(A, B'+B'', B', E'+E'', E') \neq \emptyset$.

Consider the case $\delta^2 d \geqslant 1$. Assume that $d = o\left(\frac{1}{\lambda^4}\cdot\left(\frac{n}{2^{120}\ln^2 sD}\right)^{(s+2)/2}\right)$. Since $D \leqslant n^{0.33}$ and $\lambda^2 \geqslant 100000 s\ln s$, $d = o\left(\left(\frac{n}{10^{20}D^3}\right)^{\lambda^2/200}\right)$. Hence

$$\begin{aligned}
\sum_{H_\alpha(Y)\in\mathcal{HM}_{\leqslant D}} \left(\mathop{\mathbb{E}}_{\mu} H_\alpha(Y)\right)^2 &\leqslant \sum_{0\leqslant B',B''\leqslant D/2} \sum_{0\leqslant E',E''\leqslant D} \phi(B',B'',E',E'')\left(\frac{n}{E}\right)^{E/2}\left(\frac{E}{s+2}\right)^E \\
&\qquad\cdot \left(\delta^2 d\right)^{B'}\left(\frac{2^{30}\cdot\sqrt{s\ln s}\cdot\lambda}{\sqrt{n}}\right)^{2E'} d^{B''}\left(\frac{10^5\sqrt{E}}{\sqrt{n}}\right)^{2E''} \\
&\leqslant \sum_{0\leqslant B',B''\leqslant D/2} \sum_{0\leqslant E',E''\leqslant D} \phi(B',B'',E',E'')\left(\delta^2 d\right)^{B'}\left(\frac{2^{120}\lambda^4 D\ln^2 s}{n}\right)^{E'/2} \\
&\qquad\cdot d^{B''}\left(\frac{10^{20}D^3}{n}\right)^{E''/2} \\
&\leqslant 2\sum_{B'=1}^{\infty}\left(\delta^2 d\left(\frac{2^{120}\lambda^4 D\ln^2 s}{n}\right)^{(s+2)/2}\right)^{B'} + 2\sum_{B''=1}^{\infty}\left(d\left(\frac{10^{20}D^3}{n}\right)^{\lambda^2/200}\right)^{B''} \\
&\leqslant 2\sum_{B'=1}^{\infty}\left(\lambda^4 d\left(\frac{2^{100}D\ln^2 s}{n}\right)^{(s+2)/2}\right)^{B'} + o(1) \\
&\leqslant o(1)\,.
\end{aligned}$$

Now condiser the case $\delta^2 d < 1$. Since $\lambda \geqslant 2^{10}$, $\lambda^{2s+2} > \frac{1}{\delta}$ and

$$(2s+2)\ln\lambda > \ln\left(\frac{1}{\delta}\right) \geqslant \ln\left(\frac{1}{\delta}\right) + \ln\left(\frac{1}{\delta d}\right) = \ln\left(\frac{1}{\delta^2 d}\right).$$

Since $\lambda^2 > 100000 s\ln s$, $\lambda^2/100 \geqslant \ln\left(\frac{1}{\delta^2 d}\right)$ and $B'' \leqslant E/|\ln(\delta^2 d)|$. Let $M = \max\{B', B''\}$.

Recall that $\lambda^4 D s^2 \ln^2 s = o\Big(n \log^2\big(\delta^2 d\big)\Big)$. It follows that

$$\begin{aligned}
\sum_{H_\alpha(Y)\in\mathcal{HM}_{\leqslant D}} \left(\mathop{\mathbb{E}}_{\mu} H_\alpha(Y)\right)^2 &\leqslant \sum_{0\leqslant B',B''\leqslant D/2} \sum_{0\leqslant E',E''\leqslant D} \phi(B',B'',E',E'')\left(\frac{n}{E}\right)^{E/2} M^E \\
&\qquad \cdot \big(\delta^2 d\big)^{B'} \left(\frac{2^{30}\cdot\sqrt{s\ln s}\cdot\lambda}{\sqrt{n}}\right)^{2E'} d^{B''}\left(\frac{10^5\sqrt{E}}{\sqrt{n}}\right)^{2E''} \\
&\leqslant \sum_{0\leqslant B',B''\leqslant D/2} \sum_{0\leqslant E',E''\leqslant D} \phi(B',B'',E',E'')\left(\big(\delta^2 d\big)^{\frac{2B'}{E'}} M\cdot\frac{2^{120}\lambda^4 s^2\ln^2 s}{n}\right)^{E'/2} \\
&\qquad \cdot d^{B''}\left(\frac{10^{20}D^3}{n}\right)^{E''/2} \\
&\leqslant \sum_{0\leqslant B''\leqslant D/2} \sum_{0\leqslant E',E''\leqslant D} \sum_{0\leqslant B'\leqslant E'} \phi(B',B'',E',E'')\left(\frac{2^{120}\lambda^4 D s^2\ln^2 s}{\ln^2(\delta^2 d)n}\right)^{E'/2} \\
&\qquad \cdot d^{B''}\left(\frac{10^{20}D^3}{n}\right)^{E''/2} \\
&\leqslant \sum_{E'=1}^{\infty}\left(\frac{2^{130}\lambda^4 D s^2\ln^2 s}{\ln^2(\delta^2 d)n}\right)^{E'/2} + 2\sum_{B''=1}^{\infty}\left(d\left(\frac{10^{20}D^3}{n}\right)^{\lambda^2/200}\right)^{B''} \\
&\leqslant o(1)\,.
\end{aligned}$$

By Corollary 3.39, we get the desired conclusion. □

Chapter 4

Stochastic block models with edge corruptions

In tis chapter we prove Theorem 1.2, showing that by using sophisticated algorithmic techniques it is possible to approach the Kesten-Stigum threshold even in the presence of malicious corruptions. Consistently with the discussion in Section 1.1.2, the underlying robust algorithm boils down to certifying bounds the Schatten norm of certain matrices associated with the input. We start by restating the model. The notation of the chapter is optimized for its proofs and differs from Chapter 1.

Definition 4.1 (Restatement of the model introduced in Section 1.1.3)**.** The stochastic block model describes the following joint distribution $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$ between a vector x of n binary labels and an n-vertex graph $\mathbf{G}$:

- draw a vector $\mathbf{x} \in \{\pm 1\}^n$ uniformly at random,
- for every pair of distinct vertices $i, j \in [n]$, independently create an edge $\{i, j\}$ in the graph $\mathbf{G}$ with probability $(1 + \frac{\varepsilon}{2} \cdot \mathbf{x}_i \cdot \mathbf{x}_j) \cdot \frac{d}{n}$.

Given a graph $\mathbf{G}$ sampled according to this model, the goal is to recover the (unknown) underlying vector of labels as well as possible.

For distinct vertices $i, j \in [n]$, the edge $\{i, j\}$ is present in $\mathbf{G}$ with probability $(1 + \frac{\varepsilon}{2}) \cdot \frac{d}{n}$ if the vertices have the same label $\mathbf{x}_i = \mathbf{x}_j$ and with probability $(1 - \frac{\varepsilon}{2}) \cdot \frac{d}{n}$ if the vertices have different labels $\mathbf{x}_i \neq \mathbf{x}_j$.

Recall we say that an algorithm achieves *(weak) recovery* for the stochastic block model $\{\mathsf{SBM}_n(d, \varepsilon)\}_{n \in \mathbb{N}}$ if the correlation of the algorithm's output $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ and the underlying vector $\mathbf{x}$ of labels is bounded away from zero as n grows

$$\mathop{\mathbb{E}}_{(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)} \left[\frac{1}{n} |\langle x, \hat{x}(\mathbf{G}) \rangle| \right] \geqslant \Omega_{\varepsilon, d}(1) \,. \tag{4.0.1}$$

(Here, $\Omega_{\varepsilon,d}(1)$ hides a positive number depending on ε and d but independent of n). As already discussed in the opening chapter, weak recovery is possible (also computationally

efficiently) if and only if $d > 4/\varepsilon^2$. Similarly, we say that an algorithm that given a graph $\mathbf{G}$ outputs an estimate $\hat{x}(\mathbf{G})$ for the community labels of $\mathbf{G}$ achieves ρ-robust weak recovery for $\{\mathsf{SBM}_n(d,\varepsilon)\}_{n\in\mathbb{N}}$ if

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)} \min_{G^\circ\in N_\rho(\mathbf{G})} \left[\tfrac{1}{n}|\langle \mathbf{x}, \hat{x}(G^\circ)\rangle|\right] \geqslant \Omega_{d,\varepsilon}(1)\,, \tag{4.0.2}$$

where $N_\rho(\mathbf{G})$ is the set of graphs G° that can be obtained from $\mathbf{G}$ by changing at most a ρ-fraction of its edges[1] (so that $|E(\mathbf{G}) \triangle E(G^\circ)| \leqslant \rho\cdot(|E(\mathbf{G})| + |E(G^\circ)|)$).

The main theorem of the chapter is shown next.

Theorem 4.2 (Restatement of Theorem 1.2)**.** *For every ε, d with $d > 4/\varepsilon^2$, there exists $\rho > 0$ such that ρ-robust weak recovery for $\{\mathsf{SBM}_n(d,\varepsilon)\}_{n\in\mathbb{N}}$ is possible. Moreover, the underlying algorithm runs in polynomial time.*

We present a formal statement in Corollary 4.24.

Organization

The rest of the chapter is organized as follows. In Section 4.1 we introduce the main ideas and techniques developed to prove Theorem 4.2, while also describing the shortcomings of previous approaches. Preliminary notions are discussed in Section 4.2. In Section 4.3 we present our general framework for robust algorithms, we then apply it to the stochastic block model in Section 4.4. In Section 4.5 we prove the probabilistic results needed for the algorithm to succeed. We present most of the technical probabilistic and combinatorial details through Appendix B.1, Appendix B.2, Appendix B.3 and Appendix B.4.

4.1 Techniques

To start explaining the ideas required to prove Theorem 4.2, we briefly discuss related prior approaches for weak-recovery in stochastic block models.

Basic semidefinite programming approach and robust recovery away from the Kesten-Stigum threshold. The following approach based on semidefinite programming is known to have strong robustness properties when the degree parameter d exceeds the KS threshold by a large enough constant factor [GV16].[2]

[1] That is, each G° can be obtained from $\mathbf{G}$ through a sequence of $\rho\cdot|E(\mathbf{G})|$ edits, each consisting of an addition or deletion.

[2] As discussed earlier, in the asymptotic regime $d\to\infty$ better guarantees for this approach are known [MS16]. However, these results have no bearing on constant degree parameters.

Let $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$ with $\varepsilon > 0$[3] and let $\mathbf{Y}$ be its centered adjacency matrix, so that $\mathbf{Y}_{ij} = 1 - d/n$ if $ij \in E(\mathbf{G})$, $\mathbf{Y}_{ii} = 0$, and $\mathbf{Y}_{ij} = -d/n$ otherwise. This matrix satisfies $\mathbb{E}\,\mathbf{Y} = 0$ and, up to a scaling factor, its conditional expectation agrees with $\mathbf{x}\mathbf{x}^\mathsf{T}$ on all off-diagonal entries,

$$\bar{\mathbf{Y}} := \mathbb{E}[\mathbf{Y} \mid \mathbf{x}] = \tfrac{\varepsilon d}{2n} \cdot (\mathbf{x}\mathbf{x}^\mathsf{T} - \mathrm{Id}_n)\,. \tag{4.1.1}$$

This property gives us hope that the maximizer $\hat{\mathbf{x}}$ of $\langle \mathbf{Y}, xx^\mathsf{T} \rangle$ over all $x \in \{\pm 1\}^n$ is correlated with the underlying labeling $\mathbf{x}$.[4] Concretely, the optimal value of this optimization problem is at least the value achieved by the planted labeling $\mathbf{x}$, which is $o(n)$-close to its expectation $\langle \bar{\mathbf{Y}}, \mathbf{x}\mathbf{x}^\mathsf{T} \rangle = \frac{\varepsilon d}{2} \cdot (n-1)$. Hence, the maximizer $\hat{\mathbf{x}}$ satisfies the inequality

$$\tfrac{\varepsilon d}{2} \cdot n - o(n) \leqslant \langle \mathbf{Y}, \hat{\mathbf{x}}\hat{\mathbf{x}}^\mathsf{T} \rangle = \tfrac{\varepsilon d}{2} \cdot n \cdot \left(\tfrac{1}{n^2} \langle \mathbf{x}, \hat{\mathbf{x}} \rangle^2 - \tfrac{1}{n} \right) + \langle \mathbf{Y} - \bar{\mathbf{Y}}, \hat{\mathbf{x}}\hat{\mathbf{x}}^\mathsf{T} \rangle\,.$$

This inequality allows us to conclude that $\hat{\mathbf{x}}$ achieves the desired correlation $\frac{1}{n^2}\langle \mathbf{x}, \hat{\mathbf{x}} \rangle^2 \geqslant \Omega(1)$ as long as we have an upper bound on $\langle \mathbf{Y} - \bar{\mathbf{Y}}, \hat{\mathbf{x}}\hat{\mathbf{x}}^\mathsf{T} \rangle$ smaller than $\frac{\varepsilon d}{2} \cdot n$ by a constant factor. This approach is pursued by [GV16], who proved that with high probability, for some constant factor $C > 1$,

$$\max_{x \in \{\pm 1\}^n} |\langle \mathbf{Y} - \bar{\mathbf{Y}}, xx^\mathsf{T} \rangle| \leqslant C \cdot \sqrt{d} \cdot n\,. \tag{4.1.2}$$

It follows that this estimator achieves weak recovery if $\frac{\varepsilon\sqrt{d}}{2C}$ exceeds 1 by a constant. (Since $C > 1$, this bound fails to approach the Kesten-Stigum threshold.[5])

So far, the discussed approach is not computationally efficient (the underlying optimization problem is NP-hard). To remedy this issue, we consider the following semidefinite programming relaxation of the problem,

$$\text{maximize } \langle \mathbf{Y}, X \rangle \text{ subject to } X \succeq 0\,,\ \forall i\,.\, X_{ii} = 1\,. \tag{4.1.3}$$

We refer to this relaxation as *basic SDP* (which should not be confused with the basic SDP in Chapter 3[6]).[7] To show that its optimal solution $\hat{\mathbf{X}}$ achieves constant correlation $\frac{1}{n^2}\langle \hat{\mathbf{X}}, \mathbf{x}\mathbf{x}^\mathsf{T} \rangle \geqslant \Omega(1)$, we can imitate the previous analysis (as done in [GV16]). The main difference is that we need to upper bound the deviation term $|\langle \mathbf{Y} - \bar{\mathbf{Y}}, X \rangle|$ uniformly over all feasible solutions X to the basic SDP (instead of just over all cut matrices xx^T for $x \in \{\pm 1\}^n$).

[3] We remark that it also makes sense to consider stochastic block models with negative bias parameter ε (sometimes called the anti-ferromagnetic case).

[4] This optimization problem is closely related to the likelihood maximization problem for the stochastic block model.

[5] We remark that an exact analysis of the maximum value of $|\langle \mathbf{Y} - \bar{\mathbf{Y}}, xx^\mathsf{T} \rangle| \leqslant C\sqrt{d} \cdot n$ over all $x \in \{\pm 1\}^n$ is very challenging. This maximum value is related to the coefficient of the second-order term of the maximum cut in an Erdős–Rényi graph with degree parameter d. This coefficient has been analyzed only in the asymptotic regime $d \to \infty$ [DMS17]. Even in this simplified setting, the coefficient corresponds to $C > 1$.

[6] In fact, throughout the thesis the term *basic SDP will be used to refer to the most "basic" semidefinite relaxation of the problem.*

[7] We remark that the famous Goemans–Williamson approximation algorithm for the max-cut problem [GW95] uses essentially the same basic SDP relaxation.

Here, Grothendieck's inequality [KN11, AN04] turns out to imply a bound that is at most a constant factor worse than the bound we had before (for cut matrices).

A key benefit of this kind of analysis (observed in early works on semirandom graph problems [FK01]; see also [MPW16]) is that it directly implies strong robustness guarantees. The reason is that we used only one property of the estimator $\hat{\mathbf{X}}$ (or $\hat{\mathbf{x}}$): it is a feasible solution to our optimization problem with objective value at least as high as the planted solution $\mathbf{x}\mathbf{x}^\top$ (up to potentially a small fudge factor). In other words, if $\varepsilon\sqrt{d}/2$ is a large enough constant for the above analysis of the basic SDP to succeed, then there exists some $\rho > 0$ (independent of n) such that with high probability (over $\mathbf{Y}$), every solution X to the basic SDP with objective value

$$\langle \mathbf{Y}, X \rangle \geqslant \langle \mathbf{Y}, \mathbf{x}\mathbf{x}^\top \rangle - \rho \cdot n \tag{4.1.4}$$

achieves constant correlation $\langle X, \mathbf{x}\mathbf{x}^\top \rangle \geqslant \Omega(1) \cdot n^2$.

Why does this property imply robustness? Suppose that Y' is the centered adjacency matrix of some corrupted versions G' of $\mathbf{G}$ (according to some adversarial model). Let $\hat{X}'$ be the optimal solution to the basic SDP with $\mathbf{Y}$ replaced by Y' in the objective function. Since the planted solution $\mathbf{x}\mathbf{x}^\top$ cannot have a higher objective value than $\hat{X}'$, it holds $\langle Y', \hat{X}' \rangle \geqslant \langle Y', \mathbf{x}\mathbf{x}^\top \rangle$. Hence, it suffices to verify that for the adversarial model of interest, the inequality $\langle Y', X \rangle \geqslant \langle Y', \mathbf{x}\mathbf{x}^\top \rangle$ for a feasible solution X to the basic SDP implies the previous inequality Eq. (4.1.4). For monotone adversaries, this implication holds (even without the fudge term $\rho \cdot n$) because for every monotone edge alteration,[8] the objective function increases for the planted solution by at least as much as for X. Even in the non-monotone case, every edge alteration (insertion or deletion) can change the objective function by at most 2 for the planted solution or for X. Hence, if we allow up to $\rho \cdot n/4$ edge alterations in our adversarial model, the desired implication holds.

Fragile recovery up to the Kesten–Stigum threshold using convex optimization. In the non-robust setting, several algorithms are known to achieve weak-recovery all the way up to the Kesten-Stigum threshold [MNS18, Mas14, BLM15, AS16, HS17]. The analyses of all these algorithms involve statistics of certain kinds of walks in graphs (e.g., self-avoiding, non-backtracking, or shortest walks). Among these algorithms, the technniques in this chapter are most closely related to the algorithm in [HS17], which combines walk-statistics and convex-optimization techniques similar to those discussed earlier in the context of robustness. For a parameter $s \in \mathbb{N}$, this algorithm considers a random n-by-n matrix $\mathbf{Q}^{(s)}$ such that each off-diagonal entry is obtained by evaluating a (deterministic) multivariate degree-s polynomial at the centered adjacency matrix $\mathbf{Y}$ of $\mathbf{G}$. Concretely, up to a scaling

[8] A monotone edge alteration consists of either adding an edge between vertices with the same planted label or deleting an edge between vertices with different planted labels.

factor depending only on n, d, ε, s,

$$\mathbf{Q}^{(s)}_{ij} \propto \sum_{W \in \mathrm{SAW}^s_{ij}} \mathbf{Y}_W \text{ where } \mathbf{Y}_W := \prod_{uv \in W} \mathbf{Y}_{uv} .$$

Here, SAW^s_{ij} consists of all length-s self-avoiding walks between i and j in the complete graph on n vertices. Since the entries of $\mathbf{Y}$ are independent when conditioned on $\mathbf{x}$, by our earlier observation Eq. (4.1.1) we have for all $W \in \mathrm{SAW}^s_{ij}$ with $i \neq j$,

$$\mathbb{E}[\mathbf{Y}_W \mid \mathbf{x}] = \prod_{uv \in W} \mathbb{E}[\mathbf{Y}_{uv} \mid \mathbf{x}] = (\tfrac{\varepsilon d}{2n})^s \cdot \mathbf{x}_i \mathbf{x}_j .$$

Hence, if we choose the diagonal entries of $\mathbf{Q}^{(s)}$ to be 0 (like for $\mathbf{Y}$) and the aforementioned scaling factor to be $(\frac{2n}{\varepsilon d})^s / |\mathrm{SAW}^s_{ij}|$, then $\mathbf{Q}^{(s)}$ is an unbiased estimator similar to $\mathbf{Y}$,

$$\mathbb{E}[\mathbf{Q}^{(s)} \mid \mathbf{x}] = \mathbf{x}\mathbf{x}^\intercal - \mathrm{Id}_n .$$

Note that $\mathbf{Q}^{(1)}$ is up to a scaling factor equal to $\mathbf{Y}$.

What can we gain from $\mathbf{Q}^{(s)}$ for $s > 1$? Here, it is instructive to compare the variance of entries of the matrices (conditioned on $\mathbf{x}$). In $\mathbf{Q}^{(1)}$ each entry has conditional variance about $(\frac{2n}{\varepsilon d})^2 \cdot \frac{d}{n} = \frac{4}{d\varepsilon^2} \cdot n$, substantially larger than the magnitude of its conditional expectation, which is 1. Now suppose we are barely above the Kesten–Stigum threshold so that $\frac{\varepsilon^2 d}{4} = 1 + \delta$ for some $\delta > 0$ independent of n. Then, as we increase s, the conditional variance decreases[9] (roughly like $(\frac{4\varepsilon^2}{d})^s \cdot n = n/(1+\delta)^s$) while the conditional expectation stays the same. Indeed, it turns out that we can choose $s \leqslant O_\delta(\log n)$ such that the conditional variance of an entry of $\mathbf{Q}^{(s)}$ is bounded by a constant $O_\delta(1)$ and thus has the same order as its conditional expectation.

This property implies the following inequality for a small enough $\delta' > \Omega_\delta(1)$,

$$\mathbb{E}\|\mathbf{Q}^{(s)} - \mathbf{x}\mathbf{x}^\intercal\|_F^2 \leqslant (1 - \delta') \cdot \|\mathbf{Q}^{(s)}\|_F^2 . \tag{4.1.5}$$

Based on this notion of correlation, the algorithm[10] in [HS17] considers the following (tractable) convex optimization problem for a regularization parameter $\lambda > 0$ (together with a simple rounding algorithm),

$$\text{maximize } \langle \mathbf{Q}^{(s)}, X \rangle - \lambda \cdot \|X\|_F^2 \text{ subject to } X \succeq 0 , \ \forall i \, . \, X_{ii} = 1 . \tag{4.1.6}$$

It is usedul to remark that this algorithm is an instance of a meta-algorithm for estimation problems that is based on sum-of-squares and specified in terms of low-degree polynomials

[9] The proof that the variance decreases argues that the unbiased estimators summed in an entry of $\mathbf{Q}^{(s)}$ behave similarly to pairwise independent estimators [MNS18, HS17].

[10] The description of the algorithm in [HS17] is slightly different from our description in that part of the objective function we stated appears as a constraint in [HS17]. For an appropriate choice of λ both versions are equivalent.

in two kinds of variables, "instance variables" for the input and "solution variables" for the desired output [HKP+17, HS17, RSS18].

Could this algorithm be robust? A key property of the basic SDP Eq. (4.1.3) for its robustness analysis is that every edge alteration can change the objective value (for a particular feasible solution) by at most $O_{\varepsilon,d}(1/n)$ times the original objective value of the planted solution. In contrast, a single edge alteration can change the objective value[11] of a feasible solution in Eq. (4.1.6) by as much as about $n \cdot (2/\varepsilon)^s$. Since the original objective value of the planted solution is close to its expectation $\mathbb{E}\langle \mathbf{Q}^{(s)}, \mathbf{x}\mathbf{x}^\mathsf{T}\rangle = n \cdot (n-1)$, we can afford this sensitivity of the objective function only for constant $s \leqslant O_{\varepsilon,d}(1)$. Unfortunately, the correlation Eq. (4.1.5) required for the analysis of Eq. (4.1.6) can only be achieved for s logarithmic in n (this is related to the well-known fact that constant-degree graphs have at least logarithmic mixing time).

Robust distinguishing up to the Kesten–Stigum threshold. Our discussion so far suggests a natural starting point for designing a robust algorithm that works up to the Kesten–Stigum threshold: matrices defined in terms of constant-length walks, e.g., the matrix $\mathbf{Q}^{(s)}$ for constant $s \leqslant O_{\varepsilon,d}(1)$.

Indeed, [BMR21] takes this approach in order to robustly solve the related distinguishing problem. The goal is to distinguish between an Erdős–Rényi random graph $\mathbf{G}_0 \sim \mathcal{G}(n, \frac{d}{n})$ and the stochastic block model $\mathbf{G}_1 \sim \mathsf{SBM}_n(d, \varepsilon)$ up to the Kesten–Stigum threshold $d > 4/\varepsilon^2$. Recall that $\mathbf{Q}^{(s)}$ corresponds to the polynomial $Q^{(s)}(Y) := \frac{(2n)^s}{(\varepsilon d)^s |\mathrm{SAW}^s_{ij}|} \sum_{W \in \mathrm{SAW}^s_{ij}} Y_W$. Let $\mathbf{Y}_0, \mathbf{Y}_1$ be the respective centered adjacency matrices of $\mathbf{G}_0, \mathbf{G}_1$. In order to distinguish $\mathbf{G}_0$ and $\mathbf{G}_1$, we seek an (efficiently computable) matrix norm that with high probability is much smaller for $Q^{(s)}(\mathbf{Y}_0)$ than for $Q^{(s)}(\mathbf{Y}_1)$. Both matrix norms discussed so far (the Frobenius norm in Eq. (4.1.5) and the cut norm in Eq. (4.1.2)) appear to be poor choices: The Frobenius norm of $Q^{(s)}(\cdot)$ cannot distinguish between $\mathbf{G}_0$ and $\mathbf{G}_1$ for constant s. A tight analysis of the cut-norm of $Q^{(s)}(\cdot)$ appears to be challenging due to the high amount of dependencies between the entries of the matrix. A natural alternative (actually related to the cut norm) is the spectral norm combined with some truncation step.[12] Concretely, the analysis in [BMR21] boils down[13] to showing that with high probability the spectral norm satisfies the inequality,

$$\left\|\bar{Q}^{(s)}(\mathbf{Y}_0)\right\| \leqslant o(\left\|\bar{Q}^{(s)}(\mathbf{Y}_1)\right\|)\,. \tag{4.1.7}$$

[11] For the purposes of this argument, we can think of $\mathbf{Q}^{(s)}_{ij}$ as $n \cdot (2/\varepsilon)^s$ times the average of $\mathbf{Y}_W$ over all self-avoiding walks W between i and j in $\mathbf{G}$ (as opposed to walks in the complete graph).

[12] The truncation step refers to removing vertices with unusually large degree. This truncation step is necessary because with high probability (non-regular) constant-degree random graphs have a small number of vertices with logarithmic degree that skew the spectral norm of the centered adjacency matrix in an undesirable way. This issue persists also for the matrices $Q^{(s)}(\cdot)$ when s is constant.

[13] We remark that the presentation of the algorithm and the choice of the function $\bar{Q}^{(s)}(\cdot)$ in [BMR21] is slightly different from what's described here. For example, they use non-backtracking walks instead of self-avoiding walks. However, the same proof strategy works for both versions.

Here, $\bar{Q}^{(s)}(Y)$ is a matrix-valued function obtained by composing the polynomial $Q^{(s)}(Y)$ together with a truncation step (we omit the details here). In order to prove this inequality, the authors upper-bound (the expectation of) the left-hand side using the trace-method for a parameter t logarithmic in n,

$$\mathbb{E}\left\|\bar{Q}^{(s)}(\mathbf{Y}_0)\right\| \leqslant \left(\mathbb{E}\left\|\bar{Q}^{(s)}(\mathbf{Y}_0)\right\|^t\right)^{1/t} \leqslant \left(\operatorname{Tr}\mathbb{E}\,\bar{Q}^{(s)}(\mathbf{Y}_0)^t\right)^{1/t},$$

and then lower-bound (the expectation of) the right-hand side using the planted labeling $\mathbf{x}$ as a test vector,

$$\mathbb{E}\left\|\bar{Q}^{(s)}(\mathbf{Y}_1)\right\| \geqslant \tfrac{1}{n}\,\mathbb{E}\langle \mathbf{x}, \bar{Q}^{(s)}(\mathbf{Y}_1)\,\mathbf{x}\rangle \approx \tfrac{1}{n}\,\mathbb{E}\langle \mathbf{x}, Q^{(s)}(\mathbf{Y}_1)\,\mathbf{x}\rangle = n-1\,.$$

In order to make their distinguishing algorithm robust, the authors consider the following related semidefinite program,

$$\text{maximize } \langle \bar{Q}^{(s)}(Y), X\rangle \text{ subject to } X \succeq 0\,,\ \forall i\,.\,X_{ii} = 1\,. \tag{4.1.8}$$

The optimal value for $Y = \mathbf{Y}_0$ is upper bounded by $n \cdot \left\|\bar{Q}^{(s)}(\mathbf{Y}_0)\right\|$. At the same time, the optimal value for $Y = \mathbf{Y}_1$ is lower bounded by the value of planted solution $X = \mathbf{x}\mathbf{x}^\mathsf{T}$, which is close to $n \cdot (n-1)$. This algorithm is robust because the objective function in Eq. (4.1.8) has low sensitivity to edge alterations for constant s, every edge alteration changes the objective value by at most $n \cdot (2/\varepsilon)^s$, which is a $O_{\varepsilon,s}(1/n)$ fraction of the objective value of the planted solution (also see our discussion of the sensitivity of Eq. (4.1.6)).

The push out effect and the challenges for weak-recovery using constant-length walks. *Can we use the matrix $\bar{Q}^{(s)}(\mathbf{Y})$ (with constant s) also for weak-recovery?* Ignoring the issue of robustness for now, the fact that, for the stochastic block model, the spectral norm of $\bar{Q}^{(s)}(\mathbf{Y})$ is substantially larger than for the corresponding Erdős–Rényi random graph suggests that the top eigenvector of this matrix carries useful information about the planted labeling. Its top eigenvector can be characterized in terms of the optimal solution to the following semidefinite program,

$$\text{maximize } \langle \bar{Q}^{(s)}(\mathbf{Y}), Z\rangle \text{ subject to } Z \succeq 0\,,\operatorname{Tr} Z = 1\,. \tag{4.1.9}$$

In order to prove that the optimal solution is correlated with the planted labeling, we could try to mimic the analysis of the basic SDP. To this end, we would have to prove that the spectral norm $\|\bar{Q}^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T}\|$ is smaller than the objective value of the planted solution $\frac{1}{n}\mathbf{x}\mathbf{x}^\mathsf{T}$. However, this turns out to be false for d close to the Kesten–Stigum threshold, i.e. $\langle \bar{Q}^{(s)}(\mathbf{Y}), \frac{1}{n}\mathbf{x}\mathbf{x}^\mathsf{T}\rangle < \|\bar{Q}^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T}\|$. Another ramification of the phenomenon is that for constant s, the optimal value of Eq. (4.1.9) is substantially larger than the objective value of the planted solution with high probability.

Nevertheless, it is still possible to weakly recover the hidden communities. To understand how one could overcome this challenge, suppose that we could show the following

inequality for the spectral norm (similar to the inequality Eq. (4.1.5) for the Frobenius norm, for which however this bound does not hold),

$$\|\bar{Q}^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T}\| \leqslant (1-\delta')\|\bar{Q}^{(s)}(\mathbf{Y})\|\,, \tag{4.1.10}$$

for some small enough $\delta' > \Omega_\delta(1)$. Then, even though $\mathbf{x}\mathbf{x}^\mathsf{T}$ would still be far from being the optimal solution to the program, the communities vector would partially align with the leading eigenvector of $\bar{Q}^{(s)}(\mathbf{Y})$ and thus we would still be able to weakly recover it. (This phenomenon is closely related to the push-out effect in the context of principal component analysis).

In order to actually prove Eq. (4.1.10), one needs to upper bound the spectral norm of $\bar{Q}^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T}$ and lower bound the spectral norm of $\bar{Q}^{(s)}(\mathbf{Y})$. As in [BMR21], a natural approach is that of using the trace method. Notice however that by our discussion above, compared to [BMR21], in order to achieve weak recovery the required bounds need to be significantly sharper. Furthermore, an additional technical difference is that one needs to carry out the trace method in the planted distribution $\mathsf{SBM}_n(d, \varepsilon)$ (and after an additional truncation step!). Indeed there are only few instances where the trace method is carried out on the planted distribution [BLM15, GLM16] (and [DdNS22] on which this chapter is based).

For several technical reasons we will only show

$$\|\bar{Q}^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T}\|_t \leqslant (1-\delta')\|\bar{Q}^{(s)}(\mathbf{Y})\|_t\,, \tag{4.1.11}$$

for some Schatten norm $\Omega(\log n) \leqslant t < \log n$. While for $t \gtrsim \log n$, the Schatten norm is within a constant factor of the spectral norm, in the delicate context of Eq. (4.1.11) this difference is not negligible as it might result in losing the correlation with the communities vector. A similar reasoning as the one outlined for the spectral norm carries over to the context of Schatten norm. Using the dual definition of Schatten norm we consider the following optimization problem:

$$\text{maximize } \langle \bar{Q}^{(s)}(\mathbf{Y})^2, Z\rangle \text{ subject to } Z \succeq 0\,, \operatorname{Tr} Z^{\frac{t}{t-1}} = 1\,, \tag{4.1.12}$$

(Note that we squared the matrix polynomial here since otherwise its trace may not contain meaningful information). The bound Eq. (4.1.11) implies that we can weakly recover the planted solution from the optimal solution to Eq. (4.1.12).

Achieving robustness. It is by now clear that, to enable robustness, a key property is low sensitivity of the objective function to edge alterations. For Eq. (4.1.12), however, the sensitivity of the objective function may be high because the program allows solutions that are spiky (in the sense that there could be some $i, j \in [n]$ such that $\left|Z_{ij}\right| \geqslant \frac{\omega(1)}{n^2} \sum_{i,j\in[n]} \left|Z_{ij}\right|$). In the context of the basic SDP, we could fix the sensitivity of the objective function by adding a constraint on the diagonal entries (and hence, by semidefiniteness, on all entries).

The reason we could add this constraint was because the communities matrix $\mathbf{x}\mathbf{x}^\mathsf{T}$ –which satisfied the constraint– was close to the optimal solution. However, as we approach the KS threshold this is no longer true and thus it remains unclear whether this approach could work. Indeed, the recovery analysis outlined above did not compare arbitrary solutions to the communities matrix $\mathbf{x}\mathbf{x}^\mathsf{T}$, but to an a-priori *unknown* optimal solution!

How can we fix the sensitivity? While, with high probability, $\bar{Q}^{(s)}(\mathbf{Y})$ is not positive semidefinite, its second power $\bar{Q}^{(s)}(\mathbf{Y})^2$ clearly is. A crucial consequence of this property is that the optimal solution to Eq. (4.1.12) actually has a nice analytical expression: $\bar{Q}^{(s)}(\mathbf{Y})^{2(t-1)}$. Due to the truncation, $\bar{Q}^{(s)}(\mathbf{Y})$ is approximately flat and thus, using calculations similar to the trace method, we are able to show that $\bar{Q}^{(s)}(\mathbf{Y})^{2(t-1)}$ is also approximately flat. This implies that we can control the sensitivity of the objective function by replacing the program in Eq. (4.1.12) with

$$\text{maximize } \langle \bar{Q}^{(s)}(\mathbf{Y})^2, Z\rangle \text{ subject to } Z \succeq 0\,, \operatorname{Tr} Z^{\frac{t}{t+1}} = 1\,, \forall i,\ Z_{ii} \leqslant \frac{O(1)}{n}\,, \tag{4.1.13}$$

which immediately yields a robust algorithm.

From the Schatten norm to graph counting. Proving the bound Eq. (4.1.11) turns out to be the main technical challenge behind Theorem 4.2. We outline here the main ideas. For simplicity we limit the current discussion to showing that the inequality holds in expectation (see Section 4.5). By definition

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim \mathsf{SBM}_n(d,\varepsilon)} \left[\left\| \bar{Q}^{(s)}(\mathbf{Y}) \right\|_t^t \right] = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim \mathsf{SBM}_n(d,\varepsilon)} \left[\operatorname{Tr}\left(\bar{Q}^{(s)}(\mathbf{Y}) \right)^t \right], \tag{4.1.14}$$

$$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim \mathsf{SBM}_n(d,\varepsilon)} \left[\left\| \bar{Q}^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T} \right\|_t^t \right] = \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim \mathsf{SBM}_n(d,\varepsilon)} \left[\operatorname{Tr}\left(\bar{Q}^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T} \right)^t \right] \tag{4.1.15}$$

Our approach to obtain Eq. (4.1.11) will be to reduce *both* trace computations to a graph counting problem. Consider Eq. (4.1.14), each element corresponds to a walk of the following form:

Definition (Block self-avoiding walk)**.** A closed walk of length st is a (s,t)-*block self-avoiding walk* if it can be split into t self-avoiding walks of length s.

For a (s,t)-block self avoiding walk H, we call its t self-avoiding walks $\{W_1,\dots,W_t\}$, the generating walks of H.[14] With this definition, Eq. (4.1.14) amount to computing the expected number of copies in the instance graph $\mathbf{G}$ of any (s,t)-block self-avoiding walk.

To see how to carry out this computation, and for simplicity, consider the non-truncated graph $\mathbf{G}$ with its centered adjacency matrix $\mathbf{Y}$. While this simplification will make it impossible to get a good bound, it will be useful to build some intuition. For a (s,t)-block

[14] It is possible that there are multiple choices for the set $\{W_1,\dots,W_t\}$, at the granularity of this discussion we may assume such set to be given.

self avoiding walk H, let $\mathbf{Y}_H = \prod_{i\in[t]} \mathbf{Y}_{W_i}$ where $W_1, \dots, W_t$ are the self-avoiding walks generating H. It is easy to see that

$$\begin{aligned}
\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}[\mathbf{Y}_H \mid \mathbf{x}] &= \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}\left[\prod_{ij\in E(H)} Y_{ij}^{m(ij)} \,\middle|\, \mathbf{x}\right] \\
&= \prod_{ij\in E(H)} \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}\left[Y_{ij}^{m(ij)} \,\middle|\, \mathbf{x}\right] \\
&\approx \left[\prod_{ij\in E_1(H)} \left(\frac{\varepsilon\cdot d}{2n}\mathbf{x}_i\mathbf{x}_j\right)\right] \cdot \left[\prod_{ij\in E_{\geqslant 2}(H)} \left(1+\frac{\varepsilon}{2}\mathbf{x}_i\mathbf{x}_j\right)\cdot\frac{d}{n}\right] \\
&= \left(\frac{d}{n}\right)^{|E(H)|} \cdot \left[\prod_{ij\in E_1(H)} \left(\frac{\varepsilon}{2}\mathbf{x}_i\mathbf{x}_j\right)\right] \cdot \left[\prod_{ij\in E_{\geqslant 2}(H)} \left(1+\frac{\varepsilon}{2}\mathbf{x}_i\mathbf{x}_j\right)\right], \qquad (4.1.16)
\end{aligned}$$

where $m(ij)$ is the number of times the edge ij appears in the walk H, $E_1(H)$ is the set of edges with multiplicity $m(ij) = 1$ and $E_{\geqslant 2}(H) = E(H) \setminus E_1(H)$. The main burden is then to count for each $m_1, m_2 \geqslant 0$ how many (s,t)-block self-avoiding walks we may have with m_1 edges with multiplicity one and m_2 edges of multiplicity at least 2.

The effect of centering. To understand why $\mathbf{x}\mathbf{x}^\intercal$ correlates with $\bar{Q}^{(s)}(\mathbf{Y})$, we need to observe what is the effect of subtracting the communities from $\bar{Q}^{(s)}(\mathbf{Y})$. Consider Eq. (4.1.15) and again, for simplicity, let us simply use the non-truncated adjacency matrix and the corresponding polynomial $Q^{(s)}(\mathbf{Y})$.

For any (s,t)-block self-avoiding walk H, let $W_{v_1v_2}, \dots, W_{v_tv_1}$ be its t self-avoiding walks so that for each $W_{v_iv_{i+1}}$ the vertices v_i, v_{i+1} correspond to its endpoint.[15] For every block self-avoiding walk H, there is a polynomial in Eq. (4.1.15) of the form $\prod_{i\in[t]}\left(\mathbf{Y}_{W_{v_iv_{i+1}}} - \left(\frac{\varepsilon\cdot d}{2n}\right)^s \cdot \mathbf{x}_i\cdot\mathbf{x}_{i+1}\right)$. The catch is that if one of the walks $W_{v_1v_2}, \dots, W_{v_tv_1}$ –for example $W_{v_tv_1}$– has all edges with multiplicity one in H, then by Eq. (4.1.1)

$$\begin{aligned}
0 &= \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}\left[\prod_{i\in[t-1]}\left(\mathbf{Y}_{W_{v_iv_{i+1}}} - \left(\frac{\varepsilon\cdot d}{2n}\right)^s\cdot\mathbf{x}_i\cdot\mathbf{x}_{i+1}\right)\right] \cdot \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}\left[\mathbf{Y}_{W_{v_tv_1}} - \left(\frac{\varepsilon\cdot d}{2n}\right)^s\cdot\mathbf{x}_{v_t}\mathbf{x}_{v_1}\right] \\
&= \mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}\left[\prod_{i\in[t]}\left(\mathbf{Y}_{W_{v_iv_{i+1}}} - \left(\frac{\varepsilon\cdot d}{2n}\right)^s\cdot\mathbf{x}_i\cdot\mathbf{x}_{i+1}\right)\right].
\end{aligned}$$

So, all (s,t)-block self-avoiding walks with at least one of the generating walks having all edges traversed exactly *once* in H have expectation 0 and do not contribute to

[15] To simplify the notation we write v_{t+1} for v_1.

$\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}\left[\mathrm{Tr}\left(Q^{(s)}(\mathbf{Y})-\mathbf{x}\mathbf{x}^\intercal\right)^t\right]$. This effect will approximately carry over the truncated graph $\bar{\mathbf{G}}$ (with centered adjacency matrix $\bar{\mathbf{Y}}$). Since block self-avoiding walks as the one described above turn out to have a significant contribution to Eq. (4.1.14), this observation will allow us to show

$$\mathop{\mathbb{E}}_{(x,\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}\left[\left\|\bar{Q}^{(s)}(\mathbf{Y})-\mathbf{x}\mathbf{x}^\intercal\right\|_t^t\right]\leqslant(1+\delta)^{-\Omega(t)}\cdot\mathop{\mathbb{E}}_{(\mathbf{x},\mathbf{G})\sim\mathsf{SBM}_n(d,\varepsilon)}\left[\left\|\bar{Q}^{(s)}(\mathbf{Y})\right\|_t^t\right]. \tag{4.1.17}$$

Boosting the probability of success. The approach described so far yields an algorithm that weakly recovers the communities vector with constant probability. The main reason behind this shortcoming is that we only showed the required structural inequalities hold with constant probability.[16] It turns out, however, that we can exploit robustness to argue that, indeed, we can weakly recover $\mathbf{x}$ with high probability.

This boosting argument relies on a concentration of measure inequality known as the *blowing-up lemma* (see e.g., [AGK76, Mar86, Mar96, RS+13]), which is widely used in information theory to prove strong converse results.[17] Roughly speaking, the blowing-up lemma states that if $\mathbf{z}_1,\dots,\mathbf{z}_N$ are N independent random variables taking values in a finite set, and $\mathcal{E}_N$ is an event on $\mathbf{z}_1,\dots,\mathbf{z}_N$ whose probability does not decay exponentially in N, i.e., $\lim_{N\to\infty}\frac{1}{N}\log\frac{1}{\mathbb{P}[\mathcal{E}_N]}=0$, then there exists $\ell_N=o(N)$ such that, if we "inflate" $\mathcal{E}_N$ by adding all the strings $(z_1,\dots,z_N)$ that are at a Hamming distance of at most $\ell_N=o(N)$ from $\mathcal{E}_N$, then the probability of the obtained event becomes $1-o(1)$.

In our context of interest, this means that we do not need to directly show that our graph $\mathbf{G}\sim\mathsf{SBM}_n(d,\varepsilon)$ satisfies the desired structural inequalities (such as Eq. (4.1.11)). It suffices that *some* graph $\tilde{G}\in N_{o(1)}(\mathbf{G})$ satisfies them. Since our algorithm is robust to a constant fraction ρ of adversarial changes, and since $N_{\rho-o(1)}(\mathbf{G})\subseteq N_\rho(\tilde{G})$, we can see that by paying a negligible price in the robustness and correlation guarantees, we can use the blowing-up lemma to boost the probability of success of the algorithm to $1-o(1)$.

It is worth noting that the boosting argument can give us exponentially high probability and not only $1-o(1)$. Furthermore, the argument is not unique to the stochastic block model and can be applied to a wide range of estimation problems.

4.2 Preliminaries

We present here some elementary definitions and results that we will use in the following sections. We use the notation in Chapter 2.

[16] In fact, it is possible to prove that the structural inequalities hold with high probability without invoking the boosting argument. However, the needed calculations are more complicated. In any case, the boosting argument has the advantage of achieving exponentially high probability.

[17] A strong converse result in information theory typically has the following form: If the rate of a code is above the capacity of a channel, then the probability of error of the code goes to 1 as the blocklength becomes large.

For a multigraph $H = (V, M)$, we denote by $V(H)$ the set of vertices in H, by $M(H)$ the multi-set of edges in H and by $E(H)$ the set of *distinct* edges. For $e \in E(H)$, we let $m_H(e)$ be the multiplicity of edge e in H. For $v \in V(H)$ we denote by $d^H(v)$ the number of *distinct* edges incident to v in H. Given multi-sets S_1, S_2 we denote their union by $S_1 \oplus S_2$, similarly for multi-graphs H_1, H_2 we write $H_1 \oplus H_2$ for the multi-graph H with vertex set $V(H_1) \cup V(H_2)$ and edges $M(H_1) \oplus M(H_2)$. For a set of vertices (or edges) S, we denote by $H(S)$ the subgraph of H induced by the set S. For a graph $G \subseteq H$ we denote by $H(G)$ the multigraph induced by $V(G)$. For an edge e and a multigraph H, we write $H + e$ for the multigraph obtained adding e to H. We denote by K_n be the complete graph on n vertices.

Remark 4.3. All graphs we consider in the chapter will have vertex sets that are subsets of $[n]$.

For a walk W over K_n as defined in Chapter 2, we denote by $M(W)$ the sequence of edges in W. We write $E(W)$ for the set of *distinct* edges in W and $V(W)$ for the set of *distinct* vertices. For simplicity, we also use W to refer to the multigraph with vertex set $V(W)$ and edges $M(W)$.

Definition 4.4 (Eulerian multi-graph)**.** We say that a multigraph H is Eulerian if there exists a closed walk W in H such that for any $e \in E(H)$, e appears in W exactly $m_H(e)$ times.

Definition 4.5 (Cut)**.** For a multi-graph H with vertex set V, a cut $(B, V \setminus B)$ is a partition of the vertices into two disjoint subsets. We denote with $(B, V \setminus B)$ the edges in the cut and by $H(B, V \setminus B)$ the multigraphs spanned by those edges in H . A q-multi-way cut is a partition $(B_1, \ldots, B, V \setminus (B_1 \cup \ldots \cup B_q))$ of the vertices in H in q disjoint subsets.

Definition 4.6 (Isomorphic graphs)**.** Two graphs $G = (V, E)$ and $G' = (V', E')$ are isomorphic if there exists a bijective mapping $\phi : V \to V'$ such that for any $u, v \in V$, $\{u, v\} \in E$ if and only if $\{\phi(u), \phi(v)\} \in E'$.

We formally present here the polynomials used in the subsequent sections. Let Y be the n-by-n adjacency matrix of a graph G with vertex set $[n]$, centered with respect to the Erdős-Rényi distribution with parameter d. That is for $a, b \in [n]$,

$$Y_{ab}(G) = \begin{cases} 1 - \frac{d}{n} & \text{if } ab \in E(G), \\ -\frac{d}{n} & \text{if } ab \notin E(G), \\ 0 & \text{if } a = b\,. \end{cases} \tag{4.2.1}$$

When the context is clear we write Y instead of $Y(G)$. For simplicity, for a multigraph H with vertex set $V(H) \subseteq [n]$, we write Y_H for the polynomial

$$Y_H := \prod_{ab \in E(H)} Y_{ab}^{m(ab)}\,.$$

For fixed parameters n, d, s, ε and for $i, j \in [n]$, we define the polynomial

$$Q_{ij}^{(s)}(Y) = \begin{cases} \frac{1}{|\mathrm{SAW}_{ij}^s|}\left(\frac{2n}{\varepsilon \cdot d}\right)^s \sum_{H \in \mathrm{SAW}_{ij}^s} Y_H & \text{if } i \neq j\,, \\ 0 & \text{otherwise.} \end{cases} \tag{4.2.2}$$

$Q^{(s)}(Y) : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ is then the matrix valued polynomial with entry ij equal to $Q_{ij}^{(s)}(Y)$. Notice that $\left|\mathrm{SAW}_{ij}^s\right| = (n-2)^{\underline{s-1}}$ where $(n-2)^{\underline{s-1}}$ is the falling factorial $(n-2)\cdot(n-3)\cdots(n-2-s+1)$.

As already mentioned, the polynomial Eq. (4.2.2) is an unbiased estimators of the vector of communities $\mathbf{x}$ with respect to the distribution $\mathsf{SBM}_n(d, \varepsilon)$. The following facts formally show this and other basic properties of the polynomials Eq. (4.2.1) and Eq. (4.2.2).

Fact 4.7. *Let $\mathbf{G} \sim \mathsf{SBM}_n(d, \varepsilon)(\mathbf{Y}|\mathbf{x})$ and let $\mathbf{Y}$ be its centered adjacency matrix. Let F be a simple graph. Then*

$$\mathop{\mathbb{E}}_{\mathsf{SBM}_n(d,\varepsilon)}[\mathbf{Y}_F \mid \mathbf{x}] = \left(\frac{d \cdot \varepsilon}{2n}\right)^{|E(F)|} \prod_{ab \in E(F)} \mathbf{x}_a \mathbf{x}_b\,.$$

Proof.

$$\begin{aligned} \mathbb{E}[\mathbf{Y}_F \mid \mathbf{x}] &= \prod_{ab \in E(F)} \mathbb{E}[\mathbf{Y}_{ab} \mid \mathbf{x}] \\ &= \prod_{ab \in E(F)} \left[\left(1 + \frac{\varepsilon}{2}\mathbf{x}_a\mathbf{x}_b\right)\frac{d}{n}\left(1 - \frac{d}{n}\right) - \left(1 - (1 + \frac{\varepsilon}{2}\mathbf{x}_a\mathbf{x}_b)\frac{d}{n}\right)\frac{d}{n}\right] \\ &= \prod_{ab \in E(F)} \frac{d}{n}\left[\left(1 - \frac{d}{n}\right) - \left(1 - \frac{d}{n}\right) + \frac{\varepsilon}{2}\mathbf{x}_a\mathbf{x}_b\left(1 - \frac{d}{n} + \frac{d}{n}\right)\right] \\ &= \prod_{ab \in E(F)} \frac{d \cdot \varepsilon}{2n}\mathbf{x}_a\mathbf{x}_b\,. \end{aligned}$$

□

Notice that if F is a self-avoiding walk of length-s between vertices $i, j \in [n]$ then $\mathbb{E}[\mathbf{Y}_F \mid \mathbf{x}] = \left(\frac{d\cdot\varepsilon}{2n}\right)^s \mathbf{x}_i\mathbf{x}_j$. The expectation of small powers of $\mathbf{Y}_{ab}$ is also easy to approximate. We write $f(n) = (1 \pm o(1))g(n)$ if $(1 - o(1)) \cdot g(n) \leqslant f(n) \leqslant (1 + o(1)) \cdot g(n)$.

Fact 4.8. *Let $a, b \in [n]$ and let $2 \leqslant q \lesssim \log n$ be some integer. Then*

$$\mathop{\mathbb{E}}_{\mathsf{SBM}_n(d,\varepsilon)}\left[\mathbf{Y}_{ab}^q \,\middle|\, \mathbf{x}\right] = \left(1 \pm o\left(\frac{1}{n^{0.99}}\right)\right)\left(1 + \frac{\varepsilon}{2}\mathbf{x}_a\mathbf{x}_b\right)\frac{d}{n}\,.$$

Proof. By choice of q and since d is a constant,

$$\begin{aligned}\mathbb{E}\left[\mathbf{Y}_{ab}^{q} \mid \mathbf{x}\right] &= \left(1+\frac{\varepsilon}{2}\mathbf{x}_a\mathbf{x}_b\right)\frac{d}{n}\left(1-\frac{d}{n}\right)^{q} - \left(1-(1+\frac{\varepsilon}{2}\mathbf{x}_a\mathbf{x}_b)\frac{d}{n}\right)\left(\frac{d}{n}\right)^{q} \\ &= \left(1 \pm o\left(\frac{1}{n^{0.99}}\right)\right)\left(1+\frac{\varepsilon}{2}\mathbf{x}_a\mathbf{x}_b\right)\frac{d}{n}\,.\end{aligned}$$

□

4.3 Robust recovery meta-algorithm

In this section we consider a more general problem than the stochastic block model with constant average degree. We show how to robustly recover the hidden structure even though it may have a low objective value in our optimization problem. We apply this algorithm to the stochastic block model in Section 4.4.

Problem 4.9 (Small-Correlation Robust Signal Recovery)**.** Let $v \in \{\pm 1/\sqrt{n}\}^n$ be a unit flat vector and $t = \frac{1}{C^*}\cdot \log n$ for some constant $C^* > 0$. For an *unknown* matrix $Q \in \mathbb{R}^{n\times n}$ –weakly correlated with the *unknown* signal v– and an *observed* matrix $\tilde{Q} \in \mathbb{R}^{n\times n}$ such that the tuple $(Q, \tilde{Q}, v)$ satisfies the conditions Eq. ($\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$) with parameters $\alpha, \beta, \gamma, \delta^*, \zeta, C^* \geqslant 0$. The goal is to *return* a unit vector $\hat{v}$ such that $\langle \hat{v}, v\rangle^2 \geqslant \Omega(1)$.

The set of conditions is:

$$\mathcal{A}_{\alpha,\beta,\delta^*,\gamma,\zeta,C^*}\colon \left\{\begin{array}{ll} \text{correlation:} & \operatorname{Tr} Q^{2t} = 1 \\ & \operatorname{Tr}\left(Q-\alpha\cdot vv^{\mathsf{T}}\right)^{2t} \leqslant (1-\delta^*)^{2t} \\ \text{sensitivity:} & \sum_{i\in[n]}\left(Q^{2t-2}\right)_{ii}^2 \leqslant \frac{\gamma}{n}\cdot\left(\operatorname{Tr} Q^{2t-2}\right)^2 \\ & \max_{i\in[n]}\sum_{j\in[n]}|Q_{ij}| \leqslant \zeta \\ \text{perturbations:} & \frac{1}{n}\left\|Q-\tilde{Q}\right\|_1 \leqslant \beta \\ & \max_{i\in[n]}\sum_{j\in[n]}|\tilde{Q}_{ij}| \leqslant \zeta \end{array}\right\} \qquad (\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*})$$

The correlation conditions enforce the matrix Q to contain non-trivial information about v. Here δ^* is the parameter quantifying the correlation between Q and the planted solution.[18] The sensitivity conditions (with parameters γ, ζ) are necessary so that perturbations cannot completely hide the signal. They roughly correspond to the flatness property discussed in

[18]For the stochastic block model, δ^* will be a polynomial in δ for $d \geqslant (1+\delta)\frac{4}{\varepsilon^2}$.

Section 4.1. Finally, the perturbations inequalities dictate how far the original matrix Q is from its corrupted observation $\tilde{Q}$.

To solve Problem 4.9, we will use the following algorithm.

Algorithm 4.10 (Scale-free robust recovery)**.**
Settings: Let $(Q, \tilde{Q}, v)$ be an instance of Problem 4.9 satisfying Eq. ($\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$) with parameters $\alpha, \beta, \gamma, \delta^*, \zeta \geqslant 0, C^* > 0$.
Input: $\tilde{Q} \in \mathbb{R}^{n\times n}, \gamma \geqslant 0, C^* > 0$.

1. Let $t^* = (1 - 1/t)^{-1}$. Compute a matrix $Z \in \mathbb{R}^{n\times n}$ solving the program

$$\text{maximize } \langle Z, \tilde{Q}^2 \rangle \qquad \text{subject to } \left\{ \begin{array}{r} \operatorname{Tr} Z^{t^*} \leqslant 1 \\ Z \succeq 0 \\ \forall i \in [n], \quad Z_{ii} \leqslant \dfrac{\gamma^2}{n} \end{array} \right\} \tag{P.1}$$

2. Return $M := Z\tilde{Q} + \tilde{Q}Z$.

Notice that the constraints of program P.1 are convex and can be checked in polynomial time, hence the whole algorithm runs in polynomial time.

Algorithm 4.10, will allow us to approximately recover the unknown vector v with constant probability. Concretely, we will use it to prove the central theorem below.

Theorem 4.11. *Let $Q, \tilde{Q} \in \mathbb{R}^{n\times n}$, $v \in \{\pm 1/\sqrt{n}\}^n$ form an instance of Problem 4.9. Suppose that $(Q, \tilde{Q}, v)$ satisfies Eq. ($\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$) with parameters $\alpha, \beta, \gamma, \delta^*, \zeta \geqslant 0, C^* > 0$ and that*

$$\textit{approximability:} \qquad e^{2C^*} \cdot \frac{\zeta^2}{\sqrt{\gamma}} \leqslant \frac{\delta^*}{4} \tag{4.3.1}$$

$$\textit{robustness:} \qquad 8\zeta \cdot \gamma^2 \cdot \beta \leqslant \frac{\delta^*}{4}. \tag{4.3.2}$$

Then, given $\tilde{Q}, \gamma$ there exists a polynomial time algorithm (Algorithm 4.10) that computes a matrix M with $\|M\|_{\text{nuc}} \leqslant O(1)$ such that

$$\langle M, vv^\mathsf{T} \rangle \geqslant \left(\frac{\delta^*}{2\alpha} - \frac{\gamma^2 \cdot \beta}{n} \right) \cdot \|M\|_{\text{nuc}}.$$

Before proving the theorem, let's try to understand its conditions and meaning. First, notice that if we had access to Q then simply computing the matrix maximizing the t-Schatten norm of Q^2 would result in a matrix δ-correlated with vv^T.[19] Unfortunately we only have access to an *arbitrarily perturbed* version of Q, thus the main difficulty is to use this perturbed matrix $\tilde{Q}$ as a proxy for Q and still obtain a matrix close to vv^T. In Eq. (4.3.1),

[19] We need to use Q^2 since the matrix Q may not be positive semidefinite. See Section 4.1.

the exponential term is a residue of the use of schatten norms. One side, large γ allows for stronger approximability inequalities, on the other hand the smaller γ is the more resilient the algorithm is to adversarial perturbations (as seen in Eq. (4.3.2)). On a similar note, small values ζ implies the data has low sensitivity and thus one can obtain better guarantees.

The strategy of Algorithm 4.10 to solve Problem 4.9 will be to obtain a matrix Z with small entries that is close to the matrix maximizing the t-Schatten norm of $\tilde{Q}^2$. By the sensitivity and perturbation constraints, this will also be close to the matrix maximizing the t-Schatten norm of Q^2. The approximability condition Eq. (4.3.1) ensures that the constraints of P.1 on the entries of feasible solutions are not too strong and so that $\tilde{Q}^{2t}$ is close to an optimal solution. On the other hand, the robustness condition ensures that $\tilde{Q}$ and Q are not too far from each other and so the optimal solution to P.1 non-trivially correlates with Q^2 and thus also with $vv^\top$.

Weak recovery of v then immediately follows as a corollary of the rounding below.

Lemma 4.12 (Rounding lemma)**.** *Let $\hat{\delta} > 0$. Given a matrix $M \in \mathbb{R}^{n\times n}$ and a unit vector $v \in \mathbb{R}^n$ achieving correlation in the following sense*

$$\langle M, vv^\top \rangle \geqslant \|M\|_{\text{nuc}} \cdot \delta^* ,$$

a uniformly at random choice $\hat{\mathbf{v}}$ among the top $\frac{2}{\delta^*}$ unit norm eigenvectors of M can achieve $(\delta^*)^{O(1)}$-correlation with v:

$$\mathbb{E}\langle \hat{\mathbf{v}}, v\rangle^2 \geqslant \frac{(\delta^*)^3}{8} .$$

Proof. By performing a spectral decomposition,

$$\frac{M}{\|M\|_{\text{nuc}}} = \sum_{i=[n]} \xi_i \cdot z_i z_i^\top ,$$

where z_i are orthonormal eigenvectors of M and $\xi_\ell \geqslant \xi_{\ell+1}$ for any $\ell \in [n]$ are scalars. By assumption $\delta^* = \sum_{i=[n]} \xi_i \langle z_i, v\rangle^2$. Since $\frac{M}{\|M\|_{\text{nuc}}}$ has unit nuclear norm, we have $\sum_{i=[n]} |\xi_i| = 1$. Thus for $k = \frac{2}{\delta^*}$, when $i > k$, $\xi_i \leqslant \frac{\delta^*}{2}$. It follows that $\sum_{i\in[k]} \xi_i \cdot \langle z_i, v\rangle^2 \geqslant \delta^* - \frac{\delta^*}{2} = \frac{\delta^*}{2}$ and so there must be $\mathbf{i} \in [k]$ such that $\langle z_{\mathbf{i}}, v\rangle^2 \geqslant \xi_i \cdot \langle z_i, v\rangle^2 \geqslant \frac{(\delta^*)^2}{4}$.

Hence, if we choose $\mathbf{i} \in [k]$ uniformly at random, then with probability at least $1/k$, we get a vector such that $\langle z_{\mathbf{i}}, v\rangle^2 \geqslant \frac{(\delta^*)^2}{4}$. As $k = \frac{2}{\delta^*}$, we obtain the claim. □

The rest of the section contains a proof of Theorem 4.11. We split it in two lemmas. The first lower bounds the optimal value of program P.1, the second uses such lower bound to show that any nearly optimal solution is non-trivially correlated with $vv^\top$. Together they will imply the theorem.

Lemma 4.13 (Lower bound for the Optimum)**.** *Let $(Q, \tilde{Q}, v)$ be an instance of Problem 4.9 satisfying $\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$ with parameters $\alpha, \beta, \gamma, \delta^*, \zeta \geqslant 0, C^* > 0$ and consider the program P.1. Then*

$$\text{opt}_{P.1} \geqslant 1 - e^{2C^*} \cdot \frac{\zeta^2}{\sqrt{\gamma}} - 4\zeta \cdot \gamma^2 \cdot \beta .$$

Lemma 4.14 (Correlation of nearly-optimal solutions). *Let $(Q, \tilde{Q}, v)$ be an instance of Problem 4.9 satisfying Eq. $(\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*})$ with parameters $\alpha, \beta, \gamma, \delta^*, \zeta \geqslant 0, C^* > 0$ and consider the program P.1. Suppose that*

$$e^{2C^*} \cdot \frac{\zeta^2}{\sqrt{\gamma}} + 8\zeta \cdot \gamma^2 \cdot \beta \leqslant \frac{\delta^*}{2}.$$

Then any feasible solution $Z \in \mathbb{R}^{n\times n}$ such that $\langle Z, \tilde{Q}^2\rangle \geqslant 1 - e^{2C^*} \cdot \frac{\gamma^2}{\sqrt{\zeta}} - 4\zeta \cdot \gamma^2 \cdot \beta$, satisfies

$$\langle Z\tilde{Q} + \tilde{Q}Z, vv^\mathsf{T}\rangle \geqslant \frac{\delta^*}{2\alpha} - \frac{\gamma^2 \cdot \beta}{n}.$$

We are now ready to prove Theorem 4.11.

Proof of Theorem 4.11. By hypothesis, combining Lemma 4.13 and Lemma 4.14 we get

$$\langle M, vv^\mathsf{T}\rangle \geqslant \frac{\delta^*}{2\alpha} - \frac{\gamma^2 \cdot \beta}{n}.$$

It remains to bound $\|M\|_{\mathrm{nuc}}$. By Holder's inequality with respect to the spectral norm and the nuclear norm, we have $\|Z\tilde{Q}\|_{\mathrm{nuc}} \leqslant \|Z\|_{\mathrm{nuc}}\|\tilde{Q}\|$. Since Z is a feasible solution to program P.1, we have $\|Z\|_{t*} \leqslant 1$ for $t^* = \left(1 - \frac{1}{t}\right)^{-1}$. Again by Holder's inequality, $\|Z\|_{\mathrm{nuc}} \leqslant \|Z\|_{t*}\|\mathrm{Id}_n\|_t$ and for $t = \frac{\log n}{C^*}$, $\|\mathrm{Id}_n\|_t = O(1)$. Thus $\|Z\|_{\mathrm{nuc}} = O(1)$. By the triangle inequality of trace norm, we have $\|M\|_{\mathrm{nuc}} \leqslant \|ZQ\|_{\mathrm{nuc}} + \|QZ\|_{\mathrm{nuc}} \leqslant O(1)$. The result follows. □

The rest of the section will be devoted to and. We prove Lemma 4.13 in Section 4.3.1 and Lemma 4.14 in Section 4.3.2. The central tool to both results will the following intermediate lemma, which, informally speaking, states that for any feasible solution Z to program P.1 we have $\langle Z, Q^2\rangle \approx \langle Z, \tilde{Q}^2\rangle$.

Lemma 4.15. *Let $(Q, \tilde{Q}, v)$ be an instance of Problem 4.9 satisfying Eq. $(\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*})$ with parameters $\alpha, \beta, \gamma, \delta^*, \zeta \geqslant 0, C^* > 0$. Let $Z \in \mathbb{R}^{n\times n}$ be a feasible solution to the program P.1. Then*

$$\left|\langle Z, Q^2 - \tilde{Q}^2\rangle\right| \leqslant 4\zeta \cdot \gamma^2 \cdot \beta.$$

Proof. We may rewrite

$$Q^2 - \tilde{Q}^2 = \frac{1}{2}\Big(Q - \tilde{Q}\Big)\Big(Q + \tilde{Q}\Big) + \frac{1}{2}\Big(Q + \tilde{Q}\Big)\Big(Q - \tilde{Q}\Big).$$

So let $Z' := Z \cdot \Big(\tilde{Q} + Q\Big)$ and $Z'' := \Big(\tilde{Q} + Q\Big)Z$, then we have

$$\begin{aligned}\left|\langle Z, Q^2 - \tilde{Q}^2\rangle\right| &= \left|\langle Z', Q - \tilde{Q}\rangle + \langle Z'', Q - \tilde{Q}\rangle\right| \\ &\leqslant (\|Z'\|_\infty + \|Z''\|_\infty)\left\|Q - \tilde{Q}\right\|_1 \\ &\leqslant (\|Z'\|_\infty + \|Z''\|_\infty) \cdot \beta n.\end{aligned}$$

Since by the perturbation conditions the matrix $Q - \tilde{Q}$ has bounded ℓ_1 norm, it suffices to upper bound $\|Z'\|_\infty$ and $\|Z''\|_\infty$. So notice that for any $i, j \in [n]$

$$\begin{aligned}\left|Z'_{ij}\right| &= \left|\langle Z_i, \left(Q + \tilde{Q}\right)_j \rangle\right| \\ &\leqslant \frac{\gamma^2}{n} \sum_{\ell \in [n]} \left|Q_{j\ell} - \tilde{Q}_{j\ell}\right| \\ &= 2\zeta \frac{\gamma^2}{n} .\end{aligned}$$

Applying the same reasoning to $\left|Z''_{ij}\right|$, the result follows. □

4.3.1 Lower bound for the optimum

Here we prove Lemma 4.13. Throughout the section we assume $(Q, \tilde{Q}, v)$ to be an instance of Problem 4.9 satisfying Eq. ($\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$) with parameters $\alpha, \beta, \gamma, \delta^*, \zeta \geqslant 0, C^* > 0$. We also define $t^* = (1 - 1/t)^{-1}$, where $t = \frac{\log n}{C^*}$. We will build our solution starting from a feasible solution to a simpler program than P.1 and then modifying it to satisfy the missing constraints. To this end consider the semidefinite program:

$$\text{maximize } \langle Z, Q^2 \rangle \qquad \text{subject to } \begin{Bmatrix} \operatorname{Tr} Z^{t^*} \leqslant 1 \\ Z \succeq 0 \end{Bmatrix} \tag{P.2}$$

It is easy to compute its optimum.

Lemma 4.16. *Consider the program P.2. Then* $\text{opt}_{P.2} = 1$.

Proof. Applying Hölder inequality

$$\langle Z, Q^2 \rangle \leqslant \|Z\|_{t^*} \cdot \left\|Q^2\right\|_t \leqslant 1 .$$

To see that $\text{opt}_{P.2} \geqslant 1$ instead let $Z = Q^{2t-2}$. Here Z is a feasible solution because $\operatorname{Tr} Z^{t^*} = \operatorname{Tr} Q^{(2t-2)\cdot t^*} = \operatorname{Tr} Q^{2t} = 1$ and it is clearly positive semidefinite. We also have

$$\langle Z, Q^2 \rangle = \operatorname{Tr} Q^{2t} = 1 .$$

□

We can use Lemma 4.16 to show that, for $Q = \tilde{Q}$ (that is, $\beta = 0$), we can obtain a nearly solution to program P.1 from a (specific) optimal solution to program P.2. Solutions of program P.2 may have large entries and so are not feasible solutions to program P.1 in general. Thus, the plan is to project the optimal solution Q^{2t-2} for program P.2 to the space spanned by its columns with bounded entries. Since entries of Q are also bounded, this

projection will not decrease significantly the correlation between our new matrix and Q. Consider the program:

$$\text{maximize } \langle Z, Q^2 \rangle \qquad \text{subject to } \begin{Bmatrix} \operatorname{Tr} Z^{t^*} \leqslant 1 \\ Z \succeq 0 \\ \forall i \in [n], \quad Z_{ii} \leqslant \dfrac{\gamma^2}{n} \end{Bmatrix} \tag{P.3}$$

Let $Z^* = Q^{2t-2}$ and let S be the set of large diagonal entries of Z^*, that is $S := \left\{ i \;\middle|\; Z^*_{ii} > \frac{\gamma^2}{n} \right\}$. Denote by Π_S the projector into the subspace spanned by the columns of Z^* with index in S. We first observe a simple fact concerning Z^*, Π_S.

Fact 4.17. *Consider the program P.3 and let*

$$Z := (\mathrm{Id} - \Pi_S) Z^* (\mathrm{Id} - \Pi_S) \,.$$

For any integer $t > 0$ we have

$$\operatorname{Tr}\left(Z^{*t}\right) \geqslant \operatorname{Tr}\left(Z^t\right)$$

Proof. We denote the dimension of space spanned by S as k. Then we denote i-th smallest eigenvalue of Z^* as $\lambda_i(Z^*)$ and the i-th smallest eigenvalue of the matrix Z as $\lambda_i(Z)$. We note that $\operatorname{Tr} Z^t = \sum_{i=1}^n \lambda_i(Z)^t$, and $\operatorname{Tr} Z^{*t} = \sum_{i=1}^n \lambda_i(Z^*)^t$. Furthermore, since Z^* is positive semidefinite, Z is also positive semidefinite and the smallest k eigenvalues of Z are given by 0. Thus, it's sufficient to prove that $\lambda_{k+i}(Z^*) \geqslant \lambda_{k+i}(Z)$ for any $1 \leqslant i \leqslant n-k$.

To prove this, we denote the space spanned by the top $n-k-i+1$ eigenvectors of Z as U_i. Applying Courant-Fischer min-max Lemma B.104, we have

$$\begin{aligned} \lambda_{k+i}(Z^*) &\geqslant \min_x \{ \mathbb{R}_{Z^*}(x) \mid x \in U_i \text{ and } x \neq 0 \} \\ &= \min_x \{ \mathbb{R}_{Z}(x) \mid x \in U_i \text{ and } x \neq 0 \} \\ &= \lambda_{i+k}(Z) \end{aligned}$$

This completes the proof. □

Using this fact, we prove that Z is close to an optimal solution for P.3.

Lemma 4.18. *Consider the program P.3 and let*

$$Z := (\mathrm{Id} - \Pi_S) Z^* (\mathrm{Id} - \Pi_S) \,.$$

Then

$$\langle Z, Q^2 \rangle \geqslant 1 - e^{2C^*} \cdot \frac{\zeta^2}{\sqrt{\gamma}} \,.$$

Proof. By construction Z is the projection of Z^* into the space orthogonal to the subspace spanned by columns of Z^* with index in S. The matrix Z is clearly positive semidefinite. Moreover, as shown in Fact 4.17, since it is a projection fo Z^* to a subspace, the eigenvalues of Z^* dominates the eigenvalues of Z and thus $\operatorname{Tr} Z^{t^*} \leqslant \operatorname{Tr} Z^{*t^*}$. It follows that Z is a feasible solution to Eq. (P.3).

It remains to lower bound its objective value. Now by Lemma 4.16,

$$\langle Z, Q^2 \rangle \geqslant 1 - \left| \langle Z - Z^*, Q^2 \rangle \right|,$$

hence it suffices to bound $\left| \langle Z - Z^*, Q^2 \rangle \right|$. Then

$$\begin{aligned}
\left| \langle Z - Z^*, Q^2 \rangle \right| &\leqslant \sum_{i \in S} \sum_{j \in [n]} \left| Z^*_{ij} \right| \cdot \left| \left(Q^2\right)_{ij} \right| \\
&\leqslant \sum_{i \in S} \sum_{j \in [n]} \sqrt{Z^*_{ii} \cdot Z^*_{jj}} \cdot \left| \left(Q^2\right)_{ij} \right| \\
&\leqslant \left(\sum_{i \in S} Z^*_{ii} \right)^{1/2} \cdot \left(\sum_{i \in S} \left(\sum_{j \in [n]} \sqrt{Z^*_{jj}} \left| \left(Q^2\right)_{ij} \right| \right)^2 \right)^{1/2} \\
&\leqslant \left(\sum_{i \in S} Z^*_{ii} \right)^{1/2} \cdot \left(\sum_{i \in [n]} \left(\sum_{j \in [n]} \sqrt{Z^*_{jj}} \left| \left(Q^2\right)_{ij} \right| \right)^2 \right)^{1/2} \\
&\leqslant \left(\sum_{i \in S} Z^*_{ii} \right)^{1/2} \cdot \left\| |Q^2| \right\| \cdot \left(\operatorname{Tr} Z^* \right)^{1/2}.
\end{aligned}$$

By the second sensitivity condition on Q, $\||Q|\| \leqslant \zeta$ and thus $\left\||Q^2|\right\| \leqslant \zeta^2$. By the first sensitivity condition,

$$\sum_{i \in S} Z^*_{ii} \leqslant \frac{n}{\gamma^2} \sum_{i \in [n]} \left(Z^*_{ii} \right)^2 \leqslant \frac{n}{\gamma^2} \cdot \frac{\gamma}{n} \left(\sum_{i \in [n]} Z^*_{ii} \right)^2 \leqslant \frac{1}{\gamma} \left(\operatorname{Tr} Z^* \right)^2.$$

By Lemma B.103 and choice of $t = \frac{1}{C^*} \cdot \log n$

$$\operatorname{Tr} Z^* \leqslant e^{C^*} \cdot \operatorname{Tr} Z^{*t^*}.$$

It follows that

$$\left| \langle Z - Z^*, Q^2 \rangle \right| \leqslant e^{\frac{3}{2} C^*} \cdot \frac{\zeta^2}{\sqrt{\gamma}},$$

concluding the proof. □

Using Lemma 4.18, we can prove Lemma 4.13. The idea is that since by perturbation conditions Q and $\tilde{Q}$ are close in L_1-norm, the product $\langle Z, \tilde{Q}^2 \rangle$ will not be too far from $\langle Z, Q^2 \rangle$.

Proof of Lemma 4.13. Any feasible solution for the program P.3 is a feasible solution for P.1. Choosing $Z \in \mathbb{R}^{n \times n}$ as constructed in Lemma 4.18 we get

$$\begin{aligned} \langle Z, \tilde{Q}^2 \rangle &\geqslant \langle Z, Q^2 \rangle - \langle Z, \tilde{Q}^2 - Q^2 \rangle \\ &\geqslant 1 - e^{2C^*} \cdot \frac{\zeta^2}{\sqrt{\gamma}} - 4\zeta \cdot \gamma^2 \cdot \beta \,, \end{aligned}$$

where we applied both Lemma 4.15 and Lemma 4.18. □

4.3.2 Correlation of nearly-optimal solutions

We prove here Lemma 4.14. Again, we assume $(Q, \tilde{Q}, v)$ to be an instance of Problem 4.9 satisfying Eq. ($\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$) with parameters $\alpha, \beta, \gamma, \delta^*, \zeta \geqslant 0, C^* > 0$ and we let $t^* = (1 - 1/t)^{-1}$, where $t = \frac{\log n}{C^*}$. We start by showing that any nearly optimal solution to program P.3 is non trivially correlated with the vector v in the following sense.

Lemma 4.19. *Consider the program P.3. Any feasible solution $Z \in \mathbb{R}^{n \times n}$ such that $\langle Z, Q^2 \rangle \geqslant 1 - \frac{\delta^*}{2}$ satisfies*

$$\langle ZQ + QZ, vv^\mathsf{T} \rangle \geqslant \frac{\delta^*}{2\alpha} \,.$$

Proof. Let Z be a feasible solution such that $\langle Z, Q^2 \rangle \geqslant 1 - \frac{\delta^*}{2}$. Then

$$\begin{aligned} 1 - \frac{\delta^*}{2} &\leqslant \langle Z, Q^2 \rangle \\ &= \langle Z, \left[(Q - \alpha \cdot vv^\mathsf{T}) + \alpha \cdot vv^\mathsf{T} \right]^2 \rangle \\ &= \langle Z, (Q - \alpha \cdot vv^\mathsf{T}) \rangle + \alpha \langle Z, (Q - \alpha \cdot vv^\mathsf{T}) vv^\mathsf{T} + vv^\mathsf{T} (Q - \alpha \cdot vv^\mathsf{T}) \rangle + \alpha^2 \langle Z, vv^\mathsf{T} \rangle \,. \end{aligned}$$

Now since Q satisfies the correlation conditions in Eq. ($\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$)

$$\begin{aligned} 1 - \frac{\delta^*}{2} &\leqslant 1 - \delta^* + \alpha \langle Z, (Q - \alpha \cdot vv^\mathsf{T}) vv^\mathsf{T} + vv^\mathsf{T} (Q - \alpha \cdot vv^\mathsf{T}) \rangle + \alpha^2 \langle Z, vv^\mathsf{T} \rangle \\ &\leqslant 1 - \delta^* + \alpha \langle ZQ + QZ, vv^\mathsf{T} \rangle - \alpha^2 \langle Z, vv^\mathsf{T} \rangle \\ &\leqslant 1 - \delta^* + \alpha \langle ZQ + QZ, vv^\mathsf{T} \rangle \,. \end{aligned}$$

Rearranging the result follows. □

While Lemma 4.19 shows how Z contains non-trivial information about v, from an algorithmic point of view the result is of little use since we do not have access to Q. However, as Q and $\tilde{Q}$ are close, *also* the matrix $Z\tilde{Q} + \tilde{Q}Z$ will be correlated with vv^T.

Lemma 4.20. *Let $Z \in \mathbb{R}^{n\times n}$ be a feasible solution to program P.3 such that*

$$\langle ZQ + QZ, vv^\mathsf{T}\rangle \geqslant \frac{\delta^*}{2\alpha}\,.$$

Then

$$\langle Z\tilde{Q} + \tilde{Q}Z, vv^\mathsf{T}\rangle \geqslant \frac{\delta^*}{2\alpha} - 2\frac{\gamma^2}{n}\beta\,.$$

Proof. We may rewrite the product as,

$$\begin{aligned}\langle Z\tilde{Q} + \tilde{Q}Z, vv^\mathsf{T}\rangle &= \langle Z\Big(\tilde{Q} - Q + Q\Big) + \Big(\tilde{Q} - Q + Q\Big)Z, vv^\mathsf{T}\rangle \\ &= \langle Z\Big(\tilde{Q} - Q\Big) + \Big(\tilde{Q} - Q\Big)Z, vv^\mathsf{T}\rangle + \langle ZQ + QZ, vv^\mathsf{T}\rangle\,.\end{aligned}$$

Since v is a flat unit vector and using the conditions on $Q, \tilde{Q}$,

$$\begin{aligned}\langle Z\Big(\tilde{Q} - Q\Big), vv^\mathsf{T}\rangle &\leqslant \frac{1}{n}\sum_{i,j\in[n]}\left\langle Z_i, \Big(Q - \tilde{Q}\Big)_j\right\rangle \\ &\leqslant \frac{\gamma^2}{n^2}\big\|Q - \tilde{Q}\big\|_1 \\ &\leqslant \frac{\gamma^2}{n}\beta\,.\end{aligned}$$

An analogous computation for $\langle\Big(\tilde{Q} - Q\Big)Z, vv^\mathsf{T}\rangle$ concludes the proof. □

We can now prove Lemma 4.14.

Proof of Lemma 4.14. Let $Z \in \mathbb{R}^{n\times n}$ be a feasible solution to program P.1 satisfying

$$\langle Z, \tilde{Q}^2\rangle \geqslant 1 - e^{2C^*}\cdot\frac{\zeta^2}{\sqrt{\gamma}} - 4\zeta\cdot\gamma^2\cdot\beta\,.$$

By Lemma 4.15 we have

$$\langle Z, Q^2\rangle \geqslant 1 - e^{2C^*}\cdot\frac{\zeta^2}{\sqrt{\gamma}} - 8\zeta\cdot\gamma^2\cdot\beta\,.$$

By assumptions on the parameters, and applying Lemma 4.19 and Lemma 4.20 the result follows. □

4.4 Robust recovery for stochastic block model

In this section we show how to use Algorithm 4.10 to obtain a robust algorithm for the stochastic block model and prove Theorem 4.2. First recall our settings.

Problem 4.21. Let $\delta, \rho > 0$. For a pair $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$ where $d = (1+\delta)\frac{4}{\varepsilon^2}$ we are given δ and a graph G° obtained from $\mathbf{G}$ applying at most $\rho \cdot n$ arbitrary edge edits.[20] The goal is then to return a *unit* vector $\hat{\mathbf{x}} \in \mathbb{R}^n$ that is $\delta^{O(1)}$-correlated with $\mathbf{x}$ in the sense that $\langle \hat{\mathbf{x}}, \mathbf{x} \rangle^2 \geqslant \delta^{O(1)} \|\mathbf{x}\|^2$.

Remark 4.22 (Learning the distribution parameters δ, d, ε). In principle the parameter d, ε of the distribution may not be known. However, we can easily learn good bounds on d, ε. A constant upper bound on d can be obtained from G° with a simple counting argument. Since $d \cdot \varepsilon^2 > 1$ (as otherwise the problem is impossible to solve), we can lower bound ε by $1/\sqrt{d}$. These constant approximations are precise enough for our purposes. For this reason, we will simply assume that d, ε are known.

We state here the main result of the section. Theorem 4.2 follows as a corollary.

Theorem 4.23. *Let $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$ where $d \geqslant (1+\delta)\frac{4}{\varepsilon^2}$ for some constant $\delta > 0$. Suppose G° is an arbitrary graph that differs from $\mathbf{G}$ in at most $\rho \cdot n$ edges for*

$$\rho \leqslant \left(\frac{1}{\delta} \cdot \log \frac{1}{\varepsilon} \right)^{-O(1/\delta)} .$$

Then, there exists a $n^{poly(1/\delta, \log d)}$-time algorithm (Algorithm 4.27) that, given G°, δ, computes a n-by-n matrix $\mathbf{M}$ such that

$$\langle \mathbf{M}, \mathbf{x}\mathbf{x}^\mathsf{T} \rangle \geqslant \|\mathbf{x}\|^2 \cdot \|\mathbf{M}\|_{\mathrm{nuc}} \cdot \delta^{O(1)} ,$$

with probability $1 - o(1)$.

We split the proof of Theorem 4.23 in two steps. First, we apply the results of section Section 4.3. That is, we show there are matrix polynomials $\mathbf{Q}, \tilde{Q}$, computable respectively from the adjacency matrices $\mathbf{Y}, Y^\circ$ of $\mathbf{G}$ and G° in polynomial time, such that with *constant* probability $(\mathbf{Q}, \tilde{Q}, \mathbf{x})$ satisfies $\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$ for some meaningful set of parameters. Second, in Section 4.4.2 we show how to turn this into a high probability statement.

Notice that as an immediate consequence of Theorem 4.23, applying the rounding Lemma 4.12 one gets the following corollary, which is a stronger version of the main theorem.

Corollary 4.24 (Formal version of the main theorem)**.** *For n large enough, let $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$ where $d \geqslant (1+\delta)\frac{4}{\varepsilon^2}$ for some constant $\delta > 0$. Suppose G° is an arbitrary graph that*

[20] Hence we assume $\rho \cdot n$ to be an integer

differs from $\mathbf{G}$ *in at most* $\rho \cdot n$ *edges for* ρ *as in Theorem 4.23. Then, there exists a polynomial-time algorithm that, given* G°, δ, *computes a* n*-dimensional unit vector* $\hat{\mathbf{x}}$ *such that*

$$\mathop{\mathbb{E}}_{\hat{\mathbf{x}}} \langle \hat{\mathbf{x}}, \mathbf{x} \rangle^2 \geqslant \|\mathbf{x}\|^2 \cdot \delta^{O(1)} .$$

with high probability over $(\mathbf{x}, \mathbf{G})$.

4.4.1 Applying the meta-algorithm to the stochastic block model

In this section we prove the following weaker version of Theorem 4.23.

Theorem 4.25. *The result in Theorem 4.23 holds with probability at least* 0.99.

For a graph G on n vertices, recall the definition of the centered adjacency matrix $Y(G) \in \mathbb{R}^{n \times n}$, where for $i, j \in [n]$

$$Y_{ij}(G) := \begin{cases} 1 - \frac{d}{n} & \text{if } ij \in E(G) \\ -\frac{d}{n} & \text{if } ij \notin E(G), \\ 0 & \text{if } i = j . \end{cases}$$

When the context is clear we simply write Y. As already discussed in Section 4.1 and as we will more extensively explain in Section 4.5 and Appendix B.1, we consider the following truncated version of the adjacency matrix.

Definition 4.26 (Δ-truncated adjacency matrix)**.** Let G be a graph over $[n]$ and let $\Delta \geqslant 0$ be an integer. We define $\overline{Y}(G) \in \left\{-\frac{d}{n}, 0, 1 - \frac{d}{n}\right\}^{n \times n}$ to be the matrix with entries

$$\overline{Y}_{ij}(G) := \begin{cases} 1 - \frac{d}{n} & \text{if } ij \in E(G) \text{ and } d^G(i) \leqslant \Delta,\ d^G(j) \leqslant \Delta \\ -\frac{d}{n} & \text{if } ij \notin E(G) \text{ and } d^G(i) \leqslant \Delta,\ d^G(j) \leqslant \Delta \\ 0 & \text{otherwise.} \end{cases}$$

for any $i, j \in [n]$. We we will denote the graph obtained from G by removing vertices of degree larger than Δ by $\overline{G}$.

The matrices of interest will be the block self-avoiding walk matrix polynomials as defined in Eq. (4.2.2). We restate the definition here. For a fixed integer $s > 0$, and a centered adjacency matrix Y of a graph G, let $Q^{(s)}(Y) : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ be the matrix polynomial with entries:

$$Q^{(s)}_{ij}(Y(G)) = \begin{cases} \frac{1}{|\mathrm{SAW}^s_{ij}|} \left(\frac{2n}{\varepsilon \cdot d}\right)^s \sum_{H \in \mathrm{SAW}^s_{ij}} Y_H(G) & \text{if } i \neq j , \\ 0 & \text{otherwise.} \end{cases} \tag{4.4.1}$$

For $\mathbf{G}, G^\circ$ as in Theorem 4.25, we simplify $Y(\mathbf{G})$ to $\mathbf{Y}$ and $Y(G^\circ)$ to Y°. Similarly we define $\overline{\mathbf{Y}}$ and $\overline{Y^\circ}$ to be the Δ-truncated centered adjacency matrices of $\mathbf{G}$ and G°.

The choice of the truncation threshold $\Delta > 0$ depends on the proof of the Eq. (4.1.11). As explained in Section 4.1, Δ has to be a constant if we hope to solve Problem 4.21 for constant $\rho \in [0,1]$. Indeed the smaller the choice of Δ the larger the constant fraction ρ of corruptions that our algorithm can tolerate. On the other hand, the threshold cannot be too small as otherwise we may loose too much information. With these mild conditions in mind, for technical reasons, we define

$$\Delta = \left\lceil \max\left\{ 128e^4 d^4, 40As d, 2\log(2As) + 12A\tau s^2 \cdot \log 2 + 8A^2\tau^2 s^2 \left(\log \frac{6}{\varepsilon}\right)^2 \right\} \right\rceil, \tag{4.4.2}$$

where $A \geqslant 1000 s^3$ and

$$\tau = As \log \frac{6}{\varepsilon}. \tag{4.4.3}$$

We remark that this is not a delicate choice in the sense that the results we will prove hold for many larger values of Δ, as long as the fraction of corruptions ρ is a small constant. However, it is important to observe that Δ is polynomial in (d, s), this will turn out to be especially useful in computing moments of $Q^{(s)}(\overline{\mathbf{Y}})$ and $Q^{(s)}(\overline{Y^\circ})$. Finally, notice that Δ mildly depends on ε, this suggests that the fraction of corruption that our algorithm can tolerate decreases as $\log \frac{1}{\varepsilon}$ increases. We defer a more detailed discussion to Appendix B.1.

The algorithm we use is the following.

Algorithm 4.27 (Robust Recovery for SBM)**.**
Input: An instance $(G^\circ, d, \varepsilon)$ of Problem 4.21.

1. Fix Δ, A, τ as in Eq. (4.4.2) and $t = \log n/400$.
   Let $s \geqslant 10^{10}\left(1 + \frac{1}{\delta}\right)\left(\max\left\{\log d, \log\log \frac{2}{\varepsilon}, 1, \log \frac{1}{\delta}\right\}\right)^2$.
   Compute $Q^{(s)}(\overline{Y^\circ})$ where $\overline{Y^\circ}$ is the Δ-truncation of Y°.

2. Run Algorithm 4.10 on the rescaled matrix $Q^{(s)}(\overline{Y^\circ})/\left(\mathbb{E}\,\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right)^{1/2t}$ with $C^* = 400$ and a large enough universal constant $\gamma > 0$.

The proof that Algorithm 4.27 solves Problem 4.21 essentially amounts to showing that the tuple

$$\left(Q^{(s)}(\overline{\mathbf{Y}})/\left(\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right)^{1/2t}, Q^{(s)}(\overline{Y^\circ})/\left(\mathbb{E}\,\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right)^{1/2t}, \mathbf{x}/\|\mathbf{x}\|\right)$$

satisfies $\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$ for a meaningful range of parameters, with sufficiently large probability over the realization of $\mathbf{G}, \mathbf{x}$, *for all* admissible G°. In particular, the main additional

ingredients needed to prove Theorem 4.25 consists of bounds on the moments of $Q^{(s)}(\overline{\mathbf{Y}})$ and $Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}$. We remark that, since in practice we don't have access to $\left(\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right)^{1/2t}$, we can only scale down $Q^{(s)}(\overline{Y^\circ})$ by the expectation of $\left(\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right)^{1/2t}$. It turns out however that this random variable concentrates around its expectation, so the error introduced with this rescaling will be negligible.

For simplicity let us write

$$\begin{aligned}
\mathbf{Q} &:= Q^{(s)}(\overline{\mathbf{Y}}) / \left(\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right)^{1/2t} \\
\tilde{Q} &:= Q^{(s)}(\overline{Y^\circ}) / \left(\mathbb{E}\,\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right)^{1/2t}, \\
\overline{\mathbf{x}} &:= \mathbf{x}/\|\mathbf{x}\|\,.
\end{aligned}$$

The central step in the proof of Theorem 4.25 is the statement below.

Lemma 4.28. *Consider the settings of Theorem 4.25 Then the tuple* $(\mathbf{Q}, \tilde{Q}, \overline{\mathbf{x}})$ *satisfy Eq.* $(\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*})$ *with parameters*

1. $\alpha \leqslant 2$,
2. $\delta^* = \delta^{O(1)}$,
3. $\beta = 2\rho \cdot \Delta^{s+1}$,
4. $C^* = 400$
5. $\zeta = 10\Delta^{s+1} \cdot e^{2C^*}$,

and a constant $\gamma > 0$ *depending only on* $s, C^*, \log(1/\varepsilon)$, *with probability at least* 0.99.

We prove Lemma 4.28 in Section 4.4.1.1 and directly use it here.

Proof of Theorem 4.25. In order to apply Theorem 4.11 we need to show that

$$\begin{aligned}
e^{2C^*}\frac{\zeta^2}{\sqrt{\gamma}} &\leqslant \frac{\delta^*}{4}, \\
8\zeta \cdot \gamma^2 \cdot \beta &\leqslant \frac{\delta^*}{4}.
\end{aligned}$$

By Lemma 4.28 and choice of Δ, s, A, τ for $\rho \leqslant \frac{\delta^*}{4}\left(100 \cdot \gamma^6 \cdot \Delta^{2s+2} \cdot e^{2C^*}\right)^{-1}$ a direct application of Theorem 4.11 shows that with probability at least 0.99 the algorithm finds a n-times-n matrix $\mathbf{M}$ satisfying

$$\langle \mathbf{M}, \mathbf{x}\mathbf{x}^\mathsf{T} \rangle \geqslant \left\|\mathbf{x}\mathbf{x}^\mathsf{T}\right\|_\mathrm{F} \cdot \|\mathbf{M}\|_\mathrm{nuc} \cdot \delta^{O(1)}.$$

For the running time dependence, we note that the entries in $Q^{(s)}$ are degree s polynomials in n variables, thus we can evaluate the matrix $Q^{(s)}$ in time $n^{O(s)}$. Then the convex programming, matrix powering and rounding can all be solved in $\mathrm{poly}(n)$ time as we take

$$s \geqslant 10^{10}\left(1+\frac{1}{\delta}\right)\left(\max\left\{\log d, \log\log\frac{2}{\varepsilon}, 1, \log\frac{1}{\delta}\right\}\right)^2 .$$

Therefore the running time of the algorithm can be bounded by $n^{\mathrm{poly}(1/\delta,\log d)}$ □

4.4.1.1 The stochastic block model satisfies correlation, sensitivity and perturbations constraints

In this section we prove Lemma 4.28. Our central tool will be the following concentration inequalities, which we prove in Section 4.5.

Lemma 4.29. *Consider the settings of Theorem 4.25. Let Δ, A, τ be as defined in Eq. (4.4.2) and let $t = \frac{\log n}{400}$. Then*

$$\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right]^{1/2t} = (1\pm o(1))\,\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right]^{1/2t}, \tag{4.4.4}$$

$$\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^{\mathsf{T}}\right)^{2t}\right]^{1/2t} \leqslant (1+\delta)^{-1/10}\,\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right]^{1/2t}, \tag{4.4.5}$$

$$\sum_{i\in[n]}\left(\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t-2}\right)_{ii}^2 \leqslant \frac{\gamma}{n}\cdot\left(\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t-2}\right)^2, \tag{4.4.6}$$

for a universal constant $\gamma > 0$, with probability at least 0.99. Moreover

$$\frac{n}{2} \leqslant \mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right]^{1/2t}. \tag{4.4.7}$$

Remark 4.30. The careful reader may have noticed that so far, we never explicitly compute the quantity $\left(\mathbb{E}\,\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right)^{1/2t}$. In practice, for each value of n, we need to compute this expectation to sufficiently close precision only *once* and not at every run of the algorithm. This can be done efficiently and accurately (with high probability) by sampling several graphs from $\mathsf{SBM}_n(d,\varepsilon)$, compute the corresponding $2t$-Schatten norm and take the average.

By scaling $\mathrm{Tr}\,\mathbf{Q}^{2t} = 1$. We show now that $\mathbf{Q}$ satisfy the second correlation constraint in $\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$.

Lemma 4.31. *Consider the settings of Theorem 4.25. Suppose Eq. (4.4.4) and Eq. (4.4.5) are satisfied. Then for some $0 < \alpha \leqslant 2$ and $\delta^* = \delta^{O(1)}$,*

$$\mathrm{Tr}\left(\mathbf{Q} - \alpha\cdot\overline{\mathbf{x}}\overline{\mathbf{x}}^{\mathsf{T}}\right)^{2t} \leqslant (1-\delta^*)^{-2t}.$$

Proof. By Eq. (4.4.4) and Eq. (4.4.5)

$$\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^{\mathsf{T}}\right)^{2t}\right]^{1/2t} \leqslant (1+\delta)^{-1/11}\left[\mathrm{Tr}\left(Q^{(s)}((\overline{\mathbf{Y}})\right)^{2t}\right]^{1/2t}.$$

By Eq. (4.4.7) $\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right]^{1/2t} \geqslant n/2$, so rescaling by $1/\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right]^{1/2t}$, the polynomial satisfy the correlation constraints in $\mathcal{A}_{\alpha,\beta,\gamma,\delta^*,\zeta,C^*}$ with some $0 < \alpha \leqslant 2$ and $\delta^* = \delta^{O(1)}$. □

Next we show how by construction both $\mathbf{Q}$ and $\tilde{Q}$ have columns bounded in ℓ_1-norm.

Lemma 4.32. *Consider the settings of Theorem 4.25. Suppose Eq. (4.4.4) and Eq. (4.4.5) are satisfied. Then*

$$\max_{i\in[n]} \sum_{j\in[n]} \left|\mathbf{Q}_{ij}\right| \leqslant 2\Delta^{s+1}$$
$$\max_{i\in[n]} \sum_{j\in[n]} \left|\tilde{Q}_{ij}\right| \leqslant 2\Delta^{s+1}.$$

Proof. It suffices to show a bound for any graph with no vertex with degree larger than Δ. So let $\overline{G}$ be such a graph. Let $\overline{Y}$ be its centered adjacency matrix (notice that its adjacency matrix is also its Δ-truncated adjacency matrix, so we may use both definition interchangeably). For $u, v \in V(\overline{G})$ and $\ell \in [s]$, define the set $\mathcal{W}_{u,\ell}(\overline{G}) := \left\{W \in \mathrm{SAW}_u^s \;\middle|\; \left|E(W) \setminus E(\overline{G})\right| = \ell\right\}$. That is $\mathcal{W}_{u,\ell}(\overline{G})$ contains the set of self-avoiding walks over K_n starting from u which have exactly ℓ edges not in $E(\overline{G})$. Notice that for any $W \in \mathcal{W}_{u,\ell}(\overline{G})$, $\overline{Y}_W \leqslant \left(\frac{d}{n}\right)^\ell$, and $Q_{uv}^{(s)}(\overline{Y}) \leqslant n \sum_{W\in\mathrm{SAW}_{uv}^s} \overline{Y}_W$. Recall that we denote the set of vertices at distance s from $u \in V(\overline{G})$ by $N_{\overline{G}}^s(u)$. Now, for any $u \in [n]$, by assumption $\left|N_{\overline{G}}^s(u)\right| \leqslant \Delta^s$. So

$$\begin{aligned}\sum_{v\in[n]} \left|Q_{uv}^{(s)}(\overline{Y})\right| &\leqslant n \cdot \sum_{\ell\in[s-1]} \sum_{W\in\mathcal{W}_{u,\ell}(H)} \left|\overline{Y}_W\right| \\ &\leqslant n \cdot \sum_{\ell\in[s-1]} \sum_{W\in\mathcal{W}_{u,\ell}(H)} \left(\frac{d}{n}\right)^\ell.\end{aligned}$$

For any $\ell \in [s]$, there are at most $\Delta^{s-\ell} \cdot n^\ell$ self-avoiding walks in $\mathcal{W}_{u,\ell}(\overline{G})$, thus

$$\sum_{v\in[n]} \left|Q_{uv}^{(s)}(\overline{Y})\right| \leqslant n \cdot s \cdot \Delta^s.$$

Since Eq. (4.4.4) and Eq. (4.4.5) are verified by assumption, applying the analysis above to $\overline{\mathbf{Y}}$, $\overline{Y^\circ}$ and scaling down the inequalities the result follows. □

Eq. (4.4.6) immediately implies the remaining sensitivity constraint on $\mathbf{Q}$. Thus we only need to show that $\mathbf{Q}$ and $\tilde{Q}$ are close in a ℓ_1-norm sense. To do that, we need to first argue how many different edges $\overline{\mathbf{G}}$ and $\overline{G^\circ}$ may have.

Fact 4.33. *Let $\Delta \geqslant 0$ be an integer. Let G, G° be graphs on n vertices that differs in at most $\rho \cdot n$ edges, for some $\rho \geqslant 0$. Let $\overline{G}$ and $\overline{G^\circ}$ be respectively the graphs obtained from G and G° by removing all vertices with degree larger than Δ. Then $\overline{G}$ and $\overline{G^\circ}$ differs by at most $2\rho \cdot n \cdot \Delta$ edges.*

Proof. $\overline{G}$ and G° differs by $\rho \cdot n$ edges. Each such edge changes the degree of at most 2 vertices, thus after the truncation $\overline{G}$ and $\overline{G^\circ}$ differs by at most $2\rho \cdot n \cdot \Delta$ edges. □

Next we show that if G and G° are two arbitrary graphs (with bounded maximum degree) which differ by at most one edge, than the matrices $Q^{(s)}(\overline{Y}(G))$ and $Q^{(s)}(\overline{Y}(G^\circ))$ are close.

Lemma 4.34. *Let $\Delta \geqslant 0$ be an integer. Let $\overline{G}$ be a graph on n vertices with maximum degree at most Δ. Let $\overline{G^\circ} = \overline{G} + uv$ for some $u, v \in V(\overline{G})$ such that $uv \notin E(\overline{G})$ and $d^{\overline{G^\circ}}(u), d^{\overline{G^\circ}}(v) \leqslant \Delta$. Denote respectively by $\overline{Y}, \overline{Y^\circ}$ the centered adjacency matrices of $\overline{G}$ and $\overline{G^\circ}$. Then*

$$\left\| Q^{(s)}(\overline{Y}) - Q^{(s)}(\overline{Y^\circ}) \right\|_1 \leqslant n \cdot \Delta^s \,.$$

Proof. It suffices to consider length-s self-avoiding walks traversing uv. We reuse the notation introduced in Lemma 4.32, thus for $a, b \in V(\overline{G^\circ})$ let $\mathcal{W}_{a,b,\ell}(\overline{G^\circ}) := \left\{ W \in \mathrm{SAW}^s_{ab} \;\middle|\; \left| E(W) \setminus E(\overline{G^\circ}) \right| = \ell \right\}$. Furthermore, for $\ell \leqslant s$ consider the set

$$\mathcal{W}^{uv}_{a,b,\ell}(\overline{G^\circ}) := \left\{ W \in \mathcal{W}_{a,b,\ell}(\overline{G^\circ}) \;\middle|\; uv \in E(W) \right\}.$$

In other words, we look at the subsets of self-avoiding walks in $\mathcal{W}_{a,b,\ell}$ containing the edge uv. Then,

$$\begin{aligned}
\left\| Q^{(s)}(\overline{Y}) - Q^{(s)}(\overline{Y^\circ}) \right\|_1 &\leqslant n \cdot \sum_{0 \leqslant \ell \leqslant s-1} \sum_{a,b \in [n]} \sum_{W \in \mathcal{W}^{uv}_{a,b,\ell}(\overline{G^\circ})} \left| \overline{Y}_W - \overline{Y^\circ}_W \right| \\
&\leqslant n \cdot \sum_{0 \leqslant \ell \leqslant s-1} \sum_{a,b \in [n]} \sum_{W \in \mathcal{W}^{uv}_{a,b,\ell}(\overline{G^\circ})} \left| \overline{Y}_W \right| + \left| \overline{Y^\circ}_W \right| \\
&\leqslant n \cdot \sum_{0 \leqslant \ell \leqslant s-1} \sum_{a,b \in [n]} \sum_{W \in \mathcal{W}^{uv}_{a,b,\ell}(\overline{G^\circ})} \left(\frac{d}{n} \right)^{\ell+1} + \left(\frac{d}{n} \right)^{\ell} \\
&= (1 + o(1))n \cdot \sum_{0 \leqslant \ell \leqslant s-1} \sum_{a,b \in [n]} \sum_{W \in \mathcal{W}^{uv}_{a,b,\ell}(\overline{G^\circ})} \left(\frac{d}{n} \right)^{\ell}.
\end{aligned}$$

For any $0 \leqslant \ell \leqslant s-1$ we have $\bigcup_{a,b\in[n]} \mathcal{W}^{uv}_{a,b,\ell}(\overline{G^\circ}) \leqslant s \cdot \Delta^{s-\ell-1} \cdot n^\ell$. It follows that

$$\left\| Q^{(s)}(\overline{Y}) - Q^{(s)}(\overline{Y^\circ}) \right\|_1 \leqslant n \cdot \Delta^s \,.$$

□

Now we can we show that $\mathbf{Q}$ and $\tilde{Q}$ satisfy the first perturbation constraint.

Lemma 4.35. *Consider the settings of Theorem 4.25. Suppose Eq. (4.4.4) holds. Then*

$$\frac{1}{n}\left\| \mathbf{Q} - \tilde{Q} \right\|_1 \leqslant 2\Delta^{s+1} \cdot \rho \,.$$

Proof. By Fact 4.33 there is a sequence $\left\{\overline{G}_i\right\}$ of $2\rho \cdot \Delta \cdot n$ graphs with maximum degree Δ such that $\overline{G}_{2\rho\cdot\Delta\cdot n} = \overline{G^\circ}$, $\overline{G}_1 = \overline{\mathbf{G}}$ and for any $i \in [\rho \cdot \Delta \cdot n - 1]$, $\overline{G}_i$ and $\overline{G}_{i+1}$ differ by at most one edge. For each $\overline{G}_i$ let Y_i be its centered adjacency matrix. Then

$$\begin{aligned}\left\| Q^{(s)}(\mathbf{Y}_1) - Q^{(s)}(\mathbf{Y}_{2\rho\cdot\Delta\cdot n}) \right\|_1 &\leqslant \sum_{i\in[2\rho\cdot\Delta\cdot n-1]} \left\| Q^{(s)}(Y_i) - Q^{(s)}(Y_{i+1}) \right\|_1 \\ &\leqslant 2\rho \cdot n^2 \cdot \Delta^{s+1} \,,\end{aligned}$$

where we used Lemma 4.34 in the last step. Since

$$\left\| Q^{(s)}(\overline{Y}(\mathbf{G})) - Q^{(s)}(\overline{Y}(G^\circ)) \right\|_1 = \left\| Q^{(s)}(Y(\overline{\mathbf{G}})) - Q^{(s)}(Y(\overline{G^\circ})) \right\|_1 ,$$

rescaling the lemma follows. □

Putting things together. We are ready to prove Lemma 4.28.

Proof of Lemma 4.28. We condition our analysis on the event that Eq. (4.4.4), Eq. (4.4.5) and Eq. (4.4.6) are verified, which by Lemma 4.29 happen with probability 0.99. Consider the tuple $\left(\mathbf{Q}, \tilde{Q}, \overline{\mathbf{x}}\right)$ By Lemma 4.31 the correlation constraints are satisfied for $\delta^* = \delta^{O(1)}$ and some $0 < \alpha \leqslant 2$. By Lemma 4.32 it holds that

$$\begin{aligned}\max_{i\in[n]} \sum_{j\in[n]} |\mathbf{Q}_i| &\leqslant \zeta \\ \max_{i\in[n]} \sum_{j\in[n]} \left|\tilde{Q}_i\right| &\leqslant \zeta \,,\end{aligned}$$

for any $\zeta \geqslant 2\Delta^{s+1}$. By Lemma 4.35, the tuple satisfies the remaining perturbation constraint for $\beta = 2\Delta^{s+1} \cdot \rho$. The result follows. □

4.4.2 Boosting the probability of success

In this section we conclude the proof of Theorem 4.23 showing how to increase the probability of success of Algorithm 4.27.

Definition 4.36. Let M be a function that takes as input a graph G° having $[n]$ as the set of vertices, and produces an $n \times n$ matrix $M(G^\circ)$ as output. Let $(\mathbf{G}, \mathbf{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$, and define $\mathcal{E}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho,\kappa}$ to be the event that the function M succeeds, up to robustness ρ, in providing an output that κ-correlates with the community labels.

More precisely, if $(\mathbf{G}, \mathbf{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$, then $\mathcal{E}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho,\kappa}$ is the event that for every graph G° that differs from $\mathbf{G}$ by at most ρn edges, we have

$$\langle M(G^\circ), \mathbf{x}\mathbf{x}^T \rangle \geqslant \kappa \cdot \|\mathbf{x}\|^2 \cdot \|M(G^\circ)\|_{\mathrm{nuc}}.$$

Theorem 4.25 implies that if M is Algorithm 4.27 with

$$s \geqslant 10^{10}\left(1 + \frac{1}{\delta}\right)\left(\max\left\{\log d, \log\log\frac{2}{\varepsilon}, 1, \log\frac{1}{\delta}\right\}\right)^2,$$

then if $d \geqslant (1+\delta)\frac{4}{\varepsilon^2}$ for some constant $\delta > 0$, and if $\rho \leqslant \left(\frac{1}{\delta} \cdot \log\frac{1}{\varepsilon}\right)^{-O(1/\delta)}$ and $\kappa \leqslant \delta^{O(1)}$, then for n large enough, we have

$$\mathbb{P}\left[\mathcal{E}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\right] \geqslant 0.99.$$

In this section we will prove that, at the expense of paying a negligible price in the robustness and correlation guarantees, we can boost the success probability and make it converge to 1 at a rate that is roughly exponential in $n^{\frac{1}{3}}$. In fact, we will prove a very general boosting result:

Theorem 4.37. *Let ε, d be such that $d \geqslant (1+\delta)\frac{4}{\varepsilon^2}$ for some constant $\delta > 0$. Let M be an arbitrary algorithm that takes as input a graph with $[n]$ as the set of vertices, and produces an $n \times n$ matrix as output. For every $0 < \rho' < \rho$ and every $0 < \kappa' < \kappa$, if*

$$\mathbb{P}\left[\mathcal{E}^{succ}_{n,d,\varepsilon,M,\rho,\kappa}\right] \geqslant \Omega\left(e^{-n^{\frac{1}{4}}}\right),$$

then for n is large enough, we have

$$\mathbb{P}\left[\mathcal{E}^{succ}_{n,d,\varepsilon,M,\rho',\kappa'}\right] \geqslant 1 - e^{-n^{\frac{1}{4}}}.$$

The theorem remains correct if we replace the exponent $\frac{1}{4}$ with any constant $\xi < \frac{1}{3}$.

It is worth noting that this boosting argument is not unique to the stochastic block model, and similar results hold in a wide range of estimation problems.

Combining Theorem 4.37 and Theorem 4.25, we obtain Theorem 4.23.

In order to prove Theorem 4.37, we will use a concentration of measure inequality known as the *blowing-up lemma*, and which is widely used in information theory to prove strong converse results. In order to describe this lemma, we need the following definition:

Definition 4.38. Let $\mathcal{Y}$ be an arbitrary finite set, and let $y = (y_1, \ldots, y_N) \in \mathcal{Y}^N$ and $y' = (y'_1, \ldots, y'_N) \in \mathcal{Y}^N$. The *Hamming distance* between y and y' is defined as

$$D_H(y, y') = \left|\left\{i \in [N] : \ y_i \neq y'_i\right\}\right|.$$

If $\mathcal{E}$ is a subset of $\mathcal{Y}^N$, we define the *ℓ-blowup* of $\mathcal{E}$ as:

$$\Gamma^\ell(\mathcal{E}) = \left\{y' \in \mathcal{Y}^n : \ \exists y \in \mathcal{E}, D_H(y, y') \leqslant \ell\right\}.$$

In other words, the ℓ-blowup of $\mathcal{E}$ is the set of elements of $\mathcal{Y}^n$ that are at a Hamming distance of at most ℓ from $\mathcal{E}$.

Roughly speaking, the blowing-up lemma states that if we have n independent random variables $\mathbf{y}_1, \ldots, \mathbf{y}_N$ taking values in a finite set $\mathcal{Y}$, and if $\mathcal{E} \subseteq \mathcal{Y}^n$ is an event whose probability is not too small, then if we "inflate" $\mathcal{E}$ a little by adding the elements of $\mathcal{Y}^N$ that are close to $\mathcal{E}$ in Hamming distance, then the probability of the event becomes $1 - o(1)$.

The blowing-up lemma was first introduced in [AGK76], and there are several versions of it (e.g., [Mar86] and [Mar96]). We will use the following version that was proved by Marton in [Mar96] (see also Lemma 3.6.2 in [RS+13]):

Lemma 4.39. *[Blowing up Lemma [Mar96]] Let $\mathbf{y}_1, \ldots, \mathbf{y}_N$ be N random variables taking values in the same finite set $\mathcal{Y}$ that can be of arbitrary size. If $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are independent (but not necessarily identically distributed), then for every $\mathcal{E} \subseteq \mathcal{Y}^N$ and every $\ell > \sqrt{\frac{N}{2}\log\left(\frac{1}{\mathbb{P}[\mathcal{E}]}\right)}$, we have*

$$\begin{aligned}\mathbb{P}[\Gamma^\ell(\mathcal{E})] &\geqslant 1 - \exp\left[-\frac{2}{N}\left(\ell - \sqrt{\frac{N}{2}\log\left(\frac{1}{\mathbb{P}[\mathcal{E}]}\right)}\right)^2\right] \\ &= 1 - \exp\left[-2N\left(\frac{\ell}{N} - \sqrt{\frac{1}{2N}\log\left(\frac{1}{\mathbb{P}[\mathcal{E}]}\right)}\right)^2\right],\end{aligned}$$

where $\mathbb{P}[\Gamma^\ell(\mathcal{E})] = \mathbb{P}\left[(\mathbf{y}_1, \ldots, \mathbf{y}_N) \in \Gamma^\ell(\mathcal{E})\right]$ and $\mathbb{P}[\mathcal{E}] = \mathbb{P}\left[(\mathbf{y}_1, \ldots, \mathbf{y}_N) \in \mathcal{E}\right]$.

Remark 4.40. As can be easily seen from the lemma, if $(\mathbf{y}_N)_{N\geqslant 1}$ is a sequence of independent (but not necessarily identically distributed) random variables taking values in $\mathcal{Y}$, and if $(\mathcal{E}_N)_{N\geqslant 1}$ is a sequence of events such that $\mathcal{E}_N \subseteq \mathcal{Y}^N$ and $\mathbb{P}[\mathcal{E}_N]$ is not exponentially small in the sense that $\lim_{N\to\infty} \frac{1}{N}\log \mathbb{P}[\mathcal{E}_N] = 0$, then:

- We can find a sequence of integers $(\ell_N)_{N\geqslant 1}$ such that $\lim_{N\to\infty} \frac{\ell_N}{N} = 0$ and $\lim_{N\to\infty} \mathbb{P}[\Gamma^{\ell_N}(\mathcal{E}_N)] = 1$.
- We need $\ell_N = \omega(\sqrt{N})$ in order for the lemma to guarantee that $\lim_{N\to\infty} \mathbb{P}[\Gamma^{\ell_N}(\mathcal{E}_N)] = 1$.

In the following, we will show how we can apply the blowing-up lemma to boost the success probability of a weak-recovery algorithm and make it converge to 1 at a rate that is exponential in $n^{\frac{1}{4}}$. We will do this in two steps. First, we show how to boost the conditional probability of success given the community labels $\mathbf{x}$, and then we show how to boost the success probability unconditionally.

4.4.2.1 Boosting the conditional probability of success given the community labels

Since our algorithm is robust against a linear number of adversarial edge changes, namely ρn changes, we can benefit from the blowing-up lemma to boost the success probability to $1 - e^{-n^{\frac{1}{4}}}$ by paying a negligible price in the robustness of the algorithm. However, we need to be careful how we apply the blowing-up lemma, because in $(\mathbf{G}, \mathbf{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$, we have $N = \frac{n(n-1)}{2} = \Omega(n^2)$ (conditionally[21]) independent random edges that may or may not be present in the graph, and our algorithm is only robust against up to $\rho n = \Theta(\sqrt{N})$ adversarial changes, whereas the naive and straightforward application of the blowing-up lemma needs a robustness of at least $\omega\left(\sqrt{N}\right)$ in order to guarantee the convergence of probability to 1.

In order to successfully apply the blowing-up lemma, we will faithfully reorganize the randomness of $(\mathbf{G}, \mathbf{x}) \sim \mathsf{SBM}_n(d, \varepsilon)$ in n (conditionally) independent random variables. If the new representation also faithfully captures closeness in the sense that representations that are at Hamming distance ℓ induce graphs that differ by at most $\ell \cdot o(\sqrt{n})$ edges, then our algorithm is robust against up to $\frac{\rho n}{o(\sqrt{n})} = \omega(\sqrt{n})$ adversarial changes in the random variables, and this will allow us to successfully apply the blowing-up lemma.

Lemma 4.41. *Let ε, d, M, ρ and κ be as in Theorem 4.37. For every $\rho' < \rho$, if*

$$\mathbb{P}\left[\mathcal{E}^{succ}_{n,d,\varepsilon,M,\rho,\kappa} \middle| \mathbf{x}\right] \geqslant \Omega\left(e^{-n^{\frac{1}{4}}}\right),$$

then for n large enough, we have

$$\mathbb{P}\left[\mathcal{E}^{succ}_{n,d,\varepsilon,M,\rho',\kappa} \middle| \mathbf{x}\right] \geqslant 1 - \frac{e^{-n^{\frac{1}{4}}}}{2}.$$

Proof. Roughly speaking, our plan is to define n random variables $\mathbf{z}_1, \ldots, \mathbf{z}_n$ taking values in a set $\mathcal{Z}$ such that:

(a) There is a mapping G from $\mathcal{Z}^n$ to the set of graphs having $[n]$ as the set of vertices, such that $\mathbf{G} = G(\mathbf{z}_1, \ldots, \mathbf{z}_n)$.

[21] The edges are conditionally independent given $\mathbf{x}$.

(b) For every $z, z' \in \mathcal{Z}^n$, if $D_H(z, z') \leqslant \ell$, then $G(z)$ differs from $G(z')$ by at most $2\ell \cdot n^{\frac{3}{10}}$ edges.

(c) $\mathbf{z}_1, \ldots, \mathbf{z}_n$ are conditionally independent given $\mathbf{x}$.

Property (a) means that $\mathbf{z}_1, \ldots, \mathbf{z}_n$ form a faithful representation of $\mathbf{G}$. Property (b) means that there the Hamming distance in $\mathcal{Z}^n$ is a good estimate for the number of edge changes between the induced graphs. Properties (b) and (c) will allow us to successfully apply the blowing-up lemma.

For every $x \in \{-1, +1\}^n$, define the set

$$\mathcal{Z}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}(x) = \left\{ z \in \mathcal{Z}^n : \ (G(z), x) \in \mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa} \right\}.$$

Clearly,

$$\mathbb{P}\left[\mathbf{z} \in \mathcal{Z}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}(\mathbf{x}) \middle| \mathbf{x}\right] = \mathbb{P}[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa} | \mathbf{x}].$$

If we have $\mathbb{P}[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa} | \mathbf{x}] \geqslant \Omega\left(e^{-n^{\frac{1}{4}}}\right)$, then if we take $\ell = \lceil n^{\frac{2}{3}} \rceil$, Lemma 4.39 implies that

$$\begin{aligned}
\mathbb{P}\left[\mathbf{z} \in \Gamma^\ell\left(\mathcal{Z}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}(\mathbf{x})\right) \middle| \mathbf{x}\right] &\geqslant 1 - \exp\left[-2n\left(\frac{n^{\frac{2}{3}}}{n} - \sqrt{\frac{1}{2n}\log\left(\frac{1}{\Omega\left(e^{-n^{\frac{1}{4}}}\right)}\right)}\right)^2\right] \\
&= 1 - \exp\left[-2n\left(n^{-\frac{1}{3}} - \frac{1}{\sqrt{2}} n^{-\frac{3}{8}} \sqrt{1 \pm O\left(n^{-\frac{1}{4}}\right)}\right)^2\right] \geqslant 1 - \exp\left[-n^{\frac{1}{3}}\right],
\end{aligned}$$

where the last inequality is true for n large enough.

Now assume that $(\mathbf{G}, \mathbf{x})$ and $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ are such that $\mathbf{z} \in \Gamma^\ell\left(\mathcal{Z}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}(\mathbf{x})\right)$ and $\mathbf{G} = G(\mathbf{z})$. Let $\tilde{z} = (\tilde{z}_1, \ldots, \tilde{z}_n) \in \mathcal{Z}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}(\mathbf{x})$ be such that

$$D_H(\mathbf{z}, \tilde{z}) \leqslant \ell.$$

From Property (b) we know that $G(\tilde{z})$ differs from $\mathbf{G} = G(\mathbf{z})$ by at most $2\ell \cdot n^{\frac{3}{10}} \leqslant 2\lceil n^{\frac{2}{3}} \rceil \cdot n^{\frac{3}{10}} = o(n)$ edges. On the other hand, since $G(\tilde{z}) \in \mathcal{Z}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}(\mathbf{x})$, we have

$$(G(\tilde{z}), \mathbf{x}) \in \mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}.$$

Let $\rho' < \rho$ and let G° be an arbitrary graph with $V(G^\circ) = [n]$, and which differs from $\mathbf{G} = G(\mathbf{z})$ by at most $\rho' n$ edges. Since $G(\tilde{z})$ differs from $G(\mathbf{z})$ by at most $2\ell n^{\frac{3}{10}} = o(n)$ edges, if n is large enough, the graph G° differs from $G(\tilde{z})$ by at most $\rho' n + o(n) \leqslant \rho n$ edges. Now since $(G(\tilde{z}), \mathbf{x}) \in \mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}$ and since G° differs from $G(\tilde{z})$ by at most ρn edges, we have

$$\langle M(G^\circ), \mathbf{x}\mathbf{x}^T \rangle \geqslant \kappa \cdot \|\mathbf{x}\|^2 \cdot \|M(G^\circ)\|_{\text{nuc}},$$

which implies that $(\mathbf{G}, \mathbf{x}) = (G(\mathbf{z}), \mathbf{x}) \in \mathcal{E}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho',\kappa}$ and $\mathbf{z} \in \mathcal{Z}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho',\kappa}(\mathbf{x})$. Therefore,

$$\Gamma^{\ell}\left(\mathcal{Z}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho,\kappa}(\mathbf{x})\right) \subseteq \mathcal{Z}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho',\kappa}(\mathbf{x}).$$

Now since $\mathbb{P}\left[\mathbf{z} \in \mathcal{Z}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho',\kappa}(\mathbf{x})\middle|\mathbf{x}\right] = \mathbb{P}[\mathcal{E}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho',\kappa}|\mathbf{x}]$, we conclude that

$$\mathbb{P}\left[\mathcal{E}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho',\kappa}\middle|\mathbf{x}\right] \geqslant 1 - \exp\left[-n^{\frac{1}{3}}\right]. \tag{4.4.8}$$

In the following, we will show how we can define the random variables $\mathbf{z}_1, \ldots, \mathbf{z}_n$ so that Properties (a - c) are (almost) satisfied.

We partition the set of edges $\{uv : u, v \in [n]\}$ into n subsets $B_1, \ldots, B_n$ such that each set B_i has size $\lceil \frac{n+1}{2} \rceil$ or $\lfloor \frac{n+1}{2} \rfloor$. For each $1 \leqslant i \leqslant n$, we fix an ordering $e^{(i)}_1, \ldots, e^{(i)}_{|B_i|}$ of the edges in B_i, and then define

$$\mathbf{z}_i = \left\{j \in \{1, \ldots, |B_i|\} : e^{(i)}_j \in \mathbf{G}\right\}.$$

Note that $\mathbf{z}_i$ is in one-to-one correspondence with $\{e \in B_i : e \in \mathbf{G}\}$, and so $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)$ is in one-to-one correspondence with $\mathbf{G}$. Furthermore, $\mathbf{z}_1, \ldots, \mathbf{z}_n$ take values in the power set of $\{1, \ldots, \lceil \frac{n+1}{2} \rceil\}$.

It is easy to see that $\mathbf{z}_1, \ldots, \mathbf{z}_n$ satisfy properties (a) and (c) above, but unfortunately they do not satisfy Property (b): It is possible for a change in just one random variable $\mathbf{z}_i$ to cause $\Omega(n)$ edge changes in the graph.

In order to make the above approach work, we need one more ingredient: If we look closely, we find that the main reason why Property (b) does not hold is because it is possible for a graph to contain too many edges coming from one set B_i. But since we are working in the sparse regime of stochastic block models, the probability of this happening is negligible. We can leverage this phenomenon in order to make the above approach work.

More precisely, given $\mathbf{x}$, the size $|\mathbf{z}_i|$ of the set $\mathbf{z}_i$ is the sum of $|B_i| = \Theta(n)$ Bernoulli random variables:

$$|\mathbf{z}_i| = \sum_{uv \in B_i} \mathbb{1}_{uv \in \mathbf{G}}.$$

Now for every $uv \in B_i$, we have

$$\mathbb{P}[uv \in \mathbf{G}|\mathbf{x}] = \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} \leqslant \frac{2d}{n}.$$

It follows from the Chernoff bound that

$$\mathbb{P}\left[|\mathbf{z}_i| > n^{\frac{3}{10}}\middle|\mathbf{x}\right] \leqslant \exp\left(-|B_i| \cdot D_{KL}\left(\frac{n^{\frac{3}{10}}}{|B_i|} \,\middle\|\, \frac{2d}{n}\right)\right), \tag{4.4.9}$$

where $D_{KL}(p||q) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$ is the Kullback-Leibler divergence between Bernoulli(p) and Bernoulli(q). Now notice that

$$\begin{aligned}
D_{KL}\left(\frac{n^{\frac{3}{10}}}{|B_i|}\,\middle\|\,\frac{2d}{n}\right) &= \frac{n^{\frac{3}{10}}}{|B_i|}\log\frac{\frac{n^{\frac{3}{10}}}{|B_i|}}{\frac{2d}{n}} + \left(1-\frac{n^{\frac{3}{10}}}{|B_i|}\right)\log\frac{1-\frac{n^{\frac{3}{10}}}{|B_i|}}{1-\frac{2d}{n}} \\
&= \frac{n^{\frac{3}{10}}}{\Theta(n)}\log\frac{\frac{n^{\frac{3}{10}}}{\Theta(n)}}{\frac{2d}{n}} + \left(1-\frac{n^{\frac{3}{10}}}{\Theta(n)}\right)\log\frac{1-\frac{n^{\frac{3}{10}}}{\Theta(n)}}{1-\frac{2d}{n}} \\
&= \Theta\left(n^{-\frac{7}{10}}\right)\log\Theta\left(n^{\frac{3}{10}}\right) + \left(1-\Theta\left(n^{-\frac{7}{10}}\right)\right)\log\left(1-\Theta\left(n^{-\frac{7}{10}}\right)\right) \\
&= \Theta\left(n^{-\frac{7}{10}}\right)\log\Theta\left(n^{\frac{3}{10}}\right) - \Theta\left(n^{-\frac{7}{10}}\right) \\
&\geqslant \Omega\left(n^{-\frac{7}{10}}\right).
\end{aligned} \tag{4.4.10}$$

Therefore, the probability that there is at least one $i \in [n]$ such that $|\mathbf{z}_i| \geqslant n^{\frac{3}{10}}$ can be upper bounded as

$$\mathbb{P}\left[\left\{\exists i \in [n] :\ |\mathbf{z}_i| > n^{\frac{3}{10}}\right\}\middle|\mathbf{x}\right] \leqslant n\exp\left(-\Theta(n)\cdot\Omega\left(n^{-\frac{7}{10}}\right)\right) = n\exp\left(-\Omega\left(n^{\frac{3}{10}}\right)\right) \leqslant o\left(e^{-n^{\frac{1}{4}}}\right), \tag{4.4.11}$$

where the last inequality is true for n large enough.

Let $\mathcal{E}_B$ be the event that the graph $\mathbf{G}$ contains at most $n^{\frac{3}{10}}$ edges from each set B_i. The above shows that

$$\mathbb{P}[\mathcal{E}_B|\mathbf{x}] \geqslant 1 - o\left(e^{-n^{\frac{1}{4}}}\right).$$

Now if we have $\mathbb{P}[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}|\mathbf{x}] \geqslant \Omega\left(e^{-n^{\frac{1}{4}}}\right)$, then

$$\begin{aligned}
\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\middle|\mathbf{x},\mathcal{E}_B\right] &\geqslant \mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\cap\mathcal{E}_B\middle|\mathbf{x}\right] \geqslant \mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\middle|\mathbf{x}\right] - \mathbb{P}\left[\mathcal{E}_B^c|\mathbf{x}\right] \\
&\geqslant \Omega\left(e^{-n^{\frac{1}{4}}}\right) - o\left(e^{-n^{\frac{1}{4}}}\right) = \Omega\left(e^{-n^{\frac{1}{4}}}\right).
\end{aligned}$$

If we condition on $\mathcal{E}_B$, then with probability 1, the random variables $\mathbf{z}_1,\ldots,\mathbf{z}_n$ take values in the set

$$\mathcal{Z}_{(n^{1/3})} = \left\{S \subseteq \left\{1,\ldots,\left\lceil\frac{n+1}{2}\right\rceil\right\} :\ |S| \leqslant n^{\frac{1}{3}}\right\}.$$

Now define a mapping G from $\mathcal{Z}^n_{(n^{1/3})}$ to the set of graphs having $[n]$ as the set of vertices, as follows: If $(z_1,\ldots,z_n) \in \mathcal{Z}^n_{(n^{1/3})}$, then for every $u,v \in [n]$, the edge uv is present in $G(z_1,\ldots,z_n)$ if and only if there exists $1 \leqslant i \leqslant n$ and $1 \leqslant j \leqslant \lceil\frac{n+1}{2}\rceil$ such that $uv \in B_i$, $uv = e^{(i)}_j$, and $j \in z_i$. It is easy to see that we have:

(a′) If $\mathcal{E}_B$ occurs, then $\mathbf{G} = G(\mathbf{z}_1, \dots, \mathbf{z}_m)$.

(b′) For every $z, z' \in \mathcal{Z}^n_{(n^{1/3})}$, if $D_H(z, z') \leqslant \ell$, then $G(z)$ differs from $G(z')$ by at most $2\ell n^{\frac{1}{3}}$ edges.

(c′) Given $\mathbf{x}$ and $\mathcal{E}_B$, the random variables $\mathbf{z}_1, \dots, \mathbf{z}_n$ are conditionally independent.

If we repeat the proof of Eq. (4.4.8) verbatim but instead of only conditioning on $\mathbf{x}$, we condition on $\mathbf{x}$ and $\mathcal{E}_B$, we get

$$\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho',\kappa}\middle|\mathbf{x}, \mathcal{E}_B\right] \geqslant 1 - e^{-n^{\frac{1}{3}}}.$$

Now since $\mathbb{P}[\mathcal{E}_B|\mathbf{x}] \geqslant 1 - o\left(e^{-n^{\frac{1}{4}}}\right)$, we conclude that

$$\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho',\kappa}\middle|\mathbf{x}\right] \geqslant \mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho',\kappa}\middle|\mathcal{E}_B, \mathbf{x}\right] \cdot \mathbb{P}[\mathcal{E}_B|\mathbf{x}] \geqslant 1 - e^{-n^{\frac{1}{3}}} - o\left(e^{-n^{\frac{1}{4}}}\right) \geqslant 1 - \frac{e^{-n^{\frac{1}{4}}}}{2},$$

where the last inequality is true for n large enough. □

4.4.2.2 Boosting the probability of success unconditionally

So far we managed to boost the conditional probability of success given $\mathbf{x}$, but we would like to boost the success probability unconditionally. The blowing-up lemma will help us once again in achieving that. First, we will show that if there is a lower bound $\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\right] \geqslant p$, then we can derive a lower bound on the probability that $\mathbf{x}$ satisfies $\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}|\mathbf{x}\right] \geqslant \frac{p}{2}$.

Lemma 4.42. *Let ε, d, M, ρ and κ be as in Theorem 4.37. If $p > 0$ is such that*

$$\mathbb{P}\left[\mathcal{E}^{succ}_{n,d,\varepsilon,M,\rho,\kappa}\right] \geqslant p,$$

then

$$\mathbb{P}\left[\mathbb{P}\left[\mathcal{E}^{succ}_{n,d,\varepsilon,M,\rho,\kappa}|\mathbf{x}\right] > \frac{p}{2}\right] > \frac{p}{2}.$$

Proof. Since

$$\mathbb{E}\left[\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}|\mathbf{x}\right]\right] = \mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\right] \geqslant p,$$

we have

$$\mathbb{E}\left[1 - \mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}|\mathbf{x}\right]\right] \leqslant 1 - p.$$

Now by Markov inequality, we have

$$\mathbb{P}\left[1-\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\middle|\mathbf{x}\right]\geqslant 1-\frac{p}{2}\right]\leqslant\frac{1-p}{1-\frac{p}{2}}=\frac{2-2p}{2-p}.$$

Therefore,

$$\mathbb{P}\left[\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\middle|\mathbf{x}\right]>\frac{p}{2}\right]\geqslant 1-\frac{2-2p}{2-p}=\frac{p}{2-p}>\frac{p}{2}.$$

□

Now we are ready to prove Theorem 4.37

Proof of Theorem 4.37. If $\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\right]\geqslant\Omega\left(e^{-n^{\frac{1}{4}}}\right)$ then there exist $C>0$ such that $\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\right]\geqslant Ce^{-n^{\frac{1}{4}}}$ for n large enough. Lemma 4.42 now implies that

$$\mathbb{P}\left[\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\middle|\mathbf{x}\right]>\frac{Ce^{-n^{\frac{1}{4}}}}{2}\right]>\frac{Ce^{-n^{\frac{1}{4}}}}{2}=\Omega\left(e^{-n^{\frac{1}{4}}}\right).$$

Now let ρ'' be such that $\rho'<\rho''<\rho$. We know from Lemma 4.41 we know that if $\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho,\kappa}\middle|\mathbf{x}\right]>\frac{Ce^{-n^{\frac{1}{4}}}}{2}$ and n is large enough, then

$$\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}\middle|\mathbf{x}\right]\geqslant 1-\frac{e^{-n^{\frac{1}{4}}}}{2}.$$

We conclude that

$$\mathbb{P}\left[\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}\middle|\mathbf{x}\right]\geqslant 1-\frac{e^{-n^{\frac{1}{4}}}}{2}\right]\geqslant\Omega\left(e^{-n^{\frac{1}{4}}}\right).$$

Now define

$$\mathcal{X}^{\text{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}=\left\{x\in\{-1,+1\}^n:\ \mathbb{P}\left[(\mathbf{G},\mathbf{x})\in\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}\middle|\mathbf{x}=x\right]\geqslant 1-\frac{e^{-n^{\frac{1}{4}}}}{2}\right\}.$$

We have

$$\mathbb{P}\left[\mathbf{x}\in\mathcal{X}^{\text{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}\right]=\mathbb{P}\left[\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}\middle|\mathbf{x}\right]\geqslant 1-\frac{e^{-n^{\frac{1}{4}}}}{2}\right]\geqslant\Omega\left(e^{-n^{\frac{1}{4}}}\right).$$

By the blowing-up lemma, if we take $\ell = \lceil n^{\frac{2}{3}} \rceil$, then

$$\begin{aligned}\mathbb{P}\left[\mathbf{x} \in \Gamma^{\ell}\left(\mathcal{X}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}\right)\right] &\geqslant 1 - \exp\left[-2n\left(\frac{n^{\frac{2}{3}}}{n} - \sqrt{\frac{1}{2n}\log\left(\frac{1}{\Omega\left(e^{-n^{\frac{1}{4}}}\right)}\right)}\right)^2\right] \\ &= 1 - \exp\left[-2n\left(n^{-\frac{1}{3}} - \frac{1}{\sqrt{2}}n^{-\frac{3}{8}}\sqrt{1 \pm O\left(n^{-\frac{1}{4}}\right)}\right)^2\right] \geqslant 1 - \exp\left[-n^{\frac{1}{3}}\right], \end{aligned} \tag{4.4.12}$$

where the last inequality is true for n large enough.

Let $\tilde{\mathbf{x}} \in \mathcal{X}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}$ be such that the Hamming distance $D_H(\mathbf{x}, \tilde{\mathbf{x}})$ is minimal. We break the ties using the lexicographic order of $\{-1, +1\}^n$. Define

$$V_{\mathbf{x}} = \{v \in [n] : \ \mathbf{x}_v \neq \tilde{\mathbf{x}}_v\}.$$

Clearly, $|V_{\mathbf{x}}| = D_H(\mathbf{x}, \tilde{\mathbf{x}})$. We will generate a random graph $\tilde{\mathbf{G}}$ with $V(\tilde{\mathbf{G}}) = [n]$ as follows:

- For every $u, v \in [n] \setminus V_{\mathbf{x}}$, we make the edge uv present in $\tilde{\mathbf{G}}$ if and only if it is present in $\mathbf{G}$.
- For an edge uv such that $u \in V_{\mathbf{x}}$ or $v \in V_{\mathbf{x}}$, we randomly decide its presence in $\tilde{\mathbf{G}}$ independently of $(\mathbf{G}, \mathbf{x})$ and in such a way that the conditional probability of $\tilde{\mathbf{G}}$ given $\tilde{\mathbf{x}}$ is consistent with the distribution of the stochastic block model. More precisely, we generate
$$\left\{uv \in \tilde{\mathbf{G}} : \ u \in V_{\mathbf{x}} \text{ or } v \in V_{\mathbf{x}}\right\}$$
independently of $(\mathbf{G}, \mathbf{x})$, and with the following conditional distribution: For every two sets of edges
$$\mathsf{E} \subseteq \left\{uv : \ u, v \in [n], \ u \in V_{\mathbf{x}} \text{ or } v \in V_{\mathbf{x}}\right\} \quad \text{and} \quad \mathsf{E}' \subseteq \left\{uv : \ u, v \in [n] \setminus V_{\mathbf{x}}\right\},$$
we have
$$\begin{aligned}&\mathbb{P}\left[\left\{uv \in \tilde{\mathbf{G}} : \ u \in V_{\mathbf{x}} \text{ or } v \in V_{\mathbf{x}}\right\} = \mathsf{E} \,\middle|\, \tilde{\mathbf{x}}, \left\{uv \in \tilde{\mathbf{G}} : \ u, v \in [n] \setminus V_{\mathbf{x}}\right\} = \mathsf{E}'\right] \\ &\quad = \mathbb{P}\left[\left\{uv \in \mathbf{G} : \ u \in V_{\mathbf{x}} \text{ or } v \in V_{\mathbf{x}}\right\} = \mathsf{E} \,\middle|\, \mathbf{x} = \tilde{\mathbf{x}}, \left\{uv \in \mathbf{G} : \ u, v \in [n] \setminus V_{\mathbf{x}}\right\} = \mathsf{E}'\right].\end{aligned}$$

It is easy to see that the conditional distribution of $\tilde{\mathbf{G}}$ given $\tilde{\mathbf{x}}$ is the same as the conditional distribution of $\mathbf{G}$ given $\mathbf{x} = \tilde{\mathbf{x}}$. Now since $\tilde{\mathbf{x}} \in \mathcal{X}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}$ and since the conditional distribution of $\tilde{\mathbf{G}}$ given $\tilde{\mathbf{x}}$ is that of $\mathsf{SBM}_n(d, \varepsilon)$, it follows from the definition of $\mathcal{X}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}$ that with probability at least $1 - \dfrac{e^{-n^{\frac{1}{4}}}}{2}$, we have $(\tilde{\mathbf{G}}, \tilde{\mathbf{x}}) \in \mathcal{E}^{\mathrm{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}$, which means that for every graph G° that differs from $\tilde{\mathbf{G}}$ by at most $\rho'' n$ edges, we have

$$\langle M(G^\circ), \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \rangle \geqslant \kappa \cdot \|\tilde{\mathbf{x}}\|^2 \cdot \|M(G^\circ)\|_{\mathrm{nuc}}.$$

Let $\mathcal{E}$ be the event that $\mathbf{x} \in \Gamma^{\ell}\left(\mathcal{X}^{\text{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}\right)$ and that for every graph G° that differs from $\tilde{\mathbf{G}}$ by at most $\rho'' n$ edges, we have

$$\langle M(G^{\circ}), \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \rangle \geqslant \kappa \cdot \|\tilde{\mathbf{x}}\|^2 \cdot \|M(G^{\circ})\|_{\text{nuc}}.$$

It follows from Eq. (4.4.12) and from the above discussion that

$$\mathbb{P}[\mathcal{E}] \geqslant 1 - e^{-n^{\frac{1}{3}}} - \frac{e^{-n^{\frac{1}{4}}}}{2}. \tag{4.4.13}$$

Let $\mathcal{V}$ be the event that the degree in $\mathbf{G}$ of every vertex $v \in [n]$ is at most $n^{\frac{3}{10}}$. Similarly, let $\tilde{\mathcal{V}}$ be the event that the degree in $\tilde{\mathbf{G}}$ of every vertex $v \in [n]$ is at most $n^{\frac{3}{10}}$. Since the degree of every vertex is the sum of n independent Bernoulli random variables each having a success probability of at most $\frac{2d}{n}$, then by following calculations that are very similar to Eq. (4.4.9), Eq. (4.4.10) and Eq. (4.4.11), we can deduce that

$$\mathbb{P}[\mathcal{V}] \geqslant 1 - o\left(e^{-n^{\frac{1}{4}}}\right), \tag{4.4.14}$$

and

$$\mathbb{P}[\tilde{\mathcal{V}}] \geqslant 1 - o\left(e^{-n^{\frac{1}{4}}}\right). \tag{4.4.15}$$

Now assume that $\mathcal{E} \cap \mathcal{V} \cap \tilde{\mathcal{V}}$ occurs. Observe the following:

- Since $\mathcal{E}$ occurs, we have $\mathbf{x} \in \Gamma^{\ell}\left(\mathcal{X}^{\text{succ}}_{n,d,\varepsilon,M,\rho'',\kappa}\right)$, which means that $|V_{\mathbf{x}}| = D_H(\mathbf{x}, \tilde{\mathbf{x}}) \leqslant \ell$.
- From the definition of $\tilde{\mathbf{G}}$, we can see that the graphs $\tilde{\mathbf{G}}$ and $\mathbf{G}$ can differ only in edges that are incident to vertices in $V_{\mathbf{x}}$.
- Since $\mathcal{V}$ occurs, $\mathbf{G}$ contains at most $n^{\frac{3}{10}} \cdot |V_{\mathbf{x}}| \leqslant n^{\frac{3}{10}} \cdot \lceil n^{\frac{2}{3}} \rceil = o(n)$ edges that are incident to vertices in $V_{\mathbf{x}}$.
- Since $\tilde{\mathcal{V}}$ occurs, $\tilde{\mathbf{G}}$ contains at most $n^{\frac{3}{10}} \cdot |V_{\mathbf{x}}| \leqslant n^{\frac{3}{10}} \cdot \lceil n^{\frac{2}{3}} \rceil = o(n)$ edges that are incident to vertices in $V_{\mathbf{x}}$.

Therefore, if $\mathcal{E} \cap \mathcal{V} \cap \tilde{\mathcal{V}}$ occurs, then $\tilde{\mathbf{G}}$ differs from $\mathbf{G}$ by at most $o(n) + o(n) = o(n)$ edges.

Now let G° be an arbitrary graph with $V(G^{\circ}) = [n]$ such that G° differs from $\mathbf{G}$ by at most $\rho' n$ edges. Since $\rho' < \rho''$, then if n is large enough, the graph G° differs from $\tilde{\mathbf{G}}$ by at most $\rho' n + o(n) \leqslant \rho'' n$ edges. Now since $\mathcal{E}$ occurs and since G° differs from $\tilde{\mathbf{G}}$ by at most $\rho'' n$ edges, we have

$$\langle M(G^{\circ}), \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \rangle \geqslant \kappa \cdot \|\tilde{\mathbf{x}}\|^2 \cdot \|M(G^{\circ})\|_{\text{nuc}}.$$

Now notice that

$$\langle M(G^{\circ}), \mathbf{x}\mathbf{x}^T \rangle = \langle M(G^{\circ}), \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \rangle + \langle M(G^{\circ}), \mathbf{x}\mathbf{x}^T - \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T \rangle$$

$$\geqslant \kappa \cdot \|\tilde{\mathbf{x}}\|^2 \cdot \|M(G^\circ)\|_{\text{nuc}} - |\langle M(G^\circ), \mathbf{x}\mathbf{x}^T - \tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\rangle|$$
$$\geqslant \kappa \cdot \|\mathbf{x}\|^2 \cdot \|M(G^\circ)\|_{\text{nuc}} - \left\|\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T - \mathbf{x}\mathbf{x}^T\right\| \cdot \|M(G^\circ)\|_{\text{nuc}}.$$

On the other hand, since $\mathbf{x}$ differs from $\tilde{\mathbf{x}}$ only on $V_{\mathbf{x}}$, and since $|V_{\mathbf{x}}| \leqslant \ell = \lceil n^{\frac{2}{3}} \rceil$, we have

$$\begin{aligned}\left\|\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T - \mathbf{x}\mathbf{x}^T\right\| &\leqslant \left\|\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T - \tilde{\mathbf{x}}\mathbf{x}^T\right\| + \left\|\tilde{\mathbf{x}}\mathbf{x}^T - \mathbf{x}\mathbf{x}^T\right\| = \left\|\tilde{\mathbf{x}}(\tilde{\mathbf{x}}^T - \mathbf{x}^T)\right\| + \left\|(\tilde{\mathbf{x}} - \mathbf{x})\mathbf{x}^T\right\| \\ &\leqslant \|\tilde{\mathbf{x}}\| \cdot \left\|\tilde{\mathbf{x}}^T - \mathbf{x}^T\right\| + \|\tilde{\mathbf{x}} - \mathbf{x}\| \cdot \left\|\mathbf{x}^T\right\| = \sqrt{n} \cdot 2\sqrt{|V_{\mathbf{x}}|} + 2\sqrt{|V_{\mathbf{x}}|} \cdot \sqrt{n} \\ &\leqslant 4\sqrt{n} \cdot \sqrt{n^{\frac{3}{10}}} = o(n) = o\left((\sqrt{n})^2\right) = o(\|\mathbf{x}\|^2).\end{aligned}$$

Now since $\kappa' < \kappa$, we get that for n is large enough, we have

$$\begin{aligned}\langle M(G^\circ), \mathbf{x}\mathbf{x}^T\rangle &\geqslant (\kappa - o(1)) \cdot \|\mathbf{x}\|^2 \cdot \|M(G^\circ)\|_{\text{nuc}} \\ &\geqslant \kappa' \cdot \|\mathbf{x}\|^2 \cdot \|M(G^\circ)\|_{\text{nuc}}.\end{aligned}$$

This implies that if $\mathcal{E} \cap \mathcal{V} \cap \tilde{\mathcal{V}}$ occurs and n is large enough, then $\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho',\kappa'}$ occurs as well. By combining this with Eq. (4.4.13), Eq. (4.4.14), and Eq. (4.4.15), we conclude that for n large enough, we have

$$\begin{aligned}\mathbb{P}\left[\mathcal{E}^{\text{succ}}_{n,d,\varepsilon,M,\rho',\kappa'}\right] &\geqslant 1 - e^{-n^{\frac{1}{3}}} - \frac{e^{-n^{\frac{1}{4}}}}{2} - o\left(e^{-n^{\frac{1}{4}}}\right) - o\left(e^{-n^{\frac{1}{4}}}\right) \\ &\geqslant 1 - e^{-n^{\frac{1}{4}}}.\end{aligned}$$

□

4.5 Trace bounds for stochastic block models

In this section we prove Lemma 4.29. The lemma will be a direct consequence of the following theorems: a separation in Schatten norm between $Q^{(s)}(\overline{\mathbf{Y}})$ and $Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\intercal$ and some concentration results on the diagonal entries of $Q^{(s)}(\overline{\mathbf{Y}})$.

Theorem 4.43. *Let $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$. Let Δ, A, τ be as defined in Eq. (4.4.2). Suppose $d \geqslant (1+\delta)\frac{4}{\varepsilon^2}$ for some $\delta > 0$. Let s be an integer satisfying*

$$s \geqslant 10^{10}\left(1 + \frac{1}{\delta}\right)\left(\max\left\{\log d, \log\log\frac{2}{\varepsilon}, 1, \log\frac{1}{\delta}\right\}\right)^2,$$

and let $t \in \left[\frac{1}{400}\log n\,,\, \frac{1}{100}\log n\right]$. Then for n large enough

$$\mathbb{E}\left[\text{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^t\right] \geqslant \frac{n^t}{10} \cdot n^{-\frac{2}{50A}} \cdot \left(1 - (1+\delta)^{-\sqrt{s}\cdot t} - n^{1/10}\right). \tag{4.5.1}$$

Furthermore

$$\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^{\mathsf{T}}\right)^t\right] \leqslant (1+\delta)^{-t/5} \cdot \mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^t\right]. \tag{4.5.2}$$

Let's spend a moment discussing Theorem 4.43. Consider Eq. (4.5.1), the term containing $(1+\delta)^{\sqrt{s}\cdot t}$ corresponds to random noise that is uncorrelated with the underlying partition x, and it can be made arbitrarily small by increasing the length s of the self-avoiding walks. The term $n^{-\frac{1}{100A}}$ is an approximation factor that appears for technical reasons[22]. Hhowever by choice of A it will be negligible. While we limit the constant chosen for t in a certain range, this appears to be quite flexible as long as $t < \log n$.

More interestingly, Eq. (4.5.2) shows the push-out effect of the matrix $Q^{(s)}(\overline{\mathbf{Y}})$: even though $\mathbf{x}\mathbf{x}^{\mathsf{T}}$ is not the matrix maximizing the t-Schatten norm of $Q^{(s)}(\overline{\mathbf{Y}})$, it is $\delta^{O(1)}$-correlated with it.

Theorem 4.44. *Consider the settings of Theorem 4.43. Then for any $u, v \in [n], u \neq v$*

$$\mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu}\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{vv}\right] \leqslant (1+o(1))\,\mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu}\right]^2.$$

Moreover, for any $u \in [n]$

$$\mathbb{E}\left[\left(\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu}\right)^2\right] \leqslant C \cdot \mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu}\right]^2,$$

where $C > 1$ is a universal constant.

The inequalities in Theorem 4.44 shows that both the trace itself and the diagonal entries of $\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^t$ have small variance and thus concentrate around their expectations. Together with Theorem 4.43, we can use Theorem 4.44 to prove Lemma 4.29.

Proof of Lemma 4.29. Eq. (4.4.7) follows directly by Theorem 4.43. We can bound the variance of the trace applying Theorem 4.44

$$\begin{aligned}
\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t}\right] &= \sum_{i,j\in[n]} \mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{ii}\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{jj}\right] \\
&= (1+o(1)) \sum_{i,j\in[n]} \mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{ii}\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{jj}\right] \\
&= (1+o(1)) \sum_{i,j\in[n], i\neq j} \mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{ii}\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{jj}\right]
\end{aligned}$$

[22] This approximation factor appears when trying to bound the dependencies that are caused by the truncation.

$$= (1 + o(1))\, \mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^t\right]^2,$$

where in the third step we used the fact that there is only a linear number of terms with $i = j$. Thus Eq. (4.4.4) holds with high probability through an application of Chebyshev's inequality. Then by Markov's inequality

$$\mathbb{P}\left\{\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t \geqslant (1+\delta)^{-t/10}\,\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^t\right\} \leqslant o(1)\,.$$

It remains to prove Eq. (4.4.6). By Theorem 4.44, there is a universal constant $C > 0$ such that

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in[n]}\left(\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t-2}\right)_{ii}^2\right] &\leqslant C\cdot\mathbb{E}\left[\sum_{i\in[n]}\left(\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t-2}\right)_{ii}\right]^2\\ &= C\cdot\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t-2}\right]^2\\ &= (1+o(1))\cdot C\cdot\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{2t-2}\right]^2,\end{aligned}$$

where in the last step we used concentration of the trace. By Markov's inequality, setting $\gamma = O(C)$ we obtain the desired result. □

Organization of the section. The rest of the section will contain the proofs of Theorem 4.43 and Theorem 4.44. To prove the theorems we will reduce the study of $\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^t$ and related quantities to the combinatorial problem of counting multi-graphs.

We introduce preliminary facts and a bird-eye view of the in Section 4.5.1. We prove Theorem 4.43 in Section 4.5.2 and Section 4.5.3. Finally, we obtain Theorem 4.44 in Section 4.5.4. For simplicity of the exposition, the following sections will be essentially oblivious to the technical challenges arising when studying moments of truncated graphs. We defer a high level discussion of the truncation effect to sections Appendix B.1 and Appendix B.2. We present the technical arguments in Appendix B.1, Appendix B.2, Appendix B.3 and Appendix B.4. In the forthcoming sections, we always assume the settings of Theorem 4.43 to hold.

4.5.1 Preliminary discussion

Let $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$ and let $\mathbf{Y}$ be the adjacency matrix of $\mathbf{G}$. Recall the matrix polynomial $Q^{(s)}(\mathbf{Y})$ defined in Eq. (4.2.2):

$$Q^{(s)}_{ij}(\mathbf{Y}) = \begin{cases}\frac{1}{|\mathrm{SAW}^s_{ij}|}\left(\frac{2n}{\varepsilon\cdot d}\right)^s \sum_{H\in\mathrm{SAW}^s_{ij}} Y_H & \text{if } i\neq j\,,\\ 0 & \text{otherwise.}\end{cases},$$

for all $i, j \in [n]$. By Fact 4.7, $Q^{(s)}(\mathbf{Y})$ is an unbiased estimator of $\mathbf{x}\mathbf{x}^\mathsf{T}$ (up to the diagonal entries) but, as already discussed, we will need to work with a truncated version of the graph. For a graph G, we will denote by $\overline{G}$ the graph obtained from G by deleting all the vertices which have degree more than Δ in G. Similarly, we will denote by $\overline{Y}$ its truncated adjacency matrix as defined in Definition 4.26.

4.5.1.1 From trace computations to graph counting

Next, we show how to reduce the trace computation to a multigraph counting problem. Each term in the sum $\mathrm{Tr}\left(Q^{(s)}(\overline{Y})\right)^t$ is a concatenation of self-avoiding walks and hence correspond to a multigraph over some subset of vertices $[n]$. We make this idea precise below.

Definition 4.45 (Block SAW)**.** Let $\mathrm{BSAW}_{s,t}$ be the set of multi-graphs obtained as follows. Pick $i_1, \ldots, i_t \in [n]$ distinct vertices in K_n and set $i_{t+1} := i_1$. For each $q \in [t]$, pick $W_q \in \mathrm{SAW}^s_{i_q i_{q+1}}(K_n)$. Then

$$\bigoplus_{q\in[t]} W_q \in \mathrm{BSAW}_{s,t}\,.$$

Let $H = \bigoplus_{q\in[t]} W_q$. We call $I(H) = \{i_1, \ldots, i_t\}$ the set of *pivot* vertices of H and $\mathcal{W}(H) = \{W_1, \ldots, W_t\}$ the generating self-avoiding walks of H. We denote by $M(\mathcal{W}(H))$ the sequence of edges obtained concatenating the sequences $M(W_1), \ldots, M(W_t)$. We call $\mathrm{BSAW}_{s,t}$ the set of (s,t)-block self-avoiding walks. At times, we will also use $M(\mathcal{W}(H))$ to denote the set of edges in the sequence $M(\mathcal{W}(H))$.

We can now expand $\mathrm{Tr}\left(Q^{(s)}(Y)\right)$.

Lemma 4.46. *Consider the settings of Theorem 4.43. Let G be a graph over $[n]$ with centered adjacency matrix Y. Then*

$$\mathrm{Tr}\left(Q^{(s)}(Y)^t\right) = (1 \pm o(1)) \cdot n^t \left(\frac{2}{\varepsilon \cdot d}\right)^{st} \cdot \sum_{H \in BSAW_{s,t}} Y_H\,. \tag{4.5.3}$$

Proof. By definition of trace,

$$\begin{aligned}
\mathrm{Tr}\left(Q^{(s)}(Y)^t\right) &= \sum_{i_1,\ldots,i_t\in[n]} Q^s_{i_1 i_2}(Y) \cdots Q^s_{i_{t-1} i_t}(Y) \cdot Q^s_{i_t i_1}(Y) \\
&= \frac{1}{\left((n-2)^{\underline{s-1}}\right)^t} \cdot \left(\frac{2n}{\varepsilon \cdot d}\right)^{st} \cdot \sum_{i_1,\ldots,i_t\in[n]} \left[\prod_{\ell\in[t]} \left(\sum_{W\in \mathrm{SAW}^s_{i_\ell i_{\ell+1}}} Y_W\right)\right] \\
&= \frac{1}{\left((n-2)^{\underline{s-1}}\right)^t} \cdot \left(\frac{2n}{\varepsilon \cdot d}\right)^{st} \cdot \sum_{H\in \mathrm{BSAW}_{s,t}} Y_H
\end{aligned}$$

$$= (1 \pm o(1)) \cdot n^t \left(\frac{2}{\varepsilon \cdot d}\right)^{st} \cdot \sum_{H \in \mathrm{BSAW}_{s,t}} Y_H \,.$$

□

Definition 4.45 captures all the elements in $\mathrm{Tr}\left(Q^{(s)}(Y)\right)$. Furthermore, a similar expansion can be carried out for the centered trace $\mathrm{Tr}\left(Q^{(s)}(\overline{Y}) - xx^\intercal\right)^t$. For simplicity of the notation, in the next expressions, for a given set of t vertices $\{i_1, \ldots, i_t\}$ we denote i_1 also by i_{t+1}.

Fact 4.47. *Consider the settings of Theorem 4.43. Let $M \in \mathbb{R}^{n \times n}$. Then for any graph G with centered adjacency matrix Y*

$$\mathrm{Tr}\left(Q^{(s)}(Y) - M\right)^t = (1 \pm o(1)) n^t \cdot \left(\frac{2}{\varepsilon \cdot d}\right)^{st} \cdot \sum_{i_1,\ldots,i_t \in [n]} \prod_{\ell \in [t]} \left[\sum_{W \in SAW^s_{i_\ell i_{\ell+1}}} \left(Y_W - \left(\frac{\varepsilon \cdot d}{2n}\right)^s \cdot \frac{1}{\left|SAW^s_{i_\ell i_{\ell+1}}\right|} \cdot M_{i_\ell i_{\ell+1}} \right) \right]. \quad (4.5.4)$$

Proof. Simply expanding the trace

$$\begin{aligned}
\mathrm{Tr}\left(Q^{(s)}(Y) - M\right)^t &= \sum_{i_1,\ldots,i_t \in [n]} \prod_{\ell \in [t]} \left(Q^{(s)}_{i_\ell i_{\ell+1}}(Y) - M_{i_\ell i_{\ell+1}} \right) \\
&= \sum_{i_1,\ldots,i_t \in [n]} \prod_{\ell \in [t]} \left(\frac{1}{|\mathrm{SAW}^s_{i_\ell i_{\ell+1}}|} \sum_{W \in \mathrm{SAW}^s_{i_\ell i_{\ell+1}}} \left(\frac{2n}{\varepsilon \cdot d}\right)^s Y_W - M_{i_\ell i_{\ell+1}} \right) \\
&= \left(\frac{2}{\varepsilon \cdot d}\right)^{st} \sum_{i_1,\ldots,i_t \in [n]} \prod_{\ell \in [t]} \left(\frac{n^s}{|\mathrm{SAW}^s_{i_\ell i_{\ell+1}}|} \sum_{W \in \mathrm{SAW}^s_{i_\ell i_{\ell+1}}} Y_W - \left(\frac{\varepsilon \cdot d}{2}\right)^s \cdot M_{i_\ell i_{\ell+1}} \right) \\
&= (1 \pm o(1)) n^t \cdot \left(\frac{2}{\varepsilon \cdot d}\right)^{st} \\
&\quad \cdot \sum_{i_1,\ldots,i_t \in [n]} \prod_{\ell \in [t]} \left[\sum_{W \in \mathrm{SAW}^s_{i_\ell i_{\ell+1}}} \left(Y_W - \left(\frac{\varepsilon \cdot d}{2n}\right)^s \cdot \frac{1}{|\mathrm{SAW}^s_{i_\ell i_{\ell+1}}|} \cdot M_{i_\ell i_{\ell+1}} \right) \right].
\end{aligned}$$

□

To easily refer to the elements in the sum of Fact 4.47, we additionally use the following notation. For a given walk $W \in \mathrm{SAW}^s_{ij}$ and $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$ we define the polynomial

$$\hat{\mathbf{Y}}_W := \overline{\mathbf{Y}}_W - \left(\frac{\varepsilon \cdot d}{2n}\right)^s \cdot \mathbf{x}_i \mathbf{x}_j \,. \quad (4.5.5)$$

At times, for a set of self-avoiding walks $\mathcal{W}$ and a multi-graph $H = \bigoplus_{W \in \mathcal{W}} W$, we will use the notation

$$\hat{Y}_H = \prod_{W \in \mathcal{W}} \hat{Y}_W \,. \tag{4.5.6}$$

4.5.1.2 Proof strategies

Here we outline our proof strategies.

Proof strategy for Theorem 4.43. We show the theorem in two steps, first we prove a lower bound $\left\| Q^{(s)}(\overline{\mathbf{Y}}) \right\|_t \geqslant C$ (for some meaningful quantity C), second an upper bound of the form $\left\| Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T} \right\|_t \leqslant (1+\delta)^{-\Omega(1)} \cdot C$. The two together immediately imply the theorem. Our strategy to lower bound $\left\| Q^{(s)}(\overline{\mathbf{Y}}) \right\|_t$ will be the following:

1. For any $x \in \{\pm 1\}^n$. Prove an upper bound $\overline{U}_H(x)$ for $\left| \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| x \right] \right|$ for every $H \in \mathrm{BSAW}_{s,t}$.
2. Find a large enough class $\mathrm{NBSAW}_{s,t} \subset \mathrm{BSAW}_{s,t}$ of nice and well-behaved block-self-avoiding-walks whose structure allows us to prove a lower bound $\overline{L}_H$ for $\mathbb{E}\left[\overline{\mathbf{Y}}_H \right]$ for every $H \in \mathrm{NBSAW}_{s,t}$.
3. Show that the contribution of $\mathrm{NBSAW}^c_{s,t} = \mathrm{BSAW}_{s,t} \setminus \mathrm{NBSAW}_{s,t}$ is negligible with respect to that of $\mathrm{BSAW}_{s,t}$. More precisely, we will show that

$$\sum_{H \in \mathrm{NBSAW}^c_{s,t}} \mathbb{E}[\overline{U}_H(\mathbf{x})] = o\left(\sum_{H \in \mathrm{NBSAW}_{s,t}} \overline{L}_H \right). \tag{4.5.7}$$

   This will imply that

$$(1 \pm o(1)) \cdot n^t \left(\frac{2}{\varepsilon \cdot d} \right)^{st} \cdot \sum_{H \in \mathrm{NBSAW}_{s,t}} \overline{L}_H$$

   is a good lower bound for $\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}}) \right)^t \right]$.

Next, to upper bound $\left\| Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T} \right\|_t$ we will need the following two additional ingredients:

4. Show that the class of multigraphs $\mathrm{NBSAW}_{s,t}$ is strongly correlated to $\mathbf{x}$ in the sense that for many $H \in \mathrm{NBSAW}_{s,t}$

$$\mathbb{E}\left[\hat{\mathbf{Y}}_H \right] \leqslant (1+\delta)^{-t} \cdot \overline{L}_H \,.$$

To get an intuition of why this should be true, notice that for any self-avoiding walk $W \in \mathrm{SAW}^s_{ij}$ and for any $(\mathbf{x}, \mathbf{G}) \sim \mathsf{SBM}_n(d, \varepsilon)$, we have $\mathbb{E}\left[\mathbf{Y}_W - \left(\frac{\varepsilon \cdot d}{2n}\right)^s \cdot \mathbf{x}_i \mathbf{x}_j \;\middle|\; \mathbf{x}\right] = 0$.

5. The class of multigraphs $\mathrm{NBSAW}^c_{s,t}$ is poorly correlated with $\mathbf{x}$ in the sense that for any $H \in \mathrm{NBSAW}^c_{s,t}$

$$\mathbb{E}\left[\hat{\mathbf{Y}}_H \;\middle|\; \mathbf{x}\right] \approx \overline{U}_H(\mathbf{x})\,.$$

Together with step 3 these will imply a good upper bound on $\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t\right]$.

Proof strategy for Theorem 4.44. For $u, v \in [n]$, let $\mathrm{BSAW}_{s,t,u} \subseteq \mathrm{BSAW}_{s,t}$ the set of block self-avoiding walks having u as pivot. To provide concentration we will use a similar approach to the one outlined above.

1. Find a nice set of multigraphs $\mathrm{NMULTIG}_{s,t,u,v} \subseteq \mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v}$ such that

$$\sum_{H \in \mathrm{NMULTIG}^c_{s,t,u,v}} \mathbb{E}\left[\overline{U}_H(\mathbf{x})\right] \leqslant o(1) \sum_{H \in \mathrm{NMULTIG}_{s,t,u,v}} \mathbb{E}\left[\overline{\mathbf{Y}}_H\right].$$

2. Show that for such nice multigraphs

$$\sum_{H \in \mathrm{NMULTIG}_{s,t,u,v}} \mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant (1 + o(1)) \sum_{H \in \mathrm{BSAW}_{s,t,u}} \mathbb{E}[Y_H]^2\,,$$
$$\sum_{H \in \mathrm{NMULTIG}_{s,t,u,u}} \mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant (1 + o(1)) \sum_{H \in \mathrm{BSAW}_{s,t,u}} \mathbb{E}[Y_H]^2$$

for some universal constant $C > 1$.

Combining the two we will obtain the theorem.

4.5.1.3 Additional notation

We introduce some additional definitions which will be helpful in our discussion of block self-avoiding walks. We will introduce additional notation when needed. We suggest the impatient reader to skip the section and come back here when needed.

For simplicity at times we write $\mathbf{X}$ for $\mathbf{x}\mathbf{x}^\mathsf{T}$. For a multigraph H with vertex set $V(H) \subseteq [n]$ and a vertex $v \in V(H)$ we write $d^H_1(v)$ to denote the number of edges in H of multiplicity 1 incident to v. Similarly, we write $d^H_{\geqslant 2}(v)$ to denote the number of distinct edges in H of multiplicity at least 2 incident to v. Then $d^H(v) = d^H_1(v) + d^H_{\geqslant 2}(v)$, notice that $d^H(v)$ is the number of *distinct* edges incident to v.

Definition 4.48 (Underlying graph)**.** Let $H = (V, M)$ be a multigraph. Let $G = (V, E)$ be the graph with vertex set $V(G) = V(H)$ and edge set $E(G)$ such that $\{i, j\} \in E(G)$ if there exists an edge $e \in M(H)$ with endpoints i, j. We call $G = (V, E)$ the underlying graph of H and denote it with $G(H)$.

Definition 4.49. For integers $v \leqslant m$ we denote by $\mathcal{T}(m, v)$ the set of non-isomorphic trees (picking one arbitrary representative per class) on m vertices with v leaves.

Definition 4.50 (Extension Set)**.** Let G be a graph on m vertices and r edges. For any $q \in \{r, m(m+1)/2\}$, let $\mathcal{G}(G, q)$ be the set containing a representative graph from each isomorphic class of graphs obtained from G by adding *exactly* $q - r$ edges. We call $\mathcal{G}(G, q)$ the q-extension set of G. Notice that, trivially $\mathcal{G}(G, r) = \{G\}$.

Figure 4.1: Example of tree T on 9 edges and a graph $G \in \mathcal{G}(T, 10)$.

4.5.2 Lower bound for non-centered Schatten norm

We prove here the following theorem, which implies the first half of Theorem 4.43.

Theorem 4.51. *Consider the settings of Theorem 4.43. Then*

$$\sum_{H \in BSAW_{s,t}} \mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \geqslant \frac{1}{10n^{\frac{2}{50A}}} \cdot \left(\frac{\varepsilon \cdot d}{2}\right)^{st} \cdot \left(1 - n^{-\frac{1}{10}} - (1+\delta)^{-t\sqrt{s}}\right).$$

Our proof will roughly consist of the first three steps of the strategy outlined in Section 4.5.1.2. Specifically, we show step one in Section 4.5.2.1. In section Section 4.5.2.2 we show that $\sum_{H \in \mathrm{NBSAW}^c_{s,t}} \mathbb{E}\, U_H(\mathbf{x})$ is small, hence preparing the ground for an inequality of the form Eq. (4.5.7). Finally, in Section 4.5.2.3 we will obtain a lower bound for nice block self-avoiding walks. Taken together, these results will imply Theorem 4.51.

4.5.2.1 An upper bound for every multigraph

In this section we show an upper bound on the expectation of $\mathbf{Y}_H$ for any multigraph H. In order to do this we need to introduce some definitions. Let H be a *multigraph* (hence possibly not a block self-avoiding walks) with vertex set $V(H) \subseteq [n]$.

Definition 4.52. We classify the vertices $v \in V(H)$ according to there degree-1 as follows:

- If $d_1^H(v) \leqslant \tau$, we say that v is 1-small in H. We denote the set of 1-small vertices in H as $\mathcal{S}_1(H)$.
- If $d_1^H(v) > \tau$, we say that v is 1-large in H. We denote the set of 1-large vertices in H as $\mathcal{L}_1(H)$.

Definition 4.53. We classify the vertices $v \in V(H)$ according to their degree-$\geqslant 2$ as follows:

- If $d_{\geqslant 2}^H(v) \leqslant \frac{\Delta}{2}$, we say that v is ($\geqslant 2$)-small in H. We denote the set of ($\geqslant 2$)-small vertices in H as $\mathcal{S}_{\geqslant 2}(H)$.
- If $\frac{\Delta}{2} < d_{\geqslant 2}^H(v) \leqslant \Delta$, we say that v is ($\geqslant 2$)-intermediate in H. We denote the set of ($\geqslant 2$)-intermediate vertices in H as $\mathcal{I}_{\geqslant 2}(H)$.
- If $d_{\geqslant 2}^H(v) > \Delta$, we say that v is ($\geqslant 2$)-large in H. We denote the set of ($\geqslant 2$)-large vertices in H as $\mathcal{L}_{\geqslant 2}(H)$.

Definition 4.54. We denote the set of edges of multiplicity 1 as $E_1(H)$, and denote the set of edges of multiplicity at least 2 as $E_{\geqslant 2}(H)$. An edge of multiplicity 1 is said to be *annoying* if one of its end vertices is in $\mathcal{L}_1(H)$. We denote the set of annoying edges of multiplicity 1 as $E_1^a(H)$. We partition $E_{\geqslant 2}(H)$ into two sets:

$$E_{\geqslant 2}^a(H) = \{uv \in E_{\geqslant 2}(H) : u \notin \mathcal{L}_{\geqslant 2}(H) \text{ and } v \notin \mathcal{L}_{\geqslant 2}(H)\},$$

and

$$E_{\geqslant 2}^b(H) = \{uv \in E_{\geqslant 2}(H) : u \in \mathcal{L}_{\geqslant 2}(H) \text{ or } v \in \mathcal{L}_{\geqslant 2}(H)\}.$$

Definition 4.55. For every *multigraph* H, we define the quantity

$$\begin{aligned}\overline{U}_H(x) =& 2n^{\frac{1}{50A}} \frac{1}{\prod\limits_{v \in \mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d_{\geqslant 2}^H(v) - \Delta\right)}} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{2d}{n}\right)^{|E_{\geqslant 2}^b(H)|} \\ &\cdot \prod_{v \in V(H)} \left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v) - \tau, 0\right\}} \prod_{uv \in E_{\geqslant 2}^a(H)} \left[\left(1 + \frac{\varepsilon x_u x_v}{2}\right)\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right].\end{aligned}$$

We can now present a general upper bound for the expectation of block self-avoiding walks.

Lemma 4.56. *Consider the settings of Theorem 4.43. For every multigraph H with $V(H) \subseteq [n]$, $|V(H)| \leqslant s \log n$, let $\overline{U}_H(x)$ be as in Definition 4.55. Then*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H | x\right]\right| \leqslant \overline{U}_H(x),$$

for n large enough.

We prove Lemma 4.56 in Appendix B.1. Notice that, for block self-avoiding walks, the expression

$$\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \prod_{uv\in E_{\geqslant 2}(H)} \left[\left(1+\frac{\varepsilon x_u x_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]$$

is roughly the upper bound that we can get for $|\mathbb{E}[\mathbf{Y}_H|x]|$ in the non-truncated case (see Section 4.2). Therefore, truncation has the effect of:

- An amplification by a factor $n^{\frac{1}{50A}}$.
- An amplification by a factor of $\left(\frac{6}{\varepsilon}\right)^{2d_1^H(v)-\tau}$ for every vertex in $\mathcal{L}_1(H)$.
- For every $uv \in E^b_{\geqslant 2}(H)$, $\left[\left(1+\frac{\varepsilon x_u x_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]$ is replaced by $\frac{2d}{n}$.
- For every $uv \in E^a_{\geqslant 2}(H)$, $\left[\left(1+\frac{\varepsilon x_u x_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]$ is replaced by $\left[\left(1+\frac{\varepsilon x_u x_v}{2}\right)\frac{d}{n}+\frac{3d^2}{n\sqrt{n}}\right]$.
- A reduction by a factor of $n^{\frac{1}{4}(d^H_{\geqslant 2}(v)-\Delta)}$ for every $(\geqslant 2)$-large vertex v in H.

4.5.2.2 Walks that are not nice have negligible contributions

With the tools developed in Section 4.5.2.1 we can now approach Theorem 4.51. Remember from Section 4.5.1.2 we want to show that only a specific subset of block self-avoiding walks have large contribution to the expectation Eq. (4.5.3). The next lemma formalizes this idea.

Lemma 4.57. *Consider the settings of Theorem 4.43. Let $\mathrm{NBSAW}_{s,t}$ be the set of block self-avoiding walks H with the following structure:*

- *Every edge $e \in E(H)$ satisfies $m_H(e) \leqslant 2$.*
- *Every vertex $v \in V(H)$ satisfies $d_1^H(v) \in \{0,2\}$, $d^H_{\geqslant 2}(v) \leqslant \Delta$.*
- *$E_1(H)$ is a non-empty cycle.*
- *$E_1(H)$ is a cycle on at least $t/\sqrt{A}$ edges.*
- *The edges of multiplicity 2 form a forest, i.e., $E_{\geqslant 2}(H)$ is a forest.*
- *Each connected component of the forest of edges of multiplicity 2 is connected to $E_1(H)$ through a single vertex.*

Then for n large enough

$$\sum_{H\in \mathrm{NBSAW}^c_{s,t}} \mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \left(n^{-\frac{1}{6}}+(1+\delta)^{-t\sqrt{s}}\right) \sum_{H\in \mathrm{NBSAW}_{s,t}} \mathbb{E}\,\overline{U}_H(\mathbf{x}).$$

Block self-avoiding walks in $\mathrm{NBSAW}_{s,t}$ are said to be *nice*. Block self-avoiding walks in $\mathrm{NBSAW}^c_{s,t}$ are said to be *negligible*.

Bounding negligible block self-avoiding walks with few vertices. We start our proof of Lemma 4.57 with an observation. There are many block self-avoiding walks that "*clearly*" have a negligible contribution in the expectation of Eq. (4.5.3). To get some intuition, consider the following example. Let $\mathcal{S}$ (we will use this notation *only* for this specific example) be the set of all block self-avoiding walks in $\mathrm{BSAW}_{s,t}$ which have all vertices with degree 2 and all edges with multiplicity 1. That is, each walk in $\mathcal{S}$ is a cycle. Let $\mathcal{S}'$ be instead the subset of block self-avoiding walks in which all but one vertex have degree 2, one vertex has degree 4 and all edges have multiplicity 1. Now it is immediate to see that for any $H \in \mathcal{S}$ and $H' \in \mathcal{S}'$ we have $\mathbb{E}\left[\overline{U}_H(\mathbf{x})\right] = \mathbb{E}\left[\overline{U}_{H'}(\mathbf{x})\right]$ but

$$\sum_{H' \in \mathcal{S}'} \mathbb{E}\left[\overline{U}_{H'}(\mathbf{x})\right] = (1 \pm o(1)) \cdot n^{-1} \cdot \sum_{H \in \mathcal{S}} \mathbb{E}\left[\overline{U}_H(\mathbf{x})\right],$$

where the inequality follows simply because we have $|\mathcal{S}| \approx n \cdot |\mathcal{S}'|$. In Lemma 4.60 we formalize this and similar observations. We introduce first additional tools.

Fact 4.58. *Consider the settings of Theorem 4.43 and let $\Psi \geqslant 0$. Let $H \in \mathrm{BSAW}_{s,t}$ be a multigraph on at most $O(t)$ vertices and let H^* be an induced sub-multigraph of H satisfying:*

1. *the maximum $(\geqslant 2)$-degree in H^* is $\Psi \geqslant 0$,*
2. *all the edges in the cut $H(V(H), V(H) \setminus V(H^*))$ have multiplicity one in H.*

We denote $\ell, q \geqslant 0$ as the number of multiplicity-1 edges in H^* and $H(V(H), V(H) \setminus V(H^*))$ respectively. Let Z be a set of vertices in $V(H^*)$ such that $H(V(H^*) \setminus Z)$ has no multiplicity-2 cycles. Then

$$\begin{aligned}\mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{4} n^{-1/25A} \left(\frac{6}{\varepsilon}\right)^{2\ell+2q} \mathbb{E}\,\overline{U}_{H(V, V\setminus V(H^*))}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus V(H^*))}(\mathbf{x}) \\ \cdot \left(1 + \frac{\varepsilon}{2}\right)^{|Z|\cdot\Psi} \cdot \left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|} \cdot \prod_{v \in V(H^*)} \left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v) - \tau, 0\right\}} \\ \cdot \frac{1}{\prod\limits_{v \in \mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v) - \Delta\right)}} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|} \prod_{e \in E^a_{\geqslant 2}(H^*)} \left[\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right],\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\,\overline{U}_H(\mathbf{x}) \geqslant \frac{1}{4} n^{-1/25A}\, \mathbb{E}\,\overline{U}_{H(V, V\setminus V(H^*))}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus V(H^*))}(\mathbf{x}) \\ \cdot \left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|} \cdot \prod_{v \in V(H^*)} \left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v) - \tau, 0\right\}} \\ \cdot \frac{1}{\prod\limits_{v \in \mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v) - \Delta\right)}} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|} \prod_{e \in E^a_{\geqslant 2}(H^*)} \left[\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right],\end{aligned}$$

We prove Fact 4.58 in Appendix B.4.1.

We now temporarily limit our analysis to block self-avoiding walks with bounded maximum degree-($\geqslant 2$). As observed in Section 4.5.2.1, walks with large degree-($\geqslant 2$) will have small contribution to Eq. (4.5.3) and will be easily bounded. Let

$$\mathrm{BSAW}_{s,t,d^H_{\geqslant 2}\leqslant\Delta} := \left\{ H \in \mathrm{BSAW}_{s,t} \;\middle|\; \max_{v\in v(H)} d^H_{\geqslant 2}(v) \leqslant \Delta \right\},$$

we require another definition.

Definition 4.59. Let $H \in \mathrm{BSAW}_{s,t,d^H_{\geqslant 2}\leqslant\Delta}$ and let $v \in V(H)$. We denote by $q_H(v)$ the number of connected components of the line graph with vertex set $E_H(v)$ and such that there is an edge between $e, e' \in E_H(v)$ if and only if e, e' appear in the sequence of edges $M(\mathcal{W}(H))$ consecutively. Let $q_H := \sum_{v\in V(H)} (q_H(v) - 1)$. Now, for $q \geqslant 0$ we define $\mathcal{D}_{q,s,t} \subseteq \mathrm{BSAW}_{s,t,d^H_{\geqslant 2}\leqslant\Delta}$ to be the subset of block self-avoiding walks H with $q_H = q$. We also write $\mathcal{D}_{q,s,t}(H) \subseteq V(H)$ to be the set of vertices in H with $q_H(v) \geqslant 2$. Finally we write

$$\mathcal{D}_{\geqslant 1,s,t} = \bigcup_{q\geqslant 1} \mathcal{D}_{q,s,t}.$$

When the context is clear we write $\mathcal{D}_q$ instead of $\mathcal{D}_{q,s,t}$.

We can now prove that block self-avoiding walks in $\mathcal{D}_{\geqslant 1}$ have negligible contribution to the expectation of Eq. (4.5.3).

Lemma 4.60. *Consider the settings of Theorem 4.43. Then for n large enough*

$$\sum_{H\in\mathcal{D}_{\geqslant 1}} \mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{n^{2/3}} \sum_{H\in BSAW_{s,t,d^H_{\geqslant 2}\leqslant\Delta}\setminus\mathcal{D}_{\geqslant 1}} \mathbb{E}\,\overline{U}_H(\mathbf{x}).$$

Proof. Fix $q \geqslant 1$ and consider the following procedure to obtain a block self-avoiding walk in $\mathrm{BSAW}_{s,t,d^H_{\geqslant 2}\leqslant\Delta} \setminus \mathcal{D}_q$ from a block self-avoiding walk $H \in \mathcal{D}_q$. Let $M(\mathcal{W})$ be the sequence of edges obtained concatenating the generating self-avoiding walks of H so that the subsequence $\left\{e^{(\ell-1)\cdot s+1}, \ldots, e^{(\ell-1)\cdot s+s}\right\}$ corresponds to the ℓ-th generating self-avoiding walk of H (for simplicity we let $i - 1 = st$ for $i = 1$ and analogously we let $i + 1 = 1$ for $i = st$). Let $v \in \mathcal{D}_q(H)$. Let $F_{H,v}$ be the line graph with vertex set $E_H(v)$ and edges as described in Definition 4.59. Let $E_H(v)^1$ be an arbitrary connected component of $F_{H,v}$ and let u be a vertex not in H. We construct the block self-avoiding walk $H' \in \mathcal{D}_{q-1}$ with $V(H') = V(H) \cup \{u\}$ applying the following operation on H:

- Consider the sequence of edges $M(\mathcal{W}(H))$, we replace every edge $vw \in M(\mathcal{W}(H))$ (and $wv \in M(\mathcal{W}(H))$) such that $vw \in E^1_H(v)$ with the edge uw (resp. wu).

Clearly, $H' \in \mathrm{BSAW}_{s,t,d_{\geqslant 2}^H \leqslant \Delta} \setminus \mathcal{D}_q$ and $|V(H')| - |V(H)| = 1$. Furthermore $\left|E_1^q(H')\right| \geqslant \left|E_1^q(H)\right| - \tau - 2$. Thus

$$\prod_{v \in V(H)} \left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v)-\tau,0\right\}} \leqslant \left(\frac{6}{\varepsilon}\right)^{\tau+2} \cdot \prod_{v \in V(H')} \left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^{H'}(v)-\tau,0\right\}}.$$

By Fact 4.58 it follows that

$$\frac{1}{n^{|V(H')|-|V(H)|}} \cdot \frac{\mathbb{E}\left[\overline{U}_H(\mathbf{x})\right]}{\mathbb{E}\left[\overline{U}_{H'}(\mathbf{x})\right]} \leqslant \frac{1}{n} \cdot \left(\frac{6}{\varepsilon}\right)^{2\tau} \left(1 + \frac{\varepsilon}{2}\right)^{2\Delta} \leqslant \frac{1}{n} \cdot \left(\frac{12}{\varepsilon}\right)^{3\Delta}. \tag{4.5.8}$$

To obtain a multi-graph not in $\mathcal{D}_{\geqslant 1}$ we repeatedly apply the operation above until $\mathcal{D}_q(H)$ is empty. Notice that $(st)^{O(q)}$ applications suffice. It remains to show that the contribution to the expectation of Eq. (4.5.3) of block self-avoiding walks in $\mathcal{D}_{\geqslant 1}$ is negligible. For this, observe that at each step there are at most $(st)^2$ block self-avoiding walks that can produce the same multigraph H'. So using Eq. (4.5.8), we get for any $q \geqslant 1$

$$\sum_{H \in \mathcal{D}_q} \mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{(st)^2}{n^{0.99}} \sum_{H' \in \mathcal{D}_{q-1}} \mathbb{E}\,\overline{U}_H(\mathbf{x}).$$

The result follows since the maximum degree in any block self-avoiding walk is $2t$. □

Bounding block self-avoiding walks from their shape and edges multiplicities. Next we develop a general bound on the contribution of block self-avoiding walks based on the shape of their underlying graph and the multiplicity of each edge. Together with Lemma 4.60 this will be enough to obtain Lemma 4.57. We will need the following definitions.

Definition 4.61. For a collection of disjoint connected graphs on at least two vertices $\mathcal{B} = \{B_1, \ldots, B_z\}$, we define the set $\mathcal{M}_{s,t}(B_1, \ldots, B_z)$ to be the subset of $\mathrm{BSAW}_{s,t}$ satisfying the following: if $H \in \mathcal{M}_{s,t}(B_1, \ldots, B_z)$, then for any $B \in \mathcal{B}$

(i) $B \subseteq G(H)$, we denote with B' a (arbitrary) copy of B in H and by $H(B)$ the multigraph induced by $V(B')$,

(ii) $\forall e \in E(H(B))$ that is also an edge in B, $m_H(e) \geqslant 2$,

(iii) there exists a cut $H(V(B'), V(H) \setminus V(B'))$ in H such that each edge in the cut has multiplicity 1 in H.

and furthermore

(iv) the copies $B'_1, B'_2, \ldots, B'_z$ are disjoint

(v) every edge in $H\left(V(H) \setminus \left(\bigcup_{j\in[z]} V(B'_j)\right)\right)$ has multiplicity 1.

With a slight abuse of notation, we will simply write $H(B)$ instead of $H(B')$.

Definition 4.62. Let $\mathcal{B} = \{B_1, \dots, B_z\}$ be a collection of disjoint connected graphs, let $\{\ell_i, q_i, p_i, h_i\}_{i=1}^z$ be a sequence of tuples of integers such that for all $i \in [z]$ $\ell_i, q_i, p_i, h_i \geqslant 0$. Further we denote $\mathcal{F}_i = \{\ell_i, q_i, p_i, h_i\}$. We write $\mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ for the subset of $\mathcal{M}_{s,t}(\mathcal{B})$ such that for any $i \in [z]$:

(i) the size of the cut $H(V(H) \setminus V(B_i), V(B_i))$ is ℓ_i

(ii) the number of edges in $H(B_i)$ of multiplicity one is q_i

(iii) the number of edges e in $H(B_i)$ with $m_H(e) = 2$ is h_i

(iv) the maximum degree-($\geqslant 2$) in $H(B_i)$ is Ψ_i.

(v) the edges with multiplicity larger than 2 in $H(B_i)$ satisfy

$$\sum_{\substack{e\in H(B_i)\\ m_H(e)\geqslant 3}} m_H(e) = p_i\,,$$

When the context is clear we simply write $\mathcal{M}(\mathcal{B})$ and $\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B}) \subseteq \mathcal{M}(B)$. For $\mathcal{B} = \{B\}$ we simply write $\mathcal{M}(B)$. For $\mathcal{B} = \emptyset$, the set $\mathcal{M}(\mathcal{B})$ corresponds to the set of block self-avoiding walks in $\mathrm{BSAW}_{s,t}$ where all edges have multiplicity 1. The next lemma studies the contribution of block self-avoiding walks in $\mathcal{M}(\mathcal{B})$ for all $\mathcal{B}$.

Lemma 4.63. *Consider the settings of Theorem 4.43. Let $z \geqslant 1$ and $m_1, \dots, m_z \geqslant 1$ be integers. Then for n large enough,*

$$\sum_{\substack{\text{for } i\in[z]:\\ q_i,\Psi_i,\ell_i,v_i\geqslant 0\\ T_i\in\mathcal{T}(m_i,v_i)\\ r_i\geqslant m_i-1\\ B_i\in\mathcal{G}(T_i,r_i)\\ h_i,p_i\geqslant 0}} \sum_{H\in\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\{B_1,\dots,B_z\})} \mathbb{E}\,\overline{U}_H(\mathbf{x})$$

$$\leqslant \prod_{i\in[z]}\left[(st)^{14}\cdot\left(\frac{200\Delta + s^{10} + 2^{\frac{\log n}{t}}}{(1+\delta)^{s/4}}\right)^{2m_i/s}\right]\cdot \sum_{H\in\mathcal{M}(\emptyset)} \mathbb{E}\,\overline{U}_H(\mathbf{x})\,.$$

Furthermore, restricting the sum over $r_i > m_i - 1$, or over $p_i \geqslant 1$, or $q_i \geqslant 1$, or $\Psi_i > \Delta$ or $\ell_i > 2$, for some $i \in [z]$ the inequality holds with an additional $n^{-\frac{1}{5}\left(p_i+q_i+\mathbb{1}_{[\Psi_i>\Delta]}(\Psi_i-\Delta)+\mathbb{1}_{[\ell_i>2]}(\ell_i-2)\right)}$ factor.

Lemma 4.63 formalizes the following idea: for all possible collections of z (possibly isomorphic) graphs $B_1, \dots, B_z$ respectively on $m_1, \dots, m_z$ vertices, the contribution to the expectation of Eq. (4.5.3) of block self-avoiding walks in $\mathcal{M}(\mathcal{B})$ can be upper bounded by $\sum_{H \in \mathcal{M}(\emptyset)} \mathbb{E}\,\overline{U}_H(\mathbf{x})$ times a scalar which depends on the order and the shape of the graphs. Indeed the sum on the left-hand side captures all possible choices of graphs $B_1, \dots, B_z$ and sets $\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\{B_1, \dots, B_z\})$. Moreover the lemma implies that the contribution of block self-avoiding walks in $\mathcal{M}(\mathcal{B})$ is negligible if any of the graphs in $\mathcal{B}$ contains a cycle, and that among the block self-avoiding walks in $\mathcal{M}(\mathcal{B})$, most of the mass is concentrated in a very specific subset of block self-avoiding walks. This last observation will be extremely useful in simplifying our analysis and prove Lemma 4.57. It can be observed how for certain block self-avoiding walks with few edges of multiplicity at least 2 this bound appears very rough, however, we can bound the contribution of these walks to the expectation of Eq. (4.5.3) using Lemma 4.60.

Concerning the parameters, $2(2h_j + p_j + q_j)/s + \ell_j$ is an upper bound on the maximum number of pivots of H that can be in $H(B_j)$. The parameter v_j (the number of degree 1 vertices in B_j) has a loose correspondence with the number of vertices u with $d_{\geqslant 2}^{H(B_j)}(u) = 1$ in $H(B)$, in the sense that $v_j - 2r_j \leqslant \left|\left\{u \in H(B_j) \;\middle|\; d_{\geqslant 2}^{H(B_j)}(u) = 1\right\}\right| \leqslant v$. Finally, recall from section Section 4.5.1.3 that with $\mathcal{T}(m, v)$ we denote the set of non-isomorphic (picking one arbitrary representative per class) trees on m vertices and v leaves. For a given tree T we let $\mathcal{G}(T, r)$ be the set of non-isomorphic graphs obtained from T adding $r - |E(T)|$ edges.

To prove Lemma 4.63 we need some intermediate results. First, we need to count how many block self-avoiding walks are in $\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ (for some choice of the parameters).

Lemma 4.64. *Consider the settings of Theorem 4.43. Let $\mathcal{B} = \{B_1, \dots, B_z\}$ be a collection of disjoint connected graphs each with respectively $m_1, \dots, m_z$ vertices. Let $\{\mathcal{F}_i\}_{i=1}^z$ be a sequence of tuples of integers as in Definition 4.62. Let $f_{s,t}, g_{s,t}$ be the functions*

$$f_{s,t}(m, m', \mathcal{F}, \Psi) = (\Psi)^{2h/s+10(q+\ell+p+1)+2h-2(m'-1)} \cdot (st)^{5\ell+5q+8p+4h+4-4(m'-1)},$$
$$g_{s,t}(m', \mathcal{F}) = n^{-p-\ell/2-q-2h+m'}.$$

Let $m = \sum_{j \in [z]} m_j$. Then there are at most $n^{st} \cdot \prod_{i \in [z]} f_{s,t}(m, m_i, \mathcal{F}_i, \Psi_i) \cdot g_{s,t}(m_i, \mathcal{F}_i)$ block self-avoiding walks in the set $\mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$.

We prove Lemma 4.64 in Appendix B.4.2 and directly use it here. Second we introduce bounds to split the expectation of $\overline{U}_H(\mathbf{x})$ into the expectation of its components.

Fact 4.65. *Consider the settings of Theorem 4.44. Let $\mathcal{B} = \{B_1, \dots, B_z\}$ be a collection of disjoint connected graphs on at least 2 vertices. Then for any $H \in \mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ and $i \in [z]$*

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{4} n^{-1/25A} \left(\frac{6}{\varepsilon}\right)^{2\ell_i+2q_i} \mathbb{E}\,\overline{U}_{H(V,V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(B_i)}(\mathbf{x}),$$

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \geqslant \frac{1}{4} n^{-1/25A}\,\mathbb{E}\,\overline{U}_{H(V,V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(B_i)}(\mathbf{x})\,.$$

We are now ready to prove Lemma 4.63.

Proof of Lemma 4.63. Our strategy will be the following: fix some graphs $B_1,\dots,B_{z-1}$ and some tuples $\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z-1}$. Then we will show that the contribution of block self-avoiding walks in $\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z}}(\{B_1,\dots,B_z\})$ for all possible choice of the graph B_z and the parameters $\mathcal{F}_z,\Psi_z$ is upper bounded by some function of $\sum_{H\in\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z-1}}(\{B_1,\dots,B_{z-1}\})} \mathbb{E}\,\overline{U}_H(\mathbf{x})$. This function will be easy to upper bound for the set $\mathcal{M}(\emptyset)$, thus by reiterating the analysis for $j=z,\dots,j=1$ we will obtain the desired bound.

Now, for simplicity, for any $H\in\mathcal{M}(\mathcal{B})$ let us write $H^*=H(V\setminus(B_1\cup\ldots\cup B_z))$. Define $\sum_{j\in[z]} m_j =: m$. Recall that for any m_z,v_z by Fact B.105 we have $|\mathcal{T}(m_z,v_z)|\leqslant 2v_z\cdot(8e\cdot m_z/v_z)^{2v_z}$ and that for any graph in the extension set of some tree in $T\in\mathcal{T}(m_z,v)$ we have at most m_z^2 possible choices for each additional edge. Combining these bounds with Lemma 4.64 we will be able to compute the number of block self-avoiding walks at hand.

We can start carrying out the computation. Fix some graphs $B_1,\dots,B_{z-1}$ and some tuples $\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z-1}$. By Fact 4.65, for any multigraph in $\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z}}(\{B_1,\dots,B_z\})$ for some $B_z,\mathcal{F}_z,\Psi_z$:

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \prod_{j\in[z]}\left[\left(\frac{6}{\varepsilon}\right)^{2\ell_j+2q_j}\cdot(2n^{1/50A})^{-5/2}\cdot\mathbb{E}\,\overline{U}_{H(B_j)}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(V(H^*),B_j)}(\mathbf{x})\cdot\prod_{k\in[z],k\neq j}\sqrt{\mathbb{E}\,\overline{U}_{H(B_j,B_k)}(\mathbf{x})}\right]$$
$$\cdot\,\mathbb{E}\,\overline{U}_{H^*}(\mathbf{x})\,, \tag{4.5.9}$$

where we used the squared root to avoid counting for the edges in the cut $H(B_j,B_k)$ twice and the factor $(2n^{1/50A})^{-5/2}$ appears due to the fact that we use multiple times upper bounds of the form $\overline{U}_H(\mathbf{x})$. For fixed $\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z-1}$ and fixed $m_z>0$ we can thus focus on bounding,

$$\sum_{\substack{q_z,\ell_z\geqslant 0\\ \Psi_z\geqslant 0}}\ \sum_{\substack{v_z\leqslant 2(2h_z+p_z+q_z)/s+\ell_z\\ T\in\mathcal{T}(m_z,v_z)}}\ \sum_{\substack{r\geqslant m_z-1\\ B_z\in\mathcal{G}(T,r):\\ \max\deg(B_z)=\Psi_z}}\ \sum_{\substack{h_z,p_z\geqslant 0\\ H\in\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z}}(\{B_1,\dots,B_z\})}}$$
$$\mathbb{E}\,\overline{U}_{H^*}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(B_z)}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(V(H^*),B_z)}(\mathbf{x})\cdot\prod_{k\in[z-1]}\sqrt{\mathbb{E}\,\overline{U}_{H(B_z,B_k)}(\mathbf{x})}\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_z+2q_z}. \tag{4.5.10}$$

For each $i\in[z]$, we let $\ell_i=\ell_i'+\ell_i''$ where ℓ_i' corresponds to the number of edges in $H(V(H^*),B_i)$ and ℓ_i'' corresponds to the number of remaining edges in the cut. Note that $h_z+|\{e\in H(B_z)\,:\,m_H(e\geqslant 3)\}|=r\leqslant h_z+p_z$. So, for fixed $\Psi_z,q_z,h_z,p_z,\ell_z,r,v_z$ it holds by Fact 4.58

$$\mathbb{E}\,\overline{U}_{H^*}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(B_z)}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(V(H^*),B_z)(\mathbf{x})}\cdot\prod_{k\in[z-1]}\sqrt{\mathbb{E}\,\overline{U}_{H(B_j,B_k)}(\mathbf{x})}\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_z+2q_z}$$

$$\leqslant \left[\prod_{u\in V(H^*)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^{H^*}(u)-\tau,0\right\}}\right]\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E(H^*)|}\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_z+2q_z}\cdot\left(\frac{\varepsilon d}{2n}\right)^{q_z+\ell_z'+\ell_z''/2}\cdot\left[\left(1+\frac{1}{\sqrt{n}}\right)\frac{d}{n}\right]^r$$
$$\cdot 2^{2(r-m_z-1)\cdot\Psi_z}\cdot\prod_{u\in\mathcal{L}_{\geqslant 2}(H(B_z))}\left[\left(\frac{2d}{n}\right)^{\frac{1}{4}\left(d^H_{\geqslant 2}(u)-\Delta\right)}\right]. \tag{4.5.11}$$

Notice that changing ℓ_z, q_z then the multigraph H^* changes and so does $\mathbb{E}\,\overline{U}_{H^*}(\mathbf{x})$. Moreover by Lemma 4.64, since $\Psi_z \leqslant 2t$ it follows that the contribution to Eq. (4.5.10) of block self-avoiding walks with $\Psi_z \geqslant \Delta$ will be at least a factor $n^{\frac{1}{5}}$ smaller.

By Lemma 4.60, we may assume there are no annoying edges in H^*. Thus we may upper bound Eq. (4.5.10) by

$$n^{st}\cdot\left[\prod_{j\in[z-1]} f_{s,t}\left(m,m_j,\mathcal{F}_j,\Psi_j\right)\cdot g_{s,t}(m_j,\mathcal{F}_j)\right]$$
$$\cdot\sum_{\substack{p_z,h_z,\ell_z,q_z\geqslant 0\\ r\geqslant m_z-1\\ \Psi_z\leqslant\Delta\\ v_z\leqslant 2(2h_z+p_z+q_z)/s+\ell_z}}(8e\cdot m_z/v_z)^{2v_z}\cdot f_{s,t}(m,m_z,\mathcal{F}_z,\Psi_z)\cdot g_{s,t}(m_z,\mathcal{F}_z) \tag{4.5.12}$$
$$\cdot\left(\frac{\varepsilon d}{2n}\right)^{st-\sum_{i\in[z]}(2h_i+p_i+q_i+\ell_i'+\ell_i''/2)}\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_z+2q_z}\cdot\left(\frac{\varepsilon d}{2n}\right)^{q_z+\ell_z'+\ell_z''/2}\cdot\left[\left(1+\frac{1}{\sqrt{n}}\right)\frac{d}{n}\right]^r$$
$$\cdot(2m_z)^{2(r-m_z-1)}\cdot 2^{2(r-m_z-1)\cdot\Psi_z}. \tag{4.5.13}$$

By Eq. (4.5.11) and Lemma 4.64 it is easy to see that Eq. (4.5.13) is a geometric sum. So define the quantity

$$M_{\ell_z',\ell_z''} := (st)^4\cdot\Delta^{2m_z/s+10}\cdot(100e\cdot s)^{2(m_z/s+\ell_z)}\cdot\left(\frac{\varepsilon\cdot d}{2n}\right)^{\ell_z'+\ell_z''/2}\cdot\left(\frac{d}{n}\right)^{m_z-1}\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_z}$$
$$\cdot f_{s,t}(m,m_z,\{\ell_z,0,0,m_z-1\},\Psi_z)\cdot g_{s,t}(m_z,\{\ell_z,0,0,m_z-1\}).$$

If, for fixed $\ell_z=\ell_z'+\ell''>0, h_z$ we restrict the sum Eq. (4.5.13) to the case $p_z\geqslant 1$ we get the upper bound

$$\leqslant n^{st}\cdot\left[\prod_{j\in[z-1]} f_{s,t}\left(m,m_j,\mathcal{F}_j,\Psi_j\right)\cdot g_{s,t}(m_j,\mathcal{F}_j)\right]$$
$$\cdot n^{-1/3}\sum_{\substack{\ell_z\geqslant 0\\ r\geqslant m_z-1\\ \Psi_z\leqslant\Delta\\ v_z\leqslant 2(2h_z+q_z)/s+\ell_z}} M_{\ell_z}\cdot\left(\frac{\varepsilon d}{2n}\right)^{st-\sum_{i\in[z]}(2h_i+p_i+q_i+\ell_i'+\ell_i''/2)}\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_z+2q_z}\cdot\left(\frac{\varepsilon d}{2n}\right)^{q_z+\ell_z'+\ell_z''/2}$$

$$\cdot\left[\left(1+\frac{1}{\sqrt{n}}\right)\frac{d}{n}\right]^{r}\cdot(2m_z)^{2(r-m_z-1)}\cdot 2^{2(r-m_z-1)\cdot\Psi_z}\,.$$

Similarly, restricting the sum to the case $r > m_z - 1$ (since for any additional edge in B we then have at most m_z^2 possible choices) or $q_z \geqslant 1$ we also have $n^{-\frac{1}{3}}\cdot M_{\ell_z}$. It remains to consider Eq. (4.5.13) for $q_z = 0\,, r = m_z - 1\,, p_z = 0$ for any ℓ_z. To do this we study first the behavior of M_{ℓ_z} as ℓ_z grows. We distinguish two cases depending on whether $\ell_z = 0$.

If $\ell_z = 0$, it means that for any of the graphs considered we had $H(B) = H$ and $\mathcal{B} = \{B\}$. Thus $2m_z + p_z = st$ and we get

$$\begin{aligned}M_0 \leqslant&(st)\cdot(100e\cdot s)^{4t}\cdot\left(\frac{d}{n}\right)^{st/2}\\&\cdot f_{s,t}(st/2+1, st/2+1, \{0,0,0,st/2\},\Delta)\cdot g_{s,t}(st/2+1,\{0,0,0,st/2\})\\&\leqslant\left(100\Delta+s^{10}\right)^{2t}\cdot(st)^6\cdot n\cdot d^{st/2}\,,\end{aligned}$$

which yields a ratio between Eq. (4.5.13) and $\sum_{H\in\mathcal{M}(\emptyset)}\mathbb{E}\,\overline{U}_H(\mathbf{x})$ of

$$\left(\frac{\varepsilon^2}{4d}\right)^{st/2}\cdot\left(200\Delta+s^{10}+C\right)^{2t}=\left(\frac{200\Delta+s^{10}+C}{(1+\delta)^{s/4}}\right)^{2t},$$

where $C = 2^{2\frac{\log n}{t}}$. Conversely, suppose $\ell_z \geqslant 2$ (it must be even given that the graph is Eulerian). Then

$$\begin{aligned}M_{\ell_z}\leqslant(\Delta)^{2m_z/2+10\ell_z}\cdot&(st)^{6+5\ell_z}\cdot\left(\frac{d}{n}\right)^{m_z-1}\cdot\left(\frac{d}{n}\cdot\left(\frac{6}{\varepsilon}\right)^4\right)^{\ell_z/2}\\&\cdot f_{s,t}(m,m_z,\{\ell_z,0,0,m_z-1\},\Delta)g_{s,t}(m_z,\{\ell_z,0,0,m_z-1\})\,.\end{aligned}$$

For $\ell_z = 2$ the ratio with

$$n^{st}\cdot\mathbb{E}\,\overline{U}_{H^*}(\mathbf{x})$$
$$\prod_{j\in[z-1]}\left[f_{s,t}(m,m_j,\mathcal{F}_j,\Psi_j)g_{s,t}(m_j,\mathcal{F}_j)\left(\frac{6}{\varepsilon}\right)^{2\ell_j+2q_j}\mathbb{E}\,\overline{U}_{H(B_j)}(\mathbf{x})\,\mathbb{E}\,\overline{U}_{H(H^*,B_j)}(\mathbf{x})\prod_{k\in[z],k\neq j}\sqrt{\mathbb{E}\,\overline{U}_{H(B_j,B_k)}(\mathbf{x})}\right]$$

can be bounded by

$$(st)^{14}\left(\frac{200\Delta^{10}}{(1+\delta)^{s/4}}\right)^{2m_z/s}.$$

We get an additional $n^{-\frac{\ell_z-2}{5}}$ factor if $\ell_z > 2$. We can now reiterate the analysis on $\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^{j}}(\{B_1,\ldots,B_j\})$ for $j = z-1,\ldots,1$. The result follows. □

Putting things together. We are now ready to prove Lemma 4.57.

Proof of Lemma 4.57. We argue that if a block self-avoiding walk H is not in $\mathrm{NBSAW}_{s,t}$, then it satisfies one of the following:

- $H \in \mathcal{D}_{\geqslant 1}$,
- $H \in \mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ for any non-empty $\mathcal{B}$ and tuples $\{\mathcal{F}_i\}_{i=1}^z$ with $p_i \geqslant 1$ or $q_i \geqslant 1$ or $\ell_i \geqslant 3$ or $\Psi_i > \Delta$ for some $i \in [z]$,
- $H \in \mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ for any non-empty $\mathcal{B}$ containing some B_i with a cycle,
- all edges in H have multiplicity at least 2.

Suppose this claim holds. The contribution of walks satisfying one of the first three bullets is negligible by Lemma 4.60 and Lemma 4.63. Thus we only need to consider walks with all edges with multiplicity 2. By Lemma 4.63, it suffices to show that

$$2\left(\frac{300\Delta + s^{10} + 2^{\frac{\log n}{t}}}{(1+\delta)^{s/4}}\right) \leqslant (1+\delta)^{-10\sqrt{s}}, \tag{4.5.14}$$

which, for $t \in \left[\frac{\log n}{400}, \frac{\log n}{100}\right]$, may be rewritten as

$$s \geqslant \frac{10^8}{\delta}\left(\max\left\{\log 300\Delta, \log s, 1\right\}\right)^2.$$

By assumption on s, Eq. (4.5.14) is satisfied and thus the desired inequality follows.

It remains to verify our claim. So consider some H in $\mathrm{NBSAW}^c_{s,t}$. Suppose that there is some vertex $v \in V(H)$ with $d^H_{\geqslant 2}(v) = 0$ and $d^H_1(v) \geqslant 4$, by construction then $H \in \mathcal{D}_{\geqslant 1}$. Suppose now $H \in \mathcal{M}(\mathcal{B})$ for some collection of graphs $\mathcal{B} = \{B_1, \ldots, B_z\}$. We may assume that $H \in \mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ where for all $i \in [z]$, $\ell_z = 2, q_z = p_z = 0, , h_z = m_z - 1, \Psi_z \leqslant \Delta$ as otherwise H satisfies one of the bullet points listed above. Moreover, for any $i \in [z]$ the graph B_i must be a tree. Now, the claim would follow if we can show that for any $i \in [z]$ $H(B_i)$ is connected to $H(V \setminus B_i)$ by a single vertex. Suppose this is not true, since by assumption $\ell_i = 2$, the vertex-cut between $H(B_i)$ and $H(V \setminus B_i)$ must have cardinality 2. We obtain a contradiction since this implies that $q_i + p_i \geqslant 1$ as H is a closed walk. □

4.5.2.3 Bounding the non-centered Schatten norm

The machinery of Section 4.5.2.1 allowed us to upper bound the expectation of any block self-avoiding walk under the truncated distribution. Then in Section 4.5.2.2 we used such machinery to show that certain block self-avoiding walk have small contribution to the expectation of Eq. (4.5.3). In this section we prove Theorem 4.51.

Our strategy will be the following. For most nice block self-avoiding walks we will be able to lower bound the expectation. THe remaining ones will be few and have negligible contribution to the expectation of Eq. (4.5.3). Combining these observations with Lemma 4.57 will yield reftheorem:truncation-schatten-norm-lower-bound.

Bounds on nice block self-avoiding walks. We start proving additional bounds on the expectation of nice block self-avoiding walks. We further divide the set of nice block self-avoiding walks into sets.

Definition 4.66. For $\gamma \geqslant 0$ define the set

$$\mathrm{NBSAW}_{s,t,d^H \leqslant \gamma} := \left\{ H \in \mathrm{NBSAW}_{s,t} \;\middle|\; \forall v \in V(H)\,, d^H(v) \leqslant \gamma \right\}.$$

For $m, z \geqslant 1$ let $\mathrm{NBSAW}_{s,t,m,z}$ be the subset of block self-avoiding walks $H \in \mathrm{NBSAW}_{s,t}$ such that the graph obtained from $G(H)$ by removing all edges of multiplicity 1 in H is a forest on m vertices and z components. Notice that if $H \in \mathrm{NBSAW}_{s,t,m,z}$ then $|E_{\geqslant 2}(H)| = m - z$.We also define

$$\mathrm{NBSAW}_{s,t,d^H \leqslant \gamma,m,z} := \mathrm{NBSAW}_{s,t,d^H \leqslant \gamma} \cap \mathrm{NBSAW}_{s,t,m,z}$$

.

Observe that for $m = z = 0$ we have

$$\mathrm{NBSAW}_{s,t,0,0} = \mathrm{NBSAW}_{s,t} \cap \mathcal{M}_{s,t}(\emptyset)\,.$$

For many nice block self-avoiding walks, we can lower bound the expectation.

Lemma 4.67. *Consider the settings of Theorem 4.43. Let m, z be integers such that $st - 2m - 2z \geqslant t/A$. If n is large enough, then for every $H \in NBSAW_{s,t,d^H \leqslant \Delta,m,z}$*

$$\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \geqslant \overline{L}_H := \frac{1}{3n^{\frac{1}{100A}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

We present the proof of Lemma 4.67 in Appendix B.1.3, and directly use the result. For the remaining nice block self-avoiding walks, the next two results upper bound their expectation.

Lemma 4.68. *Consider the settings of Theorem 4.43. If n is large enough, then for any $H \in NBSAW_{s,t} \setminus NBSAW_{s,t,d^H \leqslant \Delta}$*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]\right| \leqslant \frac{4}{n^{1/12}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

We prove Lemma 4.68 in Appendix B.1.3.

Fact 4.69. *Consider the settings of Theorem 4.43. Let $m, z \geqslant 0$ be integers. If n is large enough, then for any $H \in NBSAW_{s,t,m,z}$ and any $H' \in NBSAW_{s,t,0,0}$*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]\right| \leqslant 6n^{3/100A} \cdot [(1 - o(1))(1 + \delta)]^{-m+z} \cdot n^{m-z} \cdot \overline{L}_{H'}\,.$$

Proof. Notice that

$$\frac{\mathbb{E}\,\overline{U}_H(\mathbf{x})}{\mathbb{E}\,\overline{U}_{H'}(\mathbf{x})} \leqslant \left(\frac{d}{n}\right)^{z-m} \cdot \left(\frac{(1 - o(1))\varepsilon}{2}\right)^{2z-2m} = [(1 - o(1))(1 + \delta)]^{-m+z} \cdot n^{m-z}\,.$$

Thus the result follows applying Lemma 4.67 and the definition of $U_H(\mathbf{x})$. □

Counting nice block self-avoiding walks. Next we count the number of nice block self-avoiding walks with maximum degree larger than Δ or few edges of multiplicity 1.

Lemma 4.70. *Consider the settings of Theorem 4.43. Define the set*

$$NBSAW_{s,t,m,z,\ell_1,\ell_2} := \Big\{H \in NBSAW_{s,t,m,z} \;\Big|\; \big|\{v \in v(H) \;|\; d^H(v) = \Delta+1\}\big| = \ell_1\,,\ \big|\{v \in v(H) \;|\; d^H(v) = \Delta+2\}\big| = \ell_2\Big\}$$

Then for n large enough, there exists $\ell \geqslant \ell_1 + 2\ell_2$ such that

$$|NBSAW_{s,t,m,z,\ell_1,\ell_2}| \leqslant (1+o(1)) \cdot 2^t \cdot n^{-\ell} \big|NBSAW_{s,t,d^H \leqslant \Delta, m', z'}\big|\,.$$

for some $m', z' \geqslant 0$ such that $m - z = m' - z' + \ell$.

We prove Lemma 4.70 in Appendix B.4.2.

Lemma 4.71. *Consider the settings of Theorem 4.43. Let $m, z \geqslant 0$ be integers such that $st - 2m - 2z \geqslant t/10\sqrt{A}$. Then*

$$|NBSAW_{s,t,m,z}| \leqslant 2^{10t} s^t \cdot n^{-m+z} \cdot |NBSAW_{s,t,0,0}|\,.$$

We also prove Lemma 4.71 in Appendix B.4.2.

Putting things together. We can now bound the contribution to the expectation of Eq. (4.5.3) of nice block self-avoiding walks, and hence of all block self-avoiding walks of length st.

Lemma 4.72. *Consider the settings of Theorem 4.43. Then for n large enough*

$$\sum_{H \in NBSAW_{s,t}} \mathbb{E}\Big[\overline{\mathbf{Y}}_H\Big] > \frac{1}{9n^{\frac{2}{50A}}} \cdot \left(\frac{\varepsilon \cdot d}{2}\right)^{st}.$$

Proof. By Lemma 4.70 and Lemma 4.68, the contribution of nice block self-avoiding walks with maximum degree larger than Δ can be bounded by

$$\begin{aligned}
\sum_{m,z \geqslant 1} \sum_{H \in \mathrm{NBSAW}_{s,t,m,z} \setminus \mathrm{NBSAW}_{s,t,d^H \leqslant \Delta}} \mathbb{E}\Big[\overline{\mathbf{Y}}_H\Big] &\leqslant \sum_{m,z \geqslant 1} \sum_{H \in \mathrm{NBSAW}_{s,t,m,z} \setminus \mathrm{NBSAW}_{s,t,d^H \leqslant \Delta}} \mathbb{E}\Big[\overline{U}_H(\mathbf{x})\Big] \\
&\leqslant 2^t \cdot n^{-1/12} \cdot \sum_{m,z \geqslant 1} \sum_{H \in \mathrm{NBSAW}_{s,t,d^H \leqslant \Delta,m,z}} \mathbb{E}\Big[\overline{U}_H(\mathbf{x})\Big] \\
&\leqslant n^{-1/13} \cdot \sum_{m,z \geqslant 1} \sum_{H \in \mathrm{NBSAW}_{s,t,d^H \leqslant \Delta,m,z}} \mathbb{E}\Big[\overline{U}_H(\mathbf{x})\Big]\,.
\end{aligned}$$

By Lemma 4.71 and Fact 4.69, the contribution of nice block self-avoiding walks with at most $t/\sqrt{A}$ edges of multiplicity 1 can be bounded by

$$\begin{aligned}
&\sum_{\substack{m,z\geqslant 1 s.t.\\ st-2m+2z\leqslant t/\sqrt{A}}} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}} \mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \\
&\leqslant \sum_{\substack{m,z\geqslant 1 s.t.\\ st-2m+2z\leqslant t/\sqrt{A}}} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}} \mathbb{E}\left[\overline{U}_H(\mathbf{x})\right] \\
&\leqslant (6st)\cdot n^{3/100A}\cdot (2^{10}s)^t\cdot [(1-o(1))(1+\delta)]^{-st/2} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,0,0}} \overline{L}_H \\
&\leqslant (1+\delta)^{-st/3} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,0,0}} \overline{L}_H\,.
\end{aligned}$$

Combining the two bounds, it follows by Lemma 4.67

$$\sum_{H\in \mathrm{NBSAW}_{s,t}} \geqslant \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,0,0}} L_H \geqslant \frac{1}{9n^{\frac{2}{50A}}}\left(\frac{\varepsilon\cdot d}{2}\right)^{st}.$$

□

Theorem 4.51 now follows as a direct consequence.

Proof of Theorem 4.51. Combining Lemma 4.57 and Lemma 4.72, by assumption on t,Δ,s,d,τ the result follows. □

4.5.3 Upper bound on the centered Schatten norm

In this section we provide an upper bound on $\mathrm{Tr}\left(Q^{(s)}\left(\overline{\mathbf{Y}}\right)-\mathbf{x}\mathbf{x}^{\mathsf{T}}\right)^t$ to obtain Theorem 4.43. We do it through the following result.

Theorem 4.73. *Consider the settings of Theorem 4.43. Then*

$$\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})-\mathbf{x}\mathbf{x}^{\mathsf{T}}\right)^t\right] \leqslant (1+\delta)^{-t/4}\cdot \mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^t\right].$$

Theorem 4.43 immediately follows combining Theorem 4.73 with Theorem 4.51. The proof of Theorem 4.73, amounts of step 4 and 5 of the scheme outlined in Section 4.5.1.2. The main observation we will need is that the subtraction of $\mathbf{x}\mathbf{x}^{\mathsf{T}}$ has the effect of removing the contribution of many graphs and to leave the contribution of the others essentially unchanged.

Upper bound on negligible block self-avoiding walks. Here we bound from above the contribution to the expectation of Eq. (4.5.4) of negligible block self-avoiding walks. For $H \in \mathrm{BSAW}_{s,t}$, recall the definitions of $S_1(H)$ and $S_{\geqslant 2}(H)$ as in Definition 4.52, Definition 4.53.

Definition 4.74. Let $H \in \mathrm{BSAW}_{s,t}$ and let $\mathcal{W}_1(H) \subseteq \mathcal{W}(H)$ be the subset of generating self-avoiding walks W of H such that $V(W) \subseteq S_1(H) \cap S_{\geqslant 2}(H)$ and $E(W) \subseteq E_1(H)$. Then, for any $\mathbf{x} \sim \mathsf{SBM}_n(d, \varepsilon)$ define the quantity

$$\hat{U}_H(\mathbf{x}) = \frac{n^{\frac{1}{100A}} \cdot 2^{\left|E_1^q(H)\right|}}{2^{As \cdot |\mathcal{W}_1(H)|}} \cdot \overline{U}_H(\mathbf{x}) \,,$$

where A is as defined in Eq. (4.4.2).

Similarly to Lemma 4.56, for any $H \in \mathrm{BSAW}_{s,t}$ we can show an upper bound on $\hat{Y}_H$ for any $H \in \mathrm{BSAW}_{s,t}$

Lemma 4.75. *Consider the settings of Theorem 4.43. Let $H \in BSAW_{s,t}$, for n large enough and for any $\mathbf{x} \sim \mathsf{SBM}_n(d, \varepsilon)$*

$$\left| \mathbb{E}\left[\hat{\mathbf{Y}}_H \,\middle|\, \mathbf{x} \right] \right| \leqslant \hat{U}_H(\mathbf{x}) \,.$$

We defer the proof of Lemma 4.75 to Appendix B.2. Crucially, the lemma implies that $\mathbb{E}\,\hat{U}_H(\mathbf{x}) \ll \mathbb{E}\,\overline{U}_H(\mathbf{x})$ for many nice block self-avoiding walks. On the other hand for others: $\mathbb{E}\,\hat{U}_H(\mathbf{x}) \approx \mathbb{E}\,\overline{U}_H(\mathbf{x})$. Together, these two bounds will allow us to obtain Theorem 4.73. First we show that the contribution to the expectation of Eq. (4.5.4) of negligible block self-avoiding walks is *still* negligible.

Lemma 4.76. *Consider the settings of Theorem 4.43. Then for n large enough*

$$\sum_{H \in NBSAW^c_{s,t}} \mathbb{E}\,\hat{U}_H(\mathbf{x}) \leqslant \left(n^{-1/7} + (1+\delta)^{-t\sqrt{s}/2} \right) \cdot \sum_{H \in NBSAW_{s,t}} \mathbb{E}\,\overline{U}_H(\mathbf{x}) \,.$$

Proof. By Lemma 4.57, Definition 4.74 and Lemma 4.75, since $t \in \left[\frac{\log n}{400}, \frac{\log n}{100} \right]$ and $A \geqslant 1000$ for any $\{\mathcal{F}_i, \Psi_i\}_{i=1}^{z-1}$ the result follows. □

So we only need to tackle nice walks. As in the proof of Theorem 4.51 we get rid of a tiny fraction of nice block self-avoiding walks.

Lemma 4.77. *Consider the settings of Theorem 4.43. Then for n large enough*

$$\sum_{H \in NBSAW_{s,t}} \mathbb{E}\left[\hat{\mathbf{Y}}_H \right] \leqslant n^{-1/14} \sum_{\substack{m,z \geqslant 0 \\ st - 2m + 2z \geqslant t/\sqrt{A}}} \sum_{H \in NBSAW_{s,t,d^H \leqslant \Delta,m,z}} \mathbb{E}\left[\overline{U}_H(\mathbf{x}) \right] .$$

Proof. By Lemma 4.70 and Lemma 4.68, the contribution of nice block self-avoiding walks with maximum degree larger than Δ can be bounded by

$$\sum_{m,z\geqslant 1} \sum_{H\in \mathrm{NBSAW}_{s,t,m,z}\setminus \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta}} \mathbb{E}\left[\hat{\mathbf{Y}}_H\right] \leqslant \sum_{m,z\geqslant 1} \sum_{H\in \mathrm{NBSAW}_{s,t,m,z}\setminus \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right]$$
$$\leqslant 2^t\cdot n^{-1/12}\cdot \sum_{m,z\geqslant 1} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right].$$

By Lemma 4.71 and Fact 4.69, the contribution of nice block self-avoiding walks with at most $t/\sqrt{A}$ edges of multiplicity 1 can be bounded by

$$\sum_{\substack{m,z\geqslant 1 s.t.\\ st-2m+2z\leqslant t/\sqrt{A}}} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}} \mathbb{E}\left[\overline{\mathbf{Y}}_H\right]$$
$$\leqslant \sum_{\substack{m,z\geqslant 1 s.t.\\ st-2m+2z\leqslant t/\sqrt{A}}} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right]$$
$$\leqslant (6st)\cdot n^{1/100A}\cdot (2^{10}s)^t\cdot [(1-o(1))(1+\delta)]^{-st/2} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,0,0}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right]$$
$$\leqslant (1+\delta)^{-st/3} \sum_{H\in \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,0,0}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right].$$

Combining the two bounds with Lemma 4.75 the result follows. □

Upper bound on the remaining self-avoiding walks. Now we split the remaining walks in two sets. For one set, the contribution to the expectation of Eq. (4.5.4) will be roughly the same contribution to the expectation of Eq. (4.5.3). For the other, it will be considerably smaller. The catch is that in the expectation of Eq. (4.5.3) the latter set has a significantly larger contribution.

Definition 4.78. Denote by $\mathrm{NNBSAW}_{s,t}\subseteq \bigcup_{\substack{m,z\geqslant 0\\ st-2m+2z\geqslant t/\sqrt{A}}} \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}$ the set of nice self-avoiding walks H such that

$$|\mathcal{W}_1(H)|\geqslant \frac{t}{s}.$$

Denote by $\mathrm{NNBSAW}^c_{s,t}$ the set $\left(\bigcup_{\substack{m,z\geqslant 0\\ st-2m+2z\geqslant t/\sqrt{A}}} \mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}\right)\setminus \mathrm{NNBSAW}_{s,t}$.

Next we show how the contribution of walks in $\mathrm{NNBSAW}_{s,t}$, $\mathrm{NNBSAW}^c_{s,t}$ changes from the expectation of Eq. (4.5.3) to that of Eq. (4.5.4).

Lemma 4.79. *Consider the settings of Theorem 4.43. Then for n large enough*

$$\text{for every } H \in NNBSAW_{s,t} \qquad \mathbb{E}\,\hat{U}_H(\mathbf{x}) \leqslant 2^{-st} \cdot \mathbb{E}\,\overline{U}_H(\mathbf{x})\,,$$
$$\text{for every } H \in NBSAW_{s,t} \qquad \mathbb{E}\,\hat{U}_H(\mathbf{x}) \leqslant 3n^{\frac{1}{100A}} \cdot \mathbb{E}\,\overline{U}_H(\mathbf{x})\,.$$

Proof. The first inequality follows by Definition 4.78 and Lemma 4.75. The second by observing that for any $H \in \mathrm{NBSAW}_{s,t}$ we have $E_1^a(H) = \emptyset$. □

As an immediate corollary, Lemma 4.79 implies that block self-avoiding walks in $\mathrm{NNBSAW}_{s,t}$ have small expectation.

Corollary 4.80. *Consider the settings of Theorem 4.43. Then for n large enough*

$$\sum_{H \in NNBSAW_{s,t}} \mathbb{E}\left[\hat{\mathbf{Y}}_H\right] \leqslant 2^{-st/2} \sum_{H \in NNBSAW_{s,t}} L_H\,.$$

It remains to bound the contribution to nice block self-avoiding walks in $\mathrm{NNBSAW}^c_{s,t}$. We require an additional definition.

Definition 4.81. Let $w, b, m, z \geqslant 0$ be integers. Define $\mathrm{N}_{w,q,m,z} \subset \mathrm{NNBSAW}^c_{s,t}$ to be the set of nice block self-avoiding walks H such that:

- the number of vertices v with $d^H(v) \leqslant 1$ is q,
- the number of walks in $\mathcal{W}_1$ is w,
- $E_{\geqslant 2}(H)$ is a forest on m edges and z components.

Notice that by definition of $\mathrm{NNBSAW}^c_{s,t}$ the set $\mathrm{N}_{q,w,m,z}$ is empty if $w \geqslant t/s$ or $m - z \leqslant t - 2t/s$. Moreover, observe that $q < t$ since for any block self-avoiding walks all vertices with total degree 1 must be pivots.

By Lemma 4.79, we can upper bound the contribution of walks in $\mathrm{N}_{w,q,m,z}$. To upper bound their number –and hence their total contribution– we use the following two results.

Lemma 4.82. *Consider the settings of Theorem 4.43. Let $r \geqslant 0$ be an integer. For n large enough*

$$\sum_{\substack{m,z \geqslant 0 \\ m-z=r}} |NBSAW_{s,t,m,z}| \leqslant \Delta^{6t} n^{st-r}\,.$$

We prove Lemma 4.82 in Appendix B.4.2.

Lemma 4.83. *Consider the settings of Theorem 4.43. Let $m, z \geqslant 0$ be integers such that $0 \leqslant m - z < \frac{t\sqrt{s}}{2}$. Then for n large enough*

$$\left|N_{w,q,m,z}\right| \leqslant 2^{ws} \cdot (1+\delta)^{\frac{t}{10}} \cdot n^{-q} \cdot \left|NBSAW_{s,t,d^H \leqslant \Delta, m', z'}\right|$$

for some m', z' such that $m' - z' = m - z - q$. Moreover, it holds that $q > t - 2t/\sqrt{s}$.

We also defer the proof of Lemma 4.83 to Appendix B.4.2. We can finally bound the expectation of nice block self-avoiding walks in $N_{w,q,m,z}$.

Lemma 4.84. *Consider the settings of Theorem 4.43. Then for n large enough*

$$\sum_{w,q,m,z\geqslant 0}\ \sum_{H\in N_{w,q,m,z}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right] \leqslant (1+\delta)^{-t/3} \sum_{\substack{m,z\geqslant 0\\ st-2m+2z\geqslant t/\sqrt{A}}}\ \sum_{H\in NBSAW_{s,t,d^H\leqslant\Delta,m,z}} \mathbb{E}\left[\overline{U}_H(\mathbf{x})\right]$$

Proof. First consider the case $m-z\geqslant t\sqrt{s}/2$. Since

$$\bigcup_{w,q} \mathrm{N}_{w,q,m,z} \subseteq \mathrm{NBSAW}_{s,t,d^h\leqslant\Delta,m,z}\,,$$

by Lemma 4.82 and definition of $\hat{U}(\mathbf{x})\,,\overline{U}(\mathbf{x})$ we get

$$\begin{aligned}\sum_{\substack{m,z\geqslant 0\\ m-z\geqslant t/2\sqrt{s}}}\ \sum_{H\in\mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}} \mathbb{E}[\mathbf{Y}_H] &\leqslant \sum_{\substack{m-z\\ m-z\geqslant t/2\sqrt{s}}}\ \sum_{H\in\mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m,z}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right]\\ &\leqslant (1+\delta)^{-t\sqrt{s}/2}\cdot(\Delta)^{6t}\cdot(1+o(1))^{st}\cdot \sum_{H\in\mathrm{NBSAW}_{s,t,0,0}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right]\\ &\leqslant (1+\delta)^{-t/2} \sum_{H\in\mathrm{NBSAW}_{s,t,0,0}} \mathbb{E}\left[\overline{U}_H(\mathbf{x})\right].\end{aligned}$$

So consider the case $m-z<t\sqrt{s}/2$. By Lemma 4.83 then $q\geqslant t-2t/\sqrt{s}$. For any $H\in N_{w,q,m,z}$ and any $H'\in\mathrm{NBSAW}_{s,t,d^H\leqslant\Delta,m',z'}$ with $m'-z'=m-z-q$ we have

$$\begin{aligned}\mathbb{E}\left[\hat{U}_H(\mathbf{x})\right] &\leqslant 3n^{1/100A}\cdot 2^{-Asw}\cdot[(1+o(1))(1+\delta)]^{-q}\cdot n^q\cdot\mathbb{E}\left[\overline{U}_{H'}(\mathbf{x})\right]\\ &\leqslant (1+\delta)^{-t/2}\cdot n^q\cdot\mathbb{E}\left[\overline{U}_{H'}(\mathbf{x})\right].\end{aligned}$$

By Lemma 4.83 it follows

$$\sum_{H\in N_{w,q,m,z}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right] \leqslant (1+\delta)^{-t/2} \sum_{H'\in\mathrm{NBSAW}_{s,t,m',z'}} \mathbb{E}\left[\overline{U}_{H'}(\mathbf{x})\right].$$

Repeating the argument for each $w,q,m,z\geqslant 0$ such that $N_{w,q,m,z}$ is non-empty (and thus so is the corresponding $\mathrm{NBSAW}_{s,t,m',z'}$) we obtain the desired result. □

We are now ready to prove Theorem 4.73.

Proof of Theorem 4.73. By Lemma 4.76 and Lemma 4.77

$$\sum_{H\in\mathrm{BSAW}_{s,t}\setminus\left(\mathrm{NNBSAW}_{s,t}\cup\mathrm{UNNBSAW}^c_{s,t}\right)} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right] \leqslant \left(n^{-1/15}+(1+\delta)^{-t\sqrt{s}/2}\right)\sum_{H\in\mathrm{BSAW}_{s,t}} \mathbb{E}\left[\overline{U}_H(\mathbf{x})\right].$$

By Corollary 4.80 and Lemma 4.84

$$\sum_{H\in \mathrm{NNBSAW}_{s,t}\cup \mathrm{NNBSAW}_{s,t}^{c}} \mathbb{E}\left[\hat{U}_H(\mathbf{x})\right] \leqslant \left(2^{-st}+(1+\delta)^{t/3}\right) \sum_{H\in \mathrm{BSAW}_{s,t}} \mathbb{E}\left[\overline{U}_H(\mathbf{x})\right].$$

Putting the two inequalities together, by Fact 4.69 we get

$$\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})-\mathbf{x}\mathbf{x}^{\mathsf{T}}\right)^{t}\right] \leqslant (1+\delta)^{-t/4}\cdot \mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})\right)^{t}\right]$$

as desired. □

4.5.4 Concentration of block self-avoiding walks

We prove here Theorem 4.44. Our strategy will be similar to the one used for Theorem 4.43. Concretely, we will show that for any $u,v\in[n]$ most of the mass of $\mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^{t}\right)_{uu}\left(Q^{(s)}(\overline{\mathbf{Y}})^{t}\right)_{vv}\right]$ comes from a specific subset of *nice* block self-avoiding walks.

4.5.4.1 Multigraphs that are not nice have negligible contributions

In this section, for $u,v\in[n]$, we show which multigraphs have negligible contribution in $\mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^{t}\right)_{uu}\left(Q^{(s)}(\overline{\mathbf{Y}})^{t}\right)_{vv}\right]$. We need the following definitions.

Definition 4.85 (Block self-avoiding walks with fixed pivot)**.** For $u\in[n]$, let $\mathrm{BSAW}_{s,t,u}\subseteq \mathrm{BSAW}_{s,t}$ be the set of block self-avoiding walks with u as pivot. We think of u as the first (and last) vertex and we will refer to it as the *first pivot*. Notice that the set $\mathrm{BSAW}_{s,t,u}$ corresponds to the set of block self-avoiding walks arising in $\mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^{t}\right)_{uu}\right]$. For every $H\in\mathrm{BSAW}_{s,t,u}$ we write $M(\mathcal{W}(H))$ for its sequence of edges. We denote by e_1^H, e_{st}^H respectively the first and last edges in the sequence $M(\mathcal{W}(H))$ and write $M^u(H):=\left\{e_1^H,e_{st}^H\right\}$. By construction both e_1^H, e_{st}^H are incident to vertex u. Similarly, we define $\mathrm{NBSAW}_{s,t,u}=\mathrm{BSAW}_{s,t,u}\cap\mathrm{NBSAW}_{s,t}$.

Definition 4.86 (Decomposition block self-avoiding walks)**.** For $u,v\in[n]$ and a multigraph $H\in\mathrm{BSAW}_{s,t,u}\times\mathrm{BSAW}_{s,t,v}$ we denote by $H_{(1)}\in\mathrm{BSAW}_{s,t,u}$ and $H_{(2)}\in\mathrm{BSAW}_{s,t,v}$ the two block self-avoiding walks such that $H=H_{(1)}\oplus H_{(2)}$. We call $H_{(1)},H_{(2)}$ the *decomposition block self-avoiding walks* of H. We write $M_{(1)}(H)$ for $M^u(H_{(1)})$ and $M_{(2)}(H)$ for $M^v(H_{(2)})$. When the context is clear we simply write $M_{(1)},M_{(2)}$.

Our central tool will be the following lemma, which resemble Lemma 4.57.

Lemma 4.87. *Consider the settings of Theorem 4.44. Let $u,v\in[n]$ and let $\mathrm{NMULTIG}_{s,t,u,v}$ be the set of multigraphs H in $\mathrm{NBSAW}_{s,t,u}\times\mathrm{NBSAW}_{s,t,v}$ with the following structure. If $u\neq v$:*

- *$V(H_{(1)})\cap V(H_{(2)})=\emptyset$. That is, the decomposition block self-avoiding walks of H are disjoint.*

If $u = v$

- *The edges of multiplicity 2 form a forest, i.e., $E_{\geqslant 2}(H)$ is a forest.*
- *For each $v \in V(H)$, $d^H_{\geqslant 2}(v) \leqslant \Delta$.*
- *Each connected component B of $E_{\geqslant 2}(H)$ not satisfying:*

$$M_{(1)} \cup M_{(2)} \subseteq E(H(V, V \setminus B) \oplus H(B)),$$
$$u \in V(B^{uu}),$$

is connected to $E_1(H)$ through a single vertex, every edge in $M(H(B))$ satisfies $m_H(e) \leqslant 2$.

- *If there is a connected component B^{uu} of $E_{\geqslant 2}(H)$ satisfying:*

$$M_{(1)} \cup M_{(2)} \subseteq E(H(V, V \setminus B) \oplus H(B)),$$
$$u \in V(B^{uu}),$$

then $H(B^{uu})$ is connected to $E_1(H)$ by 4 edges and at most two vertices.

- *If $E_1(H)$ contains two connected components than these components are cycles. If $E_1(H)$ is connected then it is either a path, or a cycle, or a path connected to a cycle by one of its endpoints, or two cycles with a single vertex in common.*

Then for n large enough

$$\sum_{H \in NMULTIG^c_{s,t,u,v}} \mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \left(n^{-\frac{1}{6}} + (1+\delta)^{-t\sqrt{s}}\right) \sum_{H \in NMULTIG_{s,t,u,v}} \mathbb{E}\,\overline{U}_H(\mathbf{x}).$$

Multigraphs in $\mathrm{NMULTIG}_{s,t,u,v}$ are said to be *nice*. Multigraphs in $\mathrm{NMULTIG}^c_{s,t,u,v}$ are said to be *negligible*. It is important to observe that for $u = v$ the family of multigraphs we need to consider grows. However, since $\mathrm{NMULTIG}_{s,t,u,u} \subseteq \mathrm{NBSAW}_{s,t,u} \times \mathrm{NBSAW}_{s,t,u}$ we can still ensures nice multigraphs satisfy several useful properties. For example, in every $H \in \mathrm{NMULTIG}_{s,t,u,u}$ edges have multiplicity at most 4, moreover no vertex $v \in V(H)$ has degree-1 larger than 4.

Bounding negligible multigraphs with few vertices. The first step to prove Lemma 4.87 is to obtain a result similar in spirit to Lemma 4.60. For $u, v \in [n]$ define

$$(\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v})_{d^H_{\geqslant 2} \leqslant \Delta} := \left\{ H \in \mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v} \;\middle|\; \max_{v \in V(H)} d^H_{\geqslant 2}(v) \leqslant \Delta \right\}.$$

Definition 4.88. Let $u, v \in [n]$. Let $H \in (\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v})_{d^H_{\geqslant 2} \leqslant \Delta}$ and let $w \in V(H)$. We denote by $q_H(w)$ the number of connected components of the line graph with vertex

set $E_H(w)$ and such that there is an edge between $e, e' \in E_H(w)$ if and only if e, e' appear in the sequence of edges $M(\mathcal{W}(H))$ consecutively. We define

$$q_H := \left(q_H(v) + q_H(u)\mathbb{1}_{[u\neq v]} - 2\right) + \sum_{w\in V(H)\setminus\{u,v\}} (q_H(w) - 1)\,.$$

Now, for $q \geqslant 0$ we define $\mathcal{D}_{q,s,t,u,v} \subseteq (\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v})_{d^H_{\geqslant 2}\leqslant\Delta}$ to be the subset of multigraphs H with $q_H = q$. We also write $\mathcal{D}_{q,s,t,u,v}(H) \subseteq V(H)$ to be the set of vertices w in H with

$$q_H(w) \geqslant \begin{cases} 3 & \text{if } u = v = w \\ 2 & \text{otherwise}\,. \end{cases}$$

Finally we write

$$\mathcal{D}_{\geqslant 1,s,t,u,v} = \bigcup_{q\geqslant 1} \mathcal{D}_{q,s,t,u,v}\,.$$

When the context is clear we write $\mathcal{D}_{q,u,v}$ instead of $\mathcal{D}_{q,s,t,u,v}$.

We prove that multigraphs in $\mathcal{D}_{\geqslant 1,u,v}$ have negligible contribution to the expectation of $\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu}\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{vv}$.

Lemma 4.89. *Consider the settings of Theorem 4.44. Let $u, v \in [n]$. Then for n large enough*

$$\sum_{H\in\mathcal{D}_{\geqslant 1,u,v}} \mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{n^{2/3}} \sum_{H\in(BSAW_{s,t,u}\times BSAW_{s,t,v})_{d^H_{\geqslant 2}\leqslant\Delta}\setminus\mathcal{D}_{\geqslant 1,u,v}} \mathbb{E}\,\overline{U}_H(\mathbf{x})\,.$$

Proof. Fix $q \geqslant 1$ and consider the following procedure to obtain a multigraph in $(\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v})_{d^H_{\geqslant 2}\leqslant\Delta}\setminus\mathcal{D}_{q,u,v}$ from a multigraph $H \in \mathcal{D}_{q,u,v}$. Let $M^{(1)}(\mathcal{W}), M^{(2)}(\mathcal{W})$ be respectively the sequence of edges obtained concatenating the generating self-avoiding walks of $H_{(1)}$ and $H_{(2)}$. We write $\left\{e_{(1)}^{(\ell-1)\cdot s+1}, \ldots, e_{(1)}^{(\ell-1)\cdot s+s}\right\}$ for the subsequence corresponding to the ℓ-th generating self-avoiding walk of $H_{(1)}$ (for simplicity we let $i - 1 = st$ for $i = 1$ and analogously we let $i + 1 = 1$ for $i = st$). We denote the generating self-avoiding walks of $H_{(2)}$ similarly. Let $w \in \mathcal{D}_{q,u,v}(H)$. Let $F_{H,w}$ be the line graph with vertex set $E_H(w)$ and edges as described in Definition 4.88. If $w \notin \{u, v\}$, let $E_H(w)^1$ be an arbitrary connected component of $F_{H,w}$ and let z be a vertex not in H. Conversely if $w = u$ let $E_H(w)^1$ be an arbitrary connected component of $F_{H,w}$ such that $E_H(w)^1 \cap M_{(1)}(H) = \emptyset$, since $w \in \mathcal{D}_{q,u,v}(H)$ there must exists such connected component (Notice that this also covers the case $u = v$). Analogously, if $w = v$ let $E_H(w)^1$ be an arbitrary connected component of $F_{H,w}$ such that $E_H(w)^1 \cap M_{(2)}(H) = \emptyset$, since $w \in \mathcal{D}_{q,u,v}(H)$ We construct the multigraph $H' \in \mathcal{D}_{q-1,u,v}$ with $V(H') = V(H) \cup \{z\}$ applying the following operation on H:

- Consider the sequence of edges $M^{(1)}(\mathcal{W}), M^{(2)}(\mathcal{W})$, we replace every edge $w'w \in M^{(1)}(\mathcal{W})$ (and $ww' \in M^{(1)}(\mathcal{W})$) such that $w'w \in E_H^1(w)$ with the edge $w'z$ (resp. zw'). Similarly, we replace every edge $w'w \in M^{(2)}(\mathcal{W})$ (and $ww' \in M^{(2)}(\mathcal{W})$) such that $w'w \in E_H^1(w)$ with the edge $w'z$ (resp. zw').

Clearly, $H' \in (\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v})_{d_{\geqslant 2}^H \leqslant \Delta} \setminus \mathcal{D}_{q,u,v}$ and $|V(H')| - |V(H)| = 1$. Furthermore $\left|E_1^q(H')\right| \geqslant \left|E_1^q(H)\right| - \tau - 2$. Thus

$$\prod_{v \in V(H)} \left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v)-\tau,0\right\}} \leqslant \left(\frac{6}{\varepsilon}\right)^{\tau+2} \cdot \prod_{v \in V(H')} \left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^{H'}(v)-\tau,0\right\}}.$$

By Fact 4.58 it follows that

$$\frac{1}{n^{|V(H')|-|V(H)|}} \cdot \frac{\mathbb{E}\left[\overline{U}_H(\mathbf{x})\right]}{\mathbb{E}\left[\overline{U}_{H'}(\mathbf{x})\right]} \leqslant \frac{1}{n} \cdot \left(\frac{6}{\varepsilon}\right)^{2\tau} \left(1 + \frac{\varepsilon}{2}\right)^{2\Delta} \leqslant \frac{1}{n} \cdot \left(\frac{12}{\varepsilon}\right)^{3\Delta}. \tag{4.5.15}$$

To obtain a multi-graph not in $\mathcal{D}_{\geqslant 1,u,v}$ we repeatedly apply the operation above until $\mathcal{D}_{q,u,v}(H)$ is empty. Notice that $(st)^{O(q)}$ applications suffice. It remains to show that the contribution to the expectation of $\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu}\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{vv}$ of multigraphs in $\mathcal{D}_{\geqslant 1,u,v}$is negligible. For this, observe that at each step there are at most $(st)^4$ multigraphs in $(\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v})_{d_{\geqslant 2}^H \leqslant \Delta}$ that can produce the same multigraph H'. So using Eq. (4.5.15), we get for any $q \geqslant 1$

$$\sum_{H \in \mathcal{D}_{q,u,v}} \mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{(st)^4}{n^{0.99}} \sum_{H' \in \mathcal{D}_{q-1,u,v}} \mathbb{E}\,\overline{U}_H(\mathbf{x}).$$

The result follows since the maximum degree in any multigraphs in $(\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v})_{d_{\geqslant 2}^H \leqslant \Delta}$ is $4t$. □

Bounding multigraphs from their shape and edges multiplicities. Next we extend Lemma 4.63 to multigraphs in $\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v}$ for any $u, v \in [n]$. That is, we compute a general bound on multigraphs in $\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v}$ based on the shape of its underlying graph and the multiplicity of each edge. We introduce some needed definitions.

Definition 4.90. Let $u, v \in [n]$. Let $\mathcal{B} = \{B_1, \ldots, B_z\}$ be a collections of disjoint connected graphs on at least two vertices. Let B^{uv} be a connected graph on at least two vertices disjoint from any graph in $\mathcal{B}$. We define $\mathcal{M}_{s,t,u,v}(\mathcal{B}, B^{uv})$ to be the subset of $\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v}$ satisfying the following. For $H \in \mathcal{M}_{s,t,u,v}(\mathcal{B}, B^{uv})$, let $H_{(1)}, H_{(2)}$ be the decomposition block self-avoiding walks of H. For any $B \in \mathcal{B} \cup \{B^{uv}\}$:

- $B \subseteq G(H)$, we denote with B' a (arbitrary) copy of B in H and by $H(B)$ the multigraph induced by $V(B')$ (With a slight abuse of notation we will simply write $V(B)$ for $V(B')$),
- $\forall e \in E(H(B))$ that is also an edge in V, $m_H(e) \geqslant 2$,
- there exists a cut $H(V(B'), V(H) \setminus V(B'))$ in H such that each edge in the cut has multiplicity 1 in H.

Furthermore,

- every edge in $H\left(V(H) \setminus \left(\bigcup_{B \in \mathcal{B} \cup \mathcal{B}^{uv}} V(B')\right)\right)$ has multiplicity 1,
- it holds that $u, v \in V(B^{uv})$ and

$$M_{(1)}(H) \cup M_{(2)}(H) \subseteq H(B^{uv}) \cup H(B^{uv}, V \setminus B^{uv}) .$$

That is, B^{uv} contains the first pivots u, v of the decomposition block self-avoiding walks $H_{(1)} \in \mathrm{BSAW}_{s,t,u}$ and $H_{(2)} \in \mathrm{BSAW}_{s,t,u}$ of H. If no such graph B^{uv} exists we simply write $\mathcal{M}_{s,t,u,v}(\mathcal{B}, \emptyset)$. When the context is clear we drop the subscripts s, t.

Definition 4.91. Let $u, v \in [n]$. Let $\mathcal{B} = \{B_1, \ldots, B_z\}$ be a collections of disjoint connected graphs and let B^{uv} be a graph disjoint from any graph in $\mathcal{B}$. Let $\left\{\ell_i, q_i, p_i, h_i\right\}_{i=1}^{z+1}$ be a sequence of tuples of integers such that for all $i \in [z]$, $\ell_i, q_i, p_i, h_i \geqslant 0$. Let $\{\Psi_i\}_{i=1}^{z+1}$ be a sequence of positive integer. Further we denote $\mathcal{F}_i = \{\ell_i, q_i, p_i, h_i\}$. We write $\mathcal{M}_{s,t,u,v\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uv})$ for the subset of $\mathcal{M}_{s,t,u,v}(\mathcal{B}, B^{uv})$ such that for any $i \in [z+1]$

(i) the size of the cut $H(V(H) \setminus V(B_i), V(B_i))$ is ℓ_i,

(ii) the number of edges in $H(B_i)$ of multiplicity one is q_i,

(iii) the number of edges e in $H(B_i)$ with $m_H(e) = 2$ is h_i,

(iv) the maximum degree-($\geqslant 2$) in $H(B_i)$ is Ψ_i.

(v) the edges with multiplicity larger than 2 in $H(B_i)$ satisfy

$$\sum_{\substack{e \in H(B_i) \\ m_H(e) \geqslant 3}} m_H(e) = p_i .$$

Now we study the contribution of block self-avoiding walks in $\mathcal{M}_{u,v}(\mathcal{B}, B^{uv})$ for all $\mathcal{B}, B^{uv}$.

Lemma 4.92. *Consider the settings of Theorem 4.44. Let $z \geqslant 1$ and $m_1, \dots, m_z, m_{z+1}$ be nonnegative integers. Let $u, v \in [n]$. Then for n large enough,*

$$\sum_{\substack{\text{for } i\in[z+1]:\\ q_i,\Psi_i,\ell_i,v_i\geqslant 0\\ T_i\in\mathcal{T}(m_i,v_i)\\ r_i\geqslant m_i-1\\ B_i\in\mathcal{G}(T_i,r_i)\\ h_i,p_i\geqslant 0}} \sum_{H\in\mathcal{M}_{u,v,\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\{B_1,\dots,B_z\},B_{z+1})} \mathbb{E}\,\overline{U}_H(\mathbf{x})$$

$$\leqslant \prod_{i\in[z+1]}\left[(st)^{30}\cdot\left(\frac{200\Delta+s^{10}+4^{\frac{\log n}{t}}}{(1+\delta)^{s/8}}\right)^{4m_i/s}\right]\cdot\sum_{H\in\mathcal{M}(\emptyset,\emptyset)}\mathbb{E}\,\overline{U}_H(\mathbf{x})\,.$$

Furthermore:

- *restricting the sum over $r_i > m_i - 1$, or over $p_i \geqslant 1$, or $q_i \geqslant 1$, or $\Psi_i > \Delta$ or $\ell_i > 2$, for some $i \in [z]$ or $i \in [z+1]$ if $u \neq v$, the inequality holds with an additional $n^{-\frac{1}{5}\left(p_i+q_i+\mathbb{1}_{[\Psi_i>\Delta]}(\Psi_i-\Delta)+\mathbb{1}_{[\ell_i>2]}(\ell_i-2)\right)}$ factor.*
- *if $u = v$, restricting the sum over $r_{z+1} > m_{z+1} - 1$, or over multigraphs H with $m_{H_{(1)}}(e) \geqslant 3$ or $m_{H_{(2)}}(e) \geqslant 3$ for some $e \in H$, or $q_{z+1} \geqslant 1$, or $\Psi_{z+1} > \Delta$ or $\ell_{z+1} > 4$ the inequality holds with an additional $n^{-\frac{1}{5}\left(p_{z+1}/5+q_{z+1}+\mathbb{1}_{[\Psi_{z+1}>\Delta]}(\Psi_{z+1}-\Delta)+\mathbb{1}_{[\ell_{z+1}>2]}(\ell_{z+1}-2)\right)}$ factor.*

To prove Lemma 4.92 we need two intermediate steps. First, an adaptation of Lemma 4.64 to the sets in Definition 4.91. Second, a result along the lines of Fact 4.65.

Lemma 4.93. *Consider the settings of Theorem 4.44. Let $u, v \in [n]$. Let $\mathcal{B} = \{B_1, \dots, B_z\}$ be collections of disjoint connected graphs each with respectively $m_1, \dots, m_z \geqslant 2$ vertices. Let B^{uv} be a connected graph disjoint from any graph in $\mathcal{B}$ and with $m_{z+1} \geqslant 2$ vertices.Let $\{\mathcal{F}_k\}_{i=1}^{z+1}$ be a sequence of tuples of integers as in Definition 4.91. Let $f^*_{s,t}, g^*_{s,t}$ be the functions*

$$f^*_{s,t}(m, m', \mathcal{F}, \Psi) = (\Psi)^{2h/s+10(q+\ell+p+1)+2h-2(m'-1)}\cdot(st)^{5\ell+5q+8p+4h+4-4(m'-1)}\,,$$
$$g^*_{s,t}(m', \mathcal{F}) = n^{-p-\ell/2-q-2h+m'}\,.$$

Let $m = \sum_{j\in[z+1]} m_j$. Then there are at most

$$2n^{2st-2+\mathbb{1}_{[u=v]}\mathbb{1}_{[B^{uv}\neq\emptyset]}}\cdot\prod_{1\leqslant k\leqslant z'} f^*_{s,t}(m, m_k, \mathcal{F}_i, \Psi_i)\cdot g^*_{s,t}(m_i, \mathcal{F}_i, h_i)$$

block self-avoiding walk pairs in the set $\mathcal{M}_{u,v\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+z'}}(\mathcal{B})$.

We show Lemma 4.93 in Appendix B.4.2. We also extend Fact 4.65 to mutligraphs in $\mathrm{BSAW}_{s,t,u}\times\mathrm{BSAW}_{s,t,v}$.

Fact 4.94. *Consider the settings of Theorem 4.44. Let $u, v \in [n]$, let $\mathcal{B} = \{B_1, \dots, B_z\}$ be a collection of disjoint connected graphs on at least 2 vertices and let B^{uv} be a (possibly empty) graph disjoint from any graph in $\mathcal{B}$. Then for any $H \in \mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uv})$ and $i \in [z+1]$*

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{4} n^{-1/25A} \left(\frac{6}{\varepsilon}\right)^{2\ell_i+2q_i} \mathbb{E}\,\overline{U}_{H(V,V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(B_i)}(\mathbf{x}),$$

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \geqslant \frac{1}{4} n^{-1/25A}\, \mathbb{E}\,\overline{U}_{H(V,V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(B_i)}(\mathbf{x}).$$

We obtain Fact 4.65 in Appendix B.4.1 and directly apply it here. Next we prove Lemma 4.92.

Proof of Lemma 4.92. Our argument closely resembles that of Lemma 4.63. Consider first the case $u \neq v$. For any non-empty B_{z+1} the same proof as in Lemma 4.63, combined with Lemma 4.93 implies

$$\sum_{\substack{\text{for } i\in[z]:\\ q_i,\Psi_i,\ell_i,v_i\geqslant 0\\ T_i\in\mathcal{T}(m_i,v_i)\\ r_i\geqslant m_i-1\\ B_i\in\mathcal{G}(T_i,r_i)\\ h_i,p_i\geqslant 0}} \sum_{H\in\mathcal{M}_{u,v,\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z}}(\{B_1,\dots,B_z\},B_{z+1})} \mathbb{E}\,\overline{U}_H(\mathbf{x})$$

$$\leqslant \left(n^{-1/5} + (1+\delta)^{-\sqrt{st}}\right) \cdot \sum_{\substack{\text{for } i\in[z]:\\ q_i,\Psi_i,\ell_i,v_i\geqslant 0\\ T_i\in\mathcal{T}(m_i,v_i)\\ r_i\geqslant m_i-1\\ B_i\in\mathcal{G}(T_i,r_i)\\ h_i,p_i\geqslant 0}} \sum_{H\in\mathcal{M}_{u,v,\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z}}(\{B_1,\dots,B_z\},\emptyset)} \mathbb{E}\,\overline{U}_H(\mathbf{x}),$$

as it must be that either $\ell_{z+1} \geqslant 4$ or $p_{z+1} \geqslant 1$. Here we used the definition of s, Δ, t. The rest of the proof then continues as in Lemma 4.63 so we omit it. Conversely, consider the case $u = v$. Again as in Lemma 4.63, applying Lemma 4.93 we get for any B_{z+1}

$$\sum_{\substack{\text{for } i\in[z]:\\ q_i,\Psi_i,\ell_i,v_i\geqslant 0\\ T_i\in\mathcal{T}(m_i,v_i)\\ r_i\geqslant m_i-1\\ B_i\in\mathcal{G}(T_i,r_i)\\ h_i,p_i\geqslant 0}} \sum_{H\in\mathcal{M}_{u,v,\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z}}(\{B_1,\dots,B_z\},B_{z+1})} \mathbb{E}\,\overline{U}_H(\mathbf{x})$$

$$\leqslant \prod_{i\in[z]} \left[(st)^{30} \left(\frac{200\Delta + s^{10} + 2^{\frac{\log n}{t}}}{(1+\delta)^{s/8}} \right)^{4m_i/s} \right]$$

$$\cdot \sum_{\substack{q_{z+1},\Psi_{z+1},\ell_{z+1},v_{z+1}\geqslant 0\\ T_{z+1}i\in\mathcal{T}(m_{z+1}i,v_{z+1})\\ r_{z+1}\geqslant m_{z+1}-1\\ B_{z+1}i\in\mathcal{G}(T_{z+1}i,r_{z+1})\\ h_{z+1}i,p_{z+1}i\geqslant 0}} \sum_{H\in\mathcal{M}_{u,v,\{\mathcal{F}_{z+1},\Psi_{z+1}\}}(\emptyset,B_{z+1})} \mathbb{E}\,\overline{U}_H(\mathbf{x}),$$

where if we restrict the sum over $r_i > m_i - 1$ or $p_i \geqslant 1$ or $q_i \geqslant 1$ or $\Psi_i > \Delta$ or $\ell_i > 2$ for some $i \in [z]$ the inequality holds with an additional $n^{-1/5}$ factor. For simplicity of the notation let $i = z + 1$. It remains to study

$$\sum_{\substack{q_i,\Psi_i,\ell_i,v_i\geqslant 0\\ T_i\in\mathcal{T}(m_i,v_i)\\ r_i\geqslant m_i-1\\ B_i\in\mathcal{G}(T_i,r_i)\\ h_i,p_i\geqslant 0}} \sum_{H\in\mathcal{M}_{u,v,\{\mathcal{F}_i,\Psi_i\}}(\emptyset,B_i)} \mathbb{E}\,\overline{U}_H(\mathbf{x})\,.$$

For any $\{\mathcal{F}_i,\Psi_i\}, B_i, r_i, v_i$ and any $H \in \mathcal{M}_{u,u,\{\mathcal{F}_i,\Psi_i\}}(\emptyset, B_i)$ by Fact 4.94 and Fact 4.58

$$\begin{aligned}\mathbb{E}\,\overline{U}_H(\mathbf{x}) &\leqslant \left(\frac{6}{\varepsilon}\right)^{2\ell_i+2q_i}\cdot(2n^{1/50A})^{-2}\,\mathbb{E}\,\overline{U}_{H(B_i)}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(V\setminus B_i)}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(V(H\setminus B_i),B_i)}(\mathbf{x})\\ &\leqslant 2n^{1/50A}\cdot\left[\prod_{u\in V(H\setminus B_i)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^{H(V\setminus B_i)}-\tau,0\right\}}\right]\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_i+2q_i}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E(H(V\setminus B_i))|}\cdot\left(\frac{\varepsilon d}{2n}\right)^{q_i+\ell_i}\\ &\quad\cdot\left[\left(1+\frac{1}{\sqrt{n}}\frac{d}{n}\right)\right]^{r_i}\cdot 2^{2(r_i-m_i-1)\cdot\Psi_i}\cdot\prod_{u\in\mathcal{L}_{\geqslant 2}(H(B_i))}\left[\left(\frac{2d}{n}\right)^{\frac{1}{4}\left(d^H_{\geqslant 2}(u)-\Delta\right)}\right]. \qquad (4.5.16)\end{aligned}$$

Thus we can use the bound

$$\begin{aligned}&\sum_{\substack{q_i,\Psi_i,\ell_i,v_i\geqslant 0\\ T_i\in\mathcal{T}(m_i,v_i)\\ r_i\geqslant m_i-1\\ B_i\in\mathcal{G}(T_i,r_i)\\ h_i,p_i\geqslant 0}} \sum_{H\in\mathcal{M}_{u,v,\{\mathcal{F}_i,\Psi_i\}}(\emptyset,B_{z+1})} \mathbb{E}\,\overline{U}_H(\mathbf{x})\\ &\leqslant 2n^{1/50A}\sum_{\substack{q_i,\Psi_i,\ell_i,v_i\geqslant 0\\ T_i\in\mathcal{T}(m_i,v_i)\\ r_i\geqslant m_i-1\\ B_i\in\mathcal{G}(T_i,r_i)\\ h_i,p_i\geqslant 0}}\left[\prod_{u\in V(H\setminus B_i)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^{H(V\setminus B_i)}-\tau,0\right\}}\right]\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_i+2q_i}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E(H(V\setminus B_i))|}\cdot\left(\frac{\varepsilon d}{2n}\right)^{q_i+\ell_i}\\ &\cdot\left[\left(1+\frac{1}{\sqrt{n}}\frac{d}{n}\right)\right]^{r_i}\cdot 2^{2(r_i-m_i-1)\cdot\Psi_i}\cdot\prod_{u\in\mathcal{L}_{\geqslant 2}(H(B_i))}\left[\left(\frac{2d}{n}\right)^{\frac{1}{4}\left(d^H_{\geqslant 2}(u)-\Delta\right)}\right]. \qquad (4.5.17)\end{aligned}$$

By Lemma 4.93, since $\Psi_i \leqslant 4t$ it follows that the contribution to Eq. (4.5.17) of multigraphs walks with $\Psi_i \geqslant \Delta$ will be at least a factor $n^{\frac{1}{5}}$ smaller than the others. Thus by Lemma 4.89, Lemma 4.93 and Fact B.105 we may upper bound Eq. (4.5.17) by

$$2n^{1/50A}\cdot 2n^{2st-2}\cdot\sum_{\substack{p_i,h_i,\ell_i\geqslant 0\\ r_i\geqslant m_i-1\\ \Psi_i\leqslant\Delta\\ v_i\leqslant 2(2h_i+p_i+q_i)/s+\ell_i}} n^{\mathbb{1}_{[m_i\geqslant 2]}}\cdot f_{s,t}(m_i,m_i,\mathcal{F}_i,\Psi_i)\cdot g_{s,t}(m_i,\mathcal{F}_i)$$

$$\cdot(8e\cdot m_i/v_i)^{2v_i}\cdot(2m_i)^{2(r_i-m_i-1)}\cdot 2^{2(r-m_i-1)\cdot\Psi_i}$$
$$\cdot\left(\frac{6}{\varepsilon}\right)^{2\ell_i+2q_i}\cdot\left(\frac{\varepsilon d}{2n}\right)^{2st-2h_i-p_i-\ell_i-q_i}\cdot\left(\frac{\varepsilon d}{2n}\right)^{q_i+\ell_i}$$
$$\cdot\left[\left(1+\frac{1}{\sqrt{n}}\right)\frac{d}{n}\right]^{r_i}. \tag{4.5.18}$$

As for Lemma 4.63, it is easy to see that Eq. (4.5.18) is a geometric sum which can be upper bounded by

$$2n^{1/50A}\cdot 2\cdot\left(1+\frac{4}{n^{1/5}}\right)\cdot n^{2st-2}\cdot\sum_{\substack{h_i,p_i,\geqslant 0,0\leqslant\ell_i\leqslant 4\\ \Psi_i\leqslant\Delta\\ v_i\leqslant 2(2h_i)/s+\ell_i}} n^{\mathbb{1}[m_i\geqslant 2]}$$
$$\cdot f_{s,t}(m_i,m_i,\{\ell_i,0,p_i,m_i-1\},\Psi_i)\cdot g_{s,t}(m_i,\{\ell_i,0,p_i,m_i-1\})$$
$$\cdot(8e\cdot m_i/v_i)^{2v_i}$$
$$\cdot\left(\frac{\varepsilon d}{2n}\right)^{2st-2h_i-p_i}$$
$$\cdot\left[\left(1+\frac{1}{\sqrt{n}}\right)\frac{d}{n}\right]^{m_i-1}. \tag{4.5.19}$$

That is, if we restrict Eq. (4.5.18) to $r_i > m_i-1$, or $q_i>0$ or $\Psi_i>\Delta$ or $\ell_i\geqslant 5$ the contribution drops by a $n^{-1/5}$ factor. So we need only to consider the settings $q_i=0, r_i=m_i-1, \Psi_i\leqslant\Delta, \ell_i\leqslant 4$. If $\ell_i<4$ then since each multigraph considered is a product of two block self-avoiding walks it must be that $2h_i+p_i\geqslant st+1$ and thus we obtain a ratio with

$$\sum_{H\in\mathcal{M}_{u,u}(\emptyset,\emptyset)}\mathbb{E}\,\overline{U}_H(\mathbf{x}) \tag{4.5.20}$$

of at most

$$\left(\frac{\varepsilon^2}{4d}\right)^{st/2}\cdot\left(200\Delta+s^{10}+C\right)^{2t}=\left(\frac{200\Delta+s^{10}+C^2}{(1+\delta)^{s/4}}\right)^{2t},$$

where $C=2^{2\frac{\log n}{t}}$. It remains to consider the case $\ell_i=4$. Notice that for any multigraph H in the sum, this means no edge can have multiplicity larger than 4 and no edge can have multiplicity larger than 2 in $H_{(1)}$ and $H_{(2)}$. Thus we obtain a ratio with Eq. (4.5.20) of at most

$$(st)^{14}\left(\frac{200\Delta^{10}}{(1+\delta)^{s/4}}\right)^{m_i/s}.$$

Putting things together the proof follows. □

Putting things together. We are now ready to prove Lemma 4.87.

Proof of Lemma 4.87. We argue that if a multigraph $H \in (\mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v})_{d^H_{\geqslant 2} \leqslant \Delta}$ is not in $\mathrm{NMULTIG}_{s,t,u,v}$, then it satisfies one of the following:

- $H \in \mathcal{D}_{\geqslant 1,u,v}$.
- if $u \neq v$ either $E_1(H_{(1)}) = \emptyset$ or $E_1(H_{(2)}) = \emptyset$.
- if $u = v$ and $E_1(H) = 0$.
- $H \in \mathcal{M}_{u,v\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uv})$ for any non-empty $\mathcal{B} \cup \{B^{uv}\}$ containing some B_i with a cycle, for $i \in [z+1]$.
- $H \in \mathcal{M}_{u,v\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uv})$ for tuples $\{\mathcal{F}_i\}_{i=1}^{z}$ such that for some $i \in [z]$, $p_i \geqslant 1$ or $q_i \geqslant 1$ or $\Psi_i > \Delta$ or $\ell_i \geqslant 3$.
- if $u \neq v$, $H \in \mathcal{M}_{u,v\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uv})$ for any non-empty $\{B^{uv}\}$.
- if $u = v$ and $H \in \mathcal{M}_{u,v\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uv})$ for any non-empty $\{B^{uv}\}$, then $q_{z+1} \geqslant 1$ or $\ell_{z+1} \neq 4$ or there exists $e \in H(B^{uv})$ such that $m_{H_{(1)}}(e) \geqslant 3$ or $m_{H_{(2)}}(e) \geqslant 3$.

Suppose this claim holds. As for Lemma 4.57, the inequality in Lemma 4.87 immediately follows if

$$2\left(\frac{300\Delta + s^{10} + 2^{\frac{\log n}{t}}}{(1+\delta)^{s/4}}\right) \leqslant (1+\delta)^{-\sqrt{s}}, \tag{4.5.21}$$

which, for $t \in \left[\frac{\log n}{400}, \frac{\log n}{100}\right]$, may be rewritten as

$$s \geqslant \frac{10^8}{\delta}\left(\max\{\log 300\Delta, \log s, 1\}\right)^2.$$

Since by assumption Eq. (4.5.21) is satisfied, applying Lemma 4.92, Lemma 4.89 and observing that the elements in $\sum_{H \in \mathrm{NBSAW}^c_{s,t}} \mathbb{E}\,\overline{U}_H(\mathbf{x})$ form a geometric sum we obtain the desired inequality.

It remains to verify our claim. Suppose first $u \neq v$ and consider some H in $\mathrm{NMULTIG}^c_{s,t,u,v}$. If $V_{(1)} \cap V_{(2)} = \emptyset$ the claim follows as in Lemma 4.57 and the fact that $H_{(1)}, H_{(2)}$ are nice block self-avoiding walks. Otherwise either $E(H_{(1)}) \cap E(H_{(2)}) \neq \emptyset$ or $E(H_{(1)}) \cap E(H_{(2)}) = \emptyset$. In the former case $H \in \mathcal{M}_{u,v}(\mathcal{B}, B^{uv}) \cup \mathcal{M}_{uv}(\emptyset, B^{uv}) \cup \mathcal{M}_{uv}(\mathcal{B}, \emptyset)$ for some $\mathcal{B} \cup B^{uv}$ with $|(V \setminus B, V)| \geqslant 4$. In the latter $H \in \mathcal{D}_{\geqslant 1,u,v}$.

So let $u = v$, if $V_{(1)} \cap V_{(2)} = \{u\}$ then again the claim follows as in Lemma 4.57. If $\{u\} \subset V_{(1)} \cap V_{(2)}$ then $H \in \mathcal{M}_{u,u\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uu}) \cup \mathcal{M}_{u,u\{\mathcal{F}_i,\Psi_i\}}(\emptyset, B^{uu})$ for some tuples of parameters $\{\mathcal{F}_i, \Psi_i\}_{i=1}^{z+1}$ and non-empty B^{uu}. We may assume that $\mathcal{B}$ is a collection of

trees such that for each $B_i \in \mathcal{B}$, $\ell_i = 2, p_i = q_i = 0$ and $\Psi_i \leqslant \Delta$ as otherwise the claim holds. Similarly we may assume B^{uu} is a tree with $\ell_{z+1} = 4, q_{z+1} = 0$ and $\Psi_{z+1} \leqslant \Delta$. If $H(B^{uu})$ is connected to $E_1(H)$ by more than two vertices then $\ell_{z+1} \geqslant 5$ or $q_{z+1} \geqslant 1$ or there exists $e \in E(B^{uu})$ such that $\max\{m_{H_{(1)}}(e) m m_{H_{(2)}}\} \geqslant 3$ contradicting our assumption. The remaining cases are trivial. □

4.5.4.2 Bounding the variance of non-negligible multigraphs

By Lemma 4.87, to obtain Theorem 4.44 it remains to bound the variance of nice multigraphs. As for nice block self-avoiding walks, we split nice multigraphs in different sets.

Definition 4.95. For $u, v \in [n]$, define the set $\mathrm{NMULTIG}_{s,t,u,v,m,z}$ to be the set of nice multigraphs in which $E_{\geqslant 2}(H)$ is a forest on m vertices and z components. Further define

$$\begin{aligned}\mathrm{NMULTIG}^*_{s,t,u,v} :=\Big\{H \in \mathrm{NMULTIG}_{s,t,u,v} \;\Big|\; & H_{(1)} \in \mathrm{NBSAW}_{s,t,m_1,z_1} \\ & H_{(2)} \in \mathrm{NBSAW}_{s,t,m_2,z_2} \\ & st - 2m_1 - 2z_1 \geqslant t/\sqrt{A}\,, st - 2m_2 - 2z_2 \geqslant t/\sqrt{A} \;\Big\}\,,\end{aligned}$$

$$\mathrm{NMULTIG}_{s,t,u,v,d^H\leqslant\Delta} :=\left\{H \in \mathrm{NMULTIG}_{s,t,u,v} \;\middle|\; \max_{v\in V(H)} d^H(v) \leqslant \Delta\right\}.$$

That is, $\mathrm{NMULTIG}^*_{s,t,u,v}$ denotes the set of nice multigraphs in which both decomposition block self-avoiding walks have at least $t/\sqrt{A}$ edges with multiplicity 1. We also define

$$\begin{aligned}\mathrm{NMULTIG}^*_{s,t,u,v,d^H\leqslant\Delta} &:= \mathrm{NMULTIG}_{s,t,u,vd^H\leqslant\Delta} \cap \mathrm{NMULTIG}^*_{s,t,u,v}\,,\\ \mathrm{NMULTIG}^*_{s,t,u,v,d^H>\Delta} &:= \mathrm{NMULTIG}^*_{s,t,u,v} \setminus \mathrm{NMULTIG}_{s,t,u,vd^H\leqslant\Delta}\,,\\ \mathrm{NMULTIG}_{s,t,u,v,d^H\leqslant\Delta,m,z} &:= \mathrm{NMULTIG}_{s,t,u,vd^H\leqslant\Delta} \cap \mathrm{NMULTIG}_{s,t,u,v,m,z}\,.\end{aligned}$$

Variance of the products of walks with different first pivots. We consider first the settings $u, v \in [n], u \neq v$. We provide an upper bound on the expectation of nice multigraphs in $\mathrm{NMULTIG}^*_{s,t,u,u}$.

Lemma 4.96. *Consider the settings of Theorem 4.44. Let $u, v \in [n]$ and let $m \geqslant 0$ be an integer. Then for any $H \in NMULTIG^*_{s,t,u,v}$ with decomposition block self-avoiding walks $H_{(1)}, H_{(2)}$*

$$\mathbb{E}[\mathbf{Y}_H] \leqslant (1 + o(1))\,\mathbb{E}\big[\mathbf{Y}_{H_{(1)}}\big] \cdot E\big[\mathbf{Y}_{H_{(2)}}\big]\,.$$

Moreover for any $H \in NMULTIG^*_{s,t,u,v,d^H>\Delta}$

$$|\mathbb{E}[\mathbf{Y}_H]| \leqslant \frac{4}{n^{1/13}}\,\mathbb{E}\,\overline{U}_{H_{(1)}}(\mathbf{x})\,\mathbb{E}\,\overline{U}_{H_{(2)}}(\mathbf{x})\,.$$

We prove Lemma 4.96 in Appendix B.1.4. The first inequality of Theorem 4.44 follows directly.

Lemma 4.97. *Consider the settings of Theorem 4.44. Let $u, v \in [n]$, $u \neq v$. Then*

$$\mathbb{E}\Big[\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{vv}\Big] \leqslant (1+o(1))\,\mathbb{E}\Big[\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big]^2.$$

Proof. By Lemma 4.87

$$\begin{aligned}\mathbb{E}\Big[\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{vv}\Big] &\leqslant \sum_{H\in \mathrm{NMULTIG}_{s,t,u,v}} \Big[\mathbb{E}[\mathbf{Y}_H] + \Big(n^{-1/6} + (1+\delta)^t\sqrt{s}\Big)U_H(\mathbf{x})\Big] \\ &\leqslant \sum_{H\in \mathrm{NMULTIG}_{s,t,u,v,d^H\leqslant\Delta}} \Big[\mathbb{E}[\mathbf{Y}_H] + \Big(n^{-1/6} + (1+\delta)^t\sqrt{s}\Big)U_H(\mathbf{x})\Big] \\ &\quad + \sum_{H\in \mathrm{NMULTIG}_{s,t,u,v,d^H>\Delta}} \mathbb{E}[U_H(\mathbf{x})]\cdot\Big(1+n^{-1/6}+(1+\delta)^t\sqrt{s}\Big).\end{aligned}$$

By Lemma 4.68, Lemma 4.70 and Lemma 4.96

$$\begin{aligned}\sum_{H\in \mathrm{NMULTIG}_{s,t,u,v,d^H>\Delta}} \mathbb{E}[U_H(\mathbf{x})] &\leqslant \frac{1}{n^{1/13}} \sum_{H\in \mathrm{NMULTIG}_{s,t,u,v,d^H>\Delta}} \mathbb{E}\big[U_{H_{(1)}}(\mathbf{x})\big]\,\mathbb{E}\big[U_{H_{(2)}}(\mathbf{x})\big] \\ &\leqslant \frac{2^{2t}}{n^{1/13}} \sum_{m,z\geqslant 1} \sum_{H\in \mathrm{NMULTIG}_{s,t,u,v,d^H>\Delta,m,z}} \mathbb{E}\big[U_{H_{(1)}}(\mathbf{x})\big]\,\mathbb{E}\big[U_{H_{(2)}}(\mathbf{x})\big] \\ &\leqslant \frac{1}{n^{1/14}} \sum_{m,z\geqslant 0} \sum_{H\in \mathrm{NMULTIG}_{s,t,u,v,d^H\leqslant\Delta}} \mathbb{E}\big[U_{H_{(1)}}(\mathbf{x})\big]\,\mathbb{E}\big[U_{H_{(2)}}(\mathbf{x})\big].\end{aligned}$$

On the other hand by Lemma 4.67, Fact 4.69, Lemma 4.71 and Lemma 4.96

$$\begin{aligned}&\sum_{H\in \mathrm{NMULTIG}_{s,t,u,v,d^H\leqslant\Delta}\setminus \mathrm{NMULTIG}^*_{s,t,u,v,d^H\leqslant\Delta}} \mathbb{E}\Big[\overline{U}_H(\mathbf{x})\Big] \\ &\leqslant \sum_{H\in \mathrm{NMULTIG}_{s,t,u,v,d^H\leqslant\Delta}\setminus \mathrm{NMULTIG}^*_{s,t,u,v,d^H\leqslant\Delta}} \mathbb{E}\big[U_{H_{(1)}}(\mathbf{x})\big]\,\mathbb{E}\big[U_{H_{(2)}}(\mathbf{x})\big] \\ &\leqslant \big[(st)\cdot 2^{10t}\cdot s^t\cdot(1+\delta)^{-st/2}\big]^2 \sum_{H\in \mathrm{NMULTIG}^*_{s,t,u,v,d^H\leqslant\Delta}} \mathbb{E}\big[U_{H_{(1)}}(\mathbf{x})\big]\,\mathbb{E}\big[U_{H_{(2)}}(\mathbf{x})\big] \\ &\leqslant o(1) \sum_{H\in \mathrm{NMULTIG}^*_{s,t,u,v,d^H\leqslant\Delta}} L_{H_{(1)}}L_{H_{(2)}} \\ &\leqslant o(1) \sum_{H\in \mathrm{NMULTIG}^*_{s,t,u,v,d^H\leqslant\Delta}} \mathbb{E}\Big[\overline{\mathbf{Y}}_{H_{(1)}}\Big]\,\mathbb{E}\Big[\overline{\mathbf{Y}}_{H_{(2)}}\Big].\end{aligned}$$

All in all we get

$$\mathbb{E}\Big[\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{vv}\Big] \leqslant (1+o(1)) \sum_{H\in \mathrm{NMULTIG}^*_{s,t,u,v,d^H\leqslant\Delta}} \mathbb{E}\big[\mathbf{Y}_{H_{(1)}}\big]\,\mathbb{E}\big[\mathbf{Y}_{H_{(2)}}\big]$$

$$\leqslant (1+o(1)) \sum_{H \in \mathrm{NMULTIG}^*_{s,t,u,v,d^H \leqslant \Delta}} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right]$$
$$\leqslant (1+o(1)) \sum_{H \in \mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v}} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right]$$

where the last step follows by the fact that by Lemma 4.57 and Lemma 4.72

$$\sum_{\substack{m,z \geqslant 0 \\ st-2m-2z \geqslant t/\sqrt{A}}} \sum_{H \in \mathrm{NBSAW}_{s,t,u,d^H \leqslant \Delta,m,z}} \mathbb{E}[\mathbf{Y}_H] \geqslant (1-o(1)) \sum_{H \in \mathrm{BSAW}_{s,t,u}} \mathbb{E}[\mathbf{Y}_H] .$$

as for Theorem 4.51. □

Concentration of diagonal entries. It remains to prove the second inequality of Theorem 4.44. Let $u \in [n]$. Consider the sets

$$\mathrm{NMULTIG}_{s,t,u,u,\ell} := \left\{H \in \mathrm{NMULTIG}_{s,t,u,u} \;\middle|\; \left|E(H_{(1)}) \cap E(H_{(2)})\right| = \ell\right\}$$
$$\mathrm{NMULTIG}^{**}_{s,t,u,u} := \left\{H \in \mathrm{NMULTIG}^{*}_{s,t,u,u} \;\middle|\; |E_1(H)| \geqslant t/A\right\} .$$

Lemma 4.98. *Consider the settings of Theorem 4.44. Let $u, v \in [n]$ and let $m \geqslant 0$ be an integer. Then for any $H \in \mathrm{NMULTIG}^{**}_{s,t,u,u,\ell}$ with decomposition block self-avoiding walks $H_{(1)}, H_{(2)}$. If $\left|E_1(H_{(1)}) \cap E_{\geqslant 2}(H_{(2)})\right| + \left|E_1(H_{(2)}) \cap E_{\geqslant 2}(H_{(1)})\right| \leqslant \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ and $\max_{v \in V(H)} d^H(v) \leqslant \Delta$*

$$\mathbb{E}[\mathbf{Y}_H] \leqslant \left(1+(1+\delta)^{-s}\right) \cdot \left(1+\frac{\delta}{10}\right)^{-\ell} \cdot \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] E\left[\mathbf{Y}_{H_{(2)}}\right] .$$

If $\left|E_1(H_{(1)}) \cap E_{\geqslant 2}(H_{(2)})\right| + \left|E_1(H_{(2)}) \cap E_{\geqslant 2}(H_{(1)})\right| > \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ or $\max_{v \in V(H)} d^H(v) > \Delta$

$$|\mathbb{E}[\mathbf{Y}_H]| \leqslant \frac{1}{n^{1/13}} \mathbb{E}\left[\overline{U}_{H_{(1)}}(\mathbf{x})\right] \mathbb{E}\left[\overline{U}_{H_{(2)}}(\mathbf{x})\right] .$$

We prove Lemma 4.98 in Appendix B.1.4. We introduce an additional counting argument.

Lemma 4.99. *Consider the settings of Theorem 4.44. Let $u \in [n]$ and Then for $\ell \neq st$*

$$|NMULTIG_{s,t,u,u,\ell}| \leqslant \frac{\ell^{10}}{n^{\ell}} |NMULTIG_{s,t,u,u,0}| .$$

For $\ell = st$

$$|NMULTIG_{s,t,u,u,\ell}| \leqslant \frac{4}{n^{\ell-1}} |NMULTIG_{s,t,u,u,0}| .$$

We are now ready to prove the inequality.

Lemma 4.100. *Consider the settings of Theorem 4.44. Let $u \in [n]$. Then*

$$\mathbb{E}\Big[\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big] \leqslant C \cdot \mathbb{E}\Big[\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big]^2,$$

where C is a universal constant.

Proof. By Lemma 4.87

$$\begin{aligned}\mathbb{E}\Big[\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big(Q^{(s)}(\overline{\mathbf{Y}})^t\Big)_{uu}\Big] &\leqslant \sum_{H\in \mathrm{BSAW}_{s,t,u}\times \mathrm{BSAW}_{s,t,u}} \mathbb{E}[\mathbf{Y}_H] \\ &\leqslant \sum_{H\in \mathrm{NMULTIG}_{s,t,u,u}} \Big[\mathbb{E}[\mathbf{Y}_H] + \Big(n^{-1/6} + (1+\delta)^{t\sqrt{s}}\Big) U_H(\mathbf{x})\Big].\end{aligned}$$

For any $\ell \leqslant st$, we split the set $\mathrm{NMULTIG}_{s,t,u,u}$ in

$$\begin{aligned} S_{1,\ell} &:= \mathrm{NMULTIG}^{**}_{s,t,u,u,\ell} \setminus \mathrm{NMULTIG}_{s,t,u,u,d^H\leqslant\Delta} \\ S_{2,\ell} &:= \mathrm{NMULTIG}_{s,t,u,u,\ell} \setminus \mathrm{NMULTIG}^{*}_{s,t,,u,u,\ell} \\ S_{3,\ell} &:= \mathrm{NMULTIG}^{*}_{s,t,u,u,\ell} \setminus \mathrm{NMULTIG}^{**}_{s,t,u,u} \\ S_{4,\ell} &:= \mathrm{NMULTIG}^{**}_{s,t,u,u,\ell,d^H\leqslant\Delta}. \end{aligned}$$

Notice that $S_{1,\ell} \cup S_{2,\ell} \cup S_{3,\ell} \cup S_{4,\ell} = \mathrm{NMULTIG}_{s,t,u,u}$. We bound each term separately. Consider the $S_{1,\ell}$ by Lemma 4.67, Lemma 4.99 and Lemma 4.98

$$\begin{aligned}\sum_{\ell\leqslant st}\sum_{H\in S_{1,\ell}} \mathbb{E}\Big[\overline{U}_H(\mathbf{x})\Big] &\leqslant \frac{1}{n^{1/13}}\sum_{\ell\leqslant st}\frac{(1+\delta)^{-\ell}}{n^{1/50A}} n^{\ell} \sum_{H\in S_{1,\ell}} \mathbb{E}\big[U_{H_{(1)}}(\mathbf{x})\big]\,\mathbb{E}\big[U_{H_{(2)}}(\mathbf{x})\big] \\ &\leqslant \frac{1}{n^{1/14}}\sum_{H\in S_{1,0}} L_{H_{(1)}} L_{H_{(2)}} \\ &\leqslant \frac{1}{n^{1/14}}\sum_{H\in S_{1,0}} \mathbb{E}\big[\mathbf{Y}_{H_{(1)}}\big]\,\mathbb{E}\big[\mathbf{Y}_{H_{(2)}}\big].\end{aligned}$$

Next we consider $S_{2,\ell}$. Notice that if $H \in S_{2,\ell}$ then $E_1(H) \leqslant 2t/\sqrt{A}$ by Lemma 4.71 and Lemma 4.99

$$\begin{aligned}\sum_{\ell\leqslant st}\sum_{H\in S_{2,\ell}} \mathbb{E}\Big[\overline{U}_H(\mathbf{x})\Big] &\leqslant \sum_{\ell\leqslant st}\frac{(1+\delta)^{-\ell}}{n^{1/50A}} n^{\ell} \sum_{H\in S_{2,\ell}} \mathbb{E}\Big[\overline{U}_{H_{(1)}}(\mathbf{x})\Big]\,\mathbb{E}\Big[\overline{U}_{H_{(2)}}(\mathbf{x})\Big] \\ &\leqslant \frac{O(1)}{n^{1/50A}}\sum_{H\in S_{2,0}} \mathbb{E}\Big[\overline{U}_{H_{(1)}}(\mathbf{x})\Big]\,\mathbb{E}\Big[\overline{U}_{H_{(2)}}(\mathbf{x})\Big] \\ &\leqslant \frac{O(1)}{n^{1/50A}}\cdot(1+\delta)^{-t\sqrt{s}/2}\sum_{H\in \mathrm{NMULTIG}_{s,t,u,u,0,0}} \mathbb{E}\Big[\overline{U}_{H_{(1)}}(\mathbf{x})\Big]\,\mathbb{E}\Big[\overline{U}_{H_{(2)}}(\mathbf{x})\Big]\end{aligned}$$

$$\leqslant o(1) \sum_{H \in \mathrm{NMULTIG}_{s,t,u,u,0,0}} L_{H_{(1)}} L_{H_{(2)}}$$
$$\leqslant o(1) \sum_{H \in \mathrm{NMULTIG}_{s,t,u,u,0,0}} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right].$$

Notice now that $S_{3,\ell}$ is empty unless $\ell \geqslant 9t/\sqrt{A}$. Moreover by choice of A, it holds $(1+\delta)^{-9t/\sqrt{A}} n^{10/A} \leqslant (1+\delta)^{-t/\sqrt{A}}$. It follows that by Lemma 4.99 and Lemma 4.67

$$\sum_{\ell \leqslant st} \sum_{H \in S_{3,\ell}} \mathbb{E}\left[\overline{U}_H(\mathbf{x})\right] \leqslant \sum_{\ell \leqslant st} \frac{(1+\delta)^{-\ell}}{n^{1/50A}} n^{\ell} \sum_{H \in S_{3,\ell}} \mathbb{E}\left[\overline{U}_{H_{(1)}}(\mathbf{x})\right] \mathbb{E}\left[\overline{U}_{H_{(2)}}(\mathbf{x})\right]$$
$$\leqslant o(1) \sum_{H \in \mathrm{NMULTIG}_{s,t,u,u,0,0}} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right].$$

It remains to bound the contribution of nice multigraphs in $S_{4,\ell}$. We first consider the case $\ell < st$. By Lemma 4.99

$$\sum_{\ell < st} \sum_{H \in S_{4,\ell}} \mathbb{E}[\mathbf{Y}_H] \leqslant (1 + (1+\delta)^{-s}) \sum_{\ell < st} \sum_{H \in S_{4,\ell}} \left(1 + \frac{\delta}{10}\right)^{-\ell} n^{\ell} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right]$$
$$\leqslant C \sum_{H \in S_{4,0}} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right]$$

where $C > 1$ is a universal constant. For $\ell = st$, similarly we get

$$\sum_{H \in S_{4,st}} \mathbb{E}[\mathbf{Y}_H] \leqslant (1 + (1+\delta)^{-s}) \sum_{H \in S_{4,st}} \left(1 + \frac{\delta}{10}\right)^{-st} n^{st} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right]$$
$$\leqslant n \cdot (st)^{10} \cdot \left(1 + \frac{\delta}{10}\right)^{-st} \sum_{H \in S_{4,0}} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right]$$
$$\leqslant o(1) \sum_{H \in S_{4,0}} \mathbb{E}\left[\mathbf{Y}_{H_{(1)}}\right] \mathbb{E}\left[\mathbf{Y}_{H_{(2)}}\right].$$

Putting things together

$$\mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu} \left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu}\right] \leqslant C' \cdot \mathbb{E}\left[\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)_{uu}\right]^2$$

for a universal constant $C' > 1$. □

Chapter 5

Stochastic block models with node corruptions

In Chapter 4 we saw that for stochastic block models, even under a constant fraction of edge corruptions, there exists an efficient algorithm achieving weak recovery down to the KS threshold. In this chapter, based on [DdH23], we focus on node corruptions and prove Theorem 1.4. We restate the model and the results. To simplify the proofs, our notation differs from Chapter 1. In contrast to Chapter 4 we do not write random variables in bold face, but we reuse the canonical stochastic block model introduced in Definition 4.1.

Definition 5.1 (Restatement of Definition 1.3)**.** Given $\mu \in [0, 1)$ and $(x^*, G^0) \sim \mathsf{SBM}_n(d, \varepsilon)$, an adversary may choose up to μn vertices in G^0 and arbitrarily modify edges incident to at least one of them to produce the corrupted graph G.

Recall that, along the lines of Chapter 4, we say that an algorithm that given a graph G^0 outputs an estimate $\hat{x}(G^0)$ for the community labels of G^0 achieves μ-node-robust weak recovery for $\{\mathsf{SBM}_n(d, \varepsilon)\}_{n\in\mathbb{N}}$ if

$$\mathop{\mathbb{E}}_{(x^*,G^0)\sim\mathsf{SBM}_n(d,\varepsilon)} \min_{G\in V_\mu(G^0)} \left[\frac{1}{n}|\langle x^*, \hat{x}(G)\rangle|\right] \geqslant \Omega_{d,\varepsilon}(1)\,, \tag{5.0.1}$$

where $V_\mu(G^0)$ is the set of graphs that ca be obtained from G^0 as in Definition 5.1. We prove:

Theorem 5.2 (Restatement of Theorem 1.4)**.** *Let $n > 1, d > 1, \varepsilon \in (0, 1)$ and $\mu \in [0, 1)$. When $\delta := \varepsilon^2/4d - 1 \geqslant \Omega(1)$ and $\mu \leqslant \Omega_\delta(1)$, μ-node-robust weak recovery is possible. [1] Moreover, the underlying algorithm runs in polynomial time.*

This algorithm is the first one that succeeds down to the KS threshold under node corruptions, [LM22] cannot work unless δ is sufficiently large, and the algorithms discussed in Chapter 4 cannot tolerate the corruption of $\Omega_\delta(1)$ vertices.

[1] $\mu \leqslant \Omega_\delta(1)$ here means that μ is bounded by a constant depending on δ. The dependence on δ is necessary: if μ is a fixed constant, then the recovery impossible for a small enough constant δ (see Section 5.5 for details).

$\mathbb{Z}_2$ synchronization. The techniques used for the robust stochastic block model, also yield an algorithm for the closely related robust $\mathbb{Z}_2$ synchronization problem, which can be formulated as follows:

Definition 5.3 (Row/column-corrupted $\mathbb{Z}_2$ Synchronization model)**.** Given a hidden vector $x^* \in \{\pm 1\}^n$ and $\sigma > 0$, let A^0 be the uncorrupted $\mathbb{Z}_2$ synchronization matrix

$$A^0 = \sigma x^*(x^*)^\top + W$$

where $W \in \mathbb{R}^{n\times n}$ is a symmetric random matrix whose upper triangular entries are i.i.d sampled from $N(0, n)$. An adversary may select μn elements of $[n]$ and arbitrarily modify the corresponding rows and columns of A^0 to produce a corrupted matrix A that we observe.

When $\sigma \leqslant 1$, even with no corruptions (i.e. $\mu = 0$), it is information theoretically impossible to achieve weak recovery ([PWBM18]). When $\sigma \geqslant 1 + \Omega(1)$ and $\mu = 0$, a polynomial-time algorithm is known to output estimator $\hat{x} \in \{\pm 1\}^n$ such that $\langle \hat{x}, x\rangle^2 \geqslant \Omega(n^2)$ with high probability. This is due to the BBP transition phenomenon [BBAP05]. However, for reasons similar to those described in the SBM settings, when $\mu \geqslant \Omega(1)$, the analysis of known algorithms such as semidefinite programming ([MS16]) or spectral algorithm ([PWBM18]) breaks down. Here, we give an algorithm that can achieve the constant sharp threshold for robust $\mathbb{Z}_2$ synchronization:

Theorem 5.4 (Proved in Section 5.6)**.** *Given a row/column corrupted matrix A generated from Definition 5.3, when $\sigma \geqslant 1 + \Omega(1)$, there is a polynomial-time algorithm that outputs an estimator $\hat{x} \in \{\pm 1\}^n$ such that $\mathbb{E}\left[\langle \hat{x}, x^*\rangle^2\right] \geqslant \Omega(n^2)$ with high probability over x^* and A^0.*

5.1 Techniques

We outline here the main ideas behind Theorem 5.2 and Theorem 5.4. The algorithm is splitted into two components, each tailored to handle one degree regime. For the regime with average degree $d > d_\delta$, where d_δ is a constant that only depends on $\delta := d\varepsilon^2/4 - 1$, our starting point is the result of [MS16]. For the sparse regime with $d \leqslant d_\delta$, we will borrow from Chapter 4.

Push-out effect of the basic SDP. Consider the settings $d > d_\delta$ and, for simplicity of the exposition, assume $d = \omega(1)$ and $\mu = o(1)$. Recall, [MS16] proved that the following SDP program –which we refer to as the basic SDP–

$$\mathrm{SDP}(M) = \max\{\langle M, X\rangle : X \succeq 0, X_{ii} = 1 \forall i \in [n]\}$$

achieves weak recovery at the KS threshold. Concretely, for an uncorrupted graph $(x^*, G^0) \sim \mathsf{SBM}_n(d, \varepsilon)$ with centered adjacency matrix $\tilde{A}^0$, they showed for some constant $\Delta_\delta > 0$,

$$\mathrm{SDP}(\tilde{A}^0) \geqslant (2 + \Delta_\delta)n\sqrt{d}\,, \tag{5.1.1}$$

$$\mathrm{SDP}(\tilde{A}^0 - \frac{\varepsilon d}{2n} x^* x^{*\mathsf{T}}) \leqslant \left(2 + \frac{\Delta_\delta}{2}\right) n\sqrt{d}\,. \tag{5.1.2}$$

with probability $1 - \exp(-\Omega_\delta(n))$. Taken together, these inequalities highlight a *push-out effect*: a significant shift in the SDP value resulting from the subtraction of a rank-1 matrix from $\tilde{A}^0$. Additionally, the exponential concentration probability allows us to demonstrate that the *push-out effect* occurs for every principal submatrix of size $(1 - o(1))n \times (1 - o(1))n$

We already saw in Chapter 4 how this algorithm is robust against edge adversarial perturbations. A single edge alteration can change both Eq. (5.1.1) and Eq. (5.1.2) by at most 2. As $\mathrm{SDP}(\tilde{A}^0) - \mathrm{SDP}(\tilde{A}^0 - \frac{\varepsilon d}{2n} x^* x^{*\mathsf{T}}) \geqslant \Omega(n\sqrt{d})$, as long as the number of edge corruptions is bounded from above by $O(n\sqrt{d})$, the algorithm can still approximately recover the communities. However, in the node corruption model, the number of modified edges can reach $\Omega(n \cdot d)$, which is far more than $n\sqrt{d}$ when $d = \omega(1)$. The gap between nd and $n\sqrt{d}$ indicates that a fundamentally different approach is needed to handle $\Omega(n)$ node corruptions.

Push-out effect of submatrices. A priori it is not clear whether it is possible to recover the signal in presence of node corruptions, or if such an adversary has the capability of hiding all the information. A good news is that, while the basic SDP is fragile to node corruptions, it *suggests* a plausible direction to design an (inefficient!) algorithm robust to node corruptions. The key observation is that there is always a principal submatrix of size $(1 - \mu)n \times (1 - \mu)n$ free from corruption. More specifically, let $\tilde{A}$ be the adjacency matrix of the corrupted graph, the structure of node corruptions implies that the uncorrupted vertices $S^* \subseteq [n]$ satisfies $\tilde{A}_{S^*} = \tilde{A}^0_{S^*}$, where $\tilde{A}_{S^*}$ and $\tilde{A}^0_{S^*}$ denote the submatrix of $\tilde{A}$ and $\tilde{A}^0$ restricted to the set $S^* \times S^*$. Moreover, it can be shown that, with high probability, the push-out effect *still holds* for this submatrix. That is:

$$\mathrm{SDP}(\tilde{A}_{S^*}) \geqslant (2 + \Delta_\delta)(1 - \mu)n\sqrt{d}\,, \tag{5.1.3}$$

$$\mathrm{SDP}\left(\tilde{A}_{S^*} - \frac{\varepsilon d}{2n} x^*_{S^*} x^{*\mathsf{T}}_{S^*}\right) \leqslant \left(2 + \frac{\Delta_\delta}{2}\right)(1 - \mu)n\sqrt{d}\,. \tag{5.1.4}$$

In other words, if we *knew* the set of uncorrupted nodes, then we would still be able to approximately recover the communities.

Unfortunately, the set of uncorrupted nodes S^* is not immediately known. Moreover, even disregarding computational issues, it remains unclear how one could identify such a set. A rudimentary strategy to address this challenge would be to identify a subset $S \subseteq [n]$ such that the objective value of the $\mathrm{SDP}(\tilde{A}_S)$ is large and to use the optimizer X as an estimator. However, this approach presents a problem in that the selected set S may contain corrupted vertices, leading to a situation where the optimizer X may align with the corruption rather than accurately reflecting the true labels.[2]

[2]Note that even with no corruption, when δ is a small constant, the optimizer X is only weakly correlated with $x^* x^{*\mathsf{T}}$.

A natural way to circumvent this issue, is to search over *pairs* (S, X) where $S \subseteq [n]$ and X is a positive semidefinite matrix that fulfills the *submatrix push-out constraints* that is described below.

Definition 5.5. Given a corrupted graph G as described in Definition 5.1 and its centered adjacency matrix $\tilde{A}$, consider a set $S \subseteq [n]$ such that $|S| = (1-\mu)n$ and a positive semidefinite matrix X where $X_{ii} = 1$ for all $i \in [n]$, we say that the triplet $(\tilde{A}, S, X)$ satisfies *submatrix push-out constraints* if and only if for every subset $S' \subseteq S$ such that $|S'| \geqslant (1-2\mu)n$, it holds that

$$\langle \tilde{A}_{S'}, X_{S'} \rangle \geqslant (2 + \Delta_\delta)(1 - o(1))n\sqrt{d}\,.$$

Suppose one can find S, X such that $(\tilde{A}, S, X)$ satisfies the *submatrix push-out constraints*, then for $S' = S \cap S^*$, we have

$$\begin{aligned}\langle \tilde{A}_{S'}, X_{S'} \rangle &\geqslant \mathrm{SDP}\left(\tilde{A}_{S^*} - \frac{\varepsilon d}{2n}\left(x^* x^{*\mathsf{T}}\right)_{S^*}\right) + \Omega\left(n\sqrt{d}\right) \\ &\geqslant \mathrm{SDP}\left(\tilde{A}_{S'} - \frac{\varepsilon d}{2n}\left(x^* x^{*\mathsf{T}}\right)_{S'}\right) + \Omega\left(n\sqrt{d}\right).\end{aligned}$$

One can then deduce that $\langle X_{S'}, X^*_{S'} \rangle \geqslant \Omega(n^2)$. As a result, we have $\langle X, X^* \rangle \geqslant \Omega(n^2)$ as well. Subsequently, after applying the standard rounding procedure outlined in Lemma 5.19, we obtain an estimator $\hat{x} \in \{\pm 1\}^n$ with a weak recovery guarantee $\langle x^*, \hat{x} \rangle^2 \geqslant \Omega(n^2)$.

Certificates for the submatrix push-out effect. Even with the *submatrix push-out constraints*, two fundamental challenges remain. *First*, we need to prove the existence of a pair (S, X) that satisfies the submatrix push-out constraints. *Second*, we need to be able to find such a pair efficiently.

With regard to the first challenge, ideally we would like to prove that the set of uncorrupted nodes S^* and the optimizer X of $\mathrm{SDP}(\tilde{A}_{S^*})$ fulfill the submatrix push-out constraints. However, it is difficult prove this: even though we have established that $\langle \tilde{A}_{S^*}, X \rangle \geqslant (2 + \Omega(1))(1-\mu)n\sqrt{d}$, it remains unclear whether $\langle \tilde{A}_{S^*} - \tilde{A}_{S'}, X \rangle$ is small for all $S' \subseteq S^*$ of size $(1-2\mu)n$.

To overcome this barrier, we make the following crucial observation:

Lemma 5.6 (Formal statement and proof in Appendix C.3)**.** *Given S of size $(1-\mu)n$, if $\left\|\tilde{A}_S\right\|_{\mathrm{op}} \leqslant O(\sqrt{d})$, then $\mathrm{SDP}(\tilde{A}_S - \tilde{A}_{S'}) \leqslant O(\sqrt{\mu} \cdot n \cdot \sqrt{d})$ for all $S' \subseteq S$ of size at most $(1-2\mu)n$.*

This result suggests us to consider the following program:

$$\begin{aligned}\max_{X,S} \quad & \langle \tilde{A}_S, X \rangle \\ \text{s.t.} \quad & X \succeq 0 \\ & X_{ii} = 1 \quad \forall i \in [n] \\ & \left\|\tilde{A}_S\right\|_{\mathrm{op}} \leqslant O(\sqrt{d})\end{aligned} \tag{5.1.5}$$

We begin by establishing the feasibility of the program. Although the spectral norm of $\tilde{A}_{S^*}$ can potentially reach polylog(n), we can leverage the results of [FO05] and reduce it to $O(\sqrt{d})$ through the pruning of high-degree nodes. The feasibility of the program can then be confirmed by taking S as the set of uncorrputed nodes and have degree at most $O(d)$. Furthermore, by union bound and the push-out effect established in Eq. (5.1.1) and Eq. (5.1.2), we have $\mathrm{SDP}(\tilde{A}_S) \geqslant (2+\Delta_\delta)\cdot(1-2\mu)n\sqrt{d}$ with high probability. Therefore the objective value of this program is at least $(2+\Delta_\delta)\cdot(1-2\mu)n\sqrt{d}$.

The optimizer of program 5.1.5, denoted by the pair $(\hat{X}, \hat{S})$, can then be shown to satisfy the submatrix push-out constraints as defined in Definition 5.5. It follows from our previous argument that the objective value of this program is at least $(2+\Delta_\delta)\cdot(1-2\mu)n\sqrt{d}$, which implies $\langle \tilde{A}_{\hat{S}}, \hat{X}\rangle \geqslant (2+\Delta_\delta)\cdot(1-2\mu)n\sqrt{d}$. Moreover, the program constraints enforce the bound $\|\tilde{A}_{\hat{S}}\|_{\mathrm{op}} \leqslant O(\sqrt{d})$. Together with Lemma 5.6, these implies that $\mathrm{SDP}(\tilde{A}_{\hat{S}} - \tilde{A}_{S'}) \leqslant O(\sqrt{\mu}\cdot n\cdot\sqrt{d})$ for all $S' \subseteq \hat{S}$ with size at most $(1-2\mu)n$. When $\mu = o(1)$, it follows that $\mathrm{SDP}(\tilde{A}_{S'}) \geqslant \mathrm{SDP}(\tilde{A}_{\hat{S}}) + \mathrm{SDP}(\tilde{A}_{S'} - \tilde{A}_{\hat{S}}) \geqslant (2+\Delta_\delta)\cdot(1-o(1))\cdot n\sqrt{d}$ for all $S' \subseteq S_{\max}$ with size at most $(1-2\mu)n$.

As a result, due to the push-out effect, the optimizer $\hat{X}$ will now have non-trivial correlation with the ground truth x^*, that is $\langle \hat{X}, X^*\rangle \geqslant \Omega(n^2)$.

The last step is to turn this exponential-time algorithm into an efficient one. Fortunately, the above argument can be captured by the Sum-of-Squares proof system, thereby enabling us to use the Sum-of-Squares relaxation of program 5.1.5 to obtain an estimator $\hat{X}$ such that $\langle \hat{X}, X^*\rangle \geqslant \Omega(n^2)$.

Node robust algorithms for sparse graphs. In the degree regime $d \leqslant d_\delta$, a simpler approach works: *remove high-degree vertices iteratively.* Although all vertices in the graph could have degree $\omega(1)$ under corruption, a successful strategy is to limit the number of removed vertices to $O(\mu n)$ by iteratively removing the highest degree node and one of its random neighbors. In this way, in each round, the number of corrupted nodes in the remaining graph is reduced by $\Omega(1)$ in expectation, meaning that the algorithm will terminate in $O(\mu n)$ rounds in expectation. As a result, the remaining graph differs from the uncorrupted graph by $O(n)$ edges, which allows us to apply the edge robust algorithm from Chapter 4.

5.2 Preliminaries

In this section, we formally define notations and cover necessary preliminaries that will be used throughout the chapter. We use the notation in Chapter 2.

Matrix and vector notations. Given a set $S \subseteq [n]$, we use v_S to denote the subvector restricted to the set S and M_S to denote the submatrix of M where we only keep entries in

the set $S \times S$, that is $M_S = M \odot (\mathbf{1}_S \mathbf{1}_S^\top)$.

Stochastic block model notations. For a stochastic block model $(x^*, G^0) \sim \mathsf{SBM}_n(d, \varepsilon)$, at times we consider the conditional distribution of G^0 given x^*, which we denote by $\mathsf{SBM}_{d,\varepsilon}(x^*)$. In these settings, to avoid corner cases that happens with probability $o(1)$, we further assume x^* to be approximately balanced as $\langle x^*, \mathbf{1}\rangle \leqslant O(\sqrt{n})$. We oftentimes consider only the condiWe use $\delta = \varepsilon^2/4d - 1$ to denote the distance to the KS threshold, use A^0 to denote the adjacency matrix of the uncorrupted graph G^0, use A to denote the adjacency matrix of the corrupted graph G, use $X^* = x^*(x^*)^\top$ to denote the label matrix, use S^* to denote the uncorrupted set of vertices, use $\tilde{A}^0 = A^0 - \frac{d}{n}J$ to denote the centered uncorrpted adjacency matrix and use $\tilde{A} = A - \frac{d}{n}J$ to denote the centered corrupted adjacency matrix.

Basic SDP and Grothendieck norm. We define basic SDP and Grothendieck norm as follows

Definition 5.7 (Basic SDP)**.** We define basic SDP as follows

$$\mathrm{SDP}(M) = \max\{\langle M, X\rangle : X \succeq 0, X_{ii} = 1 \forall i \in [n]\} \tag{5.2.1}$$

An equivalent definition (can be easily verified using eigendecomposition of X) is

$$\mathrm{SDP}(M) = \max\left\{\sum_{i,j=1}^{n} M_{ij}\langle \sigma_i, \sigma_j\rangle : \sigma_i \sim S^{n-1}\right\} \tag{5.2.2}$$

where S^{n-1} is the n-dimensional unit sphere.

Definition 5.8 (Grothendieck norm)**.** Let matrix function $P_\Gamma : \mathbb{R}^{n\times n} \to \mathbb{R}^{2n\times 2n}$ be defined as

$$P_\Gamma(M) = \begin{bmatrix} 0 & M \\ 0 & 0 \end{bmatrix}$$

We define Grothendieck norm $\|\cdot\|_{Gr} : \mathbb{R}^{n\times n} \to \mathbb{R}$ as

$$\|M\|_{Gr} = \max\{\langle P_\Gamma(M), X\rangle : X \succeq 0, X_{ii} = 1 \forall i \in [2n]\} \tag{5.2.3}$$

An equivalent definition (the equivalence can be easily verified using eigendecomposition of X) is

$$\|M\|_{Gr} = \max\left\{\sum_{i,j=1}^{n} M_{ij}\langle \sigma_i, \delta_j\rangle : \sigma_i \sim S^{n-1}, \delta_i \sim S^{n-1}\right\} \tag{5.2.4}$$

where S^{n-1} is the n-dimensional unit sphere.

From Definition 5.7 and Definition 5.8, it is easy to get the following inequalities between the basic SDP and Grothendieck norm.

Claim 5.9 (Proved in Appendix C.3). Given matrix M, we have $\mathrm{SDP}(M) \leqslant \|M\|_{Gr}$.

Claim 5.10 (Proved in Appendix C.3). Let M be an $n \times n$ matrix whose diagonal entries are 0 and $S \subseteq [n]$ be a subset of indices, we have $\mathrm{SDP}(M_S) \leqslant \mathrm{SDP}(M)$.

Grothendieck inequality. The celebrated Grothendieck inequality relates Grothendieck norm and the $\infty \to 1$ norm.

Definition 5.11 ($\infty \to 1$ norm)**.** Let us define $\infty \to 1$ norm $\|\cdot\|_{\infty\to 1} : \mathbb{R}^{n\times n} \to \mathbb{R}$ as

$$\|M\|_{\infty\to 1} = \max\{\langle x, My\rangle : x, y \in \{\pm 1\}^n\}$$

Theorem 5.12 (Grothendieck inequality, see [AN04])**.** *Let M be a real matrix of size $n \times n$. We have*

$$\|M\|_{\infty\to 1} \leqslant \|M\|_{Gr} \leqslant K_G \|M\|_{\infty\to 1}$$

where K_G is a universal constant called the Grothendieck constant.

5.3 Reaching the KS threshold for diverging degree

In this section, we give an SoS algorithm when average degree d is larger than some constant d_δ which depends only on $\delta := \varepsilon^2 d/4 - 1$.

We begin by presenting our main technical theorem, which implies Theorem 5.2.

Theorem 5.13. *Let G be a graph as described in Definition 5.1, suppose $\delta \geqslant \Omega(1)$, there exists constants $d_\delta \leqslant O(1)$ and $\mu_\delta \geqslant \Omega(1)$ which only depend on δ, such that when $d \geqslant d_\delta$ and $\mu \leqslant \mu_\delta$, there exists a polynomial-time algorithm (Algorithm 5.14) that outputs $\hat{x} \in \{\pm 1\}^n$ satisfying*

$$\mathbb{E}\langle \hat{x}, x^*\rangle^2 \geqslant \Omega(n^2)\,.$$

Our algorithm is based on the deg-4 SoS relaxation of the following contraint set. Given a node-corrupted graph G generated according to Definition 5.1 and its centered adjacency matrix $\tilde{A} = A - \frac{d}{n}\mathbf{1}\mathbf{1}^\top$, we consider the following system of polynomial equations in PSD matrix X of size $n \times n$ and $\{0,1\}$-vector w of size n:

$$\mathcal{A} := \begin{cases} w_i^2 = w_i & \forall i \in [n] \\ \sum_i w_i = (1-\mu-\beta)n & \\ X \succeq 0 & \\ X_{ii} = 1 & \forall i \in [n] \\ \langle \tilde{A} \odot (ww^\top), X\rangle \geqslant (2+\Delta)(1-\mu-\beta)n\sqrt{d} & \\ \left\|\tilde{A} \odot (ww^\top)\right\|_{\mathrm{op}} \leqslant C_s\sqrt{d} & \end{cases} \tag{5.3.1}$$

Here Δ and C_s are constants depending on δ, and β is the small fraction of high degree nodes we need to prune to get bounded spectral norm according to Corollary C.3.

The outline of our algorithm is given below:

Algorithm 5.14 (Algorithm reaching KS threshold for diverging degree)**.**
Input: Graph G from node-corrupted SBM.

1. Run deg-4 SoS relaxation of program 5.3.1 and obtain pseudo-expectation $\tilde{\mathbb{E}}$.
2. Compute $\hat{X} := \tilde{\mathbb{E}}[X]$.
3. Apply the rounding procedure in Lemma 5.19 on $\hat{X}$ to get estimator $\hat{x}$.

The design and analysis of our SoS algorithm is based on the push-out effect of the basic SDP ([MS16]) and spectral properties of the adjacency matrix ([FO05, CRV15, LM22]) (see Appendix C.1 and Appendix C.2 for more details). Essentially, we identify a subset of the vertices whose adjacency matrix has large enough basic SDP value and is spectrally bounded. Then, we use the spectral norm bound and the Grothendieck inequality to bound the basic SDP value of the submatrix formed by corrupted vertices in the selected subset.

Theorem 5.15. *Consider the constraint set in program 5.3.1, when $\delta \geqslant \Omega(1)$, there exists functions $d_\delta \leqslant O(1)$ and $\mu_\delta \geqslant \Omega(1)$ which only depend on δ, such that when $d \geqslant d_\delta$ and $\mu \leqslant \mu_\delta$, the following holds with probability at least $1 - o(1)$*

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^* \rangle \geqslant \Omega(n^2)$$

We break down the proof of Theorem 5.15 into Lemma 5.16, Lemma 5.17 and Lemma 5.18. For simplicity, let us refer to the set of vertex i with $w_i = 1$ as set S, that is $S = \{i \in [n] | w_i = 1\}$.

In Lemma 5.16, we prove the feasibility of program 5.3.1.

Lemma 5.16 (Proof deferred to Appendix C.3)**.** *Program 5.3.1 is feasible with probability $1 - o(1)$.*

Then, in Lemma 5.17, we give a deg-4 SoS proof to show that $\langle X_{S'}, X^*_{S'} \rangle$ is large for some set S' with size at least $(1 - 2\mu - \beta)n$.

Lemma 5.17. *Consider set $S' = S \cap S^*$, which is the set of uncorrupted vertices in the set S found by the program. For X and w that satisfy the SoS program in Eq. (5.3.1), we have*

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X_{S'}, X^*_{S'} \rangle \geqslant \frac{2\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{n^2}{\varepsilon\sqrt{d}})$$

where β is the small constant fraction of high degree nodes we need to prune to get bounded spectral norm according to Corollary C.3 and $\Delta' = \Delta'(\delta)$ for some value $\Delta'(\delta)$ that only depends on δ.

Proof. We will apply the identity $\langle X_{S'}, X^*_{S'} \rangle = \langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'} \rangle - \langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'} \rangle$ and bound the value of $\langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'} \rangle$ and $\langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'} \rangle$ separately.

The value of $\langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'}\rangle$ is easy to bound. From Theorem C.1 and union bound, we can get that, with probability $1 - o(1)$, we have

$$\langle X_{S'}, \tilde{A}_{S'} - \frac{\varepsilon d}{2n} X^*_{S'}\rangle \leqslant \mathrm{SDP}(\tilde{A}_{S'} - \frac{\varepsilon d}{2n} X^*_{S'}) \leqslant (2+\rho)(1-2\mu-\beta)n\sqrt{d}$$

Now, goal is to bound $\langle X_{S'}, \tilde{A}_{S'}\rangle$. We decompose it as follows

$$\langle X_{S'}, \tilde{A}_{S'}\rangle = \langle X_S, \tilde{A}_S\rangle - \langle X_S, \tilde{A}_S - \tilde{A}_{S'}\rangle \tag{5.3.2}$$

From the constraints of Eq. (5.3.1), we have

$$\langle X_S, \tilde{A}_S\rangle \geqslant (2+\Delta)(1-\mu-\beta)n\sqrt{d} \tag{5.3.3}$$

To bound the value of $\langle X_S, \tilde{A}_S - \tilde{A}_{S'}\rangle$, we note that, by constraint $\|\tilde{A}_S\|_{\mathrm{op}} \leqslant C_s\sqrt{d}$, we can apply Lemma 5.6 to get

$$\langle X_S, \tilde{A}_S - \tilde{A}_{S'}\rangle \leqslant \mathrm{SDP}(\tilde{A}_S - \tilde{A}_{S'}) \leqslant C'_s\sqrt{\mu}n\sqrt{d} \tag{5.3.4}$$

for some constant C'_s.

Plug Eq. (5.3.3) and Eq. (5.3.4) into Eq. (5.3.2), we get

$$\langle X_{S'}, \tilde{A}_{S'}\rangle = \langle X_S, \tilde{A}_S\rangle - \langle X_S, \tilde{A}_S - \tilde{A}_{S'}\rangle \geqslant (2+\Delta)(1-\mu-\beta)n\sqrt{d} - C'_s\sqrt{\mu}n\sqrt{d}$$

Now, we can apply the identity $\langle X_{S'}, X^*_{S'}\rangle = \langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'}\rangle - \langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'}\rangle$ and get

$$\begin{aligned}
\langle X_{S'}, X^*_{S'}\rangle =& \langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'}\rangle - \langle X_{S'}, \frac{2n}{\varepsilon d}\tilde{A}_{S'} - X^*_{S'}\rangle \\
\geqslant& \frac{2n}{\varepsilon d}\Big((2+\Delta)(1-\mu-\beta)n\sqrt{d} - C'_s\sqrt{\mu}n\sqrt{d}\Big) - \frac{2n}{\varepsilon d}(2+\rho)(1-2\mu-\beta)n\sqrt{d} \\
\geqslant& \frac{\Delta'(1-\beta)4n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{4n^2}{\varepsilon\sqrt{d}})
\end{aligned}$$

□

Finally, since X is positive semidefinite and $X_{ii} = 1$ for all $i \in [n]$, we can conclude that there is a deg-4 SoS proof to show that correlation $\langle X, X^*\rangle$ is large.

Lemma 5.18 (Proof deferred to Appendix C.3)**.** *For X and w that satisfy the SoS program in Eq. (5.3.1), we have*

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \frac{\Delta'(1-\beta)2n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{2n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2$$

where β is the small constant fraction of high degree nodes we need to prune to get bounded spectral norm according to Corollary C.3 and $\Delta' = \Delta'(\delta)$ for some value $\Delta'(\delta)$ that only depends on δ.

Now, we have all the ingredients to prove Theorem 5.15

Proof of Theorem 5.15. From Lemma 5.16, we know that the SoS program in Eq. (5.3.1) is feasible with probability $1 - o(1)$. Combine this with Lemma 5.18, we know that, with probability $1 - o(1)$, the SoS program in Eq. (5.3.1) finds X and w such that they satisfy

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^* \rangle \geqslant \frac{\Delta'(1-\beta)2n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{2n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2$$

for some β that is the small constant fraction of high degree nodes we need to prune to get bounded spectral norm according to Corollary C.3 and $\Delta' = \Delta'(\delta)$ for some value $\Delta'(\delta)$ that only depends on δ.

When $\mu \leqslant \mu_\delta$ for some value μ_δ that only depends on δ and $\beta = \beta(\delta)$ for some value $\beta(\delta)$ that only depends on δ, we have:

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^* \rangle \geqslant \frac{\Delta'(1-\beta)2n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{2n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2 = \theta(\delta)n^2$$

for some $\theta(\delta)$ that only depends on δ. Thus, when $\delta \geqslant \Omega(1)$, we can get the weak recovery guarantee:

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^* \rangle \geqslant \Omega(n^2)$$

□

In order to fully establish the validity of Theorem 5.13, it remains to apply the standard rounding procedure from [HS17] on the pseudo-expectation of matrix X (as depicted in Algorithm 5.14). We will address this part in next.

Rounding. Now, we complete the proof of Theorem 5.13 by giving a rounding procedure.

Lemma 5.19 (Rounding procedure adapted from Lemma 3.5 of [HS17])**.** *Let* $\theta = \frac{1}{\|X\|_F n}\langle X, X^* \rangle$. *Let* Y *be a matrix of minimum Frobenious norm such that* $Y \succeq 0$, $\operatorname{diag} Y = 1$ *and* $\frac{1}{\|X\|_F n}\langle Y, X \rangle \geqslant \theta$. *With probability* $1 - o(1)$, *the vector* $\hat{x}$ *obtained by taking coordiate-wise sign of a Gaussian vector with mean* 0 *and covariance* Y *satisfies*

$$\mathbb{E}[\langle \hat{x}, x^* \rangle^2] \geqslant \Omega(\theta)^2 n^2$$

Proof. Apply Lemma 3.5 of [HS17] by taking $P = X$, $y = x^*$ and $\delta' = \theta$, we can get

$$\mathbb{E}[\langle \hat{x}, x^* \rangle^2] \geqslant \Omega(\theta)^2 n^2$$

Notice that, because each entry of X is within ± 1, we have $\|X\| \leqslant n$. Since $\langle X, X^* \rangle \geqslant \Omega(n^2)$ by Theorem 5.15, we have $\theta = \Omega(1)$. Thus, $\hat{x}$ weakly recovers x^*. □

Now we finish the proof of Theorem 5.13.

Proof of Theorem 5.13. By combining Theorem 5.15 and Corollary 4.24, we can compute the pseudo-expectation $\tilde{\mathbb{E}}$ for the SoS relaxtion of Eq. (5.3.1) in polynomial time. Let $\hat{X} := \tilde{\mathbb{E}}[X]$ in Eq. (5.3.1). By linearity of pseudo-expectation, we have $\hat{X} \succeq 0$, $\hat{X}_{ii} = 1$ and $\langle \hat{X}, X^* \rangle \geqslant \Omega(n^2)$ with probability $1 - o(1)$. Now applying rounding procedure in Lemma 5.19, we can then obtain $\hat{x} \in \{\pm 1\}^n$ such that $\mathbb{E}\langle \hat{x}, x^* \rangle^2 \geqslant \Omega(n^2)$. □

5.4 Reaching KS threshold for constant degree

In this section, we give an algorithm that reduces node corruption to edge corruption when $d < d_\delta$. This allows us to deal with graphs with small average degree. We will make use of Theorem 4.23 and Corollary 4.24, which we restate and combine here for convenience.

Theorem 5.20 (Combination of Theorem 4.23 and Corollary 4.24)**.** *Given a graph $G \sim \mathsf{SBM}_n(d, \varepsilon)$, suppose G' is an arbitrary graph that differs from G in at most $O(\rho n)$ edges for*

$$\rho \leqslant \left(\frac{1}{\delta} \log \frac{2}{\varepsilon}\right)^{-O(1/\delta)}$$

Then, there exists a polynomial-time algorithm that, given G' and δ, computes an n-dimensional unit vector $\hat{x}$ such that

$$\mathbb{E}[\langle \hat{x}, x^* \rangle^2] \geqslant \delta^{O(1)} n$$

5.4.1 Degree-pruning based algorithm

The algorithm is based on degree pruning. The tricky part is that node corruption can arbitrarily increase the degree of uncorrupted vertices to μn. Therefore, simply pruning high-degree vertices can be quite difficult to analyse.

Our solution is to iteratively remove the highest degree node as well as one of its neighbours that is selected uniformly at random until all vertices have small enough degree. The goal is to make sure that, in each round, we remove $\Omega(1)$ corrupted vertices in expectation.

Notice that, in each round, if the highest degree node is corrupted, then it is good. If the highest degree node is uncorrupted, then we can show, with high probability, the majority of its neighbours are corrupted vertices and we are likely to remove a corrupted vertex if we select one of its neighbours uniformly at random. A key observation is that, this approach allows us to easily bound the total number of removed vertices using a simple and standard Markov Chain drift analysis.

After the degree pruning procedure, we will invoke the edge-robust algorithm from [DdNS22] that is restated in Theorem 5.20.

Algorithm 5.21 (Algorithm reaching KS threshold for constant degree)**.**
Input: A node-corrupted stochastic block model G.

1. Set $G' \leftarrow G$
2. While there exist vertices with degree larger than $C_{\deg}(\mu)d$ in G':
   - remove the highest-degree vertex v from G',
   - remove from G' a neighbour u of v that is selected uniformly at random.
3. Run edge-robust algorithm from Theorem 5.20 on the remaining graph G'.
4. Apply the rounding procedure in Lemma 5.19 to get estimator $\hat{x}$.

In the following theorem, we will show that Algorithm 5.21 outputs an estimator $\hat{x}$ that achieves weak recovery.

Theorem 5.22. *When $d < d_\delta$ and $\delta = \Omega(1)$, for some $C_{deg}(\mu)$ that only depends on μ, Algorithm 5.21 outputs a vector $\hat{x} \in \{\pm 1\}^n$ such that*

$$\mathbb{E}[\langle \hat{x}, x^* \rangle^2] \geqslant \Omega(n^2)$$

Moreover, Algorithm 5.21 runs in polynomial time.

To prove Theorem 5.22, we will use the following two lemmas: Lemma 5.23 and Lemma 5.24. First, we prove Lemma 5.23 which says that, with probability 0.99, the pruning step of Algorithm 5.21 terminates in $O(\mu n)$ rounds. Then, in Lemma 5.24, we prove that, with probability 0.99, Algorithm 5.21 produces a graph G' that differs from G^0 by at most $O(\rho n)$ edges, such that we can apply Theorem 5.20 on G' to get an estimator $\hat{x}$ that achieves weak recovery.

Lemma 5.23. *With probability at least 0.99, for some $C_{deg}(\mu)$ that only depends on μ, step 2 of Algorithm 5.21 terminates in $O(\mu n)$ rounds.*

Proof. Let S denote the set of uncorrupted vertices and let $G[S]$ denote the induced subgraph of the uncorrupted vertices. For vertices with degree more than $C_{deg}(\mu)d$ in G, we separate them into three cases:

1. corrupted vertices,
2. uncorrupted vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$,
3. uncorrupted vertices with degree smaller than $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$.

We will prove that, with probability at least 0.99, all three cases can be eliminated in $O(\mu n)$ rounds. Therefore, with probability 0.99, step 2 of Algorithm 5.21 terminates in $O(\mu n)$ rounds.

Case 1: Since there are at most μn corrupted vertices, it takes at most μn rounds to deal with corrupted vertices with degree more than $C_{deg}(\mu)d$ in G.

Case 2: For uncorrupted vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$ and degree more than $C_{deg}(\mu)d$ in G, we bound it by the total number of vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in G^0. By Chernoff Bound, we have that, for each vertex v, the probability that v has degree more than $\frac{1}{2}C_{deg}(\mu)d$ in G^0 is roughly bounded by

$$\mathbb{P}[\deg_{G^0}(v) \geqslant \frac{1}{2}C_{deg}(\mu)d] \leqslant O(\exp(-\frac{1}{2}C_{deg}(\mu)d))$$

Let T be the set of vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in G^0. By Markov's inequality, we get that the probability that T contains more than μn vertices is roughly bounded by

$$\mathbb{P}[|T| \geqslant \mu n] \leqslant O(\frac{\exp(-\frac{1}{2}C_{deg}(\mu)d)}{\mu})$$

By setting $C_{deg}(\mu)$ to be large enough with respect to μ, we get that, with probability 0.999, there are at most μn vertices with degree larger than or equal to $\frac{1}{2}C_{deg}(\mu)d$ in G^0. Therefore, it takes at most μn rounds to remove the vertices that are uncorrupted and has degree more than $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$.

Case 3: For uncorrupted vertices that have degree smaller than $\frac{1}{2}C_{deg}(\mu)d$ in $G[S]$ but have degree more than $C_{deg}(\mu)d$ in G, the key observation is that more than half of their neighbours are corrupted vertices. Therefore, each time such a node is removed as the highest degree node, with probability more than $1/2$, the algorithm will remove a corrupted node as its random neighbour.

Now, let us only consider the rounds where the highest degree node is in case 3 and let t denote the total number of such rounds. Let X_i denote the number of corrupted vertices removed after round i and $X_0 = 0$. We know that $X_{i+1} = X_i + 1$ with probability more than $1/2$ and $X_{i+1} = X_i$ otherwise. We also know that the process has to terminate when $X_t = \mu n$. Therefore, by the standard Markov Chain drift analysis (see Lemma 1 in [HY04]), we have:

$$\mathbb{E}[t] \leqslant 2\mu n$$

and, by Markov inequality, the probability that vertices in case 3 are not eliminated after $1000\mu n$ rounds where the highest degree node is in case 3 is bounded by

$$\mathbb{P}[t \geqslant 1000\mu n] \leqslant \frac{2\mu n}{1000\mu n} = 0.002$$

Therefore, with probability at least 0.998, vertices in case 3 are eliminated in $1000n$ rounds.

Conclusion. Taking union bound over the failure probabilities, we get that, with probability at least 0.99, step 2 Algorithm 5.21 terminates in $1002\mu n = O(\mu n)$ rounds. □

Notice that Lemma 5.23 gives us an upper bound on the total number of removed vertices during the pruning step and Algorithm 5.21 guarantees that G' will have bounded degree after pruning. These two observations allow us to have the following lemma, which says that the difference between G' and G^0 is at most $O(\rho n)$ edges, where $O(\rho n)$ is the number of edges that can be tolerated by the edge-robust algorithm in Theorem 5.20.

Lemma 5.24. *Let G^0 be the uncorrupted graph, with probability 0.99, the remaining graph G' in step 3 of Algorithm 5.21 differs from G^0 by $O(\rho n)$ edges, where $\rho \leqslant \left(\frac{1}{\delta}\log\frac{2}{\varepsilon}\right)^{-O(1/\delta)}$ as defined in Theorem 5.20.*

Proof. Graph G' differs from G^0 by two types of edges:

1. corrupted edges in G',
2. uncorrupted edges that are removed from pruning.

We will bound the two cases separately.

Case 1. Algorithm 5.21 guarantees that the degree of each vertex in G' is bounded by $C_{\text{deg}}(\mu)d$. Since there are at most μn corrupted vertices in G', the maximum number of corrupted edges in G' is $C_{\text{deg}}(\mu)\mu dn$. Since $d < d_\delta$ for some d_δ, we can set $C_{\text{deg}}(\mu)$ to be small enough such that $C_{\text{deg}}(\mu)\mu dn \leqslant O(\rho n)$.

Case 2. For case 2, we consider two types of vertices that are removed in Algorithm 5.21:

- vertices with degree smaller than or equal to $C_{\text{deg}}(\mu)d$ in G^0,
- vertices with degree larger than $C_{\text{deg}}(\mu)d$ in G^0.

For the first type of vertices, we observe that, with probability 0.99, Algorithm 5.21 terminates in $O(\mu n)$ rounds by Lemma 5.23. Therefore, with probability 0.99, there can be at most $O(\mu n)$ vertices in this case. Hence, the number of uncorrupted edges that are removed from pruning the first type of vertices can be bounded by $O(C_{\text{deg}}(\mu)\mu dn)$. Similar to case 1, we can set $C_{\text{deg}}(\mu)$ properly such that $O(C_{\text{deg}}(\mu)\mu dn) \leqslant O(\rho n)$.

For the second type of vertices, we know that, for each vertex v, the probability that v has degree larger than or equal to t is bounded by

$$\mathbb{P}[\deg_{G^0}(v) \geqslant t] \leqslant O(\exp(-t))$$

Let X_t denote the number of vertices with degree t in G^0. We have

$$\mathbb{E}[X_t] \leqslant \mathbb{P}[\deg_{G^0}(v) \geqslant t] \cdot n \leqslant O(\exp(-t)n)$$

Therefore, for some properly selected $C_{\deg}(\mu)$, the expected total number of edges from vertices with degree larger than $C_{\deg}(\mu)d$ in G^0 can be bounded by

$$\mathbb{E}\left[\sum_{t=C_{\deg}(\mu)d}^{\infty} X_t t\right] = \sum_{t=C_{\deg}(\mu)d}^{\infty} \mathbb{E}[X_t]t \leqslant O\left(\sum_{t=C_{\deg}(\mu)d}^{\infty} \exp(-t)tn\right)$$

By setting $C_{\deg}(\mu)$ to be a large enough value, we have

$$\mathbb{E}\left[\sum_{t=C_{\deg}(\mu)d}^{\infty} X_t t\right] \leqslant A\rho n$$

for some universal constant A. By Markov inequality, with probability 0.99, the number of uncorrupted edges that are removed from pruning second type of vertices can be bounded by $O(\rho n)$.

Conclusion. Take union bound over failure probabilities, we get that, with probability 0.98, G' differs from G^0 by $O(\rho n)$ edges. □

Now, we prove Theorem 5.22 using Lemma 5.23 and Lemma 5.24.

Proof of Theorem 5.22. First, we prove recovery guarantees of Algorithm 5.21. From Lemma 5.24, we know that, with probability 0.98, G' differs from G^0 by $O(\rho n)$ edges. Combine the guarantees of Theorem 5.20 and the rounding procedure in Lemma 5.19, step 3 will output an estimator $\hat{x} \in \{\pm 1\}^n$ such that

$$\mathbb{E}[\langle \hat{x}, x^* \rangle^2] \geqslant 0.98 \cdot \Omega(n^2) = \Omega(n^2)$$

Now, we prove the time complexity of Algorithm 5.21. For step 2 of the algorithm, each round takes at most $O(n^2)$ time and there can be at most n rounds. Therefore, step 2 takes at most $O(n^3)$ time. For step 3, the edge-robust algorithm from Theorem 5.20 takes polynomial time. For step 4, the rounding procedure from Lemma 5.19 takes polynomial time. Therefore, Algorithm 5.21 runs in polynomial time. □

5.5 Lower bound on the corrupted fraction

As stated in Theorem 5.2, our algorithm is robust against $\Omega_\delta(1)$ fraction of corrupted nodes. One might wonder whether we can remove the dependency on $\delta = \varepsilon^2 d/4 - 1$, and find an algorithm robust against $\Omega(1)$ fraction of corrupted nodes (e.g 0.001 fraction of corrupted nodes). The following claim shows that this is impossible.

Claim 5.25. Let $n > 1, d > 1, \varepsilon \in (0,1)$ and label vector $x^* \in \{\pm 1\}^n$. Let $\delta := \varepsilon^2 d/4 - 1$ and suppose $\delta \geqslant \Omega(1)$. For $G^0 \sim \mathsf{SBM}_{d,\varepsilon}(x^*)$ and $\mu \geqslant \delta$, if an adversary removes μn vertices uniformly at random from G^0 to obtain the graph G that we observe, then it is information theoretically impossible to achieve weak recovery given G, i.e. for any estimator $\hat{x}(G) \in \{\pm 1\}^n$, we have

$$\mathbb{E}\langle \hat{x}(G), x^* \rangle^2 \leqslant o(n^2)\,.$$

Proof. Let us denote the set of remaining vertices as R. Note that the remaining graph follows distribution $\mathsf{SBM}_{d',\varepsilon'}(x^*_R)$, where $d' = (1-\mu)\cdot d$ and $\varepsilon' = \varepsilon$. When $\mu \geqslant \delta$, we have $\varepsilon'^2 d'/4 \leqslant (1-\delta)^2 \cdot (1+\delta) \leqslant 1$. According to [MNS15b], it is information theoretically impossible to achieve weak recovery when $\varepsilon'^2 d'/4 \leqslant 1$. Thus, it is information theoretically impossible to recover x^*. □

5.6 Robust synchronization

In this section, we give an algorithm to solve the row/column-corrupted $\mathbb{Z}_2$ synchronization problem using techniques from Section 5.3. The idea is similar to the robust SoS algorithm for node-corrupted stochastic block model: we find a subset of the rows/columns such that the submatrix formed by the subset has large enough basic SDP value and bounded spectral norm. Then, we use the spectral norm bound to upper bound the basic SDP value of the submatrix formed by corrupted rows/columns in the selected subset.

Phase transition for synchronization. Before introducing our algorithm, we give a small summary of the basic SDP value phase transition for $\mathbb{Z}_2$ synchronization. It is based on Theorem 5 of [MS16], where they gave a very clean result for the phase transition of deformed GOE matrices. The phase transition can naturally be extended to the $\mathbb{Z}_2$ synchronization model using a simple argument based on rotational symmetry. The following theorem informally restates Theorem 5 of [MS16] and provides the result we need for our robust $\mathbb{Z}_2$ synchronization algorithm.

Theorem 5.26 ($\mathbb{Z}_2$ synchronization phase transition [MS16])**.** *Given an uncorrupted $\mathbb{Z}_2$ synchronization matrix A^0 that is generated acoodirng to Definition 5.3,*

- *if $\sigma \in [0,1]$, then for any $\xi > 0$, we have $\mathrm{SDP}(A^0) \in [(2-\xi)n^2, (2+\xi)n^2]$ with probability $1 - o(1)$,*
- *if $\sigma > 1$, then there exists $\Delta(\sigma) > 0$ such that $\mathrm{SDP}(A^0) \geqslant (2+\Delta(\sigma))n^2$ with probability $1 - o(1)$.*

Sum-of-squares algorithm. In this corruption model, the $\mathbb{Z}_2$ synchronization is easier than the stochastic block model in the sense that, with high probability, A^0 is already bounded in spectral norm. Therefore, we can omit the pruning step that we did for robust stochastic block model. Let A be the row/column corrupted Z_2 synchronization matrix generated according to Definition 5.3, consider the following system of polynomial equations in PSD matrix X of size $n \times n$ and vector w of size n:

$$\mathcal{A} := \begin{Bmatrix} w_i^2 = w_i & \forall i \in [n] \\ \sum_i w_i = (1-\mu)n & \\ X \succeq 0 & \\ X_{ii} = 1 & \forall i \in [n] \\ \langle A \odot (ww^\top), X \rangle \geqslant (2 + \Delta(\sigma))(1-\mu)^2 n^2 & \\ \left\| A \odot (ww^\top) \right\|_{\mathrm{op}} \leqslant (\sigma + \sigma^{-1})n & \end{Bmatrix} \tag{5.6.1}$$

where $\Delta(\sigma) > 0$ is value that only depends on σ.

In Eq. (5.6.1), the vector w is a $\{0,1\}$ vector that finds a subset of the rows/columns whose submatrix behaves like uncorrupted $\mathbb{Z}_2$ synchronization matrix. Essentially, we require the submatrix $A \odot (ww^\top)$ to have two properties: *(a)* it has large enough basic SDP value and *(b)* it has small enough spectral norm. The matrix X is a PSD matrix that is a solution to the basic SDP, such that the inner product between X and $A \odot (ww^\top)$ is large enough. This certifies property (a) and also allows us to obtain our estimator X. Property (b) is easy to show because there is an SoS certificate for spectral norm.

We will prove the following theorem for the SoS relaxation of Eq. (5.6.1), which implies Theorem 5.4. It says that, with high probability, there is a deg-4 SoS proof which shows that any X which satisfies Eq. (5.6.1) has non-trivial correlation with the true labels X^*.

Theorem 5.27 (SoS proof for robust $\mathbb{Z}_2$ synchronization)**.** *When $\sigma > 1$ and $\mu \leqslant \mu^*(\sigma)$ for some value $\mu^*(\sigma)$ that only depends on σ, with probability at least $1 - o(1)$, we have:*

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^* \rangle \geqslant \Omega(n^2)$$

To prove Theorem 5.27, we need to show two things:

1. $\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^* \rangle \geqslant \Omega(n^2)$ with high probability,
2. the constraint set Eq. (5.6.1) is feasible with high probability.

Now, we prove the first property in the following lemma.

Lemma 5.28. *For X and w satisfying Eq. (5.6.1), we have, with probability $1 - o(1)$,*

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^* \rangle \geqslant \Omega(n^2)$$

Proof. Let $s \in \{0,1\}^n$ be the indicator variable for the set of uncorrupted indices. We will use the following identity to prove the lemma

$$\langle X, X^* \rangle = \langle X, X^* \odot (ww^\top) \odot (ss^\top) \rangle + \langle X, X^* \odot (J - (ww^\top) \odot (ss^\top)) \rangle \tag{5.6.2}$$

Notice that we have $\mathcal{A} \left| \frac{X,w}{4} \right. X_{ij}^2 \leqslant 1$ for each $(i,j) \in [n] \times [n]$. Therefore, we can get the following bound for the second term of Eq. (5.6.2)

$$\mathcal{A} \left| \frac{X,w}{4} \right. \langle X, X^* \odot (J - (ww^\top) \odot (ss^\top)) \rangle \geqslant -4\mu n^2 \tag{5.6.3}$$

Now, the goal is to show that $\langle X, X^* \odot (ww^\top) \odot (ss^\top) \rangle$ is large enough. To do this, we will use the following identity

$$\langle X, X^* \odot (ww^\top) \odot (ss^\top) \rangle = \frac{1}{\sigma} (\langle X, A \odot (ww^\top) \odot (ss^\top) \rangle - \langle X, (A - \sigma X^*) \odot (ww^\top) \odot (ss^\top) \rangle) \tag{5.6.4}$$

The easy part is to bound $\langle X, (A - \sigma X^*) \odot (ww^\top) \odot (ss^\top) \rangle$. We know that $(A - \sigma X^*) \odot (ww^\top) \odot (ss^\top) = (A^0 - \sigma X^*) \odot (ww^\top) \odot (ss^\top)$ since it is restricted to the set of uncorrupted rows/columns. Moreover, from Theorem 5.26, we know that, with high probability, $\mathrm{SDP}(A^0 - \sigma X^*) \leqslant (2 + \xi) n^2$ for any constant $\xi > 0$. Therefore, by monotonicity of the basic SDP from Claim 5.10, we have:

$$\begin{aligned}
\mathcal{A} \left| \frac{X,w}{4} \right. & \langle X, (A - \sigma X^*) \odot (ww^\top) \odot (ss^\top) \rangle \\
&= \langle X, (A^0 - \sigma X^*) \odot (ww^\top) \odot (ss^\top) \rangle \\
&\leqslant \mathrm{SDP}((A^0 - \sigma X^*) \odot (ww^\top) \odot (ss^\top)) \\
&\leqslant \mathrm{SDP}(A^0 - \sigma X^*) \\
&\leqslant (2 + \xi) n^2
\end{aligned}$$

The hard part is to show that $\langle X, A \odot (ww^\top) \odot (ss^\top) \rangle$ is large enough. We show this via the following identity:

$$\langle X, A \odot (ww^\top) \odot (ss^\top) \rangle = \langle X, A \odot (ww^\top) \rangle - \langle X, A \odot (ww^\top) - A \odot (ww^\top) \odot (ss^\top) \rangle$$

For the first term $\langle X, A \odot (ww^\top) \rangle$, we can simply apply the program constraint and get:

$$\mathcal{A} \left| \frac{X,w}{4} \right. \langle X, A \odot (ww^\top) \rangle \geqslant (2 + \Delta(\sigma))(1 - \mu)^2 n^2$$

For the second term $\langle X, A \odot (ww^\top) - A \odot (ww^\top) \odot (ss^\top) \rangle$, we bound it by the Grothendieck norm of $A \odot (ww^\top) - A \odot (ww^\top) \odot (ss^\top)$. Since we have $\mathcal{A} \left| \frac{X,w}{4} \right. \| A \odot (ww^\top) \|_{\mathrm{op}} \leqslant (\sigma + \sigma^{-1}) n$ from the program constraints, we can apply Lemma 5.6 to get

$$\mathcal{A} \left| \frac{X,w}{4} \right. \langle X, A \odot (ww^\top) - A \odot (ww^\top) \odot (ss^\top) \rangle \leqslant 3K_G(\sigma + \sigma^{-1}) \sqrt{\mu} n^2$$

Thus, we have:

$$\mathcal{A} \left| \frac{X,w}{4} \right. \langle X, A \odot (ww^\top) \odot (ss^\top) \rangle \geqslant (2 + \Delta(\sigma))(1 - \mu)^2 n^2 - 3K_G(\sigma + \sigma^{-1}) \sqrt{\mu} n^2$$

Plug the two parts into Eq. (5.6.4), we get:

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^* \odot (ww^\top) \odot (ss^\top)\rangle \geqslant \frac{1}{\sigma}\big((2+\Delta(\sigma))(1-\mu)^2 n^2 - 3K_G(\sigma+\sigma^{-1})\sqrt{\mu}n^2 - (2+\xi)n^2\big) = \theta(\sigma)n^2 \tag{5.6.5}$$

for some $\theta(\sigma)$ that only depends on σ.

Now, plug Eq. (5.6.3) and Eq. (5.6.5) into Eq. (5.6.2), we get

$$\begin{aligned}\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^*\rangle &= \langle X, X^* \odot (ww^\top) \odot (ss^\top)\rangle + \langle X, X^* \odot (J - (ww^\top) \odot (ss^\top))\rangle \\ &\geqslant \theta(\sigma)n^2 - 4\mu n^2\end{aligned}$$

When $\mu \leqslant \mu^*(\sigma)$ for some value $\mu^*(\sigma)$ that only depends on σ, we have

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \theta'(\sigma)n^2$$

where $\theta'(\sigma)$ is a value that only depends on σ. Thus, when $\sigma > 1$, we have, with probability $1 - o(1)$,

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \Omega(n^2)$$

□

Now, we are ready to prove Theorem 5.27. In the proof, we first prove feasibility of the contraint set in Eq. (5.6.1), then use Lemma 5.28 to complete the proof.

Proof of Theorem 5.27. The feasibility analysis is similar to the feasibility analysis in Lemma 5.16. From Theorem 5.26 and union bound, we get that the inequality $\langle A \odot (ww^\top), X\rangle \geqslant (2+\Delta(\sigma))(1-\mu)^2 n^2$ is feasible with probability $1 - o(1)$. The inequality $\|A \odot (ww^\top)\|_{\mathrm{op}} \leqslant (\sigma + \sigma^{-1})n$ is feasible with probability $1 - o(1)$ due to the famous BBP phase transition and monotonicity of spectral norm. Take union bound over failure probabilities of the two inequalities, we can conclude the feasibility analysis.

From Lemma 5.28, we get that, with probability $1-o(1)$, we have $\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \Omega(n^2)$. Therefore, we can take union bound and conclude that, with probability $1 - o(1)$, the program finds an X such that

$$\mathcal{A} \left|\frac{X,w}{4}\right. \langle X, X^*\rangle \geqslant \Omega(n^2)$$

□

Now we finish the proof of Theorem 5.4.

Proof of Theorem 5.4. By combining Theorem 5.27 and Theorem 5.13, we can compute the pseudo-expectation $\tilde{\mathbb{E}}$ for the SoS relaxtion of Eq. (5.6.1) in polynomial time. Let $\hat{X} := \tilde{\mathbb{E}}[X]$ in Eq. (5.3.1). By linearity of pseudo-expectation, we have $\hat{X} \succeq 0$, $\hat{X}_{ii} = 1$. Furthermore, we have $\langle \hat{X}, X^*\rangle \geqslant \Omega(n^2)$ with probability $1 - o(1)$. Now, applying rounding procedure in Lemma 5.19, we can then obtain $\hat{x} \in \{\pm 1\}^n$ such that $\mathbb{E}\langle \hat{x}, x^*\rangle^2 \geqslant \Omega(n^2)$. □

Chapter 6

Random CSPs with adversarial signs

We prove here the results on constraint satisfaction problems discussed in Section 1.1.4. Recall there are two algorithmic questions concerning CSPs: finding an assignment maximizing the number of satisfied constraints, or finding a strong refutations of the instance.

In Chapter 1 we saw that an important algorithmic milestone was the idea of using spectral techniques to find refutations, introduced in [FGK05] and then refined in subsequent work. This idea turned out to be useful not only for refuting random CSPs, but also for the maximization version of the problem as well, when combined with convex programming. As the results in Section 1.1.4 requires novel certificates, we follow a similar roadmap here. First we provide novel certificates, then we use them to prove tight strong refutations. Finally we apply these ideas to the maximization problem, obtaining the sharp results against adversarial perturbations of Theorem 1.5.

State-of-the-art certificates. The state of the art concerning polynomial-time computable strong refutations of random constraint satisfaction problems is [AOW15]. Allen, O'Donnell and Witmer [AOW15] show how to obtain strong refutations for random k-XOR constraint satisfaction problems on n variables and $n^{k/2}(\log n)^{O(1)}$ constraints. When k is even, $O(n^{k/2})$ constraints suffice. Thanks to a reduction from arbitrary constraint satisfaction to k-XOR (of which we provide a self-contained simpler proof in Section 6.6), similar bounds hold for any constraint satisfaction problem over k variables.

To illustrate the difference between odd k and even k, we briefly discuss how a strong refutation for random 4-XOR and random 3-XOR instances is constructed.

In general, if we have an instance of k-XOR with m constraints and n variables, a strong refutation is a certificate that

$$\max_{x\in\{-1,1\}^n} \sum_{i_1,\dots,i_k} T_{i_1,\dots,i_k} x_{i_1}\cdots x_{i_k} \leqslant \varepsilon m$$

where T is a symmetric tensor of order k such that $T_{i_1,\ldots,i_k} = 0$ if there is no constraint on the k-tuple of variables $x_{i_1}, \ldots, x_{i_k}$, and otherwise $T_{i_1,\ldots,i_k} = \pm 1$ depending on the right-hand-side of the constraint.

When $k = 4$, we can flatten the tensor to an $n^2 \times n^2$ symmetric matrix M (where $M_{(a,b),(c,d)} = T_{a,b,c,d}$) and we have

$$\max_{x\in\{-1,1\}^n} \sum_{i_1,\ldots,i_4} T_{i_1,\ldots,i_4} x_{i_1} \cdots x_{i_4} = \max_{x\in\{-1,1\}^n} (x^{\otimes 2})^{\mathsf{T}} M x^{\otimes 2}$$

Now we can relax the right-hand side to a maximization over arbitrary n^2-dimensional Boolean vectors and further relax to the ∞-to-1 norm:

$$\max_{x\in\{-1,1\}^n} (x^{\otimes 2})^{\mathsf{T}} M x^{\otimes 2} \leqslant \max_{y\in\{-1,1\}^{n^2}} y^{\mathsf{T}} M y \leqslant \max_{y,z\in\{-1,1\}^{n^2}} y^{\mathsf{T}} M z = ||M||_{\infty\to 1}$$

Finally, the last expression above can be upper bounded by εm, by using Chernoff bounds and a union bound over all the 2^{2n^2} possible choices for y and z, which is possible if m is a sufficiently large constant times n^2/ε^2. Finally, we can use Grothendieck's inequality to get us a certified upper bound of the $\infty \to 1$ norm in polynomial time up to a constant factor.

For 3-XOR, the idea is to apply a Cauchy-Schwarz step to reduce the problem of bounding a degree-4 problem, and then to flatten the resulting 4-tensor to an $n^2 \times n^2$ matrix M such that

$$\max_{x\in\{-1,1\}^n} \sum_{i_1,i_2,i_4} T_{i_1,i_2,i_3} x_{i_1} x_{i_2} x_{i_3} \leqslant \sqrt{n} \cdot \sqrt{\max_{x\in\{\pm 1\}^n} (x^{\otimes 2})^{\mathsf{T}} M x^{\otimes 2}} \leqslant \sqrt{n} \cdot \sqrt{\max_{y,z\in\{\pm 1\}^{n^2}} y^{\mathsf{T}} M z}$$

Unfortunately, now it is not possible any more to bound the maximum on the above right-hand via a union bound over 2^{2n^2} cases. Indeed, for this to be possible, we would need our distribution to have at least order of n^2 bits of entropy, and so we would need to have order of n^2 constraints.

The alternative is to obtain a bound in terms of the spectral norm of M, using the fact that

$$\max_{y,z\in\{\pm 1\}^{n^2}} y^{\mathsf{T}} M z \leqslant n^2 \cdot ||M|| \,.$$

But for a sparse matrix to have a non-trivial bound on its spectral norm, we have to have at least $\operatorname{poly}\log n$ non-zero entries per row on average[1], and for this to happen the number of constraints has to be at least of the order of $n^{1.5}\operatorname{poly}\log n$. In the regime of

[1]This is similar to the phenomenon that the quasirandomness of a $G_{n,p}$ random graphs can be certified in terms of the non-trivial eigenvalues of the adjacency matrix only if the average degree is at least logarithmic. We will return to the graph analogy shortly.

$n^{1.5}$poly log n random 3-XOR constraints, a spectral norm bound on M can be established via trace methods, and this is how the results of [AOW15] are proved in the case of odd k.

In the context of designing algorithms that satisfy as-many-as-possible clauses in the given a random CSP instance, similarly as for refutations, polynomial time approximation schemes are known [BRS11, AJT19] when the number of clauses is of the order $n^{k/2}(\log n)^{O(1)}$. Indeed the algorithmic techniques behind these PTAS are closely related to those used for refutations and, in particular, boils down to combining bounds on the spectrum of the flattened tensor representing the instance and ideas from sum-of-squares relaxations.

The groundbreaking work [GKM22], showed that a similar picture holds in the significantly more general settings of *smoothed* CSPs: where both the literal negation patterns and clauses are chosen arbitrarily, but then signs are randomly flipped with a small, yet constant, probability.[2].

Sharp strong refutations. Our first result of the chapter breaks the $n^{k/2}$poly log n barrier for strong refutations of random k-XOR instances, with odd k.

Theorem 6.1 (Strong refutations of random k-XOR)**.** *There exists an efficient algorithm that, given an instance $\boldsymbol{\mathcal{I}}$ of random k-XOR with $n^{k/2}/\varepsilon^2$ constraints, with probability at least* 0.99*, finds strong refutation of $\boldsymbol{\mathcal{I}}$, that is, a certificate that*

$$\mathrm{Opt}_{\boldsymbol{\mathcal{I}}} \leqslant \frac{1}{2} + O(\varepsilon)\,.$$

Using the known reduction of general k-CSP to k-XOR, of which we provide a simple self-contained proof, we have the following consequence.

Theorem 6.2 (Strong refutations of random CSPs)**.** *Let $P : \{-1,+1\}^k \to \{0,1\}$ be a Boolean k-ary predicate, and call $\mathbb{E}\,P$ the probability that P is satisfied by a random assignment. There exists a polynomial time algorithm that given a random instance $CSP(P)$ instances $\boldsymbol{\mathcal{I}}$, over n variables, with at least $n^{k/2}/\varepsilon^2$ constraints, with probability at least* 0.99*, finds a strong refutation of $\boldsymbol{\mathcal{I}}$, that is, a certificate that*

$$\mathrm{Opt}_{\boldsymbol{\mathcal{I}}} \leqslant \mathbb{E}\,P + O(\varepsilon)\,.$$

Robust approximation algorithms against adversarial sign patterns. As already discussed, our techniques can be further applied to design efficient algorithms finding an assignment with value Opt $- O(\varepsilon)$ beyond the $n^{k/2}$ polylog n barrier, even in the semi-random settings where: *first*, clauses are sampled randomly, and *second*, given the instance, the sign pattern of *each* clause is adversarially perturbed. Such perturbations are not captured by the smooth models of [Fei07, GKM22] and hence require different

[2]Smoothed CSPs were first introduced in [Fei07]

algorithmic challenges. In the special case of even k, [Kot22] provided a PTAS whenever $p \geqslant n^{k/2} \operatorname{polylog} n$.

Theorem 6.3 (Algorithm for k-XOR with adversarial patterns)**.** *Let n , k be positive integers, $\varepsilon > 0$, n and $n^{-k/2}/\varepsilon^2 < 1$. Let $\mathcal{I}$ be a k-XOR instance constructed through the following process:*

- *Sample a random k-XOR instance $\boldsymbol{\mathcal{I}}'$ with at least $n^{k/2}/\varepsilon^2$ constraints.*
- *Given $\boldsymbol{\mathcal{I}}'$, arbitrarily (possibly adversarially) replace the sign of each clause in $\boldsymbol{\mathcal{I}}'$.*

There exists a randomized algorithm, running in time $n^{O(k/\varepsilon^2)}$, that returns an assignment $\hat{\mathbf{x}}$ with value

$$\operatorname{Val}_{\mathcal{I}}(\hat{\mathbf{x}}) \geqslant \operatorname{Opt}_{\mathcal{I}} - O(\varepsilon) ,$$

with probability at least 0.99*.*

As in the case of strong refutations, Theorem 6.3 can be extended to k-CSPs.

Theorem 6.4 (Restatement of Theorem 1.5)**.** *Let n , k be positive integers, $\varepsilon > 0$, n and $n^{-k/2}/\varepsilon^2 < 1$. Let $P : \{-1,+1\}^k \to \{0,1\}$ be a Boolean k-ary predicate. Let $\mathcal{I}$ be a $CSP(P)$ instance constructed through the following process:*

- *Sample a random $CSP(P)$ instance $\boldsymbol{\mathcal{I}}'$ with at least $n^{k/2}/\varepsilon^2$ constraints.*
- *Given $\boldsymbol{\mathcal{I}}'$, for each clause in $\boldsymbol{\mathcal{I}}'$, replace the sign pattern with an arbitrary (possibly adversarial) sign pattern.*

There exists a randomized algorithm, running in time $n^{O(k/\varepsilon^2)}$, that returns an assignment $\hat{\mathbf{x}}$ with value

$$\operatorname{Val}_{\mathcal{I}}(\hat{\mathbf{x}}) \geqslant \operatorname{Opt}_{\mathcal{I}} - O(\varepsilon) ,$$

with probability at least 0.99*.*

6.1 Techniques

At the heart of the chapter there are new techniques to bound[3]

$$\max_{x \in \{\pm 1\}^N} x^{\mathsf{T}} \mathbf{M} x \tag{6.1.1}$$

when $\mathbf{M}$ is a random $N \times N$ matrix with only a constant expected number of non-zero entries per row and per column, and in which such entries are not independent.

[3]We use boldface to denote random variables.

A toy problem. Before explaining the main ideas, consider the following question, which models some of the difficulties that we encounter: suppose that we are given a random graph on N vertices, and such that every edge exists with probability d/N, where d is a constant, but the edges are only known to be *poly* $\log N$-wise independent, and not fully independent. Can we certify that the graph has interesting quasirandom properties, for example can we certify that the Max Cut optimum is at most a $1/2 + O(1/\sqrt{d})$ fraction of edges?

One approach could be to bound $||\mathbf{A} - \mathbb{E}\,\mathbf{A}||_{\infty\to 1}$ where $\mathbf{A}$ is the adjacency matrix of the graph. If the graph has mutually independent random edges, that is, if it is sampled from an Erdős-Reniy distribution $G_{N,\frac{d}{N}}$, then we can use a union bound over 2^{2N} cases to argue that with high probability

$$||\mathbf{A} - \mathbb{E}\,\mathbf{A}||_{\infty\to 1} \leqslant O(\sqrt{d}N)$$

which is certifiable in polynomial time, up to a constant factor loss, using Grothendieck's inequality and which certifies that the Max Cut optimum is at most $1/2 + O(1/\sqrt{d})$. Unfortunately, if the edges are only polylog N-wise independent, then it is not possible to take such union bound.

Another option in the fully independent case is to use the results of Feige and Ofek [FO05], which show that, after removing nodes of degree larger than, say, $10d$, the adjacency matrix of the residual graph has second eigenvalue at most $O(\sqrt{d})$ with high probability. Unfortunately the proof of Feige and Ofek also relies on a union bound over $2^{O(N)}$ cases, and so it cannot work in the polylog N-wise independent case.

A trace argument can be used to prove that, with high probability, we have

$$||\mathbf{A} - \mathbb{E}\,\mathbf{A}|| \leqslant O(\sqrt{d\log N})$$

which provides a polynomial time certificate that the Max Cut optimum is at most $1/2 + O(\sqrt{\log N}/\sqrt{d})$, and the trace calculation only requires $O(\log N)$-wise independence. It does, however, introduce an extra logarithmic factor, which is unavoidable because the spectral norm of $||\mathbf{A} - \mathbb{E}\,\mathbf{A}||$ is $\tilde{\Omega}(\sqrt{\log N})$ when d is constant.

It is conceivable that one could prove the result of Feige and Ofek (that the adjacency matrix has second largest eigenvalue $O(\sqrt{d})$ after the removal of high-degree vertices) through a trace bound on the adjacency matrix of the truncated graph, although it seems very difficult to deal with the conditional distribution of edges given that the edges survive the truncation.

A solution to the toy problem. Although all standard techniques fail, there is a way to combine certain recent results to solve our toy problem. The starting point is the fact that, given an undirected graph $G = (V, E)$, we can define the "non-backtracking" $2|E| \times 2|E|$ matrix B of G, and that this matrix satisfies the Ihara-Bass equation

$$\det(\mathrm{Id} - xB) = (1 - x^2)^{|E|-|V|} \cdot \det(\mathrm{Id} - xA + x^2(D - \mathrm{Id}))$$

where A is the adjacency matrix of the graph, D is the diagonal matrix of degrees, and the above equation holds as an identity of polynomials of degree $2|E|$ in x. See the survey of Horton [HST06] for an exposition of these definitions and results.

Fan and Montanari [FM17] show that bounds on the spectral radius of B imply useful PSD inequalities on A. In particular, if $\lambda_{\min}$ is the smallest real eigenvalue of B, then we have

$$A \succeq -|\lambda_{\min}| \cdot \mathrm{Id} - \frac{1}{|\lambda_{\min}|} \cdot (D - \mathrm{Id})$$

In the context of their work on the Stochastic Block Model, Bordenave, Lalarge and Massoulié [BLM15] use a trace argument to prove a result that implies that $\lambda_{\min} \geqslant -(1 + o(1)) \cdot \sqrt{d}$ in $G_{N,\frac{d}{N}}$ random graphs, and so all these results together imply that the Max Cut of a $G_{N,\frac{d}{N}}$ random graph is with high probability at most $1/2 + (1 + o(1))/\sqrt{d}$, and that this upper bound is efficiently certifiable, for example by the dual of the Goemans-Williamson relaxation.

The key point is that there was never a union bound over $2^{O(N)}$ cases in the above argument and that, in fact, everything works assuming polylog N-wise independence of the edges.[4]

From unweighted graphs to general symmetric matrices. Our goal is to develop an analog of this argument where we work with the $n^2 \times n^2$ matrix M that comes up in the analysis of 3-XOR (or, in general, with the $n^{\lceil k/2 \rceil} \times n^{\lceil k/2 \rceil}$ matrix that comes up in the analysis of k-XOR when k is odd) instead of the adjacency matrix A of the pseudorandom graph analysed above.

The first challenge in carrying out this program is that the original notion of non-backtracking matrix is defined only with respect to 0/1 Boolean symmetric matrices, while we want to study matrices with positive and negative entries that can be arbitrary integers.

A certain generalization of non-backtracking matrices was already introduced in [WF09, FM17], however for technical reasons *we* were not able to use it to carry out our program. We thus introduce a novel theory of "non-backtracking" matrices associated to any given symmetric matrix. In Section 6.3, given a symmetric $N \times N$ matrix M with Nz non-zero entries, we define an $Nz \times Nz$ "non-backtracking" matrix B_M associated to M, and we prove (see Theorem 6.6) an Ihara-Bass-type identity

$$\det(\mathrm{Id} - xB_M + x(L_M - J_M)) = (1 - x^2)^{Nz/2-N} \cdot \det(\mathrm{Id} - xM + x^2(D_M - \mathrm{Id}))$$

[4] Incidentally, this combination of Fan-Montanari ideas and Bordenave-Lalarge-Massouli's bounds, also implies that if A' is the adjacency matrix of a graph G sampled from a distribution in which edges have probability d/N and are polylog N wise independent, and then truncated by removing all vertices of degree more than, say, $10d$, then we have with high probability $A' \succeq -O(\sqrt{d}) \cdot I$, proving a one-sided version of the result of Feige and Ofek.

where D_M, L_M and J_M are certain matrices that are associated to M. When M is Boolean, $L_M = J_M$ and D_M is the diagonal matrix such that $(D_M)_{i,i} = \sum_j M_{i,j}$, so our equation becomes the standard Ihara-Bass equation in the case of Boolean M. Conveniently, closed non-backtracking walks W arising from the definition of B_M take value in $\{\pm \prod_{(i,j)\in W} M_{ij}\}$, allowing one to easily mimic arguments used for standard non-backtracking matrices.

Now, given a bound on the spectral radius of $B_M - L_M + J_M$, it is possible, with an argument in the style of Fan and Montanari, to deduce a certifiable bound on the ∞-to-1 norm of M.

Bounding the spectral radius via weighted hyper-walks. Studying the spectral radius of $B_{\mathbf{M}} - L_{\mathbf{M}} + J_{\mathbf{M}}$ –matrices associated to the matrix $\mathbf{M}$ coming from random k-XOR instances– is the main technical challenge of this work.

Our bound (Theorem 6.27) relies on a trace argument of $B_{\mathbf{M}}$. However, compared to Bordenave, Lalarge and Massoulié [BLM15] our setup presents a number of new technical challenges.

One challenge comes from the extra terms that we have in the non-Boolean case. In particular, our non-backtracking matrix $B_{\mathbf{M}}$ has entries that are the absolute values of certain entries of $\mathbf{M}$. To compute an expectation of the trace of the symmetrization of a power of $B_{\mathbf{M}}$, we replace absolute values with squares, and bound the error that we incur because of this.

Perhaps the most important challenge comes from the fact that the trace bound ultimately boils down to a weighted count of certain closed "hypergraph walks" performed on the hypergraph corresponding to constraints of the k-XOR instance. These objects arise from our notion of non-backtracking walks on the symmetric matrix $\mathbf{M}$ obtained from the instance. This count is performed by showing that such walks can be encoded with a small number of bits. It is enough to count walks in which every hyperedge is repeated at least twice, and the crux of the argument is that the second time we see a hyperedge we can encode that hyperedge in a compact way. A naive way of doing that would point back to the previous step in the walk in which that hyperedge appeared, and this costs $\log \ell$ bits where ℓ is the length of the walk. To obtain the right result, however, repeated hyperedges have to be represented with an amortized constant number of bits per occurrence. The argument of Bordenave, Lalarge and Massoulié [BLM15] relies on the assumption, which is true with high probability, that the graph is "tangle-free," meaning that small subgraphs have at most one cycle. We have to work with a looser notion of tangle-free hypergraph in order to prove that it holds with high probability, but we are still able to obtain the desired bound.

From spectral bounds to algorithms. It is clear that an algorithm certifying tight bounds on Eq. (6.1.1) for the matrix M obtained from k-XOR instances can be used for strong refutations. Instead, to obtain Theorem 6.3 additional ideas are needed.

Our starting point is the local-to-global rounding paradigm of [BRS11]. As it is often the case, the odd settings are significantly more challenging than the regimes with k even. Hence consider first a 2-XOR random instance $\boldsymbol{\mathcal{I}}$. Up to the signs of the clauses, this may be represented as a graph $\mathbf{G}$ over n vertices. Now, for a distribution ν over assignments, one may define the local and global correlations as

$$\mathrm{LC}_{\mathbf{G}}(\nu) = \mathop{\mathbb{E}}_{(\mathbf{a},\mathbf{b})\sim E(\mathbf{G})} \left|\mathrm{Cov}_{\nu}\left(\mathbf{x}^{\mathbf{a}}, \mathbf{x}^{\mathbf{b}}\right)\right|$$
$$\mathrm{GC}(\nu) = \mathop{\mathbb{E}}_{(\mathbf{a},\mathbf{b})\sim [n]\times[n]} \left|\mathrm{Cov}_{\nu}\left(\mathbf{x}^{\mathbf{a}}, \mathbf{x}^{\mathbf{b}}\right)\right|.$$

If the local correlation is bounded by ε, it is possible to obtain an assignment with value $\mathrm{Opt}_{\boldsymbol{\mathcal{I}}} - O(\varepsilon)$ simply looking at the *first moment* of ν. Moreover, one can always transform ν into a distribution with small global correlation in polynomial time.

With these observations, the argument of [BRS11] comes down to: *(i)* bounding the difference between local and global correlation in terms of the spectral radius $\rho_{\mathbf{G}}$ of the centered adjacency matrix of the graph $\mathbf{G}$, *(ii)* showing that one can always find, in time $n^{O(1/\varepsilon^2)}$, a degree $O(1)$ pseudo-distribution over the hypercube with global correlation at most ε. As we only required low-degree moments to obtain the desired assignment, the argument goes through in this case as well.

To combine this approach with the bounds previously illustrated and extend the argument to random k-XOR instances with $m \geqslant \Omega(n^{k/2}/\varepsilon^2)$ clauses, we need to introduce two novel ingredients. *First*, we need new notions of local and global correlations which difference can be bounded studying the matrix $\mathbf{M}$ arising from the instance. *Second*, we need to bound this difference not in term of the eigenvalues of $\mathbf{M}$ but rather in terms of Eq. (6.1.1).

A careful Cauchy-Schwarz application allows us to formulate notions of local and global correlations in terms of $\mathbf{M}$. Its squaring step, further allows us to get rid of absolute values, thus providing an avenue to bound the difference between local and global correlation in terms of $\max_{x\in\{\pm1\}^n} x^{\mathsf{T}}\mathbf{M}x$.

Finally, since the adversarial perturbations in Theorem 6.3 cannot alter the "hypergraph walks" required to prove our bound, we are able to generalize our result to these settings.

Perspective. Several results on the average-case complexity of Sum-of-Square relaxations rely on proving that sparse matrices with non-independent entries are “quasirandom” in an appropriate sense. At its heart this chapter introduces a new approach to prove results of this form, which applies to very sparse matrices that have only a constant expected number of non-zero entries per row and per column. These ideas may find further application, for example to the context of semi-random instances of constraint satisfaction problems [GKM22] or of higher-degree Sum-of-Square relaxations of random constraint satisfaction problems [RRS17, WEAM19].

This theory could also be useful to study problems on random weighted graphs.

The certificates in the chapter prove certain PSD inequalities, and can be seen as Semidefinite Duals of certain Sum-of-Squares relaxations, but the computation of the certificate only requires an eigenvalue computation of a certain matrix, and does not require the solution of an SDP. There might be other ways to apply this theory so that one uses SDP relaxations only in the analysis, but the algorithm itself is purely spectral.

6.2 Preliminaries

We reuse the notation introduced in Chapter 2, but introduce additional useful facts and preliminary notions. We denote random variables in **bold**. We use lower case letters $a, b, c, d, \ldots$ to denote indices or scalars (the use willl be clear from context). We use the greek letters α, β, η to denote multi-indices. The cardinality of a multi-index α is $|\alpha|$. The i-th index in α is $\alpha(i)$. We may thus write a monomial (with coefficient c) in indeterminates $x_1, \ldots, x_n$ as $c \cdot x^\alpha$. For two multi indices $\alpha, \beta \in [n]^k$ we denote by (α, β) the multi-index obtained concatenating α and β. Multi-indices $\alpha, \beta \in [n]^k$ satisfy $\alpha = \beta$ if at each position the corresponding indices are identical. We use $S(\alpha)$ to denote the unordered multi-set of indices in α. We use n to denote our ambient dimension. For functions $f, g : \mathbb{R} \to \mathbb{R}$ we write $f = o(g)$ and $g = \omega(f)$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.

Matrices. For a matrix $M \in \mathbb{R}^{n\times n}$, we denote by $\lambda_1(M) \geqslant \ldots \geqslant \lambda_n(M)$ its eigenvalues. Then $\rho(M) := \max_i |\lambda_i(M)|$ is the spectral radius of M. We write $\|M\|_{\infty\to 1} := \max_{x,y\in\{\pm 1\}^n} \langle M, xy^\mathsf{T}\rangle$ and $\|M\|_{\max} := \max_{ij} \left|M_{ij}\right|$. Furthermore, we let $\|M\|_{\mathrm{Gr}} = \max\{\langle M, X\rangle \mid X \succeq 0, X_{ii} \leqslant 1, \forall i \in [n]\}$. We denote by $|M|$ the matrix with entries $(|M|)_{ij} := \left|M_{ij}\right|$.

Graphs. For a graph G, $V(G)$ and $E(G)$ denotes respectively its set of vertices and edges. $\vec{E}(G) := \{(u, v) : u \neq v \in V(G), uv \in E(G)\}$ is the set of all its ordered pairs such that $\{u, v\} \in E(G)$. For $e \in \vec{E}(G)$, $s(e)$ and $t(e)$ are respectively the source and target of the oriented edge. We write e^{-1} for its inverse. We also write K_n for the complete graph over n vertices. For a graph G with n vertices, we write $A(G) \in \mathbb{R}^{n\times n}$ for its adjacency matrix. For a vertex $v \in V(G)$, we denote by $\deg_G(v)$ its degree. We denote by $N_{G,t}(v)$ the set of vertices in G at distance t from v. We and drop the subscript G when the context is clear. If the graph G is weighted with weights given by $w : V(G) \times V(G) \to \mathbb{R}$, then $A_{uv} = w(\{uv\})$. If $e \neq E(G)$, then we assume $w(e) = 0$.

Walks. A walk W in a graph G is a sequence of vertices $(v_1, \ldots, v_{z+1})$. We refer to the directed edges of a walk W as $e_1(W), \ldots, e_z(W)$. When the context is clear we simply write $e_1, \ldots, e_z$. We use $G(W)$ to denote the subgraph of G traversed by the walk W. The set of vertices and distinct edges in $G(W)$ are denoted by $V(W)$ and $E(W)$. The multi-set of

edges in W is denoted by $M(W)$. For $e \in E(W)$, $m_W(e)$ is the multiplicity of e in $M(W)$. We write $H(V(W), M(W))$ for the multigraph generated by W. A walk $v_1, \ldots, v_{z+1}$ is said to be non-backtracking if for any $i \leqslant z-1$, $v_i \neq v_{i+2}$. For $e, f \in \vec{E}(G)$, $\mathrm{NBW}^z_{ef}(G)$ denotes the set of non-backtracking walks in G starting with e and ending with f of length k. For simplicity we let $\mathrm{NBW}^z_{ef} := \mathrm{NBW}^z_{ef}(K_n)$.

Hypergraphs. We use the notation $H(V, E)$ for an hyper-graphs over $V(H)$ with hyper-edge set $E(H)$. We only consider hypergraphs in which edges have the same arity. The arity of the edges will be clear from context. A multi-hyper-graph is a hyper-graph where edges may have multiplicity more than 1, we denote its multi-set of hyper-edges by $M(H)$ and its set of distinct hyper-edges by $E(H)$. When clear from context, we will refer to multi-hyper-graphs simply as hyper-graphs. Given hyper-graphs H, H' we denote by $H^* = H \oplus H'$ the multi-hyper-graph obtained by taking $V(H^*)$ as the union of the vertex sets and $M(H^*)$ as the multi-set of elements either in $M(H)$ or $M(H')$.

6.2.1 CSPs, k-XOR and strong refutations

k-XOR. A random k-XOR instance $\boldsymbol{\mathcal{I}}$ with n variables and $p\binom{n}{k}(1 \pm o(1))$ clauses can be generated by picking a random symmetric tensor $\mathbf{T}$, with independent entries, such that $\mathbf{T}_\alpha = 0$ if the indices in the multi-index $\alpha \in [n]^k$ are not distinct and otherwise:

$$\mathbf{T}_\alpha = \begin{cases} 0 & \text{with probability } 1-p\,, \\ +1 & \text{with probability } p/2\,, \\ -1 & \text{with probability } p/2\,. \end{cases}$$

We denote by m the exact number of clauses in the instance. Then $\boldsymbol{\mathcal{I}}$ consists of the m k-XOR predicates $k\text{-XOR}(\alpha) = \frac{1-x^\alpha(-\mathbf{T})_\alpha}{2}$ where $\mathbf{T}_\alpha$ is non-zero. We use $\mathcal{F}_{k\text{-XOR}(n,p)}$ to denote such distribution and $\boldsymbol{\mathcal{I}} \sim \mathcal{F}_{k\text{-XOR}(n,p)}$ to denote a random instance. We let $\mathrm{Val}_{\boldsymbol{\mathcal{I}}}(x)$ be the fraction of constrained satisfied by the assignment $x \in \{\pm 1\}^n$ and $\mathrm{Opt}_{\boldsymbol{\mathcal{I}}} := \max_{x\in\{\pm1\}^n} \mathrm{Val}_{\boldsymbol{\mathcal{I}}}(x)$. For any assignment $x \in \{\pm 1\}^n$ we have

$$\mathrm{Val}_{\boldsymbol{\mathcal{I}}}(x) = \frac{1}{2} + \frac{1}{m(\boldsymbol{\mathcal{I}})} \sum_{\alpha\in[n]^k} \frac{x^\alpha \mathbf{T}_\alpha}{2}\,.$$

Notice that since m will be $(1 \pm o(1))p\binom{n}{k}$ with overwhelming probability, we blur the distinction between these parameters. Then the max k-XOR problem is that of finding an assignment with value

$$\max_{x\in\{\pm1\}^n} \sum_{\alpha\in[n]^k} \mathbf{T}_\alpha x^\alpha\,. \tag{6.2.1}$$

This is captured by the following proposition.

Proposition 6.5. *Let $\boldsymbol{\mathcal{I}} \sim \mathcal{F}_{k\text{-XOR}(n,p)}$ and let $\mathbf{T}$ be the associated k-th order tensor. Then with overwhelming probability*

$$\mathrm{Opt}_{\boldsymbol{\mathcal{I}}} \leqslant \frac{1}{2} + (1+o(1))\left(\binom{n}{k}\cdot p\right)^{-1} \cdot \sum_{\alpha\in[n]^k} \mathbf{T}_\alpha x^\alpha \,.$$

Throughout the chapter we assume k to be an *odd* integer as for the even case sharp refutation algorithms are already known (e.g see [AOW15]).

A random k-XOR instance $\boldsymbol{\mathcal{I}}$ with n variables and exactly m clauses can be generated by picking m times a clause at random out of the $\binom{n}{k}$ possible k-XOR-clause. It is possible to show that a refutation algorithm for $\boldsymbol{\mathcal{I}} \sim \mathcal{F}_{k\text{-XOR}(n,p)}$ can also be used for refutation of k-XOR instances sampled through this second process. For this reason, we blur the distinction between these two processes. We direct the reader interested in a formal reduction to [AOW15] (Appendix D).

CSPs. Given a predicate $P : \{-1,1\}^k \to \{0,1\}$, an instance $\mathcal{I}$ of the CSP(P) problem over variables $x_1,\dots,x_n$ is a multi-set of pairs (c,α) representing constraints of the form $P(c\circ x^\alpha) := P(c_1 x^{\alpha(1)},\dots,c_k x^{\alpha(k)}) = 1$ where $\alpha \in [n]^k$ is the scope and $c \in \{\pm 1\}^k$ is the negation pattern. We can represent the predicate P as a multi-linear polynomial of degree k in indeterminates $c_1 x^{\alpha(1)},\dots,c_k x^{\alpha(k)}$,

$$P(c\circ x^\alpha) = \sum_{d\leqslant k} P_d(c\circ x^\alpha)\,,$$

where P_d denotes the degree d part of the predicate. In particular $P_0 := P_0(c\circ x^\alpha)$ denotes the constant part of the polynomial, which does not depend on the assignment.

The fraction of all possible assignments that satisfy P is given by $\mathbb{E}_{\mathbf{z}\overset{u.a.r}{\sim}\{\pm1\}^k}[P(\mathbf{z})]$. For any assignment $x\in\{\pm1\}^n$ and an instance $\mathcal{I}$ over m constraints we have

$$\mathrm{Val}_{\mathcal{I}}(x) = \frac{1}{m}\sum_{(c,\alpha)\in\mathcal{I}} P(c\circ x^\alpha)$$
$$\text{and}\quad \mathrm{Opt}_{\mathcal{I}} = \max_{x\in\{\pm1\}^n} \mathrm{Val}_{\mathcal{I}}(x)\,.$$

A *random* CSP(P) instance $\boldsymbol{\mathcal{I}}$ with $(1+o(1))m = p\cdot 2^k\cdot n^k$ constraints can be generated as follows:

(i) Pick independently with probability p each pair $(\mathbf{c},\boldsymbol{\alpha})$ where $\mathbf{c}$ is a random negation pattern from $\{-1,+1\}^k$ and α is a multi-index from $[n]^k$,

(ii) For each such pair $(\mathbf{c},\boldsymbol{\alpha})$ add the constraint $P(\mathbf{c}\circ x^{\boldsymbol{\alpha}}) = 1$ to $\boldsymbol{\mathcal{I}}$.

Notice that we do not rule out predicates with same multi-index but different negation pattern as multi-indices in which an index appears multple time. We also do not assume P to be symmetric. We denote such distribution by $\mathcal{F}_{\mathrm{CSP(P)}}(n,p)$.

As in the case of k-XOR a random CSP(P) instance $\mathcal{I}$ with n variables and exactly m clauses can be generated by picking m times a clause and a negation pattern at random. Again it is possible to show that a refutation algorithm for $\mathcal{I} \sim \mathcal{F}_{\mathrm{CSP(P)}}(n, p)$ can also be used for refutation of instances sampled through this second process (see Appendix D in [AOW15]).

Refutation and certification. We say that $\mathcal{A}$ is a *δ-refutation algorithm* for random CSP(P) if $\mathcal{A}$ has the following properties:

(i) on all instances $\mathcal{I}$ the output of $\mathcal{A}$ si either $\mathrm{Opt}_{\mathcal{I}} \leqslant 1 - \delta$ or "fail",

(ii) if $\mathrm{Opt}_{\mathcal{I}} > 1 - \delta$ then $\mathcal{A}$ *never* outputs $\mathrm{Opt}_{\mathcal{I}} \leqslant 1 - \delta$.

More generally, for an set of possible inputs $\mathcal{S}$ and a property p over instances in $\mathcal{S}$, we say that an algorithm $\mathcal{A}$ *certifies* p if:

(i) on all inputs $\mathcal{I} \in \mathcal{S}$ the output of $\mathcal{A}$ is either "$\mathcal{I}$ satisfies p" or "fail",

(ii) if $\mathcal{I} \in \mathcal{S}$ does not satisfy p then $\mathcal{A}$ *never* outputs "$\mathcal{I}$ satisfies p".

In the context of random CSP(P) (and hence k-XOR), a ***strong refutation*** is a δ-refutation for $1 - \delta \leqslant \mathbb{E}_{\mathbf{x} \overset{u.a.r}{\sim} \{\pm 1\}^k}[P(\mathbf{x})] + o(1)$.

6.3 A generalized Ihara-Bass formula

In this section we present an extension of the Ihara-Bass theorem (see [HST06] and references therein) to arbitrary real symmetric matrices. We remark that our extension differs from the one in [FM17].

Throughout the section we assume to be given a *symmetric* matrix $A \in \mathbb{R}^{n \times n}$ with $2m$ non-zero entries and zeroed diagonal. We use the following notation. We will use letters u, v to denote indices in $[n]$ and e, f for indices in $[2m]$. We conveniently think of A as the adjacency matrix of a weighted undirected graph G with n vertices and $2m$ oriented edges. Then $uv \in E(G)$ if $A_{uv} \neq 0$, moreover then the inverse edge vu is also in $E(G)$ since $A_{uv} = A_{vu}$ by definition. Recall for an edge $e \in E(G)$ we write e^{-1} for its inverse and for a vertex $v \in V(G)$ we write $N^+(v)$ (respectively $N^-(v)$) for its set of outgoing (resp. incoming) oriented edges in G. We write $\sigma_{uv} = \mathrm{sign}(A_{uv})$. To reason about the spectrum of A, we introduce several matrices: the diagonal matrices

$$D(A) \in \mathbb{R}^{n \times n}, \quad \text{with } D_{uv}(A) = \begin{cases} \sum_w |A_{uw}| & u = v \\ 0 & \text{otherwise.} \end{cases}$$

$$Q(A) \in \mathbb{R}^{m \times m}, \quad \text{with } Q_{ef}(A) = \begin{cases} |A_e| & e = f \\ 0 & \text{otherwise.} \end{cases}$$

the block matrices

$$J(A) = \begin{pmatrix} 0 & \mathrm{Id}_m \\ \mathrm{Id}_m & 0 \end{pmatrix} \in \mathbb{R}^{2m \times 2m}$$
$$L(A) = \begin{pmatrix} 0 & Q(A) \\ Q(A) & 0 \end{pmatrix} \in \mathbb{R}^{2m \times 2m}$$

and the source, target and non-backtracking matrices

$$S(A) \in \mathbb{R}^{n \times 2m}, \quad \text{with } S_{ue}(A) = \begin{cases} \sigma_{uv}\sqrt{|A_{uv}|} & \text{if } u \text{ is the source of } e = uv \text{ and } u < v \\ \sqrt{|A_{uv}|} & \text{if } u \text{ is the source of } e = uv \text{ and } u > v \\ 0 & \text{otherwise.} \end{cases}$$

$$T(A) \in \mathbb{R}^{n \times 2m}, \quad \text{with } T_{ue}(A) = \begin{cases} \sigma_{uv}\sqrt{|A_{uv}|} & \text{if } u \text{ is the target of } e = vu \text{ and } u < v \\ \sqrt{|A_{uv}|} & \text{if } u \text{ is the target of } e = vu \text{ and } u > v \\ 0 & \text{otherwise.} \end{cases}$$

$$B(A) \in \mathbb{R}^{2m \times 2m}, \quad \text{with } B_{ef}(A) = \begin{cases} \sigma_e\sigma_f\sqrt{|A_eA_f|} & \text{if } ef \text{ is a non-backtracking walk} \\ & e = uv, \text{ f=}vw \text{ and } v < u, w \\ \sigma_e\sqrt{|A_eA_f|} & \text{if } ef \text{ is a non-backtracking walk} \\ & e = uv, \text{ f=}vw \text{ and } w < v < u \\ \sigma_f\sqrt{|A_eA_f|} & \text{if } ef \text{ is a non-backtracking walk} \\ & e = uv, \text{ f=}vw \text{ and } u < v < w \\ \sqrt{|A_eA_f|} & \\ & \text{if } ef \text{ is a non-backtracking walk} \\ & e = uv, \text{ f=}vw \text{ and } v > u, w \\ 0 & \text{otherwise.} \end{cases}$$

When the context is clear we simply write B for $B(A)$ (analogously for the other matrices). To gain intuition on these linear maps, it is instructive to consider the case when A is the adjacency matrix of an unweighted graph G. Then D is the degree diagonal matrix with $D_{uu} = deg_G(u)$, $L = J$ and B corresponds to the non-backtracking matrix of G.

Throughout the other sections of the chapter, for a given non-backtracking matrix $B \in \mathbb{R}^{2m \times 2m}$, we will consider the related extension matrix $B^* \in \mathbb{R}^{2n^2 \times 2n^2}$ with entries

$$B^*_{ef} = \begin{cases} B_{ef} & \text{if } e, f \in E(G) \\ 0 & \text{otherwise.} \end{cases}$$

For simplicity of the notation, we will often denote B^* simply by B. The context will always be clarified by the ambient dimension. We can now state the main result of the section.

Theorem 6.6 (Generalized Ihara-Bass Theorem)**.** *Let n, m be integers and let $A \in \mathbb{R}^{n\times n}$ be a symmetric matrix with m non-zero entries, all off-diagonal. Let B, L, J, D defined as above. Then, for any $u \in \mathbb{R}$,*

$$\det(\mathrm{Id}_{2m} - u(B + L - J)) = (1 - u^2)^{m-n}\left(\mathrm{Id}_n - uA + u^2 D - u^2 \mathrm{Id}_n\right).$$

Our proof of Theorem 6.6 closely resembles the proof of Bass [Bas92]. We first observe that the matrices above satisfy several useful identities, than tackle the theorem.

Lemma 6.7. *Using the definitions above:*

i) $SJ = T$ *and* $TJ = S$,

ii) $A = ST^{\mathsf{T}}$,

iii) $D = SS^{\mathsf{T}} = TT^{\mathsf{T}}$,

iv) $B + L = T^{\mathsf{T}}S$.

Proof. For i), notice that $SJ \in \mathbb{R}^{n\times 2m}$ and $SJ_{ue} = \langle S_{u,-}, J_{-,e}\rangle = S_{ue^{-1}} = T_{ue}$, where in the third step we used symmetry of A. A similar argument can be made to show $TJ = S$. For ii) observe that

$$A_{uv} = \langle S_{u,-}, T_{v,-}\rangle = \sum_e S_{ue}T_{ve}$$

which is nonzero only when $e = uv$. In that case, by definition $A_{uv} = \sigma_{uv}|A_{uv}| = S_{ue}T_{ve}$ since either $u < v$ or $u > v$. Consider now SS^{T}, the matrix is diagonal since each edge has at most one source vertex, then

$$(SS^{\mathsf{T}})_{uu} = \sum_e S_{ue}^2 = \sum_{v\in N^+(u)} |A_{uv}| = D_{uu}\,.$$

A symmetric derivation shows $D_{uu} = (TT^{\mathsf{T}})_{uu}$. It remains to prove iv). It is trivial to check that

$$(T^{\mathsf{T}}S)_{ee} = \langle T_{-,e}, S_{-,e}\rangle = \sum_u T_{ue}S_{ue} = 0\,,$$

since there are no self-loops in the graph. For distinct $e, f \in [2m]$

$$(T^{\mathsf{T}}S)_{ef} = \sum_u T_{ue}S_{uf}\,.$$

There is at most one non-zero element in the sum, corresponding to the case when u is the target vertex of e and the source of f, which means ef is a walk of length 2 in G. If ef is a non-backtracking walk (that is, $e \neq f^{-1}$) then $B_{ef} = (T^{\mathsf{T}}S)_{ef}$ and $L_{ef} = 0$. Conversely, if $e = f^{-1}$ then $B_{ef} = 0$ and $L_{ef} = (T^{\mathsf{T}}S)_{ef}$. Finally, signs can be checked case by case. □

We are now ready to prove Theorem 6.6.

Proof of Theorem 6.6. In the following identities all matrices are $(n+2m)\times(n+2m)$ block matrices where the first block has size $n\times n$. Let $u\in\mathbb{R}$,

$$\begin{aligned}&\begin{pmatrix}\mathrm{Id}_n & 0\\ T^\mathsf{T} & \mathrm{Id}_{2m}\end{pmatrix}\begin{pmatrix}\mathrm{Id}_n(1-u^2) & Su\\ 0 & \mathrm{Id}_{2m}-(B+L-J)u\end{pmatrix}\\ &=\begin{pmatrix}\mathrm{Id}(1-u^2) & Su\\ T^\mathsf{T}(1-u^2) & T^\mathsf{T}Su+\mathrm{Id}_{2m}-(B+L-J)u\end{pmatrix}\\ &=\begin{pmatrix}\mathrm{Id}(1-u^2) & Su\\ T^\mathsf{T}(1-u^2) & \mathrm{Id}_{2m}+Ju\end{pmatrix}.\end{aligned}\tag{6.3.1}$$

On the other hand

$$\begin{aligned}&\begin{pmatrix}\mathrm{Id}_n(1-u^2)-Au+Du^2 & Su\\ 0 & \mathrm{Id}_{2m}+Ju\end{pmatrix}\begin{pmatrix}\mathrm{Id}_n & 0\\ T^\mathsf{T}-S^\mathsf{T}u & \mathrm{Id}_{2m}\end{pmatrix}\\ &=\begin{pmatrix}\mathrm{Id}_n(1-u^2)-Au+Du^2+ST^\mathsf{T}u-SS^\mathsf{T}u^2 & Su\\ T^\mathsf{T}-S^\mathsf{T}u+JT^\mathsf{T}u-JS^\mathsf{T}u^2 & \mathrm{Id}_{2m}+Ju\end{pmatrix}\\ &=\begin{pmatrix}\mathrm{Id}_n(1-u^2) & Su\\ T^\mathsf{T}(1-u^2) & \mathrm{Id}_{2m}+Ju\end{pmatrix}.\end{aligned}\tag{6.3.2}$$

Putting Eq. (6.3.1) and Eq. (6.3.2) together and taking determinants we get

$$(1-u^2)^n\det(\mathrm{Id}_{2m}-(B+L-J)u)=\det\big(\mathrm{Id}_n(1-u^2)-Au+Du^2\big)\det(\mathrm{Id}_{2m}+Ju)\,.$$

Now notice that

$$\mathrm{Id}_{2m}+Ju=\begin{pmatrix}\mathrm{Id}_m & \mathrm{Id}_m u\\ \mathrm{Id}_m u & \mathrm{Id}_m\end{pmatrix}$$

and thus $\det(\mathrm{Id}_{2m}+Ju)=(1-u^2)^m$. Rearranging, the result follows. □

6.3.1 Norm bounds via the Ihara-Bass formula

In this section we show how Theorem 6.6 can be used to study the spectrum of a real symmetric matrix A via the spectrum of related matrices. The central tool is the theorem below.

Theorem 6.8. *Let $A\in\mathbb{R}^{n\times n}$ a symmetric matrix with zero diagonal. Let B, L, J, D be as defined in Section 6.3. Let $\lambda_{\min}$ be the smallest eigenvalue of the matrix $B+L-J\in\mathbb{R}^{2m\times 2m}$. Then for any $\lambda\leqslant\lambda_{\min}$*

$$A\succeq-|\lambda|\mathrm{Id}_n-|\lambda|^{-1}(D-\mathrm{Id}_n)\,.$$

Proof. Let $\lambda_{\min}$ be the smallest real eigenvalue of $B + L - J$. By Theorem 6.6 we know -1 is a real eigenvalue of $B + L - J$ and thus $\lambda_{\min} \leqslant -1$. Moreover, for every $\lambda < \lambda_{\min}$ we have $\det\big(\mathrm{Id}_{2m} - \lambda^{-1}B + \lambda^{-1}L - \lambda^{-1}J\big) \neq 0$ otherwise λ would be an eigenvalue smaller than $\lambda_{\min}$. Define the matrix

$$M_\lambda := \mathrm{Id}_n - \lambda^{-1}A + \lambda^{-2}(D - \mathrm{Id}_n)\,.$$

By the same reasoning as in Theorem 6.6, $\det(M_\lambda) \neq 0$ as long as $\lambda < \lambda_{\min}$. We make the stronger claim

$$\forall \lambda,\, \lambda_{\min} \,:\, M_\lambda \succ \mathbf{0}\,.$$

To prove the above claim, suppose toward a contradiction that $\lambda' < \lambda_{\min}$ is such that $M_{\lambda'}$ has a negative eigenvalue. Since M_λ tends to Id_n when $\lambda \to -\infty$, there is a value $\lambda_{PD} < \lambda'$ such that $M_{\lambda_{PD}}$ is strictly positive definite. Consider now the smallest eigenvalue of M_λ for values of λ in the range (λ_{PD}, λ'). The smallest eigenvalue of M_λ varies continuously with λ, it is positive for $\lambda = \lambda_{PD}$ and it is negative for $\lambda = \lambda'$, so it must be equal to zero for some $\lambda^* \leqslant \lambda' < \lambda_{\min}$. But this means that $\det(M_{\lambda^*}) = 0$ and so λ^* is an eigenvalue of $B + L - J$, which contradicts the definition of $\lambda_{\min}$. We have thus established our claim. Rearranging the result follows. □

A crucial consequence of Theorem 6.8 is that, exploiting the diagonal structure of the matrices D , Id_n one can bound the norm $\|A\|_{\infty\to 1}$ as a function of the smallest eigenvalue of the associated non-backtracking matrix.

Corollary 6.9. *Let $A \in \mathbb{R}^{n\times n}$ a symmetric matrix with zero diagonal. Let $\lambda_{\min}$ and $\lambda'_{\min}$ be respectively the smallest eigenvalue of the matrix $B(A)+L(A)-J(A) \in \mathbb{R}^{2m\times 2m}$ and $B(-A)+L(A)-J(A) \in \mathbb{R}^{2m\times 2m}$, for B , L , J , D as defined in Section 6.3. Then, for any $\lambda \geqslant \max\big\{|\lambda_{\min}|\,, \big|\lambda'_{\min}\big|\big\}$,*

$$\|A\|_{\infty\to 1} \leqslant 2\,\mathrm{Tr}\Big|\big(\lambda \mathrm{Id}_n + \lambda^{-1}(D(A) - \mathrm{Id}_n)\big)\Big|$$

Proof. Define

$$R := \big|\lambda \mathrm{Id}_n + \lambda^{-1}(D(A) - \mathrm{Id}_n)\big|\,.$$

By Theorem 6.8 for any $x \in \{\pm 1\}^n$ we have $\big|x^\mathsf{T}Ax\big| \leqslant \big|x^\mathsf{T}Rx\big|$. For any $y \in \{\pm 1\}^n$ we can write

$$\begin{aligned} 2\big|x^\mathsf{T}Ay\big| &\leqslant \big|(x+y)^\mathsf{T}A(x+y) - x^\mathsf{T}Ax - y^\mathsf{T}Ay\big| \\ &\leqslant \big|(x+y)^\mathsf{T}A(x+y)\big| + \big|x^\mathsf{T}Ax\big| + \big|y^\mathsf{T}Ay\big|\,. \end{aligned}$$

Now $x + y \in \{-2, 0, +2\}^n$ and thus

$$\big|(x+y)^\mathsf{T}A(x+y)\big| \leqslant 4 \max_{z\in\{\pm 1\}^n} z^\mathsf{T}Rz\,,$$

the result follows by definition of R. □

6.4 Warm-up: spectrum of binary matrices with dependencies

In preparation to a proof of Theorem 6.1, we show here how to use the ideas of Section 6.3.1 to study the spectrum of random symmetric matrices with entries in $\{-1\,,0\,,1\}$, even when dependencies between the entries appear. As we may view any symmetric matrix as the adjacency matrix of a weighted graph (up to the diagonal entries) we partially shift to the language of graphs. In particular, we study graphs sampled from the following family of distributions.

Definition 6.10 (γ-wise independent random binary-weighted graphs)**.** $\mathcal{D}_{d,\gamma}$ is the family of distributions $P_{d,\gamma}$ over graphs with vertex set $[n]$ and adjacency matrix satisfying:

1. for all $uv \in E(K_n)$, $\mathbb{P}_{\mathbf{G}\sim P_{d,\gamma}}[A_{uv}(\mathbf{G}) \neq 0] = \mathbb{P}_{\mathbf{G}\sim P_{d,\gamma}}[uv \in E(\mathbf{G})] = d/n$.
2. for all $uv \in E(K_n)$, the distribution of $A_{uv}(\mathbf{G})$ is symmetric with support $\{-1\,,0\,,+1\}$.
3. edges in $\mathbf{G}$ are γ-wise independent.

Notice that the adjacency matrix of $\mathbf{G}$ is a random binary symmetric matrix with γ-wise independent entries (up to symmetries). With a slight abuse of notation, we write $\mathbf{G} \sim P_{d,\gamma}$ and $A(\mathbf{G}) \sim P_{d,\gamma}$, respectively for a graph sampled from $P_{d,\gamma}$ in $\mathcal{D}_{d,\gamma}$ and for the adjacency matrix of a graph sampled from this distribution. We consider the associated matrices $D(\mathbf{A})\,, B(\mathbf{A})$ as defined in Section 6.3. We prove the following theorem.

Theorem 6.11. *Let n be an integer, $d > 0$ and $\gamma \geqslant C \log_d^2 n$ for a large enough universal constant $C > 0$. Consider a distribution $P_{d,\gamma} \in \mathcal{D}_{d,\gamma}$. Then for $\mathbf{A} \sim P_{d,\gamma}$*

$$\|\mathbf{A}\|_{\infty\to 1} \leqslant O\left(n\sqrt{d}\right),$$

with probability 0.99*.*

Observe that Theorem 6.11 shows how the adjacency matrix of a graph sampled from some distribution in $\mathcal{D}_{d,\gamma}$ behaves –up to a universal constant– as the adjacency matrix of an Erdős-Rényi graph with expected degree d. As we are oblivious to the specific correlations in the graph, the result only relies on local independence of the edges. The central tool behind Theorem 6.11 is the following lemma.

Lemma 6.12. *Consider the settings of Theorem 6.11. Let $z \leqslant \frac{\log_d n}{10}$. For $\mathbf{A} \sim P_{d,\gamma}$, the associated non-backtracking matrix satisfies*

$$\left\|B^{z-1}(\mathbf{A})\right\| \leqslant O(d^{z/2}) \cdot (\log n)^{O(\log\log n)}\,.$$

with probability $1 - o(1)$*.*

The choice of the non-backtracking matrix $B^{z-1}(\mathbf{A})$ stems from the observation that is contains more structural information about the graph, compared to $B(\mathbf{A})$. Indeed, the singular values of $B(\mathbf{A})$ contain only information on the degree sequence of the underlying graph. Another perspective on this can be obtained recalling that for any natural norm $\|\cdot\|_*$, by Gelfand's formula Fact 2.1, $\rho(B(\mathbf{A})) = \lim_{z\to\infty}\|B(\mathbf{A})^z\|_*^{1/z}$. Hence one can expect that a careful analysis on non-backtracking matrices of sufficiently large length, would yield a tighter bound on the spectral radius (see [BLM15] for a more in-depth discussion). Indeed, by Fact 2.1, a consequence of Lemma 6.12 is the following result regarding the spectral radius of $B(\mathbf{A})$.

Corollary 6.13. *Consider the settings of Theorem 6.11. For $\mathbf{A} \sim P_{d,\gamma}$, the associated non-backtracking matrix satisfies*

$$\rho(B(\mathbf{A})) \leqslant O(\sqrt{d})\,.$$

with probability $1 - o(1)$.

We can combine Corollary 6.13 with Corollary 6.9 to obtain the theorem.

Proof of Theorem 6.11. By Corollary 6.13 for both $\mathbf{A}$ and $-\mathbf{A}$ we get $\rho(B(\mathbf{A})) \leqslant O(\sqrt{d})$ and $\rho(B(-\mathbf{A})) \leqslant O(\sqrt{d})$. By Lemma D.2 we have $\operatorname{Tr} D(\mathbf{A}) \leqslant O(nd)$ with probability at least 0.999, where $D(\mathbf{A})$ is the associated degree matrix as defined in Section 6.3. Finally applying Corollary 6.9 the result follows. □

The rest of the section is devoted to the proof of Lemma 6.12. We will prove the result via the trace method.

Proposition 6.14. *Let $\mathbf{M}$ be an n-by-n random matrix and $c > 0$. Then*

$$\mathop{\mathbb{E}}_{\mathbf{M}}\left[\operatorname{Tr}\left(\mathbf{M}\mathbf{M}^{\mathsf{T}}\right)^q\right] \leqslant \gamma \Longrightarrow \mathbb{P}\left(\|\mathbf{M}\| \geqslant c\cdot\gamma^{1/2q}\right) \leqslant c^{-2q}\,.$$

In Section 6.4.1 we illustrate how the trace computation can be reduced to a path counting problem. In Section 6.4.2 we compute the expectation of such paths. Finally we put things together in Section 6.4.3.

6.4.1 Powers of non-backtracking matrices

For simplicity, we will write B in place of $B(A)$ for an adjacency matrix A. Moreover we will consider the $2n^2$-by-$2n^2$ extension of $B(A)$ as described in Section 6.3. Indeed this extension has the same eigenvalues and singular values (up to some additional zero-value ones). We will overload our notation and simply denote it by B. We now explain how to reduce the computation of the trace of powers of non-backtracking walks to a graph counting argument. We start by introducing additional notions.

For a graph G and $e, f \in \vec{E}(G)$, we write NBW^z_{ef} to denote the set of length z non-backtracking walks in G starting with e and ending with f. Then $\mathrm{NBW}^z(G) = \bigcup_{e,f\in\vec{E}(G)} \mathrm{NBW}^z_{ef}(G)$. For simplicity we let $\mathrm{NBW}^z_{ef} := \mathrm{NBW}^z_{ef}(K_n)$. A walk $W \in \mathrm{NBW}^{2q,z}$ over e distinct edges is said *t-tangle*-free if the number of vertices in the walk is at least $e - t + 1$. That is, any minimum spanning tree of the subgraph traversed by the walk has $e - t$ edges. For $t = 0$ the definition implies the walk is a path. We remark that this definition differs from the one in [BLM15] (which corresponds to $t < 2$). We adopt this different definition in preparation of our arguments in Section 6.5. A graph G is said *t-tangle*-free if every walk in $\mathrm{NBW}^{2q,z}(G)$ is t-tangle-free. We use $\mathrm{TGF}^{z,t}_{ef}(G) \subseteq \mathrm{NBW}^z_{ef}(G)$ to denote the subset of t-tangle-free walks. When the value of t is clear from context we will simply say W is tangle-free and write $\mathrm{TGF}^z_{ef}(G) \subseteq \mathrm{NBW}^z_{ef}(G)$ for the corresponding set.

We can extend the notion of closed non-backtracking matrices by introducing closed block non-backtracking matrices. $\mathrm{BNBW}^{q,z}_e$ is the set of walks W of length $q \cdot z$ with starting edge $e \in \vec{E}(G)$ and ending edge $e^{-1} \in \vec{E}(G)$ satisfying:

(i) W can be partitioned into q non-backtracking walks $W_1, \ldots, W_q$ of length z.

(ii) For each pair of walks W_i , W_{i+1} (with the convention $W_{q+1} = W_1$) the starting edge of W_{i+1} is the inverse of the ending edge of W_i. I.e. we have $e_z(W_i) = e_1^{-1}(W_{i+1})$.

Similarly, $\mathrm{BTGF}^{q,z,t}_e \subseteq \mathrm{BNBW}^{q,z}_e$ is the set of closed block non-backtracking walks such that each block is t-tangle-free. Given $W \in \mathrm{BNBW}^{q,z}_e$, we use $W_i \in \mathrm{NBW}^z$ to denote the i-th block of W (where the first block maybe chosen arbitrarily among the ones starting at e). For $uv \in \vec{E}(G)$, recall we write $\sigma_{uv} = \mathrm{sign}(A_{uv})$, we further let

$$\tilde{\sigma}_{uv} = \begin{cases} \sigma_{uv} & \text{if } u < v\,, \\ 1 & \text{otherwise.} \end{cases}$$

This notation will be used to describe the k-th power of the non-backtracking matrix.

Fact 6.15. *Let G be a weighted graph with m edges. Let $A \in \mathbb{R}^{n\times n}$ be its adjacency matrix and $B \in \mathbb{R}^{2n^2\times n^2}$ the associated non-backtracking matrix. Let z be a positive integer and let $e, f \in \vec{E}(G)$. Then for $e_1 = e$, $e_z = f$*

$$(B^{z-1})_{ef} = \sum_{W\in NBW^z_{ef}} \tilde{\sigma}_{e_1^{-1}} \cdot \tilde{\sigma}_{e_z} \cdot \sqrt{|A_{e_1} \cdot A_{e_z}|} \cdot \prod_{s=2}^{z-1} A_{e_s}\,.$$

Fact 6.15 can be checked simply by expanding B^{z-1}. We can now define the tangle-free $2n^2$-by-$2n^2$ non-backtracking matrix $B^{(z-1)_t}$ associated to B^{z-1} (we will drop the subscript t when the context is clear). For $e, f \in [2n^2]$,

$$(B^{(z-1)_t})_{ef} := \sum_{W\in \mathrm{TGF}^{z,t}_{ef}} \tilde{\sigma}_{e_1^{-1}} \cdot \tilde{\sigma}_{e_z} \cdot \sqrt{|A_{e_1} \cdot A_{e_z}|} \cdot \prod_{s=2}^{z-1} A_{e_s}\,.$$

It will be convenient to decompose the terms $(B^{z-1})_{ef}$ as follows:

$$\begin{aligned}
(B^{z-1})_{ef} &= \sum_{W \in \mathrm{NBW}^z_{ef}} \tilde{\sigma}_{e_1^{-1}} \cdot \tilde{\sigma}_{e_z} \sqrt{|A_{e_1} \cdot A_{e_z}|} \cdot \prod_{s=1}^{z-1} A_{e_s} \\
&= \sum_{W \in \mathrm{TGF}^{z,t}_{ef}} \tilde{\sigma}_{e_1^{-1}} \cdot \tilde{\sigma}_{e_z} \sqrt{|A_{e_1} \cdot A_{e_z}|} \cdot \prod_{s=2}^{z-1} A_{e_s} \\
&\quad + \sum_{W \in \mathrm{NBW}^z_{ef} \setminus \mathrm{TGF}^{z,t}_{ef}} \tilde{\sigma}_{e_1^{-1}} \cdot \tilde{\sigma}_{e_z} \sqrt{|A_{e_1} \cdot A_{e_z}|} \cdot \prod_{s=2}^{z-1} A_{e_s} \\
&= (B^{(z-1)_t})_{ef} \\
&\quad + \sum_{W \in \mathrm{NBW}^z_{ef} \setminus \mathrm{TGF}^{z,t}_{ef}} \tilde{\sigma}_{e_1^{-1}} \cdot \tilde{\sigma}_{e_z} \sqrt{|A_{e_1} \cdot A_{e_z}|} \cdot \prod_{s=2}^{z-1} A_{e_s} \,.
\end{aligned}$$

Notice that for a t-tangle-free graph, $B^{z-1} = B^{(z-1)_t}$. The trace of powers of B^{z-1} can also be written as a sum over block non-backtracking walks. In particular, we can obtain the fact below observing that even though B is not a normal matrix, it has some symmetry as

$$\left[\left(B^{(z-1}\right)^{\mathsf{T}}\right]_{ef} = (B^z)_{fe} = (B^z)_{e^{-1}f^{-1}} \,.$$

Conveniently, the factors $\tilde{\sigma}_{e_1^{-1}} \cdot \tilde{\sigma}_{e_z}$ disappear.

Fact 6.16. *Let G be a weighted graph with m edges. Let $A \in \mathbb{R}^{n\times n}$ be its adjacency matrix and $B \in \mathbb{R}^{2n^2\times 2n^2}$ the associated non-backtracking matrix. Let $q, z \geqslant 2$ and t be integers. Then*

$$\mathrm{Tr}\left[B^{z-1}\left(B^{z-1}\right)^{\mathsf{T}}\right]^q = \sum_{W \in BNBW^{2q,z}} \prod_{i=1}^{2q} \left[\sqrt{|A_{e_1(W_i)} \cdot A_{e_z(W_i)}|} \cdot \left(\prod_{s=2}^{z-1} A_{e_s(W_i)}\right)\right]$$

and

$$\mathrm{Tr}\left[B^{(z-1)_t}\left(B^{(z-1)_t}\right)^{\mathsf{T}}\right]^q = \sum_{W \in BTGF^{2q,z,t}} \prod_{i=1}^{2q} \left[\sqrt{|A_{e_1(W_i)} \cdot A_{e_z(W_i)}|} \cdot \left(\prod_{s=2}^{z-1} A_{e_s(W_i)}\right)\right] .$$

6.4.2 Expectation of block non-backtracking walks

By Proposition 6.14 and Fact 6.16 we need to study the expectation of closed block non-backtracking walks under distributions in $\mathcal{D}_{d,\gamma}$. First we observe that, if an edge in a walk in $\mathrm{BNBW}^{2q,z}$ has multiplicity one, then by symmetry the expectation of the whole walk is zero.

Fact 6.17. *Let n be an integer, $d > 0$,,$0 < \gamma \leqslant n$. Consider a distribution $P_{d,\gamma} \in \mathcal{D}_{d,\gamma}$. Let $W \in BNBW^{2q,z}$ with $z > 1$ and $2q \cdot z \leqslant \gamma$. If there exists $e \in E(W)$ with multiplicity $m_W(e) = 1$, then*

$$\mathop{\mathbb{E}}_{\mathbf{A}\sim P_{d,\gamma}} \prod_{i=1}^{2q} \left[\sqrt{\left|\mathbf{A}_{e_1(W_i)} \cdot \mathbf{A}_{e_z(W_i)}\right|} \cdot \left(\prod_{s=2}^{z-1} \mathbf{A}_{e_s(W_i)} \right) \right] = 0 \,. \tag{6.4.1}$$

Proof. By assumption there must be $e \in E(W)$ with $m_W(e) = 1$ and by construction this cannot be any of the starting edges $e_{t\cdot z+1}$, with $0 \leqslant t \leqslant 2q-1$ in W. Let's denote this edge by $f \in \mathbb{E}(W)$. Then we may write

$$\begin{aligned}
&\mathop{\mathbb{E}}_{\mathbf{A}\sim P_{d,\gamma}} \prod_{i=1}^{2q} \left[\sqrt{\left|\mathbf{A}_{e_1(W_i)} \cdot \mathbf{A}_{e_z(W_i)}\right|} \cdot \left(\prod_{s=2}^{z-1} \mathbf{A}_{e_s(W_i)} \right) \right] \\
&\quad = \mathop{\mathbb{E}}_{\mathbf{A}\sim P_{d,\gamma}} \left[\prod_{i=1}^{2q} \left[\sqrt{\left|\mathbf{A}_{e_1(W_i)} \cdot \mathbf{A}_{e_z(W_i)}\right|} \cdot \left(\prod_{s=2}^{z-1} \mathbf{A}_{e_s(W_i)} \right) \right] \cdot \frac{1}{\mathbf{A}_f} \right] \cdot \mathop{\mathbb{E}}_{\mathbf{A}\sim P_{d,\gamma}} \mathbf{A}_f \,,
\end{aligned}$$

where we used γ-wise independence of the edges. By symmetry of the distribution $\mathbb{E}\,\mathbf{A}_f = 0$. The result follows. □

For any $t \geqslant 0$, we say that a tangle-free walk $W \in \mathrm{BTGF}^{2q,z,t}$ is *interesting* if there is no edge in W with multiplicity one. We denote the set of interesting walks in $\mathrm{BTGF}^{2q,z,t}$ by $\mathrm{IBTGF}^{2q,z,t}$. The next fact bounds the expectation of interesting closed block non-backtracking walks.

Fact 6.18. *Let n be an integer, $d > 0$, and $0 < \gamma \leqslant n$. Consider a distribution $P_{d,\gamma} \in \mathcal{D}_{d,\gamma}$. Let $W \in IBTGF^{2q,z,t}$ with $z > 1, t \geqslant 0$ and $2q \cdot z \leqslant \gamma$. Then*

$$\mathop{\mathbb{E}}_{\mathbf{A}\sim P_{d,\gamma}} \prod_{i=1}^{2q} \left[\sqrt{\left|\mathbf{A}_{e_1(W_i)} \cdot \mathbf{A}_{e_z(W_i)}\right|} \cdot \left(\prod_{s=2}^{z-1} \mathbf{A}_{e_s(W_i)} \right) \right] \leqslant \left(\frac{d}{n}\right)^{|E(W)|} .$$

Proof. By γ-wise independence,

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{A}\sim P_{d,\gamma}} \prod_{i=1}^{2q} \left[\sqrt{\left|\mathbf{A}_{e_1(W_i)} \cdot \mathbf{A}_{e_z(W_i)}\right|} \cdot \left(\prod_{s=2}^{z-1} \mathbf{A}_{e_s(W_i)} \right) \right] &\leqslant \prod_{e\in E(W)} \mathop{\mathbb{E}}_{\mathbf{A}\sim P_{d,\gamma}} |\mathbf{A}_e|^{m_W(e)} \\
&= \prod_{e\in E(W)} \mathop{\mathbb{E}}_{\mathbf{A}\sim P_{d,\gamma}} |\mathbf{A}_e| \\
&= \prod_{e\in E(W)} \left(\frac{d}{n}\right).
\end{aligned}$$

□

6.4.3 Bound on the spectrum of non-backtracking matrices

We are ready to tackle Lemma 6.12. Our proof consists of two main ingredients. First we shows that, for a wide range of parameters, graphs sampled from distributions in $\mathcal{D}_{d,\gamma}$ are tangle-free. Then we prove that the number of closed, tangle-free, block non-backtracking walks is not large. Combining these results with Fact 6.18 will conclude the proof.

Lemma 6.19. *Let n be a large enough integer. Let $d > 0$, $z \leqslant \log_d n \leqslant \gamma \leqslant n$. Consider a distribution $P_{d,\gamma} \in \mathcal{D}_{d,\gamma}$. Then for $t \geqslant 100 \log\log_d n$, $\mathbf{G} \sim P_{d,\gamma}$ is t-tangle-free with probability at least $1 - o(1)$.*

Proof. For $t' \geqslant t$, the number of non-backtracking walks over v vertices of length z with t'-tangles is upper bounded by $n^v v^z$. Each such walk W appears in the graph with probability $\left(\frac{d}{n}\right)^e \leqslant \left(\frac{d}{n}\right)^{t'+v-1}$. For $t' \geqslant t$, by union bound the probability that such a walk appears in the graph is at most $(vd)^{\log_d n} n^{-t+1} = o(1)$. □

The fact that our graphs of interest are with high probability $O(\log\log_d n)$-tangle-free means we need only to focus on this small subset of closed block non-backtracking walks. For the rest of the section we will fix

$$t = 100 \log\log_d n$$

and drop the superscript t. We will refer to $O(\log\log n)$-tangle-free walks simply by tangle-free walks and write $\mathbf{B}^{(z-1)}$ in place of $\mathbf{B}^{(z-1)_t}$. We say that a walk $W \in \mathrm{IBTGF}^{2q,z}$ over v vertices is *canonical* if its set of vertices is $[v] \subseteq [n]$ and the vertices are first visited in order. The use of canonical paths is convenient as, by construction, no two canonical paths are isomorphic (so we are in fact choosing an arbitrary element for each equivalence class). We denote the set of canonical paths by $\mathcal{W}^{2q,z} \subseteq \mathrm{IBTGF}^{2q,z}$. Notice that for $P \in \mathcal{W}^{2q,z}$, there are $\binom{n}{v}(v)!$ isomorphic walks in $\mathrm{IBTGF}^{2q,z}$.

Lemma 6.20 (Enumeration of canonical paths)**.** *Let $\mathcal{W}^{2q,z}(v,e)$ be the set of canonical paths with v vertices and e distinct edges. We have*

$$\left|\mathcal{W}^{2q,z}(v,e)\right| \leqslant z^{4tq} \cdot (2zq)^{6tq\cdot(e-v+1)}.$$

We defer the proof of Lemma 6.20 to Section 6.4 and use it here to prove the main lemma of the section.

Proof of Lemma 6.12. First notice that by Lemma 6.19, with high probability $\mathbf{B}^{z-1} = \mathbf{B}^{(z-1)}$, thus it suffices to bound the spectral norm of $\mathbf{B}^{(z-1)}$. Then

$$\begin{aligned}
&\mathbb{E}\left[\mathrm{Tr}\left[\mathbf{B}^{(z-1)}\left(\mathbf{B}^{(z-1)}\right)^{\mathsf{T}}\right]^q\right] \\
&\quad= \sum_{W\in\mathrm{IBTGF}^{2q,z}} \mathbb{E}\left[\prod_{i=1}^{2q}\left[\sqrt{\left|\mathbf{A}_{e_1(W_i)}\cdot\mathbf{A}_{e_z(W_i)}\right|}\cdot\left(\prod_{s=2}^{z-1}\mathbf{A}_{e_s(W_i)}\right)\right]\right]
\end{aligned}$$

$$\leqslant \sum_{W \in \mathrm{IBTGF}^{2q,z}} \mathbb{E} \prod_{i=1}^{2q} \left[\sqrt{\left| \mathbf{A}_{e_1(W_i)} \cdot \mathbf{A}_{e_z(W_i)} \right|} \cdot \left(\prod_{s=2}^{z-1} \mathbf{A}_{e_s(W_i)} \right) \right].$$

Now by Fact 6.18 and Lemma 6.20

$$\begin{aligned}
&\sum_{W \in \mathrm{IBTGF}^{2q,z}} \mathbb{E} \prod_{i=1}^{2q} \left[\sqrt{\left| \mathbf{A}_{e_1(W_i)} \cdot \mathbf{A}_{e_z(W_i)} \right|} \cdot \left(\prod_{s=2}^{z-1} \mathbf{A}_{e_s(W_i)} \right) \right] \\
&\qquad \leqslant \sum_{W \in \mathrm{IBTGF}^{2q,z}} \left(\frac{d}{n} \right)^{E(W)} \\
&\qquad \leqslant \sum_{v \geqslant 3}^{z \cdot q+1} \sum_{e=v-1}^{z \cdot q} \left| \mathcal{W}^{2q,z}(v,e) \right| \cdot \binom{n}{v} (v)! \cdot \left(\frac{d}{n} \right)^{e} \\
&\qquad \leqslant \sum_{v \geqslant 3}^{z \cdot q+1} \sum_{e=v-1}^{z \cdot q} (3 \cdot n)^{v} \cdot \left(2^{2t} z \right)^{2qt} \cdot (2zq)^{6tq \cdot (e-v+1)} \cdot \left(\frac{d}{n} \right)^{e} \\
&\qquad \leqslant \sum_{v \geqslant 3}^{z \cdot q+1} \sum_{e=v-1}^{z \cdot q} \left(2^{2t} z \right)^{2qt} \cdot d^{e} \cdot n \cdot \left(\frac{(6zq)^{6tq}}{n} \right)^{e-v+1} \\
&\qquad \leqslant \left(2^{2t} z \right)^{2qt} \cdot d^{zq} \cdot n \cdot \sum_{v \geqslant 3}^{z \cdot q+1} \sum_{e=v-1}^{z \cdot q} \left(\frac{(6zq)^{6tq}}{n} \right)^{e-v+1} \\
&\qquad \leqslant z^{O(qt)} \cdot d^{zq} \cdot n \cdot \sum_{v \geqslant 3}^{z \cdot q+1} \sum_{e=v-1}^{z \cdot q} \left(\frac{(6zq)^{6tq}}{n} \right)^{e-v+1}.
\end{aligned}$$

For $t = 100 \log \log_d n$, $q = \frac{\log n}{10^3 \log^2 \log_d n}$ and $z \leqslant \frac{\log_d n}{6}$ the series converges. Thus

$$\mathbb{E} \left[\mathrm{Tr} \left[\mathbf{B}^{(z-1)} \left(\mathbf{B}^{(z-1)} \right)^{\mathsf{T}} \right]^{q} \right] \leqslant O\left(z^{O(qt)} \cdot d^{zq} \cdot n \right)$$

Finally by Proposition 6.14 with probability at least $1 - o(1)$

$$\left\| \mathbf{B}^{z-1} \right\| \leqslant O(d^{z/2}) \cdot \left(z^{2t} \cdot n^{1/q} \right) \leqslant O(d^{z/2}) \cdot (\log n)^{\mathrm{poly}(\log \log n)},$$

concluding the proof. □

6.5 Strong refutations for random k-XOR

We prove here a result on strong refutation of random k-XOR instances.

Theorem 6.21. *Consider a random k-XOR instance $\boldsymbol{\mathcal{I}} \sim \mathcal{F}_{k\text{-XOR}(n,p)}$ for $k \geqslant 3$. For n large enough, there exists a universal constant $C > 0$ and a polynomial time algorithm that, if*

$$p \geqslant \frac{C \cdot n^{-k/2}}{\varepsilon^2}$$

certifies with probability at least 0.99

$$\mathrm{Opt}_{\boldsymbol{\mathcal{I}}} \leqslant \frac{1}{2} + O(\varepsilon)\,.$$

An efficient algorithm for k-XOR refutation was already known to hold for $p \geqslant \Omega\left(n^{-k/2}\log^{3/2} n\right)$ or when k is even [AOW15]. Thus we only need to design an algorithm for the settings $p \leqslant O\left(n^{-k/2}\,\mathrm{polylog}(n)\right)$ and k odd. We restrict our analysis to those. Following our discussion in Section 6.2.1 and using Proposition 6.5, the theorem can be directly obtained as a corollary to the result below.

Theorem 6.22 (Sharp bounds for random polynomials)**.** *Let $\mathbf{T}$ be a random tensor, with independent entries, such that $\mathbf{T}_\alpha = 0$ if the indices in the multi-index $\alpha \in [n]^k$ are not distinct and otherwise:*

$$\begin{aligned}\mathbb{E}[\mathbf{T}_\alpha] &= 0\\ \mathbb{P}[\mathbf{T}_\alpha \neq 0] &\leqslant p\\ \mathbb{P}[\Omega(1) \leqslant |\mathbf{T}_\alpha| \leqslant 1 \mid \mathbf{T}_\alpha \neq 0] &= 1\end{aligned}$$

For $k \geqslant 3$ and n large enough, there exists a universal constant $C > 0$ and a polynomial time algorithm that, if

$$C \cdot n^{-k/2} \leqslant p \leqslant n^{-k/2} \cdot O\left(\log^{10} n\right),$$

certifies with probability larger than 0.99

$$\max_{x\in\{\pm1\}^n} \sum_{\alpha\in[n]^k} \mathbf{T}_\alpha \cdot x^\alpha \leqslant O\left(\sqrt{p}\cdot n^{3k/4}\right).$$

Remark 6.23. Theorem 6.22 is not strictly about strong refutations. In fact, it states that the value of a random polynomial evaluated over the hypercube is concentrated around its expectation.

For the reminder of the section, we assume without loss of generality that $\mathbf{T}$ is symmetric. Indeed we may have this assumption without loss of generality by Fact D.6. For any assignment of $x \in \{\pm1\}^n$, by Cauchy-Schwarz,

$$\sum_{\alpha\in[n]^k} \mathbf{T}_\alpha \cdot x^\alpha \leqslant \left(\sum_{i\in[n]} x_i^2\right)^{1/2}\left(\sum_{\ell\in[n]}\left(\sum_{\alpha'\in[n]^{k-1}} \mathbf{T}_{(\alpha',\ell)}\cdot x^{\alpha'}\right)^2\right)^{1/2}$$

$$\leqslant \sqrt{n} \cdot \left(\sum_{\ell \in [n]} \left(\sum_{\alpha' \in [n]^{k-1}} \mathbf{T}_{(\alpha',\ell)} \cdot x^{\alpha'} \right)^2 \right)^{1/2} .$$

Consider the n^{k-1}-by-n^{k-1} matrix with entries, for $\alpha_1, \alpha_2, \beta_1, \beta_2 \in [n]^{(k-1)/2}$,

$$\mathbf{A}_{(\alpha,\beta),(\alpha',\beta')} := \sum_{\ell \in [n]} \mathbf{T}_{(\alpha\alpha'\ell)} \cdot \mathbf{T}_{(\beta\beta'\ell)} \tag{6.5.1}$$

we may use the rewriting

$$\begin{aligned} \sum_{\ell \in [n]} \left(\sum_{\alpha,\beta \in [n]^{k-1}} \mathbf{T}_{(\alpha,\beta,\ell)} x^{(\alpha,\beta)} \right)^2 &= \sum_{\ell \in [n]} \sum_{\alpha,\alpha',\beta,\beta' \in [n]^{k-1}} \mathbf{T}_{\alpha,\beta,\ell} \cdot \mathbf{T}_{\alpha',\beta',\ell} \cdot x^{(\alpha,\beta,\alpha',\beta')} \\ &= \langle x^{\otimes k-1}, \mathbf{A} x^{\otimes k-1} \rangle \\ &\leqslant \max_{z,y \in \{\pm 1\}^{n^{k-1}}} \langle z, \mathbf{A} y \rangle = \|\mathbf{A}\|_{\infty \to 1} . \end{aligned}$$

Thus we obtain

$$\max_{x \in \{\pm 1\}^n} \sum_{\alpha \in [n]^k} \mathbf{T}_\alpha x^\alpha \leqslant \sqrt{n \cdot \|\mathbf{A}\|_{\infty \to 1}} ,$$

which means that any algorithm computing an upper bound to $\|\mathbf{A}\|_{\infty \to 1}$ *immediately yields* a refutation algorithm for k-XOR. In particular, by Fact 2.15 we get that there is an algorithm (based on sum-of-squares) that certifies

$$\max_{x_1,\ldots,x_k \in \{\pm 1\}^n} \langle \mathbf{T}, x_1 \otimes \cdots \otimes x_k \rangle \leqslant \sqrt{K_G \cdot n \cdot \|\mathbf{A}\|_{\infty \to 1}} , \tag{6.5.2}$$

in time $O(n^{O(1)})$, where K_G is Grothendieck's constant. Now, Theorem 6.22 follows combining the above reasoning with the result below.

Lemma 6.24. *Consider the settings of Theorem 6.22. Consider the n^{k-1}-by-n^{k-1} matrix defined in Eq. (6.5.1).Then with probability at least* 0.998

$$\|\mathbf{A}\|_{\infty \to 1} \leqslant n^{k-1} \cdot O\left(p \cdot n^{k/2} \right) .$$

We may rewrite $\mathbf{A} = \mathbf{A}' + \mathbf{A}''$ where for any $\alpha_1, \alpha_2, \beta_1, \beta_2 \in [n]^{(k-1)/2}$

$$\mathbf{A}'_{(\alpha_1,\beta_1),(\alpha_2,\beta_2)} := \begin{cases} \mathbf{A}_{(\alpha_1,\beta_1),(\alpha_2,\beta_2)} & \text{if } \left| S(\alpha_1,\alpha_2) \cap S(\beta_1,\beta_2) \right| = 0 \\ 0 & \text{otherwise} \end{cases} \tag{6.5.3}$$

and $\mathbf{A}'' = \mathbf{A} - \mathbf{A}'$. This decomposition is convenient as, for example, now $\mathbb{E}\, \mathbf{A}'_{(\alpha_1,\beta_1),(\alpha_2,\beta_2)} = 0$ for all multi-indices $\alpha_1, \alpha_2, \beta_1, \beta_2$. By triangle inequality, to prove Lemma 6.24 it suffices to bound the norm $\|\cdot\|_{\infty \to 1}$ of both $\mathbf{A}'$ and $\mathbf{A}''$.

Lemma 6.25. *Consider the settings of Lemma 6.24. Let* $\mathbf{A}'$ *as defined in Eq. (6.5.3) and let* $\mathbf{A}'' = \mathbf{A} - \mathbf{A}'$. *Then with probability* $1 - o(1)$

$$\|\mathbf{A}''\|_{\infty\to 1} \leqslant n^{k-1-\Omega(1)} \cdot (k-1)! \cdot O(p \cdot n^{k/2}) .$$

The proof of Lemma 6.25 is straightforward as the required bound is very loose. We defer it to Appendix D.1. The next result bounds $\|\mathbf{A}'\|_{\infty\to 1}$, this is the main technical challenge of this work and the subsequent sections are dedicated to its proof.

Lemma 6.26. *Consider the settings of Lemma 6.24. Let* $\mathbf{A}'$ *as defined in Eq. (6.5.3) Then with probability at least* 0.999

$$\|\mathbf{A}'\|_{\infty\to 1} \leqslant n^{k-1} \cdot O(p \cdot n^{k/2}) .$$

6.5.1 Bounding the norm of A′

Using a reasoning similar in spirit to that shown in Section 6.4, we prove Lemma 6.26 studying the associated non-backtracking matrix.

Theorem 6.27. *Consider the settings of Lemma 6.26. Let* $\mathbf{B}, \mathbf{L}, \mathbf{J}$ *be the matrices associated to* $\mathbf{A}'$ *as defined in Section 6.3. Then with probability* $1 - o(1)$

$$\rho(\mathbf{B} + \mathbf{L} - \mathbf{J}) \leqslant O\left(p \cdot n^{k/2}\right) .$$

We can use Theorem 6.27 to prove Lemma 6.26. The argument closely resembles that used for Theorem 6.11

Proof of Lemma 6.26. By Theorem 6.27 for both $\mathbf{A}'$ and $-\mathbf{A}'$ we get $\rho(B(\mathbf{A}')) \leqslant O\left(p \cdot n^{k/2}\right)$ and $\rho(B(-\mathbf{A}')) \leqslant O\left(p \cdot n^{k/2}\right)$. By Lemma D.4 we have $\operatorname{Tr} D(\mathbf{A}') \leqslant O\left(p^2 \cdot n^{2k-1}\right)$ with probability at least $1 - 10^4$, where $D(\mathbf{A}')$ is the associated degree matrix as defined in Section 6.3. Finally applying Corollary 6.9 the result follows. □

By definition of spectral radius, there exists a unit vector $v \in \mathbb{R}^m$ such that

$$\rho(\mathbf{B} + \mathbf{L} - \mathbf{J}) \leqslant v^{\mathsf{T}}(\mathbf{B} + \mathbf{L} - \mathbf{J})v \leqslant \rho(\mathbf{B}) + \|\mathbf{L}\| + \|\mathbf{J}\|.$$

Thus, we can use next two results to obtain Theorem 6.27.

Lemma 6.28. *Consider the settings of Theorem 6.27 and suppose* $p \leqslant n^{-1}$. *Then with probability* $1 - o(1)$

$$\|\mathbf{L}\| + \|\mathbf{J}\| \leqslant O(1) .$$

Proof. By construction $\|\mathbf{J}\| \leqslant 1$ and $\|\mathbf{L}\| \leqslant \max_{\alpha_1,\beta_1,\alpha_2,\beta_2\in[n]^{k-1}} \left|\mathbf{A}'_{(\alpha_1,\beta_1)(\alpha_2,\beta_2)}\right|$. Now, for fixed $\alpha_1,\beta_1,\alpha_2,\beta_2 \in [n]^{k-1}$ and $t \geqslant 1$

$$\begin{aligned}
\mathbb{P}\left(\left|\mathbf{A}'_{(\alpha_1,\beta_1)(\alpha_2,\beta_2)}\right| \geqslant t\right) &= \mathbb{P}\left(\left|\sum_{\ell\in[n]} \mathbf{T}_{(\alpha_1\alpha_2\ell)}\mathbf{T}_{(\beta_1\beta_2\ell)}\right| \geqslant t\right) \\
&\leqslant \sum_{q\geqslant t} \mathbb{P}\left(\sum_{\ell\in[n]} \left|\mathbf{T}_{(\alpha_1\alpha_2\ell)}\mathbf{T}_{(\beta_1\beta_2\ell)}\right| = q\right) \\
&\leqslant \sum_{q\geqslant t} \binom{n}{q} \cdot p^{2q} \cdot (1-p)^{n-2q} \\
&\leqslant \sum_{q\geqslant t} \binom{n}{q} \cdot p^{2q} \\
&\leqslant O(1) \cdot \left(\frac{e\cdot n\cdot p^2}{t}\right)^t.
\end{aligned}$$

Thus, as $p \leqslant n^{-k/2}\operatorname{poly}\log(n)$ and $\mathbf{A}'$ has n^{2k-2} entries, by union bound $\max_{\alpha_1,\alpha_2,\beta_1,\beta_2\in[n]^{k-1}} \left|\mathbf{A}'_{(\alpha_1,\beta_1)(\alpha_2,\beta_2)}\right| \leqslant 1000$, with probability $1 - n^{-\Omega(1)}$. □

Lemma 6.29. *Consider the settings of Theorem 6.27. Then with probability* $1 - o(1)$

$$\rho(\mathbf{B}) \leqslant O\left(p \cdot n^{k/2}\right).$$

The proof of Lemma 6.29 follows a recipe similar to the one used in Section 6.4. We consider the extended non-backtracking $(2n^{k-1})$-by-$(2n^{k-1})$ matrix $\mathbf{B}$. We use the trace method Proposition 6.14 to bound the spectral norm of the closely related non-backtracking matrix $\mathbf{B}^{z-1}$ for large enough z, and then use Gelfand's formula Fact 2.1 to relate the result to the spectral radius of $\mathbf{B}$. In particular, Lemma 6.29 is an immediate consequence of the following result.

Lemma 6.30. *Consider the settings of Theorem 6.27. Let* $q \leqslant \frac{\log n}{(10^3 \log\log n)^2}$ *and* $z \leqslant \frac{\log n}{50}$. *Then with probability* $1 - o(1)$

$$\operatorname{Tr}\left[\left(\mathbf{B}^{z-1}\right)\left(\mathbf{B}^{z-1}\right)^{\mathsf{T}}\right]^q \leqslant \left(p\cdot n^{k/2}\right)^{2qz} \cdot O\left(z^{\frac{\log n}{10\log\log n}+2}\right).$$

The rest of the section is dedicated to proving Lemma 6.30.

6.5.1.1 From block non-backtracking walks to block hyper non-backtracking walks

We start our analysis by opening up the terms in $\left[\left(\mathbf{B}^{z-1}\right)\left(\mathbf{B}^{z-1}\right)^{\mathsf{T}}\right]^q$. By Fact 6.15 and Fact 6.16 we know how to represent these terms using the entries of $\mathbf{A}'$. As these terms contain

absolute values of sums, it will be more convenient to work with simpler products in the entries of $\mathbf{T}$. In order to achieve this we need to manipulate our walks further. For each walk in $W \in \mathrm{BNBW}^{2q,z}$ with $W = (\alpha_1^1, \beta_1^1), (\alpha_2^1, \beta_2^1), \ldots, (\alpha_{z+1}^1, \beta_{z+1}^1), \ldots, (\alpha_{z+1}^{2q}, \beta_{z+1}^{2q})$ (using multi-indices in $\alpha_j^i, \beta_j^i \in [n]^{(k-1)/2}$), we use the rewriting

$$\begin{aligned}
&\prod_{i=1}^{2q}\left[\sqrt{\left|\mathbf{A}'_{(\alpha_1^i,\beta_1^i),(\alpha_2^i,\beta_2^i)}\mathbf{A}'_{(\alpha_z^i,\beta_z^i),(\alpha_{z+1}^i,\beta_{z+1}^i)}\right|}\cdot\left(\prod_{s=2}^{z-1}\mathbf{A}'_{(\alpha_s^i,\beta_s^i),(\alpha_{s+1}^i,\beta_{s+1}^i)}\right)\right]\\
&=\prod_{i=1}^{2q}\left[\left|\mathbf{A}'_{(\alpha_1^i,\alpha_2^i),(\beta_1^i,\beta_2^i)}\right|\cdot\left(\prod_{s=2}^{z-1}\mathbf{A}'_{(\alpha_s^i,\beta_s^i),(\alpha_{s+1}^i,\beta_{s+1}^i)}\right)\right]\\
&=\sum_{\ell_2^1,\ldots,\ell_z^1,\ldots,\ell_z^{2q}\in[n]}\prod_{i=1}^{2q}\left[\left|\sum_{\ell_1^i\in[n]}\mathbf{T}_{(\alpha_1^i\alpha_2^i\ell_1^i)}\mathbf{T}_{(\beta_1^i\beta_2^i\ell_1^i)}\right|\left(\prod_{s=2}^{z-1}\mathbf{T}_{(\alpha_s^i\alpha_{s+1}^i\ell_s^i)}\mathbf{T}_{(\beta_s^i\beta_{s+1}^i\ell_s^i)}\right)\right],
\end{aligned}\tag{6.5.4}$$

where in the first step we used the fact that the last edge of each W_i in W is the first edge of W_{i+1}. Elements in the sum of the form

$$\prod_{i=1}^{2q}\left[\left|\sum_{\ell_1^i\in[n]}\mathbf{T}_{(\alpha_1^i\alpha_2^i\ell_1^i)}\mathbf{T}_{(\beta_1^i\beta_2^i\ell_1^i)}\right|\left(\prod_{s=2}^{z-1}\mathbf{T}_{(\alpha_s^i\alpha_{s+1}^i\ell_s^i)}\mathbf{T}_{(\beta_s^i\beta_{s+1}^i\ell_s^i)}\right)\right]$$

such that every term with odd degree in

$$\prod_{i=1}^{2q}\left[\left(\prod_{s=2}^{z-1}\mathbf{T}_{(\alpha_s^i\alpha_{s+1}^i\ell_s^i)}\mathbf{T}_{(\beta_s^i\beta_{s+1}^i\ell_s^i)}\right)\right]$$

also appears in

$$\prod_{i=1}^{2q}\left|\sum_{\ell_1^i\in[n]}\mathbf{T}_{(\alpha_1^i\alpha_2^i\ell_1^i)}\mathbf{T}_{(\beta_1^i\beta_2^i\ell_1^i)}\right|$$

are called *annoying*. We denote the set of annoying terms by $\mathrm{AN}(W)$. If a term is not annoying it is said to be *nice*. $\overline{\mathrm{AN}}(W)$ is the set of nice terms in Eq. (6.5.4). We can upper bound the expectation of nice terms with the expectation of a related polynomial. The next result formalizes this idea, we defer its proof to Appendix D.1.

Lemma 6.31. *Consider the settings of Theorem 6.27. Let $W \in BNBW^{2q,z}$. Then for any term in $\overline{AN}(W)$*

$$\mathbb{E}\prod_{i=1}^{2q}\left[\left|\sum_{\ell_1^i}\mathbf{T}_{(\alpha_1^i\alpha_2^i\ell_1^i)}\mathbf{T}_{(\beta_1^i\beta_2^i\ell_1^i)}\right|\left(\prod_{s=2}^{z-1}\mathbf{T}_{(\alpha_s^i\alpha_{s+1}^i\ell_s^i)}\mathbf{T}_{(\beta_s^i\beta_{s+1}^i\ell_s^i)}\right)\right]$$

$$\leqslant O(1)^{2q} \cdot \mathbb{E} \prod_{i=1}^{2q} \left[\left(\sum_{\ell_1^i} \mathbf{T}_{(\alpha_1^i \alpha_2^i \ell_1^i)} \mathbf{T}_{(\beta_1^i \beta_2^i \ell_1^i)} \right)^2 \left(\prod_{s=2}^{z-1} \mathbf{T}_{(\alpha_s^i \alpha_{s+1}^i \ell_s^i)} \mathbf{T}_{(\beta_s^i \beta_{s+1}^i \ell_s^i)} \right) \right]. \tag{6.5.5}$$

Instead, for each annoying term in $\mathrm{AN}(W)$ we bound the expectation as

$$\begin{aligned}
&\mathbb{E} \prod_{i=1}^{2q} \left[\left| \sum_{\ell_1^i} \mathbf{T}_{(\alpha_1^i \alpha_2^i \ell_1^i)} \mathbf{T}_{(\beta_1^i \beta_2^i \ell_1^i)} \right| \left(\prod_{s=2}^{z-1} \mathbf{T}_{(\alpha_s^i a_{s+1}^i \ell_s^i)} \mathbf{T}_{\beta_s^i b_{s+1}^i \ell_s^i} \right) \right] \\
&\leqslant \sum_{\ell_1^i, \ldots, \ell_1^{2q}} \mathbb{E} \left| \prod_{i=1}^{2q} \left[\mathbf{T}_{(\alpha_1^i \alpha_2^i \ell_1^i)} \mathbf{T}_{(\beta_1^i \beta_2^i \ell_1^i)} \left(\prod_{s=2}^{z-1} \mathbf{T}_{(\alpha_s^i \alpha_{s+1}^i \ell_s^i)} \mathbf{T}_{(\beta_s^i \beta_{s+1}^i \ell_s^i)} \right) \right] \right| \\
&\leqslant \sum_{\ell_1^i, \ldots, \ell_1^{2q}} \mathbb{E} \prod_{i=1}^{2q} \left[\left| \mathbf{T}_{(\alpha_1^i \alpha_2^i \ell_1^i)} \right| \left| \mathbf{T}_{(\beta_1^i \beta_2^i \ell_1^i)} \right| \left(\prod_{s=2}^{z-1} \left| \mathbf{T}_{(\alpha_s^i \alpha_{s+1}^i \ell_s^i)} \right| \left| \mathbf{T}_{(\beta_s^i \beta_{s+1}^i \ell_s^i)} \right| \right) \right] \\
&\leqslant O(1)^{2q} \sum_{\ell_1^i, \ldots, \ell_1^{2q}} \mathbb{E} \prod_{i=1}^{2q} \left[\mathbf{T}^2_{(\alpha_1^i \alpha_2^i \ell_1^i)} \mathbf{T}^2_{(\beta_1^i \beta_2^i \ell_1^i)} \left(\prod_{s=2}^{z-1} \left| \mathbf{T}_{(\alpha_s^i \alpha_{s+1}^i \ell_s^i)} \right| \left| \mathbf{T}_{(\beta_s^i \beta_{s+1}^i \ell_s^i)} \right| \right) \right]
\end{aligned} \tag{6.5.6}$$

We will encode every term in Eq. (6.5.5) and Eq. (6.5.6) as a sequence Z of multi-indices $(\alpha_1^1, \beta_1^1, \ell_1^1, \alpha_1^2, \beta_2^1, \ell_2^1, \ldots, \ell_z^{2q}, \alpha_{z+1}^{2q}, \beta_{z+1}^{2q})$, where the ℓ_s^i's are indices and α_s^i, β_s^i's are multi-indices of cardinality $(z-1)/2$. For a subsequence Z, we will use $S(Z)$ to denote the set of its indices (without repetitions). Lemma 6.31 together with these observations allows us to upper bound the contribution of each walk $W \in \mathrm{BNBW}^{2q,z}$ to the expectation of the trace of $\left[\left(\mathbf{B}^{z-1} \right) \left(\mathbf{B}^{z-1} \right)^\intercal \right]^q$ as

$$\begin{aligned}
&\mathbb{E} \prod_{i=1}^{2q} \left[\left| \mathbf{A}'_{(\alpha_1^i, \beta_2^i), (\alpha_2^i, \beta_2^i))} \right| \cdot \left(\prod_{s=2}^{z-1} \mathbf{A}'_{(\alpha_s^i \beta_s^i), (\alpha_{s+1}^i \beta_{s+1}^i)} \right) \right] \\
&\quad \leqslant O(1)^{2q} \cdot \mathbb{E} \sum_{\substack{\alpha_1^1, \ldots, \alpha_{z+1}^1, \ldots \alpha_{z+1}^{2q}, \\ \beta_1^1, \ldots, \beta_{z+1}^1, \ldots \beta_{z+1}^{2q}, \\ \ell_1^1, \ldots, \ell_z^{2q} \\ \text{in } C}} \mathbf{T}_{(\alpha_1^1 \alpha_2^1 \ell_1^1)} \cdots \mathbf{T}_{(\beta_z^{2q} \beta_{z+1}^{2q} \ell_z^{2q})}, \qquad (6.5.7) \\
&\quad + O(1)^{2q} \cdot \mathbb{E} \sum_{\substack{\alpha_1^1, \ldots, \alpha_{z+1}^1, \ldots \alpha_{z+1}^{2q}, \\ \beta_1^1, \ldots, \beta_{z+1}^1, \ldots \beta_{z+1}^{2q}, \\ \ell_1^1, \ldots, \ell_z^{2q} \\ \text{in } C^*}} \left| \mathbf{T}_{(\alpha_1^1 \alpha_2^1 \ell_1^1)} \right| \cdots \left| \mathbf{T}_{(\beta_z^{2q} \beta_{z+1}^{2q} \ell_z^{2q})} \right|. \qquad (6.5.8)
\end{aligned}$$

Here C is the set of conditions (using the convention $i+1=1$ for $i=2q$)

$$\begin{cases} \text{pre-processing:} & \left|S(\alpha_j^i\alpha_{j+1}^i)\cap S(\beta_j^i\beta_{j+1}^i)\right|=0 \\ & \text{all distinct in } (\alpha_j^i,\alpha_{j+1}^i,\ell_j^i) \\ & \text{all distinct in } (\beta_j^i,\beta_{j+1}^i,\ell_j^i) \\ \text{block:} & \alpha_z^i=\alpha_2^{i+1} \\ & \alpha_{z+1}^i=\alpha_1^{i+1} \\ & \beta_{z+1}^i=\beta_1^{i+1} \\ & \beta_z^i=\beta_2^{i+1} \\ \text{non-backtracking:} & \alpha_j^i\neq\alpha_{j+2}^i \text{ or } \beta_j^i\neq\beta_{j+2}^i \end{cases} \tag{C}$$

and C^* is the set of conditions in C with the addition of $\left\{\forall i\in[2q]\,,\ell_z^i=\ell_1^{i+1}\right\}$ and

- If a tuple Z' in $\left\{(\alpha_j^i,\alpha_{j+1}^i,\ell_j^i)\,,(\beta_j^i,\beta_{j+1}^i,\ell_j^i)\right\}$ appears an odd number of times in the whole sequence then there exists i' such that $\left|S(Z)\cap S(\alpha_1^{i'},\alpha_1^{i'+1},\ell_1^{i'})\right|\geqslant k-1$ or $\left|S(Z)\cap S(\beta_1^{i'},\beta_2^{i'},\ell_1^{i'})\right|\geqslant k-1$.

Let's try to unravel the meaning of C and C^*. The set of conditions C are meant to captures products arising from terms in $\overline{\mathrm{AN}}(W)$ for $W\in\mathrm{BNBW}^{2q,z}$. Notice the *block* condition correspond to the observation that in each $W\in\mathrm{BNBW}^{2q,z}$ partitioned in non-backtracking walks $W_1,\ldots,W_{2q}$, for any consecutive W_i,W_{i+1} the first edge traversed in W_{i+1} is the last edge traversed in W_i. The pre-processing condition allows to exclude corner cases[5] as these where handled with $\mathbf{A}''$. A consequence of this condition is that the underlying hyper-graph cannot have self-loops. Finally, the crucial non-backtracking conditions will help us upper bound the number of terms in the sum.

The terms appearing in C^* correspond to the ones generated by $\mathrm{AN}(W)$ for $W\in\mathrm{BNBW}^{2q,z}$. As such the first additional condition captures the fact that we don't take the square of the first and last edges in each non-backtracking walk W_i in W. The second condition captures the fact that these terms are annoying, and thus only few specific elements in the product may have odd degree. While our bound for annoying terms is tough, the number of such terms will be small. We remark that every sequence satisfying C^* also satisfies C.

Block non-backtracking hyper walks. Recall we can encode every term in Eq. (6.5.7) and Eq. (6.5.8) as a sequence Z of multi-indices $(\alpha_1^1,\beta_1^1,\ell_1^1,\alpha_2^2,\beta_2^1,\ell_2^1,\ldots,\ell_z^{2q},\alpha_{z+1}^{2q},\beta_{z+1}^{2q})$ where α_j^i,β_j^i's have cardinality $k-1$ and ℓ_j^i's are indices. We denote the set of sequences over $2q$ blocks of size z satisfying C as the set of block non-backtracking hyper-walks Z and denote

[5]The experienced reader may recognize the *pre-processing* condition in C implies the *no self-loop* condition from [AOW15].

it by $\mathrm{HBNBW}^{2q,z}$. We denote by $\mathrm{HBNBW}^{2q,z}(C^*) \subseteq \mathrm{HBNBW}^{2q,z}$ the subset of sequences in $\mathrm{HBNBW}^{2q,z}$ which also satisfies C^*. We will interchangeably use the terms hyper-walk and sequence. Notice that each hyper-walk $Z \in \mathrm{HBNBW}^{2q,z}$ generates a multi-hyper-graph[6] $H(Z)$ over vertices in $[n]$. For simplicity we will say "hyper-graph" instead of "multi-hyper-graph". We then call $H(Z)$ the underlying hyper-graph of Z.We denote with $V(Z)$ its set of vertices (corresponding to the set of distinct indices in the sequence). An hyper-edge $S(\alpha, \alpha', \ell)$ is in $H(Z)$ if $\mathbf{T}_{(\alpha,\alpha',\ell)}$ appears in the product encoded by Z. We use $E(H(Z))$ to denote the set of distinct hyper-edges in $H(Z)$. For each $e \in E(Z)$we denote by $m_e(H(Z))$ the multiplicity of e in $H(Z)$. We say $H(Z)$ is the underlying multi-hyper-graph of Z. For $Z \in \mathrm{HBNBW}^{2q,z}$ we can graphically represent each such $H(Z)$ as depicted in Fig. 6.1.

Furthermore, each $Z \in \mathrm{HBNBW}^{2q,z}$ can be further split into subsequences $Z_1, \ldots, Z_{2q}$ such that $Z_i = (\alpha_1^i, \beta_1^i, \ell_1^i, \alpha_2^i, \beta_2^i, \ldots, \ell_z^i, \alpha_{z+1}^i, \beta_{z+1}^i)$ for all $i \in [n]$. We say $Z_1, \ldots, Z_{2q}$ are non-backtracking hyper-walks and denote the set of such walks as $\mathrm{HNBW}^{2q,z}$. We use $\mathrm{HNBW}^{2q,z}(Z)$ to denote the non-backtracking hyper-walks corresponding to Z (picking one arbitrary such partition) and $H(Z_i)$ to denote the underlying hyper-graph of each such sequence. We refer to $Z_1, \ldots, Z_{2q}$ simply as the subsequences of Z.

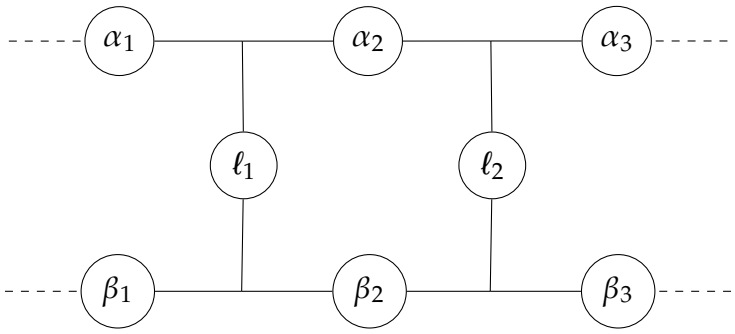


Figure 6.1: Representation of part of a hyper-walk satisfying C for 3-XOR.

Using our newly introduced notation, we can thus write

$$\begin{aligned}
&\mathbb{E} \sum_{W \in \mathrm{BNBW}^{2q,z}} \prod_{i=1}^{2q} \prod_{s=1}^{z} \mathbf{A}'_{e_s(W_i)} \\
&\leqslant O(1)^{2q} \cdot \mathbb{E} \sum_{Z \in \mathrm{HBNBW}^{2q,z}} \mathbf{T}_{(\alpha_1^1(Z)\alpha_2^1(Z)\ell_1^1(Z))} \cdots \mathbf{T}_{(\beta_z^{2q}(Z)\beta_{z+1}^{2q}(Z)\ell_z^{2q}(Z))} \\
&\quad + O(1)^{2q} \cdot \mathbb{E} \sum_{Z \in \mathrm{HBNBW}^{2q,z}(C^*)} \left| \mathbf{T}_{(\alpha_1^1(Z)\alpha_2^1(Z)\ell_1^1(Z))} \cdots \mathbf{T}_{(\beta_z^{2q}(Z)\beta_{z+1}^{2q}(Z)\ell_z^{2q}(Z))} \right|. \qquad (6.5.9)
\end{aligned}$$

A useful observation, formalized in the following fact, is that for any $Z \in \mathrm{HBNBW}^{2q,z} \setminus \mathrm{HBNBW}^{2q,z}(C^*)$, if the underlying hyper-graph has a hyper-edge with multiplicity 1 then its expectation is 0.

[6]That is, a hyper-graph in which hyper-edges may have multiplicity larger than 1.

Fact 6.32. *Consider the settings of Theorem 6.27. Let $Z \in HBNBW^{2q,z}$ and let H be the underlying hyper-graph of Z. If H contains a hyper-edge e of multiplicity 1, then*

$$\mathbb{E} \prod_{e \in E(H(Z))} \mathbf{T}_e^{m_e(H(Z))} = 0 \,.$$

Proof. By independence of hyper-edges and symmetry of their distribution

$$\mathbb{E} \prod_{e \in E(H(Z))} \mathbf{T}_e^{m_e(H(Z))} = \prod_{e \in E(H(Z))} \mathbb{E}\, \mathbf{T}_e^{m_e(H(Z))} = 0 \,.$$

□

6.5.1.2 Encoding tangle-free block non-backtracking hyper walks

Our goal in this section is to define a meaningful notion of tangle-freeness for hyper-walks in $\mathrm{HBNBW}^{2q,z}$ and then obtain results along the lines of Lemma 6.19 and Lemma 6.20. For this we need to introduce additional notions.

Recall each Z in $\mathrm{HBNBW}^{2q,z}$ is a sequence $(\alpha_1^1, \beta_1^1, \ell_1^1, \ldots, \ell_z^{2q}, \alpha_{z+1}^{2q}, \beta_{z+1}^{2q})$, which we can decompose it into subsequences $Z_1, \ldots, Z_{2q} \in \mathrm{HNBW}^{2q,z}$. We can imagine each Z_i being revealed through the following discovering process:

1. At time $t_i = 0$ we are given multi-indices α_1^i, β_1^i,
2. At time $t_i = j > 0$ we reveal the multi-indices $\ell_j^i \alpha_{j+1}^i \beta_{j+1}^i$.

In other words, having α_1^i, β_1^i as our starting indices, we reveal at each time the multi-indices involved in the next two hyper-edges of the underlying hypergraph $H(Z_i)$ of the subsequence Z_i. For each time $1 \leqslant j \leqslant z$, we denote by $R_{ij} \subseteq [n]^3$ the sequence of multi-indices in Z_i revealed at time j, we also use R_{ij} to denote the corresponding multi-set. We will also refer to R_{ij} as a tuple (of size k). We denote the two hyper-edges revealed with R_{ij} by $E_{ij} = \{e_{i,j,1}, e_{i,j,2}\}$ and by $H(Z_i, j)$ the underlying hyper-graph of Z_i revealed up to time j. We can now extend this idea to Z. We discover Z by revealing in order $Z_1, \ldots, Z_{2q}$ as described above. Notice that due to the block condition in C we have $\bigcup_{i \in [2q], j \in [z]} R_{ij} = V(Z)$. We define the collection of sets $R_{11}, \ldots, R_{2q,z}$ as $\mathcal{R}(Z)$. We provides two partitions of $\mathcal{R}(Z)$.

- For $0 \leqslant s \leqslant k$, $R_{ij} \in \mathcal{G}_s$ if exactly s indices in R_{ij} did not appear in the sequence Z before. For a subsequence Z_i of Z we write $\mathcal{G}_s(Z_i) \subseteq \mathcal{G}_s$ for the subset of tuples in $\mathcal{G}_s$ corresponding to reveals in Z_i.

The second partition is:

- For $0 \leqslant s \leqslant 2$, $R_{ij} \in \mathcal{P}_s$ if for the hyper-graph $\bigoplus_{i' < i} H(Z_i) \oplus H(Z_i, j-1)$ with hyper-edge set E we have $\left|E \cap E_{ij}\right| = 2 - s$. That is, $R_{ij} \in \mathcal{P}_s$ if it reveals s new hyper-edges.

We are now ready to define *t-tangle-free* non-backtracking hyper-walk. Given $Z \in \mathrm{HBNBW}^{2q,z}$ we say that the subsequence $Z_i \in \mathrm{HNBW}^{2q,z}$ is t-tangle free if

$$\left|\mathcal{T}_i := \left\{R_{ij} \in (\mathcal{P}_2 \cup \mathcal{P}_1) \cap \mathcal{G}_0(Z_i)\right\}\right| \leqslant t\,. \tag{6.5.10}$$

In words, Eq. (6.5.10) is saying that for each Z_i the number of tuples R_{ij} that reveal multi-indices containing indices all already seen in Z_i, but which reveal at least one new hyper-edge, is at most t. We denote by $\mathrm{HTGF}^{2q,z,t} \subseteq \mathrm{HNBW}^{2q,z,t}$ the set of t-tangle-free non-backtracking hyper-walks. We also use $\mathrm{HBTGF}^{2q,z,t} \subseteq \mathrm{HBNBW}^{2q,z}$ to denote the set of block non-backtracking hyper-walks in which each Z_i is in $\mathrm{HTGF}^{2q,z,t}$. If the block non-backtracking walks in $\mathrm{BNBW}^{2q,z}$ over $\mathbf{A}'$ only yields t-tangle free sequences, we say $\mathbf{A}'$ is t-tangle-free. Using a similar approach as the one shown in the context of sparse graphs in Lemma 6.19, we will only need to consider walks in $\mathrm{HBTGF}_t^{2q,z,t}$ for $t \leqslant 100 \log\log n$.

Lemma 6.33 (Sparse hyper-graphs are tangle-free)**.** *Consider the settings of Theorem 6.27. Let $n^{-k/2} \leqslant p \leqslant n^{-k/2}\log^{10} n$ and $z \leqslant \frac{\log n}{50}$. Then for $t \geqslant 100\log\log n$, $\mathbf{A}'$ is t-tangle free with probability $1 - o(1)$.*

Proof. Let $Z \in \mathrm{HNBW}^{2q,z}$ be t-tangled and let $v = |V(H(Z))|, e = |E(H(Z))|$. The probability that $\prod_{e\in E(H(Z))} \mathbf{T}_e \neq 0$ is p^e. Now by definition

$$\begin{aligned} e &\geqslant t + \sum_{1\leqslant s<(k+1)/2} |\mathcal{G}_1(Z)| + 2\sum_{(k+1)/2\leqslant s\leqslant k} |\mathcal{G}_s(Z)|\,,\\ v &\leqslant 2 + \sum_{s\leqslant k} s|\mathcal{G}_s(Z)|\,. \end{aligned}$$

Combining the two we get $e \geqslant t - 3 + \frac{2}{k}v$. The number of sequences in $\mathrm{HNBW}^{2q,z}$ over v indices is at most $n^v \cdot v^{kz+(k-1)}$. Thus by union bound

$$\begin{aligned} \mathbb{P}\left\{\exists Z \in \mathrm{HNBW}^{2q,z} \setminus \mathrm{HTGF}^{2q,z,t} \;:\; \prod_{e\in E(H(Z))} \mathbf{T}_e \neq 0\right\} &\leqslant \sum_{t'\leqslant t}\sum_{v\leqslant k(z+1)-1} n^v \cdot v^{k(z+1)-1}\cdot p^{t-3+\frac{2}{k}v}\\ &\leqslant \sum_{t'\leqslant t}\sum_{v\leqslant k(z+1)-1} \left(np^{2/k}\right)^v \cdot v^{k(z-1)-1}\cdot p^{t'-3}\\ &\leqslant \sum_{t'\leqslant t}\left(np^{2/k}v\right)^{2kz}\cdot p^{t'/2}\\ &\leqslant 2\cdot\left(k\log n^{11}\right)^{\frac{k\log n}{11}}\cdot 2^{25k\cdot\log(n)\cdot\log\log n}\\ &\leqslant 2^{1+\frac{k\log n}{10}\log\log n - 25k\log(n)\cdot\log\log n}\\ &\leqslant o(1)\,. \end{aligned}$$

□

For the remainder of the section we set $t = 100 \log\log n$. For simplicity we will say that t-tangle free sequences in $\mathrm{HNBW}^{2q,z}$ are simply tangle-free, we will drop the superscript t.

Next, we define canonical tangle-free block non-backtracking hyper walks. We use lexicographic order for the multi-indices: $\alpha \preceq \beta$ if at each position $i \leqslant \min\{|\alpha|, |\beta|\}$ we have $\alpha(i) \leqslant \beta(i)$. We say that $Z \in \mathrm{HBTGF}^{2q,z}$ is *canonical* if $\alpha_1 = (1, \ldots, (k-1)/2)$ and for every other multi-index β (including ℓ's of multiplicity 1) in the sequence one the following applies:

(i) all the indices in β already appeared in the sequence,

(ii) indices in β are ordered and consecutive (that is for $i \leqslant |\beta| - 1$, $\beta(i) = \beta(i+1) - 1$), moreover the index $\beta(0) - 1$ already appeared in the sequence.

We use $\mathcal{H}^{2q,z}$ to denote the set of canonical sequences. As in the context of graphs, canonical hyper-walks are convenient as each correspond to a representative for a class of isomorphic hyper-walks. For any canonical sequence over v vertices there are at most $\binom{n}{v} v! \leqslant n^v$ isomorphic sequences in $\mathrm{HBTGF}^{2q,z}$. We use tuples of the form $\Psi = (v, g_0, \ldots, g_k, r_0, r_1, r_2, d_1, d_2, e^*)$ to encode parameters. Then $\mathcal{H}^{2q,z}(e, \Psi)$ is the set of canonical sequences Z such that the underlying hyper-graph $H(Z)$ has e distinct hyper-edges, e^* hyper-edges of multiplicity 1 and:

$$\begin{aligned}
\sum_{s \leqslant k} s|\mathcal{G}|_s(Z) &= v \\
0 \leqslant s \leqslant k\,, |\mathcal{G}_s(Z)| &= g_s \\
0 \leqslant s \leqslant 2\,, |\mathcal{P}_s(Z)| &= r_s \\
\left| \bigcup_{1 \leqslant s \leqslant 2} \mathcal{P}_s(Z) \setminus \left(\bigcup_{1 \leqslant s' \leqslant k} \mathcal{G}_{s'}(Z) \right) \right| &= d_1 \\
\left| \bigcup_{1 \leqslant s \leqslant 2} \mathcal{P}_s(Z) \setminus \mathcal{G}_k(Z) \right| &= d_2\,.
\end{aligned}$$

The next result, similar in spirit to Lemma 6.20, upper bounds the number of canonical sequences.

Lemma 6.34 (Enumeration of canonical sequences)**.** *We have*

$$\left|\mathcal{H}^{2q,z}(e, \Psi)\right| \leqslant \left[\prod_{1 \leqslant s \leqslant k-1} \left(\binom{k}{s} v^s \right)^{g_s} \right] \cdot v^{kd_1} \cdot k^{4qt} \cdot z^{16qt(d_2+4)+d_1}$$

Proof. We can encode each $Z \in \mathcal{H}^{2q,z}(e, \Psi)$ by encoding each of the tuple $R_{11}, \ldots, R_{2q,z}$. Notice that since Z is canonical we can encode Z simply by encoding reveals in $\bigcup_{s \leqslant k-1} \mathcal{G}_s$ (and their position) as for each $R_{ij} \in \mathcal{G}_s$ we only have one choice for the indices. We can encode tuples as follows:

- For $R_{ij} \in \mathcal{G}_s$ for $1 \leqslant s < k$ we have $\binom{k}{s} v^s$ choices for the indices that appeared already. We need not to specify other indices as the sequence is canonical. There are z ways to position such tuples in Z_i so overall $\binom{k}{s} v^s z$ possibilities for each such R_{ij}. In conclusion fixing the cardinalities of $\mathcal{G}_1, \dots, \mathcal{G}_k$ there are $\prod_{1 \leqslant s \leqslant k-1} \left(\binom{k}{s} v^s z \right)^{|\mathcal{G}_s|}$ choices over Z.

- For $R_{ij} \in \mathcal{T}_i$, there are kz possible choices for the vertices in R_{ij}, there are z ways to position such tuples. So overall there are at most (kz^2) choices. By tangle-freeness there are at most t such R_{ij}'s in each Z_i and thus $\left(kz^2\right)^{2qt}$ possibilities over Z.

- For $R_{ij} \in \left| \bigcup_{1 \leqslant s \leqslant 2} \mathcal{P}_s \setminus \left(\bigcup_{1 \leqslant s' \leqslant k} \mathcal{G}_{s'} \right) \right|$ we have v^k candidate indices and $z \cdot 2q$ ways to position the tuple. Overall there are $\left(v^k z 2 \cdot q\right)^{\left| \bigcup_{1 \leqslant s \leqslant 2} \mathcal{P}_s \setminus \left(\bigcup_{1 \leqslant s' \leqslant 3} \mathcal{G}_{s'} \right) \right|} = \left(2 v^k z q\right)^{d_1}$ possibilities over Z.

- For any subsequence $R_{ij}, \dots, R_{i(j+r)}$ of tuples in $\mathcal{P}_0$, we only need to encode the tuples $R_{i(j+r')}$ with $0 \leqslant r' \leqslant r$ for which at least one of the hyper-edges in $E_{i(j+r')}$ appeared for the first time in the sequence in some $R_{i'j'} \in \bigcup_{1 \leqslant s \leqslant 2} \mathcal{P}_s \setminus \mathcal{G}_k$. In fact we can reconstruct the whole subsequence $R_{ij}, \dots, R_{i(j+r)}$ by the position and the hyper-edges of these tuples. Each of these $R_{i'j'}$ can only appear at most $2t$ times in the subsequence $R_{ij}, \dots, R_{i(j+r)}$ since by assumption Z_i is t-tangle-free. For convention if the subsequence ends in R_{iz} we also specify the hyper-edges revealed by this tuple. Each subsequence $R_{ij}, \dots, R_{i(j+r)}$ can be position in z different ways in Z_i and has length at most z. Since there are $2q$ such Z_i's in Z, overall we have $z^{8qt \cdot 2\left(\left|\bigcup_{1 \leqslant s \leqslant 2} \mathcal{P}_r \setminus \mathcal{G}_k\right| + 1\right)} = z^{16qt \left|\bigcup_{1 \leqslant s \leqslant 2} \mathcal{P}_r \setminus \mathcal{G}_k\right| + 2} = z^{16qt(d_2+2)}$ distinct choices.

We deduce that for any valid $\Psi = (v, g_0, \dots, g_k, r_0, r_1, r_2, d_1, d_2, e^*)$

$$\begin{aligned} |\mathcal{H}(e, \Psi)| &\leqslant \left[\prod_{1 \leqslant s \leqslant 2} \left(\binom{k}{s} v^s z \right)^{g_s} \right] \cdot (kz^2)^{2qt} \left(2 v^k z q\right)^{d_1} z^{16qt(d_2+2)} \\ &= \left[\prod_{1 \leqslant s \leqslant 2} \left(\binom{k}{s} v^s \right)^{g_s} \right] \cdot v^{kd_1} \cdot k^{4qt} \cdot z^{16qt(d_2+4)+d_1} . \end{aligned}$$

□

6.5.1.3 Putting things together

We are finally ready to prove Lemma 6.30.

Proof of Lemma 6.30. For ZHBNBW^{2q,z} we write α^i_j in place of $\alpha^i_j(Z)$ to denote multi-indices in the sequence Z. By definition

$$\mathrm{Tr}\left[\left(\mathbf{B}^{z-1} \right) \left(\mathbf{B}^{z-1} \right)^{\mathsf{T}} \right]^q$$

$$= \sum_{Z \in \mathrm{HBNBW}^{2q,z}} \prod_{i=1}^{2q} \left[\left| \sum_{\ell_1^i} \mathbf{T}_{(\alpha_1^i \alpha_2^i \ell_1^i)} \mathbf{T}_{(\beta_1^i \beta_2^i \ell_1^i)} \left| \left(\prod_{s=2}^{z-1} \mathbf{T}_{(\alpha_s^i \alpha_{s+1}^i \ell_s^i)} \mathbf{T}_{(\beta_s^i, \beta_{s+1}^i \ell_s^i)} \right) \right| \right] .$$

By Lemma 6.33 with probability $1 - o(1)$,

$$\sum_{Z \in \mathrm{HBNBW}^{2q,z}} \prod_{i=1}^{2q} \left[\left| \sum_{\ell_1^i} \mathbf{T}_{(\alpha_1^i \alpha_2^i \ell_1^i)} \mathbf{T}_{(\beta_1^i \beta_2^i \ell_1^i)} \left| \left(\prod_{s=2}^{z-1} \mathbf{T}_{(\alpha_s^i \alpha_{s+1}^i \ell_s^i)} \mathbf{T}_{(\beta_s^i, \beta_{s+1}^i \ell_s^i)} \right) \right| \right]$$
$$= \sum_{Z \in \mathrm{HBTGF}^{2q,z}} \prod_{i=1}^{2q} \left[\left| \sum_{\ell_1^i} \mathbf{T}_{(\alpha_1^i \alpha_2^i \ell_1^i)} \mathbf{T}_{(\beta_1^i \beta_2^i \ell_1^i)} \left| \left(\prod_{s=2}^{z-1} \mathbf{T}_{(\alpha_s^i \alpha_{s+1}^i \ell_s^i)} \mathbf{T}_{(\beta_s^i, \beta_{s+1}^i \ell_s^i)} \right) \right| \right] .$$

Now as shown in Eq. (6.5.9)

$$\mathbb{E} \sum_{Z \in \mathrm{HBTGF}^{2q,z}} \prod_{i=1}^{2q} \left[\left| \sum_{\ell_1^i} \mathbf{T}_{(\alpha_1^i \alpha_2^i \ell_1^i)} \mathbf{T}_{(\beta_1^i \beta_2^i \ell_1^i)} \left| \left(\prod_{s=2}^{z-1} \mathbf{T}_{(\alpha_s^i \alpha_{s+1}^i \ell_s^i)} \mathbf{T}_{(\beta_s^i, \beta_{s+1}^i \ell_s^i)} \right) \right| \right]$$
$$\leqslant O(1)^{2q} \cdot \mathbb{E} \sum_{Z \in \mathrm{HBTGF}^{2q,z}} \mathbf{T}_{(\alpha_1^1, \alpha_2^1, \ell_1^1)} \cdots \mathbf{T}_{(\beta_z^{2q}, \beta_{z+1}^{2q}, \ell_z^{2q})} \tag{6.5.11}$$
$$+ O(1)^{2q} \cdot \mathbb{E} \sum_{Z \in \mathrm{HBTGF}^{2q,z}(C^*)} \left| \mathbf{T}_{(\alpha_1^1, \alpha_2^1, \ell_1^1)} \cdots \mathbf{T}_{(\beta_z^{2q}, \beta_{z+1}^{2q}, \ell_z^{2q})} \right| . \tag{6.5.12}$$

We start by studying Eq. (6.5.11). By Fact 6.32 we only need to consider hyper-walks $Z \in \mathrm{HBTGF}^{2q,z}$ with at most $2qk$ distinct hyper-edges since each hyper-edge must have multiplicity at least 2 in the underying hyper-graph $H(Z)$. Thus we can upper bound Eq. (6.5.11) with

$$\sum_{\substack{e \leqslant 2qz, v \geqslant 0 \\ g_k, \ldots, g1 \geqslant 0 \\ r_2, r_1, r_0 \geqslant 0 \\ d_1, d_2 \geqslant 0}} n^v \cdot p^e \cdot |\mathcal{H}(e, \Psi)| = \sum_{e \leqslant 2qz, \Psi} n^v \cdot p^e \cdot |\mathcal{H}(e, \Psi)| ,$$

where the right-hand side is a simple rewriting for compactness. Since for any $Z \in \mathrm{HBTGF}^{2q,z}$ we have

$$v = \sum_{i \leqslant k} s |\mathcal{G}_s|$$
$$e = 2 \sum_{(k+1)/2 \leqslant s \leqslant k} |\mathcal{G}_s| + \left| \bigcup_{1 \leqslant s \leqslant 2} \mathcal{P}_s \leqslant \left(\bigcup_{(K+1)/2 \leqslant s' \leqslant 3} \mathcal{G}_{s'} \right) \right| ,$$

it follows using Lemma 6.34

$$\sum_{e \leqslant 2qz, \Psi} n^v \cdot p^e \cdot |\mathcal{H}(e, \Psi)|$$

$$\leqslant \sum_{e\leqslant 2qz,\Psi} n^v \cdot p^e \cdot \left[\prod_{1\leqslant s\leqslant k-1}\left(\binom{k}{s}v^s\right)^{g_s}\right] \cdot v^{kd_1}\cdot k^{4qt}\cdot z^{16qt(d_2+4)+d_1}$$

$$\leqslant 2\sum_{e\leqslant 2qz,\Psi} \underbrace{\prod_{(k+1)/2\leqslant s\leqslant k}\left(\binom{k}{s}v^s n^s p^2\right)^{g_s}\cdot \prod_{1\leqslant s<(k+1)/2}\left(\binom{k}{s}v^s n^s p\right)^{g_s}\cdot\left(v^k p\right)^{d_1}\cdot k^{4qt}\cdot z^{16qt(d_2+4)+d_1}}_{=:L(e,\Psi)}.$$

It is easy to see that this is a geometric sum. Using the assumptions $p \geqslant n^{-3/2}n$, $z \leqslant \log n$, $q \leqslant \frac{\log n}{(10^3\log\log n)^2}$ and $t = 100\log\log n$, as $k \leqslant O(\log n)$ we have that decreasing g_s in Ψ and increasing g_{s-1} with $(k+1)/2 < s \leqslant k$ to obtain Ψ' we have

$$\frac{L(e,\Psi)}{L(e,\Psi')} \geqslant \frac{n}{v\cdot k\cdot z^{8qt}} \geqslant \omega(1).$$

Similarly, decreasing g_s in Ψ and increasing g_{s-1} with $1 < s < (k+1)/2$ to obtain Ψ' we have

$$\frac{L(e,\Psi)}{L(e,\Psi')} \geqslant \frac{n}{v\cdot k\cdot z^{8qt}} \geqslant \omega(1).$$

Finally, decreasing g_k in Ψ and increasing any g_s with $1 \leqslant s < (k+1)/2$ to obtain Ψ' we get

$$\frac{L(e,\Psi)}{L(e,\Psi')} \geqslant \frac{n^{k-(k-1)/2}p}{\binom{k}{\frac{k-1}{2}}\cdot v^k\cdot k\cdot z^{8qt}} \geqslant \frac{\sqrt{n}}{(2v)^k\cdot z^{8qt}} \geqslant \omega(1).$$

Thus as $\sum_{s\leqslant k}|\mathcal{G}_k| \leqslant qz$ we have

$$\sum_{e\leqslant 2qz,\Psi} n^v\cdot p^e\cdot|\mathcal{H}(e,\Psi)| \leqslant \left(n^k p^2\right)^{qz}\cdot O\left(z^{65qt}\right),$$

where we also accounted for the $O(1)^{2q}$ multiplicative factor. The argument for Eq. (6.5.12) is similar. We can upper bound Eq. (6.5.12) by

$$\sum_{e^*\leqslant 2qz}\sum_{e\leqslant 2qz+e^*,\Psi} n^{v-ke^*}\cdot p^e\cdot|\mathcal{H}(e,\Psi)|\cdot\left(2(k!)qkn\right)^{e^*}$$

where we used the crucial fact that for any $Z \in \mathrm{HBTGF}^{2q,z}(C^*)$ any tuple (α,β,ℓ) appearing an odd number of times must satisfy $\left|S(\alpha,\beta,\ell)\cap S(\alpha_1^i,\alpha_2^i,\ell_1^i)\right| \geqslant k-1$ or $\left|S(\alpha,\beta,\ell)\cap S(\beta_1^i,\beta_2^i,\ell_1^i)\right| \geqslant k-1$ and thus the number of possible choices for those indices is at most $(2(k!)qkn)$. Repeating the analysis above the result follows. □

6.6 Strong refutations for random CSPs

Theorem 6.22 can also be used to obtain strong refutations for random CSPs. Similar reductions appeared already in the literature (e.g. [AOW15]), however we crucially exploits the sharp novel bound in Theorem 6.22 to obtain stronger results as in Theorem 6.2. We use the notation introduced in Section 6.2.1.

Theorem 6.35. *Let $P : \{-1,+1\}^k \to \{0,1\}$ be a predicate. Consider a random instance $\boldsymbol{\mathcal{I}} \sim \mathcal{F}_{\mathrm{CSP(P)}}(n,p)$ for odd $k \geqslant 3$. For n large enough, there exists a universal constant $C > 0$ and a polynomial time algorithm that, if*

$$p \geqslant \frac{C \cdot n^{-k/2}}{\varepsilon^2}$$

certifies with probability 0.99 that

$$\mathrm{Opt}_{\boldsymbol{\mathcal{I}}} \leqslant \mathop{\mathbb{E}}_{\mathbf{z} \overset{u.a.r}{\sim} \{\pm 1\}^k} [P(\mathbf{z})] + O(\varepsilon)\,.$$

Recall from Section 6.2.1 that we can represent the predicate P as a multi-linear polynomial of degree k,

$$P(c \circ x^\alpha) = \sum_{d \leqslant k} P_d(c \circ x^\alpha)\,,$$

where P_d denotes the degree d part of the predicate. In particular $P_0 := P_0(c \circ x^\alpha)$ denotes the constant part of the polynomial, which does not depend on the assignment and the negative pattern. For a instance $\mathcal{I}$ with m constraints, we define

$$S_{\mathcal{I}}(x) := \sum_{(c,\alpha)\in\mathcal{I}} \sum_{1\leqslant d\leqslant k} P_d(c \circ x^\alpha)\,.$$

Notice that by definition, for any fixed $x \in \{\pm 1\}^n$ it holds

$$S_{\mathcal{I}}(x) = m \cdot (\mathrm{Val}_{\mathcal{I}}(x) - P_0) = m \cdot \left(\mathrm{Val}_{\mathcal{I}}(x) - \mathop{\mathbb{E}}_{\mathbf{z} \overset{u.a.r}{\sim} \{\pm 1\}^k} [P(\mathbf{z})]\right),$$

where we used the fact that for every $1 \leqslant d \leqslant k$, $\mathbb{E}_{\mathbf{z} \overset{u.a.r}{\sim} \{\pm 1\}^k}[P_d(\mathbf{z})] = 0$ by symmetry. Thus to obtain Theorem 6.35 it suffices to obtain a tight upper bound on $\max_{x\in\{\pm 1\}^n} S_{\boldsymbol{\mathcal{I}}}(x)$ for $\boldsymbol{\mathcal{I}} \sim \mathcal{F}_{\mathrm{CSP(P)}}(n,p)$. We further write $S_{\mathcal{I},d}(x)$ to denote the degree $d \leqslant k$ part of $S_{\mathcal{I}}(x)$. We then have $S_{\mathcal{I}}(x) = \sum_{d\leqslant k} S_{\mathcal{I},d}(x)$. We can write $S_{\mathcal{I},d}(x)$ as a $n^{\lfloor d/2\rfloor}$-by-$n^{\lceil d/2\rceil}$ matrix $M_{\mathcal{I},d}$ such that $S_{\mathcal{I},d}(x) = \langle x^{\otimes\lfloor d/2\rfloor}, M_{\mathcal{I},d} x^{\otimes\lceil d/2\rceil}\rangle$ for any $x \in \mathbb{R}^n$. The next lemma, provides a rough bound on the spectral norm of each $M_{\boldsymbol{\mathcal{I}},d}$ when $\boldsymbol{\mathcal{I}} \sim \mathcal{F}_{\mathrm{CSP(P)}}(n,p)$. We remark that these bounds on $S_{\boldsymbol{\mathcal{I}},d}(x)$ are significantly less sharp than Theorem 6.22, nevertheless they will be good enough for our needs.

Lemma 6.36. *Consider the settings of Theorem 6.35. For some $d < k$ let $M_{\boldsymbol{I},d}$ be the $n^{\lfloor d/2 \rfloor}$-by-$n^{\lceil d/2 \rceil}$ matrix representing $S_{\boldsymbol{I},d}$. Then with probability $1 - n^{\Omega(1)}$,*

$$\left\| M_{\boldsymbol{I},d} \right\| \leqslant O\left(\sqrt{p \cdot 2^k \cdot n^{k - \lfloor d/2 \rfloor}} + 1 \right) \sqrt{\log n}, .$$

Proof. Let $0 \leqslant \tau \leqslant O(1)$ be the largest coefficient in absolute value in the polynomial $P(z)$, for any $z \in \{\pm 1\}^k$. For each $\alpha \in [n]^k$ and $c \in \{\pm 1\}^k$ let $\mathbf{M}_{(c,\alpha)}$ be the $n^{\lfloor d/2 \rfloor} \times n^{\lceil d/2 \rceil}$ matrix flattening of $P_d(c \circ x^\alpha)$ so that

$$P_d(c \circ x^\alpha) = \langle x^{\otimes \lfloor d/2 \rfloor}, \mathbf{M}_{(c,\alpha)} x^{\otimes \lceil d/2 \rceil} \rangle ,$$

for any $x \in \mathbb{R}^n$. We use the decomposition

$$M_{\boldsymbol{I},d} = \sum_{(c,\alpha) \in \{\pm 1\}^k \times [n]^k} \mathbf{M}_{(c,\alpha)} .$$

Now each $\mathbf{M}_{(c,\alpha)}$ satisfies, for any $\alpha_1, \alpha_1' \in [n]^{\lfloor d/2 \rfloor}$ and $\alpha_2, \alpha_2' \in [n]^{\lceil d/2 \rceil}$

$$\begin{aligned}
\mathbb{E}\left[\left(\mathbf{M}_{(c,\alpha)} \right)_{\alpha_1 \alpha_2} \right] &= 0 \\
\text{if } (\alpha_1, \alpha_2) \neq \alpha , \quad \mathbb{P}\left[\left(\mathbf{M}_{(c,\alpha)} \right)_{\alpha_1 \alpha_2} \neq 0 \right] &= 0 \\
\text{if } (\alpha_1, \alpha_2) = \alpha , \quad \mathbb{P}\left[\left(\mathbf{M}_{(c,\alpha)} \right)_{\alpha_1 \alpha_2} \neq 0 \right] &\leqslant p \\
\left| \left(\mathbf{M}_{(c,\alpha)} \right)_{\alpha_1 \alpha_2} \right| &\leqslant \tau \\
\text{for } \alpha_1 \neq \alpha_1' \text{ and } \alpha_2 \neq \alpha_2' , \quad \mathbb{E}[\left(\mathbf{M}_{(c,\alpha)} \right)_{\alpha_1 \alpha_2} \left(\mathbf{M}_{(c,\alpha)} \right)_{\alpha_1' \alpha_2'}] &= 0 .
\end{aligned}$$

We can thus apply Bernstein's inequality for matrices as in Theorem D.7. We have

$$\begin{aligned}
\sigma^2 &:= \max\left\{ \left\| \sum_{(c,\alpha)} \mathbb{E}\, \mathbf{M}_{(c,\alpha)} \mathbf{M}_{(c,\alpha)}^\mathsf{T} \right\| , \left\| \sum_{(c,\alpha)} \mathbb{E}\, \mathbf{M}_{(c,\alpha)}^\mathsf{T} \mathbf{M}_{(c,\alpha)} \right\| \right\} \\
&\leqslant p \cdot \tau^2 \cdot 2^k \cdot n^{k - \lfloor d/2 \rfloor} .
\end{aligned}$$

Thus choosing $t = O\left(\sqrt{p \cdot \tau^2 \cdot 2^k \cdot n^{k - \lfloor d/2 \rfloor}} + \tau \right) \sqrt{\log n}$ the result follows. □

We can now prove the main theorem of the section.

Proof of Theorem 6.35. As for Theorem 6.21, we focus on the case $p \leqslant n^{-k/2} \operatorname{polylog}(n)$ as for larger p results are known already [AOW15]. Moreover with high probability the number of clauses in $\boldsymbol{I}$ is $m = (1 + o(1)) p n^k$. For any $x \in \{\pm 1\}^n$ we have

$$\operatorname{Val}_{\boldsymbol{I}}(x) = \mathop{\mathbb{E}}_{\mathbf{z} \overset{u.a.r}{\sim} \{\pm 1\}^k} [P(\mathbf{z})] + \frac{1}{m} \sum_{1 \leqslant d \leqslant k} S_{\boldsymbol{I},d}(x) .$$

By Lemma 6.36, with probability at least $1 - kn^{-\Omega(1)} = 1 - o(1)$,

$$\begin{aligned}
\max_{x\in\{\pm1\}^n} \frac{1}{m}\sum_{1\leqslant d<k} S_{\mathcal{I},d}(x) &\leqslant \frac{1}{m}\sum_{1\leqslant d<k} n^{d/2}\cdot\left\|M_{\mathcal{I},d}\right\| \\
&\leqslant \frac{1}{m}\sum_{1\leqslant d<k} n^{d/2}\cdot O\left(\sqrt{p\cdot 2^k\cdot n^{k-\lfloor d/2\rfloor}}+1\right)\sqrt{\log n} \\
&\leqslant \sum_{1\leqslant d<k} n^{-k/2+d/2-k/4+k/2-\frac{\lfloor d/2\rfloor}{2}}\cdot 2^{k/2}\cdot \mathrm{polylog}(n) \\
&\leqslant O\left(n^{-\frac{k}{4}+\frac{\lceil d/2\rceil}{2}}\right)\mathrm{polylog}(n) \\
&\leqslant o(1)\,.
\end{aligned}$$

Applying Theorem 6.22 to $\max_{x\in\{\pm1\}^n} S_{\mathcal{I},k}(x)$ the desired bound on $\mathrm{Opt}_{\mathcal{I}}$ follows. Finally, we also immediately obtain a polynomial time applying any efficient method to certify the spectral norm of $\sum_{d<k} M_{\mathcal{I},d}$. □

6.7 Algorithm for k-XOR with adversarial signs

In this section, we exploit the bounds obtained in Section 6.5 to design a polynomial time algorithm that, given a semi-random k-XOR instance $\mathcal{I}$ with n variables and *arbitrary* (possibly adversarial) signs, achieves a $(1-\varepsilon)$-approximation in time $n^{O(k/\varepsilon^2)}$. We start by defining the semi-random model of interest.

Definition 6.37 (Semi-random k-XOR)**.** A semi-random k-XOR instance $\mathcal{I}$ with n variables and $m := p\binom{n}{k}(1\pm o(1))$ clauses can be generated through the following process:

(i) Pick a random symmetric tensor $\mathbf{T}'$, with independent entries, such that $\mathbf{T}'_\alpha = 0$ if the indices in the multi-index $\alpha \in [n]^k$ are not distinct and otherwise:

$$\mathbf{T}'_\alpha = \begin{cases} 0 & \text{with probability } 1-p\,, \\ +1 & \text{with probability } p/2\,, \\ -1 & \text{with probability } p/2\,. \end{cases}$$

(ii) Given $\mathbf{T}'$, pick an arbitrary binary function (possibly chosen adversarially from $\mathbf{T}'$) $\sigma : [n]^k \to \{\pm1\}$ and let T be the tensor with entries $T_\alpha := \sigma(\alpha)\mathbf{T}'_\alpha$.

$\mathcal{I}$ consists of the k-XOR predicates $k\text{-XOR}(\alpha) = \frac{1-x^\alpha(-T)_\alpha}{2} = \frac{1-\sigma(\alpha)\cdot x^\alpha(-\mathbf{T}')_\alpha}{2}$ where $\mathbf{T}'_\alpha$ is non-zero.

For a semi-random instance $\mathcal{I}$, we denote by $m(\mathcal{I})$ the exact number of clauses in the instance. For convenience, we denote by $\mathbf{T}'_{\mathcal{I}}$ and $\sigma_{\mathcal{I}}$ respectively the random tensor and the adversarial binary function associated to $\mathcal{I}$. We also use $T_{\mathcal{I}}$ to denote the tensor with

entries $(T_{\mathcal{I}})_\alpha := \sigma(\alpha)(\mathbf{T}'_{\mathcal{I}})_\alpha$. When the context is clear we drop the subscripts. The max k-XOR problem is that of finding an assignment with value

$$\mathrm{Opt}_{\mathcal{I}} := \frac{1}{m(\mathcal{I})} \max_{x\in\{\pm 1\}^n} \sum_{\alpha\in[n]^k} T_\alpha x^\alpha = \frac{1}{m(\mathcal{I})} \max_{x\in\{\pm 1\}^n} \sum_{\alpha\in[n]^k} \sigma(\alpha)\cdot \mathbf{T}'_\alpha x^\alpha .$$

With the above objective and adversarial model and in mind, we prove the following theorem, which implies Theorem 6.3.

Theorem 6.38. *Let n, k be positive integers, $\varepsilon > 0$, n and $n^{-k/2}/\varepsilon^2 < p(n) := p < 1$. Let $\mathcal{I}$ be a semi-random k-XOR instance with parameters n, p as in Definition 6.37. There exists a randomized algorithm (Algorithm 6.41), running in time $n^{O(k/\varepsilon^2)}$, that returns an assignment $\hat{\mathbf{x}}$ satisfying*

$$\mathrm{Val}_{\mathcal{I}}(\hat{\mathbf{x}}) \geqslant \mathrm{Opt}_{\mathcal{I}} - O(\varepsilon),$$

with probability at least 0.99.

Before proving the result, we introduce some additional notation. We focus on the settings with odd k, since for even k Theorem 6.38 is implied by [AJT19]. We denote by $\mathcal{Q}_t$ the set of degree-t pseudo-distributions in indeterminates $x_1, \dots, x_n$ satisfying

$$\left\{x_i^2 = 1\,, \forall i \in [n]\right\}.$$

We write $x = (x_1, \dots, x_n)$. Let $\mathcal{I}$ be a k-XOR instance for odd $k \geqslant 3$ and let $T_{\mathcal{I}} \in \mathbb{R}^{n^{\otimes k}}$ be the associated tensor. As usual, when the context is clear we drop the subscript. Throughout the rest of the section we also use the symbols α, β to denote multi-indices in $[n]^q$, for some $q \in \mathbb{N}$. For an index $\ell \in [n]$ and multi-indices $\alpha, \alpha' \in [n]^{\frac{k-1}{2}}$, we write $(\boldsymbol{\alpha}, \boldsymbol{\alpha}', \boldsymbol{\ell}) \sim T$ to denote an entry picked uniformly at random among the *non-zero* entries of T. For an assignment $x \in \{\pm 1\}^n$, $\mathbb{E}_{(\boldsymbol{\alpha},\boldsymbol{\alpha}',\boldsymbol{\ell})\sim T} x^{(\boldsymbol{\alpha},\boldsymbol{\alpha}',\boldsymbol{\ell})}$ is the expectation of the monomial $x^{(\boldsymbol{\alpha},\boldsymbol{\alpha}',\boldsymbol{\ell})}$ for the uniform distribution over all *non-zero* entries of T. For a fixed $\ell \in [n]$, we write $(\boldsymbol{\alpha}, \boldsymbol{\alpha}') \sim T_\ell$ to denote a non-zero entry in T picked uniformly at random among those containing index ℓ. Furthermore, we denote by $D(T)$ the distribution over $[n]$ such that, for each $\ell \in [n]$, its probability is proportional to the number of non-zero entries in T with index ℓ. Therefore, we have

$$\mathop{\mathbb{E}}_{(\boldsymbol{\alpha},\boldsymbol{\alpha}',\boldsymbol{\ell})\sim T} x^{(\boldsymbol{\alpha},\boldsymbol{\alpha}',\boldsymbol{\ell})} = \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim D(T)} \mathop{\mathbb{E}}_{(\boldsymbol{\alpha},\boldsymbol{\alpha}')\sim T_{\boldsymbol{\ell}}} x^{(\boldsymbol{\alpha},\boldsymbol{\alpha}',\boldsymbol{\ell})}.$$

Finally, we introduce the following crucial definitions.

Definition 6.39 (Local correlation)**.** Let $t \geqslant 2$. Let $T \in \mathbb{R}^{n^{\otimes k}}$ be a symmetric tensor. Let $\mu \in \mathcal{Q}_t$, we define the local correlation of μ on T to be

$$\mathrm{LC}_T(\mu) := \mathop{\mathbb{E}}_{\boldsymbol{\ell}\sim D(T)} \mathop{\mathbb{E}}_{\substack{(\boldsymbol{\alpha},\boldsymbol{\alpha}')\sim T_{\boldsymbol{\ell}} \\ (\boldsymbol{\beta},\boldsymbol{\beta}')\sim T_{\boldsymbol{\ell}}}} \tilde{\mathbb{E}}\left(x^{(\boldsymbol{\alpha},\boldsymbol{\alpha}',\boldsymbol{\beta},\boldsymbol{\beta}')} - 2x^{(\boldsymbol{\alpha},\boldsymbol{\alpha}')}\tilde{\mathbb{E}}x^{\boldsymbol{\beta}}\tilde{\mathbb{E}}x^{\boldsymbol{\beta}'} + \tilde{\mathbb{E}}x^{\boldsymbol{\alpha}}\tilde{\mathbb{E}}x^{\boldsymbol{\alpha}'}\tilde{\mathbb{E}}x^{\boldsymbol{\beta}}\tilde{\mathbb{E}}x^{\boldsymbol{\beta}'}\right)$$

For an instance $\mathcal{I}$ as in Definition 6.37 we interchangeably use $\mathrm{LC}_{\mathcal{I}}(\mu)$, $\mathrm{LC}_{\mathbf{T}'}(\mu)$ and $\mathrm{LC}_{T}(\mu)$.

Definition 6.40 (Global correlation)**.** Let $t \geqslant 2$. Let $\mu \in \mathcal{Q}_t$, we define the global correlation of μ to be

$$\mathrm{GC}(\mu) := \mathop{\mathbb{E}}_{\ell \sim [n]} \mathop{\mathbb{E}}_{\substack{(\alpha,\alpha') \sim [n]^{k-1} \\ (\beta,\beta') \in [n]^{k-1}}} \left| \mathrm{Cov}_{\mu}\left(x^{(\alpha,\alpha')}, x^{(\beta,\beta')}\right) \right| + \left| \mathrm{Cov}_{\mu}\left(x^{\alpha}, x^{\beta}\right) \right| + \left| \mathrm{Cov}_{\mu}\left(x^{\alpha'}, x^{\beta'}\right) \right| .$$

Our proof of Theorem 6.38 will be inspired by the *local correlation to global correlation* approach of [BRS11]. First, we show that if we can find a pseudo-distribution in $\mathcal{Q}_t$ with local correlation at most ε^2, then we can obtain the desired approximation. Second, we argue that we can always find a pseudo-distribution in $\mathcal{Q}_t$ with low global correlation. Finally, using Theorem 6.22, we show that high local correlation implies high global correlation. This will yield the desired result.

We start by present the algorithm behind Theorem 6.38. Its correctness is then analyzed in the subsequent sections.

Algorithm 6.41 (Algorithm for semi-random k-XOR)**.**
Input: A k-XOR instance $\mathcal{I}$ as in Definition 6.37, $\varepsilon > 0$, $t \geqslant \Omega(k/\varepsilon^2)$.
Output: assignment $\hat{\mathbf{x}} \in \{\pm 1\}^n$
Operations:

1. Find a pseudo-distribution $\mu \in \mathcal{Q}_t$ maximizing $\mathbb{E}_{(\alpha,\alpha',\ell)\sim T} \tilde{\mathbb{E}}_{\mu} x^{(\alpha,\alpha',\ell)}$, for large enough $t \geqslant \Omega(k/\varepsilon^2)$.
2. If $\mathrm{GC}(\mu) > \varepsilon^2$, let μ' be the pseudo-distribution returned by Algorithm 6.43 on input μ.
3. For each $i \in [n]$, set $\hat{\mathbf{x}}_i = 1$ with probability $\frac{1+\tilde{\mathbb{E}}_{\mu'} x_i}{2}$ and -1 otherwise. Return $\hat{\mathbf{x}}$.

Remark 6.42 (Running time). Finding a pseudo-distribution in $\mathcal{Q}_t$ requires time $n^{O(k/\varepsilon^2)}$. Step 2 is repeated at most $O(1)$ times. The first part can be checked in time $n^{O(k/\varepsilon^2)}$ and the second depends on the running time R of Algorithm 6.43. Finally, step 3 requires time $O(n)$. All in all the algorithm takes time $n^{O(k/\varepsilon^2)} + R$.

A crucial building block of Algorithm 6.41 is the subroutine below, used to find a pseudo-distribution in $\mathcal{Q}_t$ with low global correlation. The idea behind its approach is that, conditioning our pseudo-distribution, we can obtain a new pseudo-distribution "significantly closer" to a product distribution. We analyze the guarantees of Algorithm 6.43 in Section 6.7.2.

Algorithm 6.43 (Driving down global correlation)**.**
Input: pseudo-distribution $\mu \in \mathcal{Q}_t$
Output: pseudo-distribution $\mu' \in \mathcal{Q}_{t-k/\varepsilon^2}$
Operations:

0. Set $\mu' = \mu$. Sequentially repeat for both $c \in \left\{k-1, \frac{k-1}{2}\right\}$ on input μ'.

   (a) Pick uniformly at random $\alpha_1, \ldots, \alpha_{C/\varepsilon^2} \in [n]^c$, for a large enough constant C.

   (b) Sequentially set

$$x^{\alpha_i} = \begin{cases} 1\,, & \text{with prob. } \frac{1+\tilde{\mathbb{E}}_\mu[x^{\alpha_i} | x^{\alpha_1}, \ldots, x^{\alpha_{i-1}}]}{2} \\ -1\,, & \text{otherwise.} \end{cases}$$

   (c) For $i \in [C/\varepsilon^2]$, let μ_i be the pseudo-distribution obtained from μ' conditioning on the sampled values of $x^{\alpha_1}, \ldots, x^{\alpha_{i-1}}$.

   (d) Find μ' among $\{\mu_i\}_{i \in [C/\varepsilon^2]}$ minimizing

$$\mathop{\mathbb{E}}_{\alpha, \alpha' \in [n]^c} \left[\mathrm{Cov}_{\mu'} \left(x^\alpha, x^{\alpha'} \right)^2 \right].$$

1. Return μ'.

Remark 6.44 (Running time). It suffices to consider steps (a)-(d). The first two steps of the algorithm require time $O(n^{O(k)} C/\varepsilon^2)$. In the third step we can compute each μ_i in time $n^{O(k/\varepsilon^2)}$ and there are $O(C/\varepsilon^2)$ of them. The last step also takes time $n^{O(k/\varepsilon^2)}$.

6.7.1 Rounding with low local correlation

We show here that, given a pseudo-distribution with low local correlation, the rounding step provides nearly optimal guarantees.

Lemma 6.45 (Low local correlation rounding)**.** *Consider the settings of Theorem 6.38. Let $t \geqslant 2k$ and $\mu \in \mathcal{Q}_t$. Let $\mathcal{I}$ be a semi-random k-XOR instance with parameters n, p as in Definition 6.37. Let T be the associated tensor. Suppose $\mathrm{LC}_T(\mu) \leqslant \varepsilon^2$. Then the last step in Algorithm 6.41 outputs an assignment $\hat{\mathbf{x}}$ satisfying*

$$\mathrm{Val}_{\mathcal{I}}(\hat{\mathbf{x}}) \geqslant \mathrm{Opt}_{\mathcal{I}} - O(\varepsilon)\,,$$

with probability at least 0.99.

Proof. By Markov's inequality, it will suffices to show that

$$\mathop{\mathbb{E}}_{(\alpha,\alpha',\ell)\sim T} \sigma(\alpha,\alpha',\ell)\left(\tilde{\mathbb{E}}x^{(\alpha,\alpha',\ell)} - \tilde{\mathbb{E}}x^{\alpha}\tilde{\mathbb{E}}x^{\alpha'}\tilde{\mathbb{E}}x^{\ell}\right) \leqslant O(\varepsilon)\,.$$

So applying Fact 2.12, since $t \geqslant 2k$ and for any $0 \leqslant c \leqslant 1$ we have $c^2 \leqslant c$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(\alpha,\alpha',\ell)\sim T} \sigma(\alpha,\alpha',\ell)\left(\tilde{\mathbb{E}}x^{(\alpha,\alpha',\ell)} - \tilde{\mathbb{E}}x^{\alpha}\tilde{\mathbb{E}}x^{\alpha'}\tilde{\mathbb{E}}x^{\ell}\right)\\
&\quad= \tilde{\mathbb{E}} \mathop{\mathbb{E}}_{\ell\sim D(T)} \mathop{\mathbb{E}}_{(\alpha,\alpha')\sim T_\ell} \sigma(\alpha,\alpha',\ell)\left(x^{(\alpha,\alpha',\ell)} - \tilde{\mathbb{E}}x^{\alpha}\tilde{\mathbb{E}}x^{\alpha'}\tilde{\mathbb{E}}x^{\ell}\right)\\
&\quad= \tilde{\mathbb{E}} \mathop{\mathbb{E}}_{\ell\sim D(T)} \mathop{\mathbb{E}}_{(\alpha,\alpha')\sim T_\ell} \sigma(\alpha,\alpha',\ell)x^{\ell}\left(x^{(\alpha,\alpha')} - \tilde{\mathbb{E}}x^{\alpha}\tilde{\mathbb{E}}x^{\alpha'}\right)\\
&\quad\leqslant \sqrt{\tilde{\mathbb{E}} \mathop{\mathbb{E}}_{\ell\sim D(T)}\left(\mathop{\mathbb{E}}_{(\alpha,\alpha')\sim T_\ell} \sigma(\alpha,\alpha',\ell)x^{\ell}\right)^2}\sqrt{\tilde{\mathbb{E}} \mathop{\mathbb{E}}_{\ell\sim [n]}\left(\mathop{\mathbb{E}}_{(\alpha,\alpha')\sim T_\ell} x^{(\alpha,\alpha')} - \tilde{\mathbb{E}}x^{\alpha}\tilde{\mathbb{E}}x^{\alpha'}\right)^2}\\
&\quad\leqslant \sqrt{\tilde{\mathbb{E}} \mathop{\mathbb{E}}_{\ell\sim D(T)}\left(\mathop{\mathbb{E}}_{(\alpha,\alpha')\sim T_\ell} x^{(\alpha,\alpha')} - \tilde{\mathbb{E}}x^{\alpha}\tilde{\mathbb{E}}x^{\alpha'}\right)^2}\\
&\quad= \mathrm{LC}_T(\mu)^{1/2}\,.
\end{aligned}$$

□

6.7.2 Driving down global correlation

Here we show that, via Algorithm 6.43, we can always efficiently obtain a pseudo-distribution in $\mathcal{Q}_t$ with low global correlation. Concretely, we prove the following statement.

Lemma 6.46 (Driving down global correlation)**.** *Consider the settings of Theorem 6.38. Let $t \geqslant C \cdot k/\varepsilon^2$, for a large enough constant C. Let $\mu \in \mathcal{Q}_t$ be the pseudo-distribution in input for Step 2 of Algorithm 6.41 and let $\mu' \in \mathcal{Q}_k$ be its output. Then with probability at least 0.998, it holds that $\mathrm{GC}(\mu') \leqslant \varepsilon^2$.*

We need an intermediate observation to prove the lemma: a simple, yet crucial, statement about the covariance of (pseudo)distributions after conditioning [BRS11, RT12, MR16].

Lemma 6.47 (See [Sch22])**.** *Let $0 \leqslant q < t-2$ and let $0 < c < (t-q)/2$. Let μ be a pseudodistribution in $\mathcal{Q}_t$. Let $\mathbf{i}_1, \ldots, \mathbf{i}_\ell$ be indices sampled uniformly at random from $[n]$ without replacement. Suppose we sequentially set, for all $q \leqslant \ell$,*

$$\mathbf{x}_{\mathbf{i}_q} = \begin{cases} 1\,, & \text{with prob. } \frac{1+\tilde{\mathbb{E}}_\mu\left[\mathbf{x}_{\mathbf{i}_q}\middle|\mathbf{x}_{\mathbf{i}_1},\ldots,\mathbf{x}_{\mathbf{i}_{q-1}}\right]}{2} \\ -1\,, & \text{otherwise.}\end{cases}$$

Then there exists some $q \leqslant \ell$ such that

$$\mathop{\mathbb{E}}_{\mathbf{x}_{\mathbf{i}_1},\ldots,\mathbf{x}_{\mathbf{i}_q}} \mathop{\mathbb{E}}_{\alpha,\alpha'\in[n]^c}\left(\tilde{\mathbb{E}}_\mu\left[\mathbf{x}^{\alpha}\mathbf{x}^{\alpha'} - \tilde{\mathbb{E}}\mathbf{x}^{\alpha}\tilde{\mathbb{E}}\mathbf{x}^{\alpha'} \;\middle|\; \mathbf{x}_{\mathbf{i}_1},\ldots\mathbf{x}_{\mathbf{i}_q}\right]^2\right) \leqslant \frac{2\log 2}{\ell}\,.$$

We can use this result to argue that Algorithm 6.43 decreases global correlation. Thus proving Lemma 6.46.

Proof of Lemma 6.46. Consider without loss of generality the iteration of Algorithm 6.43 with $c = k-1$. The argument for $c = \frac{k-1}{2}$ is analogous. Let $\mu \in Q_t$ be the pseudo-distribution in input and let $\mu' \in Q_{t-k/\varepsilon^2}$ be its output. It suffices to show that for a large enough constant $C > 1$, chosen by Algorithm 6.43, with probability at least 0.999, it holds:

$$\mathop{\mathbb{E}}_{\alpha,\alpha'\sim[n]^{k-1}} \left[\mathrm{Cov}_{\mu'} \left(x^{\alpha}, x^{\alpha'} \right)^2 \right] \leqslant \frac{1}{10\varepsilon^2} .$$

This follows immediately by Lemma 6.47 and Markov's inequality.By union bound over the two iterations of the algorithm, the result follows. □

6.7.3 From local correlation to global correlation

In this section we obtain Theorem 6.38. We start introducing some additional notation. Let $q \in \mathbb{N}$, let $\mu \in Q_t$ for some $t \geqslant 2q$. For each $\alpha, \alpha' \in [n]^q$, let $y^{(\alpha,\alpha')} = x^{(\alpha,\alpha')} - \tilde{\mathbb{E}}_{\mu} x^{\alpha} \tilde{\mathbb{E}}_{\mu} x^{\alpha'}$. Furthermore, recall that we may specify any semi-random instance $\mathcal{I}$ by the pair $(\sigma, \mathbf{T}')$ where $\sigma : [n]^k \to \{\pm 1\}$ and $\mathbf{T}'$ is a random symmetric tensor in $\mathbb{R}^{n^{\otimes q}}$, as specified in Definition 6.37. Then we may write

$$\begin{aligned}
\mathrm{LC}_{\mathbf{T}'}(\mu) &= \mathop{\mathbb{E}}_{\ell\in D(\mathbf{T}')} \mathop{\mathbb{E}}_{\substack{(\alpha,\alpha')\sim\mathbf{T}'_\ell \\ (\beta,\beta')\sim\mathbf{T}'_\ell}} \tilde{\mathbb{E}} \left(x^{(\alpha,\alpha',\beta,\beta')} - 2x^{(\alpha,\alpha')} \tilde{\mathbb{E}} x^{\beta} \tilde{\mathbb{E}} x^{\beta'} + \tilde{\mathbb{E}} x^{\alpha} \tilde{\mathbb{E}} x^{\alpha'} \tilde{\mathbb{E}} x^{\beta} \tilde{\mathbb{E}} x^{\beta'} \right) \\
&= \mathop{\mathbb{E}}_{\ell\sim D(\mathbf{T}')} \tilde{\mathbb{E}} \left(\mathop{\mathbb{E}}_{(\alpha,\alpha')\sim\mathbf{T}'_\ell} y^{(\alpha,\alpha')} \right)^2 \\
&= \mathop{\mathbb{E}}_{\ell\sim D(\mathbf{T}')} \mathop{\mathbb{E}}_{\substack{(\alpha,\alpha')\sim\mathbf{T}'_\ell \\ (\beta,\beta')\sim\mathbf{T}'_\ell}} \tilde{\mathbb{E}} y^{(\alpha,\alpha')} y^{(\beta,\beta')} . \qquad (6.7.1)
\end{aligned}$$

Now, Eq. (6.7.1) can be rewritten as the matrix product $\langle \mathbf{A}, X \rangle$ where $X \in \mathbb{R}^{n^{k-1} \times n^{k-1}}$ is the positive semidefinite matrix with entries $X_{(\alpha,\beta),(\alpha',\beta')} := \tilde{\mathbb{E}} y^{(\alpha,\alpha')} y^{(\beta,\beta')}$ and $\mathbf{A}$ is a random matrix with entries

$$\mathbf{A}_{(\alpha,\beta),(\alpha',\beta')} \propto \sum_{\ell} \mathbf{T}'_{(\alpha,\alpha',\ell)} \mathbf{T}'_{(\beta,\beta',\ell)} . \qquad (6.7.2)$$

Similarly, we may rewrite

$$\mathop{\mathbb{E}}_{\ell\sim[n]} \mathop{\mathbb{E}}_{\substack{(\alpha,\alpha')\sim[n]^{k-1} \\ (\beta,\beta')\sim[n]^{k-1}}} \tilde{\mathbb{E}} \left(x^{(\alpha,\alpha',\beta,\beta')} - 2x^{(\alpha,\alpha')} \tilde{\mathbb{E}} x^{\beta} \tilde{\mathbb{E}} x^{\beta'} + \tilde{\mathbb{E}} x^{\alpha} \tilde{\mathbb{E}} x^{\alpha'} \tilde{\mathbb{E}} x^{\beta} \tilde{\mathbb{E}} x^{\beta'} \right) = \langle \bar{J}, X \rangle ,$$

where $\bar{J}$ is the normalization of the all-ones matrix J. We can now use this notation to prove the following two statements, which will allow us to relate local and global correlation.

Lemma 6.48. *Consider the settings of Theorem 6.38. Let $X \in \mathbb{R}^{n^{k-1}\times n^{k-1}}$ be a positive semi-definite matrix satisfying $\|X\|_{\max} \leqslant O(1)$. For a semi-random instance $\mathcal{I}$ as in Definition 6.37, let $\mathbf{A}$ be the associate matrix as defined in Eq. (6.7.2). Then, with probability at least 0.999*

$$\langle \mathbf{A} - \bar{J}, X\rangle \leqslant O(\varepsilon^2)\,.$$

Proof. Recall, for any matrix M we used the notation $\|M\|_{\mathrm{Gr}} = \max\{\langle M, X\rangle \mid X \succeq 0, X_{ii} \leqslant 1\,, \forall i \in [n]\}$. Observe that by Cauchy-Schwarz and Grothendieck's inequality

$$\langle \mathbf{A} - \bar{J}, X\rangle \leqslant \|X\|_{\max} \cdot \left\|\mathbf{A} - \bar{J}\right\|_{\mathrm{Gr}} \leqslant O(\left\|\mathbf{A} - \bar{J}\right\|_{\infty\to 1})\,.$$

Thus it remains to bound $\left\|\mathbf{A} - \bar{J}\right\|_{\infty\to 1}$. Define $\mathbf{A}^*$ to the be matrix with entries

$$\mathbf{A}^*_{(\alpha,\beta),(\alpha',\beta')} = \begin{cases} (\mathbf{A} - \bar{J})_{(\alpha,\beta),(\alpha,\beta')} & \text{if } |S(\alpha,\alpha') \cap S(\beta,\beta')| = \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

By the triangle inequality

$$\left\|\mathbf{A} - \bar{J}\right\|_{\infty\to 1} \leqslant \|\mathbf{A}^*\|_{\infty\to 1} + \left\|\mathbf{A} - \bar{J} - \mathbf{A}^*\right\|_{\infty\to 1}\,.$$

Now

$$\begin{aligned}\left\|\mathbf{A} - \bar{J} - \mathbf{A}^*\right\|_{\infty\to 1} &\leqslant \sum_{\alpha,\beta,\alpha',\beta} \left|\mathbf{A} - \bar{J} - \mathbf{A}^*\right|_{(\alpha,\beta),(\alpha',\beta')} \\ &\leqslant \sum_{\substack{\alpha,\beta,\alpha',\beta \\ \text{s.t. } S(\alpha,\alpha')\cap S(\beta,\beta')\neq\emptyset}} \left|\mathbf{A}_{(\alpha,\beta),(\alpha',\beta')}\right| + \left|\bar{J}_{(\alpha,\beta),(\alpha',\beta')}\right|\,.\end{aligned}$$

The first term can be shown to be $o(1)$ repeating the argument of Lemma 6.25. The second term is a sum of at most n^{2k-3} elements of value at most $O(n^{-2k+2})$ and thus also $o(1)$. It remains to bound $\|\mathbf{A}^*\|_{\infty\to 1}$. Here $\mathbf{A}^*$ satisfies the premises of Lemma 6.24 and thus for $p \geqslant n^{-k/2}/\varepsilon^2$ we immediately get

$$\left\|\mathbf{A} - \bar{J} - \mathbf{A}^*\right\|_{\infty\to 1} \leqslant O(\varepsilon^2) + o(1)$$

with probability at least 0.999 over the draw of $\mathbf{T}$. □

Next we show that $\langle \bar{J}, X\rangle$ is a lower bound to the global correlation.

Lemma 6.49. *For a semi-random instance $\mathcal{I}$ as in Definition 6.37, let $\mathbf{A}$ be the associate matrix as defined in Eq. (6.7.2). Let μ be a pseudo-distribution of degree at least $2k$, and let $X \in \mathbb{R}^{n^{k-1}\times n^{k-1}}$ be the positive semi-definite matrix satisfying*

$$\mathrm{LC}_{\mathbf{T}}(\mu) = \langle \mathbf{A}, X\rangle\,.$$

Then

$$\langle \bar{J}, X\rangle = \mathop{\mathbb{E}}_{\ell\sim[n]} \mathop{\mathbb{E}}_{\substack{(\boldsymbol{\alpha},\boldsymbol{\alpha}')\sim[n]^{k-1} \\ (\boldsymbol{\beta},\boldsymbol{\beta}')\sim[n]^{k-1}}} \tilde{\mathbb{E}}\left(x^{(\boldsymbol{\alpha},\boldsymbol{\alpha}',\boldsymbol{\beta},\boldsymbol{\beta}')} - 2x^{(\boldsymbol{\alpha},\boldsymbol{\alpha}')}\tilde{\mathbb{E}}x^{\boldsymbol{\beta}}\tilde{\mathbb{E}}x^{\boldsymbol{\beta}'} + \tilde{\mathbb{E}}x^{\boldsymbol{\alpha}}\tilde{\mathbb{E}}x^{\boldsymbol{\alpha}'}\tilde{\mathbb{E}}x^{\boldsymbol{\beta}}\tilde{\mathbb{E}}x^{\boldsymbol{\beta}'}\right) \leqslant \mathrm{GC}(\mu)\,.$$

Proof. For fixed $\ell, \alpha, \alpha', \beta, \beta'$ we may rewrite

$$\begin{aligned}
&\tilde{\mathbb{E}}\Big(x^{(\alpha,\alpha',\beta,\beta')} - 2x^{(\alpha,\alpha')}\tilde{\mathbb{E}}x^{\beta}\tilde{\mathbb{E}}x^{\beta'} + \tilde{\mathbb{E}}x^{\alpha}\tilde{\mathbb{E}}x^{\alpha'}\tilde{\mathbb{E}}x^{\beta}\tilde{\mathbb{E}}x^{\beta'}\Big) \\
&\quad\leqslant \tilde{\mathbb{E}}\Big(x^{(\alpha,\alpha',\beta,\beta')} - x^{(\alpha,\alpha')}\tilde{\mathbb{E}}x^{\beta}\tilde{\mathbb{E}}x^{\beta}\Big) - \mathrm{Cov}_{\mu}\Big(x^{\alpha}, x^{\alpha'}\Big) \\
&\quad\leqslant \tilde{\mathbb{E}}\Big(x^{(\alpha,\alpha',\beta,\beta')} - x^{(\alpha,\alpha')}\tilde{\mathbb{E}}x^{\beta}\tilde{\mathbb{E}}x^{\beta}\Big) + \Big|\mathrm{Cov}_{\mu}\Big(x^{\alpha}, x^{\alpha'}\Big)\Big|.
\end{aligned}$$

Furthermore

$$\begin{aligned}
&\tilde{\mathbb{E}}\Big(x^{(\alpha,\alpha',\beta,\beta')} - x^{(\alpha,\alpha')}\tilde{\mathbb{E}}x^{\beta}\tilde{\mathbb{E}}x^{\beta}\Big) \\
&\quad= \tilde{\mathbb{E}}\Big(x^{(\alpha,\alpha',\beta,\beta')} - x^{(\alpha,\alpha')}\tilde{\mathbb{E}}x^{\beta}\tilde{\mathbb{E}}x^{\beta}\Big) + \tilde{\mathbb{E}}x^{(\alpha,\alpha')}\tilde{\mathbb{E}}x^{(\beta,\beta')} - \tilde{\mathbb{E}}x^{(\alpha,\alpha')}\tilde{\mathbb{E}}x^{(\beta,\beta')} \\
&\quad= \mathrm{Cov}_{\mu}\Big(x^{(\alpha,\alpha'),x^{(\beta,\beta')}}\Big) - \tilde{\mathbb{E}}x^{(\alpha,\alpha')}\mathrm{Cov}_{\mu}\Big(x^{\beta}, x^{\beta'}\Big) \\
&\quad\leqslant \Big|\mathrm{Cov}_{\mu}\Big(x^{(\alpha,\alpha'),x^{(\beta,\beta')}}\Big)\Big| + \Big|\mathrm{Cov}_{\mu}\Big(x^{\beta}, x^{\beta'}\Big)\Big|.
\end{aligned}$$

The result follows. □

We are finally ready to prove the theorem.

Proof of Theorem 6.38. Let $\boldsymbol{\mu}$ be the pseudo-distribution used by Algorithm 6.41 in the last step. By Lemma 6.46, with probability at least 0.999, it satisfies $\mathrm{GC}(\boldsymbol{\mu}) \leqslant \varepsilon^2$. Combining Lemma 6.48 and Lemma 6.49 it follows that $\mathrm{LC}_{\mathcal{I}}(\boldsymbol{\mu}) \leqslant O(\varepsilon^2)$. We obtain the desired result applying Lemma 6.45. □

6.8 Algorithm for CSPs with adversarial signs patterns

We *sketch* here how to extend Theorem 6.38 to arbitrary predicates –with adversarial sign patterns– on k Boolean variables. We start by introducing the model.

Definition 6.50. Let $P : \{-1,1\}^k \to \{0,1\}$. A semi-random k-CSP instance $\mathcal{I}$ with n variables and $m := p \cdot 2^k \cdot n^k(1 \pm o(1))$ constraints can be generated as follows.

(i) Pick independently with probability p each pair $(\mathbf{c}', \boldsymbol{\alpha})$ where $\mathbf{c}'$ is a random negation pattern from $\{\pm 1\}^k$ and $\boldsymbol{\alpha}$ is a multi-index from $[n]^k$,

(ii) Given the m pairs, replace each such $\mathbf{c}'$ with an arbitrary, possibly adversarially chosen, negation pattern c.

(iii) For each pair $(c, \boldsymbol{\alpha})$ add the constraint $P(c \circ x^{\boldsymbol{\alpha}}) = 1$ to $\mathcal{I}$.

We prove the following theorem, which implies Theorem 6.4.

Theorem 6.51. *Let n, k be positive integers, $\varepsilon > 0, n$ and $n^{-k/2}/\varepsilon^2 < p(n) := p < 1$. Let $P : \{\pm 1\}^k \to \{0,1\}$ be a predicate. Let $\mathcal{I}$ be a CSP(P) instance with parameters n, p as in Definition 6.50. There exists a randomized algorithm (Algorithm 6.52), running in time $n^{O(k^2/\varepsilon^2)}$, that returns an assignment $\hat{\mathbf{x}}$ satisfying*

$$\mathrm{Val}_{\mathcal{I}}(\hat{\mathbf{x}}) \geqslant \mathrm{Opt}_{\mathcal{I}} - O(\varepsilon),$$

with probability at least 0.99.

Given a predicate $P : \{\pm 1\}^k \to 0,1$, for each $c \in \{\pm 1\}^k$ and $\alpha \in [n]^k$ we may rewrite

$$P(c \circ \alpha) = \sum_{\beta \subseteq \alpha} q_\beta \cdot \sigma(\beta) \cdot x^\beta,$$

where q_β is a constant coefficient and $\sigma(\beta) \in \{\pm 1\}$. For an instance $\mathcal{I}$, and $d \leqslant k$, let $M(d, \mathcal{I}) \in \mathbb{R}^{n^{\otimes d}}$ be the tensor with entries $M_\beta(d, \mathcal{I}) = q$ where q is the *number* of pairs such that $(c, \alpha) \in \mathcal{I}$ with $\beta \subseteq \alpha$ and the corresponding coefficient q_β in $P(c \circ \alpha)$ is non-zero. Furthermore, let $D(\mathcal{I}, d)$ be the distribution over indices in $[n]$ such that the probability of $\ell \in [n]$ is proportional to the fraction of pairs $(c, \alpha) \in \mathcal{I}$ with $\ell \subseteq \alpha$. Then we write $T(d, \mathcal{I})_\ell$ for the set of multi-indeces (α, ℓ) corresponding to non-zero in $T(d, \mathcal{I})$. We can now introduce new notions of local and global correlation. Let $\mu \in \mathcal{Q}_t$ for some $t \geqslant 2k$. For even $d \leqslant k$ let

$$\mathrm{LC}_{\mathcal{I}}(\mu, d) := \left(\mathop{\mathbb{E}}_{(\boldsymbol{\alpha}, \boldsymbol{\alpha}') \sim T(d, \mathcal{I})} \left| \mathrm{Cov}_\mu \left(x^{\boldsymbol{\alpha}}, x^{\boldsymbol{\alpha}'} \right) \right| \right)^2$$

$$\mathrm{GC}(\mu, d) := \left(\mathop{\mathbb{E}}_{\boldsymbol{\alpha}, \boldsymbol{\alpha}' \sim [n]^{d/2}} \left| \mathrm{Cov}_\mu \left(x^{\boldsymbol{\alpha}}, x^{\boldsymbol{\alpha}'} \right) \right| \right)^2.$$

For odd $d \leqslant k$ let

$$\mathrm{LC}_{\mathcal{I}}(\mu, d) := \mathop{\mathbb{E}}_{\boldsymbol{\ell} \sim D(\mathcal{I}, d)} \mathop{\mathbb{E}}_{\substack{(\boldsymbol{\alpha}, \boldsymbol{\alpha}') \sim T(d, \mathcal{I})_\ell \\ (\boldsymbol{\beta}, \boldsymbol{\beta}') \sim T(d, \mathcal{I})_\ell}} \tilde{\mathbb{E}} \left(x^{(\boldsymbol{\alpha}, \boldsymbol{\alpha}', \boldsymbol{\beta}, \boldsymbol{\beta}')} - 2 x^{(\boldsymbol{\alpha}, \boldsymbol{\alpha}')} \tilde{\mathbb{E}} x^{(\boldsymbol{\beta}, \boldsymbol{\beta}')} + \tilde{\mathbb{E}} x^{\boldsymbol{\alpha}} \tilde{\mathbb{E}} x^{\boldsymbol{\alpha}'} \tilde{\mathbb{E}} x^{\boldsymbol{\beta}} \tilde{\mathbb{E}} x^{\boldsymbol{\beta}'} \right)$$

$$\mathrm{GC}(\mu d) := \mathop{\mathbb{E}}_{\boldsymbol{\ell} \in [n]} \mathop{\mathbb{E}}_{\substack{(\boldsymbol{\alpha}, \boldsymbol{\alpha}') \sim [n]^{d-1} \\ (\boldsymbol{\beta}, \boldsymbol{\beta}') \sim [n]^{d-1}}} \left| \mathrm{Cov}_\mu \left(x^{\boldsymbol{\alpha}}, x^{\boldsymbol{\alpha}'} \right) \right| + \left| \mathrm{Cov}_\mu \left(x^{\boldsymbol{\beta}}, x^{\boldsymbol{\beta}'} \right) \right| + \left| \mathrm{Cov}_\mu \left(x^{(\boldsymbol{\alpha}, \boldsymbol{\alpha}')}, x^{(\boldsymbol{\beta}, \boldsymbol{\beta}')} \right) \right|.$$

Both the algorithm and the proof structure closely resemble the ones used for Theorem 6.38, so we only discuss the steps that differ. We start by presenting the algorithm.

Algorithm 6.52 (Algorithm for semi-random k-XOR)**.**
Input: A k-CSP instance $\mathcal{I}$ as in Definition 6.50, ε.
Output: assignment $\hat{\mathbf{x}} \in \{\pm 1\}^n$
Operations:

1. Find a pseudo-distribution $\mu \in \mathcal{Q}_t$ maximizing $\mathbb{E}_{(c,\boldsymbol{\alpha})\sim\mathcal{I}} \tilde{\mathbb{E}}_\mu P(c \circ \boldsymbol{\alpha})$, for large enough $t \geqslant \Omega(k/\varepsilon^2)$.
2. If $\mathrm{GC}(\mu) > \varepsilon^2$, let $\boldsymbol{\mu}'$ be the pseudo-distribution returned by Algorithm 6.53 on input μ.
3. For each $i \in [n]$, set $\hat{\mathbf{x}}_i = 1$ with probability $\frac{1+\tilde{\mathbb{E}}_{\mu'} x_i}{2}$ and -1 otherwise. Return $\hat{\mathbf{x}}$.

Algorithm 6.53 (Driving down global correlation)**.**
Input: pseudo-distribution $\mu \in \mathcal{Q}_t$
Output: pseudo-distribution $\boldsymbol{\mu}' \in \mathcal{Q}_{t-k/\varepsilon^2}$
Operations:

0. Set $\mu' = \mu$. Sequentially repeat for $d \in [k]$ on input μ'.
   (a) Let C be a large enough constant. Let $q = C \cdot k/\varepsilon^2$. Pick uniformly at random $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_q \in [n]^d$.
   (b) Sequentially set
   $$\mathbf{x}^{\boldsymbol{\alpha}_i} = \begin{cases} 1\,, & \text{with prob. } \frac{1+\tilde{\mathbb{E}}_\mu[x^{\boldsymbol{\alpha}_i} | \mathbf{x}^{\boldsymbol{\alpha}_1}, \ldots, \mathbf{x}^{\boldsymbol{\alpha}_{i-1}}]}{2} \\ -1\,, & \text{otherwise.} \end{cases}$$
   (c) For each $i \in [q]$, let $\boldsymbol{\mu}_i$ be the pseudo-distribution obtained from $\boldsymbol{\mu}'$ conditioning on the sampled values of $\mathbf{x}^{\boldsymbol{\alpha}_1}, \ldots, \mathbf{x}^{\boldsymbol{\alpha}_{i-1}}$.
   (d) Find $\boldsymbol{\mu}'$ among $\{\boldsymbol{\mu}_i\}_{i\in[q]}$ minimizing
   $$\mathop{\mathbb{E}}_{\boldsymbol{\alpha},\boldsymbol{\alpha}'\in[n]^d} \left[\mathrm{Cov}_{\mu'} \left(x^{\boldsymbol{\alpha}}, x^{\boldsymbol{\alpha}'} \right)^2 \right].$$
1. Return $\boldsymbol{\mu}'$.

Remark 6.54 (Running time). The running time of steps (a)-(d) is at most $O(n^{k^2/\varepsilon^2})$. The steps are called k times so overall the running time is $O(n^{k^2/\varepsilon^2})$.

Rounding with low local correlation. The next result is the CSP version of Lemma 6.45.

Lemma 6.55 (Low local correlation rounding)**.** *Consider the settings of Theorem 6.51. Let $t \geqslant 2k$ and $\mu \in \mathcal{Q}_t$. Let $\mathcal{I}$ be a semi-random k-CSP instance with parameters n, p as in Definition 6.50. Let T be the associated tensor. Suppose $\sum_{d \leqslant k} \mathrm{LC}_{\mathcal{I}}(\mu, d) \leqslant \varepsilon^2$. Then the last step in Algorithm 6.52 outputs an assignment $\hat{\mathbf{x}}$ satisfying*

$$\mathrm{Val}_{\mathcal{I}}(\hat{\mathbf{x}}) \geqslant \mathrm{Opt}_{\mathcal{I}} - O(\varepsilon)\,,$$

with probability at least 0.99.

Proof. By Markov's inequality it suffices to show that

$$\mathop{\mathbb{E}}_{(\mathbf{c},\boldsymbol{\alpha})\sim\mathcal{I}} \left|\tilde{\mathbb{E}}_{\mu} P(\mathbf{c}\circ\boldsymbol{\alpha}) - P\big(\mathbf{c}\circ\boldsymbol{\alpha},\mu\big)\right| \leqslant O\left(\sum_{d\leqslant k}\sqrt{\mathrm{LC}_{\mathcal{I}}(\mu,d)}\right) \leqslant O(\varepsilon)\,.$$

To do so we may rewrite

$$\mathop{\mathbb{E}}_{(\mathbf{c},\boldsymbol{\alpha})\sim\mathcal{I}} \left|\tilde{\mathbb{E}}_{\mu} P(\mathbf{c}\circ\boldsymbol{\alpha}) - P\big(\mathbf{c}\circ\boldsymbol{\alpha},\mu\big)\right| \leqslant \mathop{\mathbb{E}}_{(\mathbf{c},\boldsymbol{\alpha})\sim\mathcal{I}} \sum_{\beta\subseteq\boldsymbol{\alpha}} O(1)\left|\tilde{\mathbb{E}} x^{\beta} - \prod_{b\in\beta}\tilde{\mathbb{E}} x_b\right|.$$

Now, using a derivation as in Lemma 6.45 the result follows. □

Driving down global correlation. Next we analyze the guarantees of Algorithm 6.53 and obtain a statement resembling Lemma 6.46.

Lemma 6.56 (Driving down global correlation)**.** *Consider the settings of Theorem 6.51. Let $t \geqslant C\cdot k^2/\varepsilon^2$, for a large enough constant C. Let $\mu \in \mathcal{Q}_t$ be the pseudo-distribution in input for Step 2 of Algorithm 6.52 and let $\boldsymbol{\mu}' \in \mathcal{Q}_k$ be its output. Then with probability at least 0.998, it holds that $\sum_{d\leqslant k} \mathrm{GC}(\boldsymbol{\mu}', d) \leqslant \varepsilon^2$.*

Proof. For the iteration with $d = k$, we know by the proof of Lemma 6.46 that with probability at least 0.999 it holds:

$$\mathop{\mathbb{E}}_{\boldsymbol{\alpha},\boldsymbol{\alpha}'\sim[n]^{k-1}}\left[\mathrm{Cov}_{\mu'}\left(x^{\boldsymbol{\alpha}}, x^{\boldsymbol{\alpha}'}\right)^2\right] \leqslant \frac{1}{10\varepsilon^2}\,.$$

So consider now the other iterations. Repeating the analysis as in the previous case, we have that in expectation, for each $d < k$

$$\mathop{\mathbb{E}}_{\boldsymbol{\alpha},\boldsymbol{\alpha}'\sim[n]^{d}}\left[\mathrm{Cov}_{\mu'}\left(x^{\boldsymbol{\alpha}}, x^{\boldsymbol{\alpha}'}\right)^2\right] \leqslant \frac{1}{10000k\varepsilon^2}\,.$$

By linearity of expectations, applying Markov's inequality we get

$$\sum_{d\leqslant k}\mathop{\mathbb{E}}_{\boldsymbol{\alpha},\boldsymbol{\alpha}'\sim[n]^{d}}\left[\mathrm{Cov}_{\mu'}\left(x^{\boldsymbol{\alpha}}, x^{\boldsymbol{\alpha}'}\right)^2\right] \leqslant \frac{1}{10\varepsilon^2}\,,$$

with probability at least 0.998. □

From local correlation to global correlation. As in the case of k-XOR, for odd $d \leqslant k$ we may rewrite

$$\mathrm{LC}_{\mathcal{I},d}(\mu) = \langle \mathbf{A}_d, X_d \rangle$$

for a positive semidefinite matrix X_d and a matrix $\mathbf{A}$ with entries

$$\mathbf{A}_{(\alpha,\beta),(\alpha',\beta')} \propto \sum_{\ell} M_{(\alpha,\alpha,\ell)}(d, \mathcal{I}') \cdot M_{(\beta,\beta',\ell)}(d, \mathcal{I}') \,.$$

Similarly we may write

$$\mathop{\mathbb{E}}_{\ell \sim [n]} \mathop{\mathbb{E}}_{\substack{(\alpha,\alpha') \sim [n]^{k-1} \\ (\beta,\beta') \sim [n]^{k-1}}} \tilde{\mathbb{E}} \Big(x^{(\alpha,\alpha',\beta,\beta')} - 2x^{(\alpha,\alpha')} \tilde{\mathbb{E}} x^{(\beta,\beta')} + \tilde{\mathbb{E}} x^{\alpha} \tilde{\mathbb{E}} x^{\alpha'} \tilde{\mathbb{E}} x^{\beta} \tilde{\mathbb{E}} x^{\beta'} \Big) = \langle \bar{J}, X_d \rangle \,.$$

Moreover, by the analysis in Lemma 6.49, we have for all odd $d \leqslant k$

$$\langle \bar{J}, X_d \rangle \leqslant \mathrm{GC}(\mu, d) \,.$$

It remains to prove an analogue of Lemma 6.48.

Lemma 6.57. *Consider the settings of Theorem 6.38. Let μ be a pseudo-distribution in $\mathcal{Q}_t$ for some $t \geqslant 2k$. Suppose $\mathrm{GC}(\mu, d) \leqslant \varepsilon^2$ for all $d \leqslant k$. Then, with probability at least 0.998 over the randomness of the instance $\mathcal{I}$, for all $d \leqslant k$,*

$$\mathrm{LC}_{\mathcal{I}}(\mu, d) \leqslant O(\varepsilon^2) \,. \tag{6.8.1}$$

Proof. By Lemma 6.57, Eq. (6.8.1) holds with probability 0.999 for $d = k$. Now, for odd $d < k$, a similar analysis combined with the bounds of Lemma 6.36 implies Eq. (6.8.1) with probability $1 - o(1)$. For even d we instead combine Eq. (6.8.1) with the standard local-to-global correlation result (e.g. see Lemma 4.1 in [BRS11]). Taking a union bound over all $d \leqslant k$ the result follows. □

Finally, Theorem 6.51 immediately follows combining Lemma 6.55, Lemma 6.56 and Lemma 6.57.

Part II

Privacy from robustness

Chapter 7

Private algorithms for stochastic block models and mixture models

In this chapter, based on [CCAd+23], we continue our discussion around the relationship between privacy and robustness, proving Theorem 1.9 and Theorem 1.8. We restate the theorems and the state-of-the-art to provide a more detailed comparison.

Stochastic Block model. We consider the two-community model already discussed in Chapter 1, Chapter 4 and Chapter 5. We restate it here in the form of a marginal distribution over graphs, given the vector of communities.[1]

Model 7.1 (Marginal distribution of stochastic block model)**.** The stochastic block model describes the distribution[2] of an n-vertex graph $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, x)$, where x is a vector of n binary labels, $d \in \mathbb{N}$, $\gamma > 0$, and for every pair of distinct vertices $i, j \in [n]$ the edge $\{i, j\}$ is independently added to the graph $\mathbf{G}$ with probability $(1 + \gamma \cdot x_i \cdot x_j)\frac{d}{n}$.

Note that for distinct vertices $i, j \in [n]$, the edge $\{i, j\}$ is present in $\mathbf{G}$ with probability $(1 + \gamma)\frac{d}{n}$ if the vertices have the same label $x_i = x_j$ and with probability $(1 - \gamma)\frac{d}{n}$ if the vertices have different labels $x_i \neq x_j$.[3] In Chapter 4 and Chapter 5 we studied the *weak recovery* problem, which amounts to finding a partition $\hat{x}(\mathbf{G})$ correlated with the true partition. Here we mostly focus on *exact recovery*, where the goal is to actually recover the true partition with high probability. Recall the statistical and computational landscape of these objectives:

[1] Note that, we changed the notation of the parameters, as we reserve ε for differential privacy.

[2] We use **bold** characters to denote random variables.

[3] At times we may write d_n, γ_n to emphasize that these may be functions of n. We write $o(1), \omega(1)$ for functions tending to zero (resp. infinity) as n grows.

	Objective	can be achieved (and efficiently so) *iff*
weak recovery	$\mathbb{P}_{\mathbf{G}\sim\mathsf{SBM}_n(d,\gamma,x)}\left(\frac{1}{n}\lvert\langle x,\hat{x}(\mathbf{G})\rangle\rvert \geqslant \Omega_{d,\gamma}(1)\right) \geqslant 1-o(1)$	$\gamma^2\cdot d \geqslant 1$
exact recovery	$\mathbb{P}_{\mathbf{G}\sim\mathsf{SBM}_n(d,\gamma,x)}\left(\hat{x}(\mathbf{G})\in\{x,-x\}\right) \geqslant 1-o(1)$	$\frac{d}{\log n}\left(1-\sqrt{1-\gamma^2}\right) \geqslant 1$

The next result is a formal version of Theorem 1.9

Theorem 7.2 (Restatement of Theorem 1.9)**.** *Let $x\in\{\pm1\}^n$ be balanced[4]. For any $\gamma, d, \varepsilon, \delta > 0$ satisfying*

$$\frac{d}{\log n}\left(1-\sqrt{1-\gamma^2}\right)\geqslant\Omega(1)\quad\text{and}\quad\frac{\gamma d}{\log n}\geqslant\Omega\left(\frac{1}{\varepsilon^2}\cdot\frac{\log(1/\delta)}{\log n}+\frac{1}{\varepsilon}\right),$$

there exists an (ε,δ)-differentially private[5] algorithm that, on input $\mathbf{G}\sim\mathsf{SBM}_n(d,\gamma,x)$, returns $\hat{x}(\mathbf{G})\in\{x,-x\}$ with probability $1-o(1)$. Moreover, the algorihtm runs in polynomial time.

For any constant $\varepsilon > 0$, Theorem 7.2 states that (ε,δ)-differentially private exact recovery is possible, in polynomial time, already a constant factor close to the non-private threshold. Previous results [MNVT22] could only achieve comparable guarantees in time $O(n^{O(\log n)})$. It is also important to observe that the theorem provides a trade-off between signal-to-noise ratio of the instance (captured by the expression on the left-hand side with γ, d) and the privacy parameter ε. In particular, we distinguish two regimes: for $d\geqslant\Omega(\log n)$ one can achieve exact recovery with high probability and privacy parameters $\delta = n^{-\Omega(1)}$, $\varepsilon = O(1/\gamma + 1/\gamma^2)$. For $d\geqslant\omega(\log n)$ one can achieve exact recovery with high probability and privacy parameters $\varepsilon = o(1)$, $\delta = n^{-\omega(1)}$.

Further, we present a second, exponential-time, algorithm based on the exponential mechanism [MT07] which improves over the above in two regards: First, it gives *pure* privacy guarantees, i.e., $\delta = 0$, and second, provides strong privacy guarantees for a larger range of graph parameters. In fact, we will also prove a lower bound which shows that its privacy guarantees are information theoretically optimal.[6] All hidden constants are absolute and do not depend on any graph or privacy parameters unless stated otherwise. In what follows we denote by $\mathrm{err}(\hat{x}, x)$ the minimum of the hamming distance of $\hat{x}$ and x, and the one of $-\hat{x}$ and x, divided by n.

Theorem 7.3 (Informal, see Theorem 7.42)**.** *Let $\gamma\sqrt{d}\geqslant\Omega(1)$, $x\in\{\pm1\}^n$ be balanced, and $\zeta\geqslant\exp\left(-\Omega\left(\gamma^2 d\right)\right)$. For any $\varepsilon\geqslant\Omega\left(\frac{\log(1/\zeta)}{\gamma d}\right)$, there exists an algorithm which on input $\mathbf{G}\sim\mathsf{SBM}_n(\gamma,d,x)$ outputs an estimate $\hat{x}(\mathbf{G})\in\{\pm1\}^n$ satisfying $\mathrm{err}(\hat{x}(\mathbf{G}),x)\leqslant\zeta$ with probability at least $1-\zeta$. In addition, the algorithm is ε-private. Further, we can achieve error $\Theta\left(1/\sqrt{\log(1/\zeta)}\right)$ with the increased success probability $1-e^{-n}$.*

[4] A vector $x\in\{\pm1\}^n$ is said to be balanced if $\sum_{i=1}^n x_i = 0$.

[5] See Definition 7.26 for a precise definition of adjacent graphs.

[6] Optimality holds in the "small error" regime, otherwise it is almost optimal. See the lower bound for more detail.

A couple of remarks are in order. First, the algorithm works across all degree-regimes in the literature and matches known non-private thresholds and rates up to constants.[7] In particular, for $\gamma^2 d = \Theta(1)$, we achieve weak/partial recovery with either constant or exponentially high success probability. Recall that the optimal non-private threshold is $\gamma^2 d > 1$. For the regime, where $\gamma^2 d = \omega(1)$, it is known that the optimal error rate is $\exp\big(-(1-o(1))\gamma^2 d\big)$ [ZZ16] even non-privately which we match up to constants - here $o(1)$ denotes a function that tends to zero as $\gamma^2 d$ tends to infinity. Moreover, our algorithm achieves exact recovery as soon as $\gamma^2 d = \Omega\big(\log n\big)$ since then $\zeta < \frac{1}{n}$. This also matches known non-private thresholds up to constants [ABH15, MNS15a]. We remark that [MNVT22] gave an ε-DP exponential time algorithm which achieved exact recovery and has inverse polynomial success probability in the utility case as long as $\varepsilon \geqslant \Omega\Big(\frac{\log n}{\gamma d}\Big)$. We recover this result as a special case.[8] In fact, their algorithm is also based on the exponential mechanism, but their analysis only applies to the setting of exact recovery, while our result holds much more generally. Another crucial difference is that we show how to privatize a known boosting technique frequently used in the non-private setting, allowing us to achieve error guarantees which are optimal up to constant factors.

As discussed in Chapter 1, our next result is an information theoretic lower bound which shows that the trade-off of our algorithms is, to some extent, inherent.

Theorem 7.4 (Restatement of Theorem 1.10)**.** *Suppose there exists an ε-differentially private algorithm such that for any balanced $x \in \{\pm 1\}^n$, on input $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, x)$, outputs $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ satisfying*

$$\mathbb{P}(\mathrm{err}(\hat{x}(\mathbf{G}), x) < \zeta) \geqslant 1 - \eta\,.$$

Then,

$$\varepsilon \geqslant \Omega\left(\frac{\log(1/\zeta)}{\gamma d} + \frac{\log(1/\eta)}{\zeta n \gamma d}\right). \tag{7.0.1}$$

Notice that this is a lower bound for a large range of error rates (partial to exact recovery). For failure probability $\eta = \zeta$, the lower bound simplifies to $\varepsilon \geqslant \Omega\Big(\frac{\log(1/\zeta)}{\gamma d}\Big)$ and hence matches Theorem 7.3 up to constants. For exponentially small failure probability, $\eta = e^{-n}$, it becomes $\varepsilon \geqslant \Omega\Big(\frac{1}{\zeta \gamma d}\Big)$. To compare, using the substitution $\sqrt{\log(1/\zeta)} \to \zeta$, Theorem 7.3 requires $\varepsilon \geqslant \Omega\Big(\frac{1}{\zeta^2 \gamma d}\Big)$ in this regime.

Further, as shown in Chapter 1, this lower bound also suggests that the guarantees obtained by our efficient algorithm in Theorem 7.2 might be close to optimal.

Learning mixtures of spherical Gaussians. Recall the standard Gaussian mixture model.

Model 7.5 (Restatement of Model 1.7)**.** Let $D_1, \ldots, D_k$ be Gaussian distributions on $\mathbb{R}^d$ with covariance Id and means $\mu_1, \ldots, \mu_k$ satisfying $\|\mu_i - \mu_j\| \geqslant \Delta$ for any $i \neq j$. Given a

[7] For ease of exposition we did not try to optimize these constants.

[8] With slightly worse constants.

set $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\}$ of n samples from the uniform mixture over $D_1, \ldots, D_k$, estimate $\mu_1, \ldots, \mu_k$.

It is known that when the minimum separation is $\Delta = o(\sqrt{\log k})$, superpolynomially many samples are required to estimate the means up to small constant error [RV17]. Just above this threshold, at separation $k^{O(1/\gamma)}$ for any constant γ, there exist efficient algorithms based on the sum-of-squares hierarchy recovering the means up to accuracy $1/\mathrm{poly}(k)$ [HL18, KSS18, ST21]. In the regime where $\Delta = O(\sqrt{\log k})$ these algorithms yield the same guarantees but require quasipolynomial time. Recently, [LL22] showed how to efficiently recover the means as long as $\Delta = O(\log(k)^{1/2+c})$ for any constant $c > 0$.

Our algorithm for privately learning mixtures of k spherical Gaussians provides statistical guarantees matching those of the best known non-private algorithms. Gaussian Mixture Models have also already been studied in the context of differential privacy by [KSSU19, CKM+21, TCK+22] using the privacy framework first introduced in [NRS07] (see also recent work for robust moment estimation in the differential-privacy setting [KMV22, AL22]). The works of [KSSU19, CKM+21] require an explicit bound R on the euclidean norm of the centers as the sample complexity of these algorithms depends on this bound. For a mixture of k Gaussians, if there is a non-private algorithm that requires the minimum distance between the centers to be at least Δ, then [CKM+21, TCK+22] can transform this non-private algorithm into a private one that needs the minimum distance between the centers to be at least $\Delta + \sqrt{\log n}$, where n is the number of samples.

Theorem 7.6 (Restatement of Theorem 1.8)**.** *Consider an instance of Model 7.5. Let $t > 0$ be such that $\Delta \geqslant O\left(\sqrt{t}k^{1/t}\right)$. For $n \geqslant \Omega\left(k^{O(1)} \cdot d^{O(t)}\right)$, $k \geqslant (\log n)^{1/5}$, there exists an algorithm, running in time $(nd)^{O(t)}$, that outputs vectors $\hat{\boldsymbol{\mu}}_1, \ldots, \hat{\boldsymbol{\mu}}_k$ satisfying*

$$\max_{\ell \in [k]} \left\| \hat{\boldsymbol{\mu}}_\ell - \mu_{\pi(\ell)} \right\|_2 \leqslant O(k^{-12}),$$

with high probability, for some permutation $\pi : [k] \to [k]$. Moreover, for $\varepsilon \geqslant k^{-10}$, $\delta \geqslant n^{-10}$, the algorithm is (ε, δ)-differentially private[9] for any input Y.

Prior to this work, known differentially private algorithms could learn a mixture of k-spherical Gaussian either if: (1) they were given a ball of radius R containing all centers [KSSU19, CKM+21];[10] or (2) the minimum separation between centers needs an additional additive $\Omega(\sqrt{\log n})$ term[11].

Theorem 7.6 is the first to simultaneously break both barriers. That is, the algorithm requires no explicit upper bounds on the means (this also means the sample complexity does not depend on R) and only minimal separation assumptions $O(\sqrt{\log k})$. While previous results only focused on mixtures of Gaussians, our algorithm also works for the

[9] Our notion of adjacent databases here is the obvious one. See Definition 7.51.

[10] In [KSSU19, CKM+21] the sample complexity of the algorithm depends on this radius R.

[11] For $k \leqslant n^{o(1)}$ our algorihtm provides a significant improvement as $\sqrt{\log k} = o(\sqrt{\log n})$.

significantly more general class of mixtures of Poincaré distributions. Concretely, in the high dimensional regime $k \geqslant \sqrt{\log d}$, our algorithm recovers the state-of-the-art guarantees provided by non-private algorithms which are based on the sum-of-squares hierarchy [KSS18, HL18, ST21]:[12]

- If $\Delta \geqslant k^{1/t^*}$ for some $t^* \in \mathbb{N}$, then by choosing $t \geqslant \Omega(t^*)$ the algorithm recovers the centers, up to a $1/\mathrm{poly}(k)$ error, in time $\mathrm{poly}(k, d)$ and using only $\mathrm{poly}(k, d)$ samples.
- If $\Delta \geqslant \Omega(\sqrt{\log k})$ then choosing $t = O(\log k)$ the algorithm recovers the centers, up to a $1/\mathrm{poly}(k)$ error, in quasi-polynomial time $\mathrm{poly}(k^{O(t)}, d^{O(t^2)})$ and using a quasi-polynomial number of samples $\mathrm{poly}(k, d^{O(t)})$.

For simplicity of exposition we will limit the presentation to mixtures of spherical Gaussians. We reiterate that separation $\Omega(\sqrt{\log k})$ is information-theoretically necessary for algorithms with polynomial sample complexity [RV17].

7.1 Techniques

We introduce here the main ideas behind the tools used in the chapter. The algorithms we design have the following structure in common: First, we solve a convex optimization problem with constraints and objective function depending on our input Y. Second, we round the optimal solution computed in the first step to a solution X for the statistical estimation problem at hand.

We organize our privacy analyses according to this structure. In order to analyze the first step, we prove a simple sensitivity bound for strongly convex optimization problems, which bounds the ℓ_2-sensitivity of the optimal solution in terms of a uniform sensitivity bound for the objective function and the feasible region of the optimization problem.

For bounded problems –such as recovery of stochastic block models– we use this sensitivity bound, in the second step, to show that introducing small additive noise to standard rounding algorithms is enough to achieve privacy.

For unbounded problems –such as learning GMMs– we use this sensitivity bound to show that on adjacent inputs, either most entries of X only change slightly, as in the bounded case, or few entries vary significantly. We then combine different privacy techniques to hide both type of changes.

Privacy from sensitivity of strongly convex optimization problems. Before illustrating our techniques with some examples, it is instructive to explicit our framework. Here we have a set of inputs $\mathcal{Y}$ and a family of strongly convex functions $\mathcal{F}(\mathcal{Y})$ and convex sets

[12] We remark that [LL22] give a polynomial time algorithm for separation $\Omega(\log(k)^{1/2+c})$ for constant $c > 0$ in the non-private setting but for a less general class of mixture distributions.

$\mathcal{K}(\mathcal{Y})$ parametrized by these inputs. The generic *non-private* algorithm based on convex optimization we consider works as follows:

1. Compute $\hat{X} := \operatorname{argmin}_{X \in \mathcal{K}(Y)} f_Y(X)$;
2. Round $\hat{X}$ into an integral solution.

For an estimation problem, a distributional assumption on $\mathcal{Y}$ is made. Then one shows how, for typical inputs $\mathbf{Y}$ sampled according to that distribution, the above scheme recovers the desired structured information. Indeed many of the algorithms seen throughout the previous chapters adhere to this description (e.g. Algorithm 4.10).

We can provide a privatized version of this scheme by arguing that, under reasonable assumptions on $\mathcal{F}(Y)$ and $\mathcal{K}(\mathcal{Y})$, the output of the function $\operatorname{argmin}_{X \in \mathcal{K}(Y)} f_Y(X)$ has low ℓ_2-sensitivity. The consequence of this crucial observation is that one can combine the rounding step 2 with some standard privacy mechanism and achieve differential privacy. That is, the second step becomes:

2. Add random noise $\mathbf{N}$ and round $\hat{X} + \mathbf{N}$ into an integral solution.

Our sensitivity bound is simple, yet it generalizes previously known bounds for strongly convex optimization problems (we provide a detailed comparison later in the section). For adjacent $Y, Y' \in \mathcal{Y}$, it requires the following ingredients:

(i) For each $X \in \mathcal{K}(Y) \cap \mathcal{K}(Y')$ it holds $|f_Y(X) - f_{Y'}(X)| \leqslant \alpha$;

(ii) For each $X \in \mathcal{K}(Y)$ its projection Z onto $\mathcal{K}(Y) \cap \mathcal{K}(Y')$ satisfies $|f_Y(X) - f_Y(Z)| \leqslant \alpha$.

Here we think of α as some small quantity (relatively to the problem parameters). Notice, we may think of *(i)* as Lipschitz-continuity of the function $g(Y, X) = f_Y(X)$ with respect to Y and of *(ii)* as a bound on the change of the constrained set on adjacent inputs. In fact, these assumptions are enough to conclude low ℓ_2 sensitivity. If $\hat{X}$ and $\hat{X}'$ are the outputs of the first step on inputs Y, Y', then there exists $Z \in \mathcal{K}(Y) \cap \mathcal{K}(Y')$ such that

$$|f_Y(\hat{X}) - f_Y(Z)| + |f_{Y'}(\hat{X}') - f_{Y'}(Z)| \leqslant O(\alpha)\,.$$

By strong convexity of f_Y , $f_{Y'}$ this implies

$$\left\|\hat{X} - Z\right\|_2^2 + \left\|\hat{X}' - Z\right\|_2^2 \leqslant O(\alpha)$$

which ultimately means $\|\hat{X} - \hat{X}'\|_2^2 \leqslant O(\alpha)$. Thus, starting from our assumptions on the point-wise distance of f_Y , $f_{Y'}$ we were able to conclude low ℓ_2-sensitivity of our output!

A simple application: weak recovery of stochastic block models. The ideas introduced above, combined with existing algorithms for weak recovery of stochastic block models, immediately imply a private algorithm for the problem. To illustrate this, consider Model 7.1 with parameters $\gamma^2 d \geqslant C$, for some large enough constant $C > 1$. Let $x \in \{\pm 1\}^n$ be balanced. Here the input Y is an n-by-n matrix corresponding to the rescaled centered adjacency matrix of the graph:

$$Y_{ij} = \begin{cases} \frac{1}{\gamma d}\left(1 - \frac{d}{n}\right) & \text{if } ij \in E(G) \\ -\frac{1}{\gamma n} & \text{otherwise.} \end{cases}$$

The basic semidefinite program [GV16, MS16] can be recast as the strong constrained optimization question of finding the orthogonal projection of the matrix Y onto the set $\mathcal{K} := \{X \in \mathbb{R}^{n \times n} \mid X \succeq \mathbf{0}\,, \|X\|_\infty \leqslant 1/n\}$. That is:

$$\hat{X} := \operatorname{argmin}_{X \in \mathcal{K}} \|Y - X\|_{\mathrm{F}}^2 .$$

It is a standard fact that, if our input was $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, x)$, then with high probability $X(\mathbf{G}) = \operatorname{argmin}_{X \in \mathcal{K}} f_{Y(\mathbf{G})}(X)$ would have leading eigenvalue, eigenvector pair satisfying

$$\lambda_1(\mathbf{G}) \geqslant 1 - O(1/\gamma^2 d) \quad \text{and} \quad \langle v_1(\mathbf{G}), x/\|x\|\rangle^2 \geqslant 1 - O\left(1/\gamma^2 d\right).$$

This problem fits perfectly the description of the previous paragraph. In fact, it stands to reason that the closeness of the projections $\hat{X}, \hat{X}'$ of inputs Y, Y' should be proportional to the distance between Y and Y'. Our sensitivity argument above formalizes this simple intuition. Concretely, observe that the constrained set $\mathcal{K}$ is fixed and that for each $X \in \mathcal{K}$ it holds $|f_Y(X) - f_{Y'}(X)| \leqslant O\left(\|Y - Y'\|_{\mathrm{F}}^2 + \|Y - Y'\|_1\right)$. It is easy to see that on adjacent input we have $\|Y - Y'\|_{\mathrm{F}}^2 + \|Y - Y'\|_1 \leqslant O(1/n\gamma d)$ and thus this immediately yields $\|\hat{X} - \hat{X}'\|_{\mathrm{F}}^2 \leqslant O(1/n\gamma d)$.

The rounding step is now straightforward. Using the Gaussian mechanism we return the leading eigenvector of $\hat{X} + \mathbf{N}$ where $\mathbf{N} \sim N\left(0, \frac{1}{n\gamma d} \cdot \frac{\log(1/\delta)}{\varepsilon^2}\right)^{n \times n}$. This matrix has Frobeinus norm significantly larger than $\hat{X}$ but its spectral norm is only

$$\|\mathbf{N}\| \leqslant \frac{\sqrt{n \log(1/\delta)}}{\varepsilon} \cdot \sqrt{\tfrac{1}{n\gamma d}} \leqslant \frac{1}{\varepsilon} \cdot \sqrt{\frac{\log(1/\delta)}{\gamma d}}\,.$$

Thus by standard linear algebra, for typical instances $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, x)$, the leading eigenvector of $\hat{X}(\mathbf{G}) + \mathbf{N}$ will be highly correlated with the true community vector x whenever the average degree d is large enough. In conclusion, a simple randomized rounding step is enough!

Remark 7.7 (From weak recovery to exact recovery). In the non-private setting, given a weak recovery algorithm for the stochastic block model, one can use this as an initial estimate for a boosting procedure based on majority voting to achieve exact recovery. We show that this can be done privately. See Section 7.4.2.

An advanced application: learning mixtures of Gaussians. In the context of stochastic block models our argument greatly benefited from two key properties: first, on adjacent inputs the difference $\|Y - Y'\|_F$ was bounded; and second, the convex set $\mathcal{K}$ was fixed. In the context of learning mixtures of spherical Gaussians as in Model 7.5, *both* these properties are *not* satisfied (notice how one of this second properties would be satisfied assuming bounded centers!). So additional ingredients are required.

The first observation, useful to overcome the first obstacle, is that before finding the centers, one can first find the n-by-n membership matrix $W(Y)$ where $W(Y)_{ij} = 1$ if i, j where sampled from the same mixture component and 0 otherwise. The advantage here is that, on adjacent inputs, $\|W(Y) - W(Y')\|_F^2 \leqslant 2n/k$ and thus one recovers the first property.[13] Here early sum-of-squares algorithms for the problem [HL18, KSS18] turns out to be convenient as they rely on minimizing the function $\|W\|_F^2$ subject to the following system of polynomial inequalities in variables $z_{11}, \dots, , z_{1k}, \dots, z_{nk}$, with $W_{ij} = \sum_\ell z_{i\ell} z_{j\ell}$ for all $i, j \in [n]$ and a parameter $t > 0$.

$$\left\{\begin{array}{ll}
z_{i\ell}^2 = z_{i\ell} & \forall i \in [n], \ell \in [k] \quad \text{(indicators)} \\
\sum_{\ell \in [k]} z_{i\ell} \leqslant 1 & \forall i \in [n] \quad \text{(cluster membership)} \\
z_{i\ell} \cdot z_{i\ell'} = 0 & \forall i \in [n], \ell \in [k] \quad \text{(unique membership)} \\
\sum_i z_{i\ell} = n/k & \forall \ell \in [k] \quad \text{(size of clusters)} \\
\mu'_\ell = \frac{k}{n} \sum_i z_{i\ell} \cdot y_i & \forall \ell \in [k] \quad \text{(means of clusters)} \\
\frac{k}{n} \sum_i z_{i\ell} \langle y_i - \mu'_\ell, u \rangle^{2t} \leqslant (2t)^t \cdot \|u\|_2^t & \forall u \in \mathbb{R}^d, \ell \in [k] \quad \text{(subgaussianity of } t\text{-moment)}
\end{array}\right\} \quad (\mathcal{P}(Y))$$

For the scope of this discussion,[14] we may disregard computational issues and assume we have access to an algorithm returning a point from the convex hull $\mathcal{K}(Y)$ of all solutions to our system of inequalities.[15] Here each indicator variable $z_{i\ell} \in \{0, 1\}$ is meant to indicate whether sample y_i is believed to be in cluster C_ℓ. In the non-private settings, the idea behind the program is that –for typical $\mathbf{Y}$ sampled according to Model 7.5 with minimum separation $\Delta \geqslant k^{1/t}\sqrt{t}$– any solution$W(\mathbf{Y}) \in \mathcal{K}(\mathbf{Y})$ is close to the ground truth matrix $W^*(\mathbf{Y})$ in Frobenius norm: $\|W(\mathbf{Y}) - W^*(\mathbf{Y})\|_F^2 \leqslant 1/\mathrm{poly}(k)$. Each row $W(\mathbf{Y})_i$ may be seen as

[13] Notice for typical inputs $\mathbf{Y}$ from Model 7.5 one expect $\|W(\mathbf{Y})\|_F \approx n^2/k$.

[14] While this is far from being true, it turns out that having access to a pseudo-distribution satisfying $\mathcal{P}(Y)$ is enough for our subsequent argument to work, albeit with some additional technical work required.

[15] We remark that a priori it is also not clear how to encode the subgaussian constraint in a way that we could recover a degree-t pseudo-distribution satisfying $\mathcal{P}(Y)$ in polynomial time. By now this is well understood, we discuss this in Section 7.2.

inducing a uniform distribution over a subset of $\mathbf{Y}$.[16]. Thus, combining the above bound with the fact that subgaussian distributions at small total variation distance have means that are close, we can conclude the algorithm recovers the centers of the mixture.

While this program suggests a path to recover the first property, it also possesses a fatal flaw: the projection W' of $W \in \mathcal{K}(Y)$ onto $\mathcal{K}(Y) \cap \mathcal{K}(Y')$ may be *far* in the sense that $|\|W\|_F^2 - \|W'\|_F^2| \geqslant \Omega(\|W\|_F^2 + \|W'\|_F^2) \geqslant \Omega(n^2/k)$. The reason behind this phenomenon can be found in the constraint $\sum_i z_{i\ell} = n/k$. The set indicated by the vector $(z_{1\ell} \ \dots \ , z_{n\ell})$ may be subgaussian in the sense of $\mathcal{P}(Y)$ for input Y but, upon changing a single sample, this may no longer be true. We work around this obstacle in two steps:

1. We replace the above constraint with $\sum_i z_{i\ell} \leqslant n/k$.
2. We compute $\hat{W} := \operatorname{argmin}_{W \text{ solving } \mathcal{P}(Y)} \|J - W\|_F^2$, where $J \in \mathbb{R}^{n\times n}$ is the all-ones matrix.[17]

The catch now is that the program is satisfiable for *any* input Y. Moreover, we can guarantee property *(ii)* (required by our sensitivity argument) for $\alpha \leqslant O(n/k)$, since we can obtain $W' \in \mathcal{K}(Y) \cap \mathcal{K}(Y')$ simply zeroing out the row/column in W corresponding to the sample differing in Y and Y'. Then for typical inputs $\mathbf{Y}$, the correlation with the true solution is guaranteed by the new strongly convex objective function.

From low sensitivity of the indicators to low sensitivity of the estimates. For adjacent inputs Y, Y' let $\hat{W}, \hat{W}'$ be respectively the matrices computed by the above strongly convex programs. Our discussion implies that, applying our sensitivity bound, we can show $\|\hat{W} - \hat{W}'\|_F^2 \leqslant O(n/k)$. The problem is that simply applying a randomized rounding approach here cannot work. The reason is that even tough the vector $\hat{W}_i$ induces a subgaussian distribution, the vector $\hat{W}_i + v$ for $v \in \mathbb{R}^n$, *might not*. Without the subgaussian constraint we cannot provide any meaningful utility bound. In other words, the root of our problem is that there exists heavy-tailed distributions that are arbitrarily close in total variation distance to any given subgaussian distribution.

On the other hand, our sensitivity bound implies $\|\hat{W} - \hat{W}'\|_1^2 \leqslant o(\|\hat{W}\|_1)$ and thus, all but a vanishing fraction of rows $i \in [n]$ must satisfy $\|\hat{W}_i - \hat{W}'_i\|_1 \leqslant o(\|\hat{W}_i\|_1)$. For each row i, let μ_i, μ'_i be the means of the distributions induced respectively by $\hat{W}_i, \hat{W}'_i$. We thus find ourselves in the following settings:

1. For a set of $(1 - o(1)) \cdot n$ good rows $\left\|\mu_i - \mu'_i\right\|_2 \leqslant o(1)$,
2. For the set $\mathcal{B}$ of remaining bad rows, the distance $\left\|\mu_i - \mu'_i\right\|_2$ may be unbounded.

[16]More generally, we may think of a vector $v \in \mathbb{R}^n$ as the vector inducing the distribution given by $v/\|v\|_1$ onto the set Y of n elements.

[17]We remark that for technical reasons our function in Section 7.5.1 will be slightly different. We do not discuss it here to avoid obfuscating our main message.

We hide differences of the first type as follows: pick a random subsample $\boldsymbol{S}$ of $[n]$ of size n^c, for some small $c > 0$, and for each picked row use the Gaussian mechanism. The subsampling step is useful as it allows us to decrease the standard deviation of the entry-wise random noise by a factor n^{1-c}. We hide differences of the second type as follows: use the classic high dimensional (ε, δ)-private histogram learner on $\boldsymbol{S}$ and for the k largest bins of highest count privately return their average. The crux of the argument here is that the cardinality of $\mathcal{B} \cap \boldsymbol{S}$ is sufficiently small that the privacy guarantees of the histogram learner can be extended even for inputs that differ in $|\mathcal{B} \cap \boldsymbol{S}|$ many samples. Finally, standard composition arguments will guarantee privacy of the whole algorithm.

Comparison with the framework of Kothari-Manurangsi-Velingker. Solving an optimization problem whose constraints depend on the input can cause privacy leaks, since there could exist two adjacent inputs such that the constraints are satisfiable for one but not for the other. [KMV22] deals with this issue via their "stable outlier rate selection" procedure which is an instantiation of exponential mechanisms with a complicated score function. We avoid this issue by casting the problem as a strongly convex program with constraints that are satisfiable on *any* input. This shift in paradigm brings several advantages. First, it provides a recipe that can be easily extended to other high dimensional problems of interest: recast the program as strongly convex optimization over the whole space and add noise to the output. For example, in the context of SBM, the framework of [KMV22] would require one to sample from an exponential distribution over matrices. Constructing and sampling from such distributions is an expensive operation. However, it is well-understood that an optimal fractional solution to the basic the basic SDP relaxation we consider can be found in *near quadratic time* using the standard matrix multiplicative weight method [AK07, Ste10a]. Making the whole algorithm run in near-quadratic time. Whether our algorithm can be speed up to near-linear time, as in [AK07, Ste10a], remains a fascinating open question.

Comparison with previous works on empirical risk minimization. Results along the lines of the sensitivity bound described at the beginning of the section (see Lemma 7.25 for a formal statement) have been extensively used in the context of empirical risk minimization [CMS11, KST12, SCS13, BST14, WYX17, MSVV21]. Most results focus on the special case of unconstrained optimization of strongly convex functions. In contrast, our sensitivity bound applies to the significantly more general settings where both the objective functions and the constrained set may depend on the input.[18] Most notably for our settings of interest, [CMS11] studied unconstrained optimization of (smooth) strongly convex functions depending on the input, with bounded gradient.We recover such a result for $X' = X$ in *(ii)*. In [MSVV21],

[18]The attentive reader may argue that one could cast convex optimization over a constrained domain as unconstrained optimization of a new convex function with the appropriate penalty terms. In practice however, this turns out to be hard to do for constraints such as Definition 7.21.

the authors considered constraint optimization of objective functions where the domain (but *not* the function) may depend on the input data. They showed how one can achieve differential privacy while optimize the desired objective function by randomly perturbing the constraints. It is important to remark that, in [MSVV21], the notion of utility is based on the optimization problem (and their guarantees are tight only up to logarithmic factors). In the settings we consider, even in the special case where f does not depend on the input, this notion of utility may not correspond to the notion of utility required by the estimation problem, and thus, the corresponding guarantees may turn out to be too loose to ensure the desired error bounds.

Exponential time pure-DP algorithm for SBM. Our exponential time algorithm is based on the exponential mechanism [MT07]. In particular, for a given graph G, recall that $Y = \frac{1}{\gamma d}\left(A(G) - \frac{d}{n}J\right)$, where $A(G)$ is the adjacency matrix of G and J the all-ones matrix. Define the function $s\colon \{\pm 1\}^n \to \mathbb{R}$ as $s(x) = \langle x, Yx\rangle$ and $\Delta = \frac{2}{\gamma d}$. In privacy terms, these are called the *score function* and the *sensitivity* - the maximum amount s can change on adjacent graphs - which can be readily seen to be $\frac{2}{\gamma d}$. The exponential mechanism then amounts to outputting a sample from the distribution with density

$$p(x) \propto \exp\left(\tfrac{\varepsilon}{2\Delta}s(x)\right). \tag{7.1.1}$$

Standard arguments show that this procedure is ε-private. Note, that it is well-known that if $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, x^*)$ and $\gamma^2 d$ is larger than some universal constant, Y is close to $\frac{1}{n}x^*(x^*)^\top$ in cut-norm (or $(\infty \to 1)$-norm). That is, the quadratic form $Y - \frac{1}{n}x^*(x^*)^\top$ is close to zero over the hypercube [GV16]. It follows, that the maximizer of score function s over the hypercube is close to $\frac{1}{\sqrt{n}}x^*$. It is not too hard to show that with exponentially high probability (in n), this remains true also for samples from the above distribution. On an intuitive level this follows since we assign exponentially larger mass to points achieving comparable scores as the maximizer than to points achieving smaller scores (see Section 7.4.3).

While this algorithm matches known non-private thresholds and rates up to constants and has close to optimal privacy guarantees (see the discusion at th ebeginning of the chapter), there are several obstacles to making it efficient. We discuss several approaches: First, one could try to sample from the distribution in Eq. (7.1.1) directly. Note that this corresponds to an Ising model over the hypercube with interaction matrix $J := \frac{\varepsilon}{\delta}Y$. However, known samplers, e.g. [EKZ22, KLR22], require strong assumptions on the spectrum of J which are not satisfied in our setting - in particular, J could have arbitrarly many eigenvalues of magnitude larger than 1. A second approach would be to relax the support of the distribution to all positive semi-definite matrices with diagonal entries equal to $1/n$ - similar to the set $\mathcal{K}$ considered for our approximate DP algorithm - and the score function to the inner product of Y with such matrices. Although such "convexification" techniques of the exponential mechanism have recently seen success in the design of pure-DP algorithms [HKM22] and this particular relaxation is known to work in the non-private

setting [GV16, FC20], it fails in this case: The volume of the set of matrices achieving large enough score is smaller by a factor of $\exp(-n^2)$ than the set of all feasible matrices. Hence, the exponential boost by the reweighing of Eq. (7.1.1) is not enough to ensure outputting such a candidate. A third strategy would be the following: In the non-private setting, for the restricted regime of $\gamma^2 d \geqslant C \log n$, for a large enough constant $C > 0$, standard matrix concentration bounds show that PCA can recover the label of a large constant fraction of the vertices. However, known lower bounds for pure-DP PCA algorithms [KT13], prevent us from recovering this result in the private setting: In particular, define two matrices to be adjacent if there difference has spectral norm at most 1. In this setting, any ε-DP algorithm, which outputs a vector achieving constant correlation with the top eigenvector of an $n \times n$ input matrix needs to have spectral norm at least $\Omega(\frac{n}{\varepsilon})$. Translated to our scaling used above, this would mean $\|Y\| \geqslant \Omega(\frac{n\gamma d}{\varepsilon})$, whereas we have $\|Y\| \leqslant O(1)$.

Information theoretic privacy lower bounds for stochastic block models. Our information theoretic lower bound for stochastic block models is based on the following idea. Suppose we have an ε-differentially private exact recovery algorithm of SBM such that, over the randomness of the algorithm and the input $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, x)$, the algorithm outputs $\hat{x}(\mathbf{G}) \in \{\pm x\}$ with probability at least $2/3$. Note for any $x \in \{\pm 1\}^n$, $\mathsf{SBM}_n(d, \gamma, x)$ is just a product distribution of $\binom{n}{2}$ Bernoulli distributions. Fixing an arbitrary balanced $y \in \{\pm 1\}^n$, there exist balanced $y_1, \ldots, y_n \in \{\pm 1\}^n$ such that $\mathrm{Ham}(y, y_i) = 2$ for $i \in [n]$. For each $i \in [n]$, one may find a coupling ω_i of distributions $\mathsf{SBM}_n(d, \gamma, y)$ and $\mathsf{SBM}_n(d, \gamma, y_i)$ such that, if $(\mathbf{G}, \mathbf{G}') \sim \omega_i$ then $\mathbf{G}$ and $\mathbf{G}'$ typically differ by only $O(\gamma d)$ edges. Then by the assumption our algorithm is ε-differentially private, it follows that, on input $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, y)$, our private algorithm outputs $\pm y_i$ with probability at least $e^{-\varepsilon \cdot O(\gamma d)} \cdot \frac{2}{3}$ for each $i \in [n]$. Since the sum of probabilities of disjoint events does not exceed one, we get $n \cdot e^{-O(\varepsilon \gamma d)} \cdot \frac{2}{3} \leqslant 1$, which implies $\varepsilon \geqslant \Omega(\frac{\log n}{\gamma d})$.

7.2 Preliminaries

We reuse the notation of Chapter 2. We use **boldface** characters for random variables. Often times we use the letter C do denote universal constants independent of the parameters at play. We write $o(1), \omega(1)$ for functions tending to zero (resp. infinity) as n grows.

Vectors, matrices, tensors. For a matrix M, we denote its eigenvalues by $\lambda_1(M), \ldots, \lambda_n(M)$, we simply write λ_i when the context is clear. We denote by $\|M\|$ the spectral norm of M. We denote by $\mathbb{R}^{d^{\otimes t}}$ the set of real-valued order-t tensors. for a $d \times d$ matrix M, we denote by $M^{\otimes t}$ the *t-fold Kronecker product* $\underbrace{M \otimes M \otimes \cdots \otimes M}_{t \text{ times}}$. We define the *flattening*, or *vectorization*, of M to be the d^t-dimensional vector, whose entries are the

entries of M appearing in lexicographic order. With a slight abuse of notation we refer to this flattening with M, ambiguities will be clarified form context. We denote by $N\left(0,\sigma^2\right)^{d^{\otimes t}}$ the distribution over Gaussian tensors with d^t entries with standard deviation σ. Given $u,v \in \{\pm 1\}^n$, we use $\mathrm{Ham}(u,v) := \sum_{i=1}^n \mathbb{1}_{[u_i \neq v_i]}$ to denote their Hamming distance. Given a vector $u \in \mathbb{R}^n$, we let $\mathrm{sign}(u) \in \{\pm 1\}^n$ denote its sign vector. A vector $u \in \{\pm 1\}^n$ is said to be *balanced* if $\sum_{i=1}^n u_i = 0$.

Graphs. We consider graphs on n vertices and let $\mathcal{G}_n$ be the set of all graphs on n vertices. For a graph G on n vertices we denote by $A(G) \in \mathbb{R}^{n\times n}$ its adjacency matrix. When the context is clear we simply write A. Given two graphs G, H on the same vertex set V, let $G \setminus H := (V, E(G) \setminus H(G))$. Given a graph H, $H' \subseteq H$ means H' is a subgraph of H such that $V(H') = V(H)$ and $E(H) \subseteq E(H)$. The Hamming distance between two graphs G, H is defined to be the size of the symmetric difference between their edge sets, i.e. $\mathrm{Ham}(G,H) := |E(G) \triangle E(H)|$.

7.2.1 Differential privacy

In this section we introduce standard notions of differential privacy [DMNS06].

Definition 7.8 (Differential privacy)**.** An algorithm $\mathcal{M} : \mathcal{Y} \to \mathcal{O}$ is said to be (ε,δ)-differentially private for $\varepsilon, \delta > 0$ if and only if, for every $S \subseteq \mathcal{O}$ and every neighboring datasets $Y, Y' \in \mathcal{Y}$ we have

$$\mathbb{P}[\mathcal{M}(Y) \in S] \leqslant e^{\varepsilon} \cdot \mathbb{P}[\mathcal{M}(Y') \in S] + \delta\,.$$

To avoid confusion, for each problem we will exactly state the relevant notion of neighboring datasets. Differential privacy is closed under post-processing and composition.

Lemma 7.9 (Post-processing)**.** *If $\mathcal{M} : \mathcal{Y} \to \mathcal{O}$ is an (ε,δ)-differentially private algorithm and $\mathcal{M}' : \mathcal{Y} \to \mathcal{Z}$ is any randomized function. Then the algorithm $\mathcal{M}'(\mathcal{M}(Y))$ is (ε,δ)-differentially private.*

In order to talk about composition it is convenient to also consider DP algorithms whose privacy guarantee holds only against subsets of inputs.

Definition 7.10 (Differential Privacy Under Condition)**.** An algorithm $\mathcal{M} : \mathcal{Y} \to \mathcal{O}$ is said to be (ε,δ)-differentially private under condition Ψ (or (ε,δ)-DP under condition Ψ) for $\varepsilon, \delta > 0$ if and only if, for every $S \subseteq \mathcal{O}$ and every neighboring datasets $Y, Y' \in \mathcal{Y}$ both satisfying Ψ we have

$$\mathbb{P}[\mathcal{M}(Y) \in S] \leqslant e^{\varepsilon} \cdot \mathbb{P}[\mathcal{M}(Y') \in S] + \delta\,.$$

It is not hard to see that the following composition theorem holds for privacy under condition.

Lemma 7.11 (Composition for Algorithm with Halting, [KMV22])**.** *Let $\mathcal{M}_1 : \mathcal{Y} \to O_1 \cup \{\perp\}$, $\mathcal{M}_2 : O_1 \times \mathcal{Y} \to O_2 \cup \{\perp\}$, $\dots$, $\mathcal{M}_t : O_{t-1} \times \mathcal{Y} \to O_t \cup \{\perp\}$ be algorithms. Furthermore, let $\mathcal{M}$ denote the algorithm that proceeds as follows (with O_0 being empty): For $i = 1 \dots, t$ compute $o_i = \mathcal{M}_i(o_{i-1}, Y)$ and, if $o_i = \perp$, halt and output $\perp$. Finally, if the algorithm has not halted, then output o_t. Suppose that:*

- *For any $1 \leqslant i \leqslant t$, we say that Y satisfies the condition Ψ_i if running the algorithm on Y does not result in halting after applying $\mathcal{M}_1, \dots, \mathcal{M}_i$.*
- *$\mathcal{M}_1$ is $(\varepsilon_1, \delta_1)$-DP.*
- *$\mathcal{M}_i$ is $(\varepsilon_i, \delta_i)$-DP (with respect to neighboring datasets in the second argument) under condition Ψ_{i-1} for all $i = \{2, \dots, t\}$.*

Then $\mathcal{M}$ is $(\sum_i \varepsilon_i, \sum_i \delta_i)$-DP.

7.2.1.1 Basic differential privacy mechanisms

The Gaussian and the Laplace mechanism are among the most widely used mechanisms in differential privacy. They work by adding a noise drawn from the Gaussian (respectively Laplace) distribution to the output of the function one wants to privatize. The magnitude of the noise depends on the sensitivity of the function.

Definition 7.12 (Sensitivity of function)**.** Let $f : \mathcal{Y} \to \mathbb{R}^d$ be a function, its ℓ_1-sensitivity and ℓ_2-sensitivity are respectively

$$\Delta_{f,1} := \max_{\substack{Y, Y' \in \mathcal{Y} \\ Y, Y' \text{ are adjacent}}} \left\| f(Y) - f(Y') \right\|_1 \qquad \Delta_{f,2} := \max_{\substack{Y, Y' \in \mathcal{Y} \\ Y, Y' \text{ are adjacent}}} \left\| f(Y) - f(Y') \right\|_2 .$$

For function with bounded ℓ_1-sensitivity the Laplace mechanism is often the tool of choice to achieve privacy.

Definition 7.13 (Laplace distribution)**.** The Laplace distribution with mean μ and parameter $b > 0$, denoted by $\mathrm{Lap}(\mu, b)$, has PDF $\frac{1}{2b} e^{-|x-\mu|/b}$. Let $\mathrm{Lap}(b)$ denote $\mathrm{Lap}(0, b)$.

A standard tail bound concerning the Laplace distribution will be useful throughout the chapter.

Fact 7.14 (Laplace tail bound)**.** *Let $\boldsymbol{x} \sim Lap(\mu, b)$. Then,*

$$\mathbb{P}\left[|\boldsymbol{x} - \mu| > t \right] \leqslant e^{-t/b} .$$

The Laplace distribution is useful for the following mechanism

Lemma 7.15 (Laplace mechanism)**.** *Let $f : \mathcal{Y} \to \mathbb{R}^d$ be any function with ℓ_1-sensitivity at most $\Delta_{f,1}$. Then the algorithm that adds $Lap\left(\frac{\Delta_{f,1}}{\varepsilon}\right)^{\otimes d}$ to f is $(\varepsilon, 0)$-DP.*

It is also useful to consider the "truncated" version of the Laplace distribution where the noise distribution is shifted and truncated to be non-positive.

Definition 7.16 (Truncated Laplace distribution)**.** The (negatively) truncated Laplace distribution w with mean μ and parameter b on $\mathbb{R}$, denoted by $\mathrm{tLap}(\mu, b)$, is defined as $\mathrm{Lap}(\mu, b)$ conditioned on the value being non-positive.

Lemma 7.17 (Truncated Laplace mechanism)**.** *Let $f : \mathcal{Y} \to \mathbb{R}$ be any function with ℓ_1-sensitivity at most $\Delta_{f,1}$. Then the algorithm that adds $tLap\left(-\Delta_{f,1}\left(1 + \frac{\log(1/\delta)}{\varepsilon}\right), \Delta_{f,1}/\varepsilon\right)$ to f is (ε, δ)-DP.*

The following tail bound is useful when reasoning about truncated Laplace random variables.

Lemma 7.18 (Tail bound truncated Laplace)**.** *Suppose $\mu < 0$ and $b > 0$. Let $\boldsymbol{x} \sim tLap(\mu, b)$. Then,for $y < \mu$ we have that*

$$\mathbb{P}\left[\boldsymbol{x} < y\right] \leqslant \frac{e^{(y-\mu/b)}}{2 - e^{\mu/b}}\,.$$

In constrast, when the function has bounded ℓ_2-sensitivity, the Gaussian mechanism provides privacy.

Lemma 7.19 (Gaussian mechanism)**.** *Let $f : \mathcal{Y} \to \mathbb{R}^d$ be any function with ℓ_2-sensitivity at most $\Delta_{f,2}$. Let $0 < \varepsilon\,, \delta \leqslant 1$. Then the algorithm that adds $N\left(0, \frac{\Delta_{f,2}^2 \cdot 2\log(2/\delta)}{\varepsilon^2} \cdot \mathrm{Id}\right)$ to f is (ε, δ)-DP.*

7.2.1.2 Private histograms

Here we present a classical private mechanism to learn a high dimensional histogram.

Lemma 7.20 (High-dimensional private histogram learner, see [KV17])**.** *Let $q, b\,, \varepsilon > 0$ and $0 < \delta < 1/n$. Let $\{I_i\}_{i=-\infty}^{\infty}$ be a partition of $\mathbb{R}$ into intervals of length b, where $I_i := \left\{x \in \mathbb{R} \;\middle|\; q + (i-1)\cdot b \leqslant x < q + i \cdot b\right\}$. Consider the partition of $\mathbb{R}^d$ into sets $\left\{B_{i_1,\ldots,i_d}\right\}_{i_1,\ldots,i_d=1}^{\infty}$ where*

$$B_{i_1,\ldots,i_d} := \left\{x \in \mathbb{R}^d \;\middle|\; \forall j \in [d]\,, x_j \in I_{i_j}\right\}$$

Let $Y = \left\{y_1, \ldots, y_n\right\} \subseteq \mathbb{R}^d$ be a database of n points. For each $B_{i_1,\ldots,i_d}$, let $p_{i_1,\ldots,i_d} = \frac{1}{n}\left|\left\{j \in [n] \;\middle|\; y_j \in B_{i_1,\ldots,i_d}\right\}\right|$. For $n \geqslant \frac{8}{\varepsilon\alpha} \cdot \log\frac{2}{\delta\beta}$, there exists an efficient (ε, δ)-differentially private algorithm that returns $\hat{\boldsymbol{p}}_{1,\ldots,1}, \ldots, \hat{\boldsymbol{p}}_{i_1,\ldots,i_d}, \ldots$ satisfying

$$\mathbb{P}\left[\max_{i_1,\ldots,i_d \in \mathbb{N}} |p_{i_1,\ldots,i_d} - \hat{\boldsymbol{p}}_{i_1,\ldots,i_d}| \geqslant \alpha\right] \leqslant \beta\,.$$

Proof. We consider the following algorithm, applied to each $i_1, \ldots, i_d \in \mathbb{N}$ on input Y:

1. If $p_{i_1,\ldots,i_d} = 0$ set $\hat{p}_{i_1,\ldots,i_d} = 0$, otherwise let $\hat{\boldsymbol{p}}_{i_1,\ldots,i_d} = p_{i_1,\ldots,i_d} + \boldsymbol{\tau}$ where $\boldsymbol{\tau} \sim \mathrm{Lap}\left(0, \frac{2}{n\varepsilon}\right)$.
2. If $\hat{\boldsymbol{p}}_{i_1,\ldots,i_d} \leqslant \frac{3\log(2/\delta)}{\varepsilon n}$ set $\hat{\boldsymbol{p}}_{i_1,\ldots,i_d} = 0$.

First we argue utility. By construction we get $\hat{\boldsymbol{p}}_{i_1,\ldots,i_d} = 0$ whenever $p_{i_1,\ldots,i_d} = 0$, thus we may focus on non-zero $p_{i_1,\ldots,i_d}$. There are at most n non zero $p_{i_1,\ldots,i_d}$. By choice of n, δ and by Fact 7.14 the maximum over n independent trials $\boldsymbol{\tau} \sim \mathrm{Lap}\left(0, \frac{2}{n\varepsilon}\right)$ is bounded by α in absolute value with probability at least β.

It remains to argue privacy. Let $Y = \left\{y_1, \ldots, y_n\right\}, Y' = \left\{y'_1, \ldots, y'_n\right\}$ be adjacent databases. For $i_1, \ldots, i_d \in \mathbb{N}$, let

$$p_{i_1,\ldots,i_d} = \left|\left\{j \in [n] \;\middle|\; y_j \in B_{i_1,\ldots,i_d}\right\}\right|$$
$$p'_{i_1,\ldots,i_d} = \left|\left\{j \in [n] \;\middle|\; y'_j \in B_{i_1,\ldots,i_d}\right\}\right|.$$

Since Y, Y' are adjacent there exists only two set of indices $\mathcal{I} := \{i_1, \ldots, i_d\}$ and $\mathcal{J} := \left\{j_1, \ldots, j_d\right\}$ such that $p_{\mathcal{I}} \neq p'_{\mathcal{I}}$ and $p_{\mathcal{J}} \neq p'_{\mathcal{J}}$. Assume without loss of generality $p_{\mathcal{I}} > p'_{\mathcal{I}}$. Then it must be $p_{\mathcal{I}} = p'_{\mathcal{I}} + 1/n$ and $p_{\mathcal{J}} = p'_{\mathcal{J}} - 1/n$. Thus by the standard tail bound on the Laplace distribution in Fact 7.14 and by Lemma 7.15, we immediately get that the algorithm is (ε, δ)-differentially private. □

7.2.2 Explictly bounded distributions

This sections builds on Section 2.2. We will consider a subset of subgaussian distributions denoted as certifiably subgaussians. Many subgaussians distributions are known to be certifiably subgaussian (see [KSS18]).

Definition 7.21 (Explicitly bounded distribution)**.** Let $t \in \mathbb{N}$. A distribution D over $\mathbb{R}^d$ with mean μ is called $2t$-explicitly σ-bounded if for each even integer s such that $1 \leqslant s \leqslant t$ the following equation has a degree s sum-of-squares proof in the vector variable u

$$\left|\frac{2s}{u}\right\} \left\{\mathop{\mathbb{E}}_{\mathbf{x}\sim D} \langle \mathbf{x} - \mu, u\rangle^{2s} \leqslant (\sigma s)^s \cdot \|u\|_2^{2s}\right\}$$

Furthermore, we say that D is explicitly bounded if it is $2t$-explicitly σ-bounded for every $t \in \mathbb{N}$. A finite set $X \subseteq \mathbb{R}^d$ is said to be $2t$-explicitly σ-bounded if the uniform distribution on X is $2t$-explicitly σ-bounded.

Sets that are $2t$-explicitly σ-bounded with large intersection satisfy certain key properties. Before introducing them we conveniently present the following definition.

Definition 7.22 (Weight vector inducing distribution)**.** Let Y be a set of size n and let $p \in [0,1]^n$ be a vector satisfying $\left\|p\right\|_1 = 1$. We say that p induces the distribution D with support Y if

$$\mathbb{P}_{\mathbf{y}\sim D}\left[\mathbf{y} = y_i\right] = p_i \,.$$

Theorem 7.23 ([KSS18, HL18]). *Let $Y \subseteq \mathbb{R}^d$ be a set of cardinality n. Let $p, p' \in [0,1]^n$ be weight vectors satisfying $\|p\|_1 = \|p'\|_1 = 1$ and $\|p - p'\|_1 \leqslant \beta$. Suppose that p (respectively p') induces a $2t$-explicitly σ_1-bounded (resp. σ_2) distribution over Y with mean $\mu_{(p)}$ (resp. $\mu_{(p')}$). There exists an absolute constant β^* such that, if $\beta \leqslant \beta^*$, then for $\sigma = \sigma_1 + \sigma_2$:*

$$\|\mu_{(p)} - \mu_{(p')}\| \leqslant \beta^{1-1/2t} \cdot O\left(\sqrt{\sigma t}\right).$$

In the context of learning Gaussian mixtures, we will make heavy use of the statement below.

Theorem 7.24 ([KSS18, HL18]). *Let Y be a $2t$-explicitly σ-bounded set of size n. Let $p \in \mathbb{R}^n$ be the weight vector inducing the uniform distribution over Y. Let $p' \in \mathbb{R}^n$ be a unit vector satisfying $\|p - p'\|_1 \leqslant \beta$ for some $\beta \leqslant \beta^*$ where β^* is a small constant. Then p' induces a $2t$-explicitly $(\sigma + O(\beta^{1-1/2t}))$-bounded distribution over Y.*

7.3 Stability of strongly-convex optimization

In this section, we prove ℓ_2 sensitivity bounds for the minimizers of a general class of (strongly) convex optimization problems. In particular, we show how to translate a uniform point-wise sensitivity bound for the objective functions into a ℓ_2 sensitivity bound for the minimizers.

Lemma 7.25 (Stability of strongly-convex optimization). *Let $\mathcal{Y}$ be a set of databases. Let $\mathcal{K}(\mathcal{Y})$ be a family of closed convex subsets of $\mathbb{R}^m$ parametrized by $Y \in \mathcal{Y}$ and let $\mathcal{F}(\mathcal{Y})$ be a family of functions $f_Y : \mathcal{K}(Y) \to \mathbb{R}$, parametrized by $Y \in \mathcal{Y}$, such that:*

(i) for adjacent databases $Y, Y' \in \mathcal{Y}$ and $X \in \mathcal{K}(Y)$ there exist $Z \in \mathcal{K}(Y) \cap \mathcal{K}(Y')$ satisfying $\left|f_Y(X) - f_{Y'}(Z)\right| \leqslant \alpha$ and $\left|f_Y(Z) - f_{Y'}(Z)\right| \leqslant \alpha$.

(ii) f_Y is κ-strongly convex in $X \in \mathcal{K}(Y)$.

Then for $Y, Y' \in \mathcal{Y}$, $\hat{X} := \arg\min_{X \in \mathcal{K}(Y)} f_Y(X)$ and $\hat{X}' := \arg\min_{X' \in \mathcal{K}(Y')} f_{Y'}(X')$, it holds

$$\left\|\hat{X} - \hat{X}'\right\|_2^2 \leqslant \frac{12\alpha}{\kappa}.$$

Proof. Let $Z \in \mathcal{K}(Y) \cap \mathcal{K}(Y')$ be a point such that $\left|f_Y(\hat{X}) - f_{Y'}(Z)\right| \leqslant \alpha$ and $\left|f_Y(Z) - f_{Y'}(Z)\right| \leqslant \alpha$. By κ-strong convexity of f_Y and $f_{Y'}$ (Proposition E.13) it holds

$$\begin{aligned}\left\|\hat{X} - \hat{X}'\right\|_2^2 &\leqslant 2\left\|\hat{X} - X'\right\|_2^2 + 2\left\|X' - \hat{X}'\right\|_2^2 \\ &\leqslant \frac{4}{\kappa}\left(f_Y(Z) - f_Y(\hat{X}) + f_{Y'}(Z) - f_{Y'}(\hat{X}')\right).\end{aligned}$$

Suppose w.l.o.g. $f_Y(\hat{X}) \leqslant f_{Y'}(\hat{X}')$, for a symmetric argument works in the other case. Then

$$f_Y(Z) \leqslant f_{Y'}(Z) + \alpha \leqslant f_Y(\hat{X}) + 2\alpha$$

and

$$f_Y(\hat{X}) \leqslant f_{Y'}(\hat{X}') \leqslant f_{Y'}(Z) \leqslant f_Y(\hat{X}) + \alpha\,.$$

It follows as desired

$$f_Y(Z) - f_Y(\hat{X}) + f_{Y'}(Z) - f_{Y'}(\hat{X}') \leqslant 3\alpha\,.$$

□

7.4 Private recovery for stochastic block models

In this section, we present how to achieve exact recovery in stochastic block models privately and thus prove Theorem 7.2. To this end, we first use the stability of strongly convex optimization (Lemma 7.25) to obtain a private weak recovery algorithm in Section 7.4.1. Then we show how to privately boost the weak recovery algorithm to achieve exact recovery in Section 7.4.2. In Section 7.4.4, we complement our algorithmic results by providing an almost tight lower bound on the privacy parameters. We start by defining the relevant notion of adjacent databases.

Definition 7.26 (Adjacent graphs)**.** Let $G\,, G'$ be graphs with vertex set $[n]$. We say that $G\,, G'$ are adjacent if $|E(G) \triangle E(G')| = 1\,.$

Remark 7.27 (Parameters as public information)*.* We remark that we assume the parameters $n\,, \gamma\,, d$ to be *public information* given in input to the algorithm.

7.4.1 Private weak recovery for stochastic block models

In this section, we show how to achieve weak recovery privately via stability of strongly convex optimization (Lemma 7.25). We first introduce one convenient notation. The error rate of an estimate $\hat{x} \in \{\pm 1\}^n$ of the true partition $x \in \{\pm 1\}^n$ is defined as $\mathrm{err}(\hat{x}, x) := \frac{1}{n} \cdot \min\{\mathrm{Ham}(\hat{x}, x), \mathrm{Ham}(\hat{x}, -x)\}$.[19] Our main result is the following theorem.

Theorem 7.28. *Suppose $\gamma\sqrt{d} \geqslant 12800\,, \varepsilon, \delta \geqslant 0$. There exists an (Algorithm 7.29) such that, for any $x \in \{\pm 1\}^n$, on input $\mathbf{G} \sim \mathsf{SBM}_n(\gamma\,, d\,, x)$, outputs $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ satisfying*

$$\mathrm{err}(\hat{x}(\mathbf{G}), x) \leqslant O\left(\frac{1}{\gamma\sqrt{d}} + \frac{1}{\gamma d} \cdot \frac{\log(2/\delta)}{\varepsilon^2}\right)$$

with probability $1 - \exp(-\Omega(n))$. Moreover, the algorithm is (ε, δ)-differentially private for any input graph and runs in polynomial time.

[19] Note $|\langle \hat{x}, x\rangle| = (1 - 2\,\mathrm{err}(\hat{x}, x)) \cdot n$ for any $\hat{x}, x \in \{\pm 1\}^n$.

Before presenting the algorithm we introduce some notation. Given a graph G, let $Y(G) := \frac{1}{\gamma d}(A(G) - \frac{d}{n}J)$ where $A(G)$ is the adjacency matrix of G and J denotes all-one matrices. Define $\mathcal{K} := \left\{X \in \mathbb{R}^{n\times n} \;\middle|\; X \succeq 0\,, X_{ii} = \frac{1}{n}\ \forall i\right\}$. The algorithm starts with projecting matrix $Y(G)$ to set $\mathcal{K}$. To ensure privacy, then it adds Gaussian noise to the projection X_1 and obtains a private matrix X_2. The last step applies a standard rounding method.

Algorithm 7.29 (Private weak recovery for SBM)**.**
Input: Graph G.
Operations:

1. Projection: $X_1 \leftarrow \operatorname{argmin}_{X\in\mathcal{K}}\|Y(G) - X\|_F^2$.
2. Noise addition: $\mathbf{X}_2 \leftarrow X_1 + \mathbf{W}$ where $\mathbf{W} \sim \mathcal{N}\left(0, \frac{24}{n\gamma d}\frac{\log(2/\delta)}{\varepsilon^2}\right)^{n\times n}$.
3. Rounding: Compute the leading eigenvector $\mathbf{v}$ of $\mathbf{X}_2$ and return $\operatorname{sign}(\mathbf{v})$.

In the rest of this section, we will show Algorithm 7.29 is private in Lemma 7.31 and its utility guarantee in Lemma 7.32. Then Theorem 7.28 follows directly from Lemma 7.31 and Lemma 7.32.

Privacy analysis. Let $\mathcal{Y}$ be the set of all matrices $Y(G) = \frac{1}{\gamma d}(A(G) - \frac{d}{n}J)$ where G is a graph on n vertices. We further define $q : \mathcal{Y} \to \mathcal{K}$ to be the function

$$q(Y) := \operatorname{argmin}_{X\in\mathcal{K}}\|Y - X\|_{\mathrm{F}}^2. \tag{7.4.1}$$

We first use Lemma 7.25 to prove that function q is stable.

Lemma 7.30 (Stability)**.** *The function q as defined in Eq. (7.4.1) has ℓ_2-sensitivity $\Delta_{q,2} \leqslant \sqrt{\frac{24}{n\gamma d}}$.*

Proof. Let $g : \mathcal{Y}\times\mathcal{K} \to \mathbb{R}$ be the function $g(Y, X) := \|X\|_{\mathrm{F}}^2 - 2\langle Y, X\rangle$. Applying Lemma 7.25 with $f_Y(\cdot) = g(Y, \cdot)$, it suffices to prove that g has ℓ_1-sensitivity $\frac{4}{n\gamma d}$ with respect to Y and that it is 2-strongly convex with respect to X. The ℓ_1-sensitivity bound follows by observing that adjacent Y, Y' satisfy $\|Y - Y'\|_1 \leqslant \frac{2}{\gamma d}$ and that any $X \in \mathcal{K}$ satisfies $\|X\|_\infty \leqslant \frac{1}{n}$. Thus it remains to prove strong convexity with respect to $X \in \mathcal{K}$. Let $X, X' \in \mathcal{K}$ then

$$\begin{aligned}
\|X'\|_{\mathrm{F}}^2 &= \|X\|_{\mathrm{F}}^2 + 2\langle X' - X, X\rangle + \|X - X'\|_{\mathrm{F}}^2 \\
&= \|X\|_{\mathrm{F}}^2 + 2\langle X' - X, X + Y - Y\rangle + \|X - X'\|_{\mathrm{F}}^2 \\
&= g(Y, X) + \langle X' - X, \nabla g(X, Y)\rangle + 2\langle X', Y\rangle + \|X - X'\|_{\mathrm{F}}^2\,.
\end{aligned}$$

That is $g(Y, X)$ is 2-strongly convex with respect to X. Note any $X \in \mathcal{K}$ is symmetric. Then the result follows by Lemma 7.25. □

Then it is easy to show the algorithm is private.

Lemma 7.31 (Privacy). *The weak recovery algorithm (Algorithm 7.29) is (ε, δ)-DP.*

Proof. Since any $X \in \mathcal{K}$ is symmetric, we only need to add a symmetric noise matrix to obtain privacy. Combining Lemma 7.30 with Lemma 7.19, we immediately get that the algorithm is (ε, δ)-private. □

Utility analysis. Now we show the utility guarantee of our priavte weak recovery algorithm.

Lemma 7.32 (Utility). *For any $x \in \{\pm 1\}^n$, on input $\mathbf{G} \sim \mathsf{SBM}_n(\gamma, d, x)$, Algorithm 7.29 efficiently outputs $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ satisfying*

$$\operatorname{err}(\hat{x}(\mathbf{G}), x) \leqslant \frac{6400}{\gamma\sqrt{d}} + \frac{7000}{\gamma d} \cdot \frac{\log(2/\delta)}{\varepsilon^2},$$

with probability $1 - \exp(-\Omega(n))$.

To prove Lemma 7.32, we need the following lemma which is an adaption of a well-known result in SBM [GV16, Theorem 1.1]. Its proof is deferred to Appendix E.1.

Lemma 7.33. *Consider the settings of Lemma 7.32. With probability $1 - \exp(-\Omega(n))$,*

$$\left\| X_1(\mathbf{G}) - \frac{1}{n} x x^\top \right\|_F^2 \leqslant \frac{800}{\gamma\sqrt{d}}.$$

Proof of Lemma 7.32. By Lemma 7.33, we have

$$\left\| X_1(\mathbf{G}) - \frac{1}{n} x x^\top \right\| \leqslant \left\| X_1(\mathbf{G}) - \frac{1}{n} x x^\top \right\|_F \leqslant \sqrt{\frac{800}{\gamma\sqrt{d}}} =: r(\gamma, d)$$

with probability $1 - \exp(-\Omega(n))$. We condition our following analysis on this event happening.

Let $\mathbf{u}$ be the leading eigenvector of $X_1(\mathbf{G})$. Let λ_1 and λ_2 be the largest and second largest eigenvalues of $X_1(\mathbf{G})$. By Weyl's inequality (Lemma E.10) and the assumption $\gamma\sqrt{d} \geqslant 12800$, we have

$$\lambda_1 - \lambda_2 \geqslant 1 - 2r(\gamma, d) \geqslant \frac{1}{2}.$$

Let $\mathbf{v}$ be the leading eigenvector of $X_1(\mathbf{G}) + \mathbf{W}$. By Davis-Kahan's theorem (Lemma E.11), we have

$$\|\mathbf{u} - \mathbf{v}\| \leqslant \frac{2\|\mathbf{W}\|}{\lambda_1 - \lambda_2} \leqslant 4\|\mathbf{W}\|,$$
$$\left\|\mathbf{u} - x/\sqrt{n}\right\| \leqslant 2\left\| X_1(\mathbf{G}) - \frac{1}{n} x x^\top \right\| \leqslant 2r(\gamma, d).$$

Putting things together and using Fact E.5, we have

$$\left\|\mathbf{v} - x/\sqrt{n}\right\| \leqslant \|\mathbf{u} - \mathbf{v}\| + \left\|\mathbf{u} - x/\sqrt{n}\right\| \leqslant \frac{24\sqrt{6}}{\sqrt{\gamma d}} \frac{\sqrt{\log(2/\delta)}}{\varepsilon} + 2r(\gamma, d)$$

with probability $1 - \exp(-\Omega(n))$.

Observe $\mathrm{Ham}(\mathrm{sign}(y), x) \leqslant \|y - x\|^2$ for any $y \in \mathbb{R}^n$ and any $x \in \{\pm 1\}^n$. Then with probability $1 - \exp(-\Omega(n))$,

$$\frac{1}{n} \cdot \mathrm{Ham}(\mathrm{sign}(\mathbf{v}), x) \leqslant \left\|\mathbf{v} - x/\sqrt{n}\right\|^2 \leqslant \frac{6400}{\gamma\sqrt{d}} + \frac{7000}{\gamma d} \cdot \frac{\log(2/\delta)}{\varepsilon^2}.$$

□

Proof of Theorem 7.28. By Lemma 7.31 and Lemma 7.32. □

7.4.2 Private exact recovery for stochastic block models

In this section, we prove Theorem 7.2. We show how to achieve exact recovery in stochastic block models privately by combining the private weak recovery algorithm we obtained in the previous section and a private majority voting scheme.

Since exact recovery is only possible with logarithmic average degree (just to avoid isolated vertices), it is more convenient to work with the following standard parameterization of stochastic block models. Let $\alpha > \beta > 0$ be fixed constants. The intra-community edge probability is $\alpha \cdot \frac{\log n}{n}$, and the inter-community edge probability is $\beta \cdot \frac{\log n}{n}$. In the language of Model 7.1, it is $\mathsf{SBM}_n(\frac{\alpha+\beta}{2} \cdot \log n, \frac{\alpha-\beta}{\alpha+\beta}, x)$. Our main result is the following theorem.

Theorem 7.34 (Private exact recovery of SBM, restatement of Theorem 7.2)**.** *Let $\varepsilon, \delta \geqslant 0$. Suppose α, β are fixed constants satisfying*[20]

$$\sqrt{\alpha} - \sqrt{\beta} \geqslant 16 \quad \textit{and} \quad \alpha - \beta \geqslant \Omega\left(\frac{1}{\varepsilon^2} \cdot \frac{\log(2/\delta)}{\log n} + \frac{1}{\varepsilon}\right), \tag{7.4.2}$$

Then there exists an algorithm (Algorithm 7.36) such that, for any balanced[21] *$x \in \{\pm 1\}^n$, on input $\mathbf{G} \sim \mathsf{SBM}_n(\frac{\alpha+\beta}{2} \cdot \log n, \frac{\alpha-\beta}{\alpha+\beta}, x)$, outputs $\hat{x}(\mathbf{G}) \in \{x, -x\}$ with probability $1 - o(1)$. Moreover, the algorithm is (ε, δ)-differentially private for any input graph and runs in polynomial time.*

Remark 7.35. In a standard regime of privacy parameters where $\varepsilon \leqslant O(1)$ and $\delta = 1/\mathrm{poly}(n)$, the private exact recovery threshold Eq. (7.4.2) reads

$$\sqrt{\alpha} - \sqrt{\beta} \geqslant 16 \quad \text{and} \quad \alpha - \beta \geqslant \Omega\big(\varepsilon^{-2} + \varepsilon^{-1}\big),$$

Recall the non-private exact recovery threshold is $\sqrt{\alpha} - \sqrt{\beta} > \sqrt{2}$. Thus the non-private part in Eq. (7.4.2), i.e. 16, is close to optimal.

[20] In the language of Model 7.1, for any t we have $\sqrt{\alpha} - \sqrt{\beta} \geqslant t$ if and only if $\frac{d}{\log n}(1 - \sqrt{1 - \gamma^2}) \geqslant \frac{t^2}{2}$.

[21] Recall a vector $x \in \{\pm 1\}^n$ is said to be balanced if $\sum_{i=1}^n x_i = 0$.

Algorithm 7.36 starts with randomly splitting the input graph G into two subgraphs $\mathbf{G}_1$ and $\mathbf{G}_2$. Setting the graph-splitting probability to $1/2$, each subgraph will contain about half of the edges of G. Then we run an (ε, δ)-DP weak recovery algorithm (Algorithm 7.29) on $\mathbf{G}_1$ to get a rough estimate $\tilde{x}(\mathbf{G}_1)$ of accuracy around 90%. Finally, we boost the accuracy to 100% by doing majority voting (Algorithm 7.37) on $\mathbf{G}_2$ based on the rough estimate $\tilde{x}(\mathbf{G}_1)$. That is, if a vertex has more neighbors from the opposite community (according to $\tilde{x}(\mathbf{G}_1)$) in $\mathbf{G}_2$, then we assign this vertex to the opposite community. To make the majority voting step private, we add some noise to the vote.

Algorithm 7.36 (Private exact recovery for SBM)**.**
Input: Graph G
Operations:

1. Graph-splitting: Initialize $\mathbf{G}_1$ to be an empty graph on vertex set $V(G)$. Independently put each edge of G in $\mathbf{G}_1$ with probability $1/2$. Let $\mathbf{G}_2 = G \setminus \mathbf{G}_1$.
2. Rough estimation on $\mathbf{G}_1$: Run the (ε, δ)-DP partial recovery algorithm (Algorithm 7.29) on $\mathbf{G}_1$ to get a rough estimate $\tilde{x}(\mathbf{G}_1)$.
3. Majority voting on $\mathbf{G}_2$: Run the $(\varepsilon, 0)$-DP majority voting algorithm (Algorithm 7.37) with input $(\mathbf{G}_2, \tilde{x}(\mathbf{G}_1))$ and get output $\hat{\mathbf{x}}$.
4. Return $\hat{\mathbf{x}}$.

Algorithm 7.37 (Private majority voting)**.**
Input: Graph G, rough estimate $\tilde{x} \in \{\pm 1\}^n$
Operations:

1. For each vertex $v \in V(G)$, let $\mathbf{Z}_v = \mathbf{S}_v - \mathbf{D}_v$ where
   - $\mathbf{D}_v = \sum_{\{u,v\} \in E(G)} \mathbb{1}_{[\tilde{x}_u \neq \tilde{x}_v]}$,
   - $\mathbf{S}_v = \sum_{\{u,v\} \in E(G)} \mathbb{1}_{[\tilde{x}_u = \tilde{x}_v]}$.

   Set $\hat{\mathbf{x}}_v = \mathrm{sign}(\mathbf{Z}_v + \mathbf{W}_v) \cdot \tilde{x}(\mathbf{G}_1)_v$ where $\mathbf{W}_v \sim \mathrm{Lap}(2/\varepsilon)$.
2. Return $\hat{\mathbf{x}}$.

In the rest of this section, we will show Algorithm 7.36 is private in Lemma 7.39 and it recovers the hidden communities exactly with high probability in Lemma 7.41. Then Theorem 7.34 follows directly from Lemma 7.39 and Lemma 7.41.

Privacy analysis. We first show the differential privcay of the majority voting algorithm (Algorithm 7.37) with respect to input graph G (i.e. assuming fixed the input rough

estimate).

Lemma 7.38. *Algorithm 7.37 is $(\varepsilon, 0)$-DP with respect to input G.*

Proof. Observing the ℓ_1-sensitivity of the degree count function Z in step is 2, the $(\varepsilon, 0)$-DP follows directly from Laplace mechanism (Lemma 7.19) and post-processing (Lemma 7.9). □

Then the privacy of the private exact recovery algorithm (Algorithm 7.36) is a consequence of composition.

Lemma 7.39 (Privacy)**.** *Algorithm 7.36 is (ε, δ)-DP.*

Proof. Let $\mathcal{A}_1 : \mathcal{G}_n \to \{\pm 1\}^n$ denote the (ε, δ)-DP recovery algorithm in step 2. Let $\mathcal{A}_2 : \mathcal{G}_n \times \{\pm 1\}^n \to \{\pm 1\}^n$ denote the $(\varepsilon, 0)$-DP majority voting algorithm in step 3. Let $\mathcal{A}$ be the composition of $\mathcal{A}_1$ and $\mathcal{A}_2$.

We first make several notations. Given a graph H and an edge e, H_e is a graph obtained b adding e to H. Given a graph H, $\mathbf{G}_1(H)$ is a random subgraph of H by keeping each edge of H with probability $1/2$ independently.

Now, fix two adjacent graphs G and G_e where edge e appears in G_e but not in G. Also, fix two arbitrary possible outputs $x_1, x_2 \in \{\pm 1\}^n$ of algorithm $\mathcal{A}$.[22] It is direct to see,

$$\mathbb{P}(\mathcal{A}(G) = (x_1, x_2)) = \sum_{H \subseteq G} \mathbb{P}(\mathcal{A}_1(H) = x_1)\, \mathbb{P}(\mathcal{A}_2(G \setminus H, x_1) = x_2)\, \mathbb{P}(\mathbf{G}_1(G) = H). \tag{7.4.3}$$

Since $\mathbb{P}(\mathbf{G}_1(G) = H) = \mathbb{P}(\mathbf{G}_1(G_e) = H) + \mathbb{P}(\mathbf{G}_1(G_e) = H_e)$ for any $H \subseteq G$, we have

$$\begin{aligned}\mathbb{P}(\mathcal{A}(G_e) = (x_1, x_2)) = \sum_{H \subseteq G} &\mathbb{P}(\mathcal{A}_1(H) = x_1)\, \mathbb{P}(\mathcal{A}_2(G_e \setminus H, x_1) = x_2)\, \mathbb{P}(\mathbf{G}_1(G_e) = H) \\ &+ \mathbb{P}(\mathcal{A}_1(H_e) = x_1)\, \mathbb{P}(\mathcal{A}_2(G_e \setminus H_e, x_1) = x_2)\, \mathbb{P}(\mathbf{G}_1(G_e) = H_e)\end{aligned} \tag{7.4.4}$$

Since both $\mathcal{A}_1$ and $\mathcal{A}_2$ are (ε, δ)-DP, we have for each $H \subseteq G$,

$$\mathbb{P}(\mathcal{A}_1(H_e) = x_1) \leqslant e^{\varepsilon}\, \mathbb{P}(\mathcal{A}_1(H) = x_1) + \delta, \tag{7.4.5}$$

$$\mathbb{P}(\mathcal{A}_2(G_e \setminus H, x_1) = x_2) \leqslant e^{\varepsilon}\, \mathbb{P}(\mathcal{A}_2(G \setminus H, x_1) = x_2) + \delta. \tag{7.4.6}$$

Plugging Eq. (7.4.5) and Eq. (7.4.6) into Eq. (7.4.4), we obtain

$$\begin{aligned}\mathbb{P}(\mathcal{A}(G_e) = (x_1, x_2)) &\leqslant \sum_{H \subseteq G} [e^{\varepsilon}\, \mathbb{P}(\mathcal{A}_1(H) = x_1)\, \mathbb{P}(\mathcal{A}_2(G \setminus H, x_1) = x_2) + \delta]\, \mathbb{P}(\mathbf{G}_1(G) = H) \\ &= e^{\varepsilon}\, \mathbb{P}(\mathcal{A}(G) = (x_1, x_2)) + \delta.\end{aligned}$$

Similarly, we can show

$$\mathbb{P}(\mathcal{A}(G) = (x_1, x_2)) \leqslant e^{\varepsilon}\, \mathbb{P}(\mathcal{A}(G_e) = (x_1, x_2)) + \delta. \tag{7.4.7}$$

□

[22] We can imagine that algorithm $\mathcal{A}$ first outputs (x_1, x_2) and then outputs x_2 as a post-processing step.

Utility analysis. We first show the utility guarantee of the priavte majority voting algorithm.

Lemma 7.40. *Suppose* $\mathbf{G}$ *is generated by first sampling* $\mathbf{G} \sim \mathsf{SBM}_n(\frac{\alpha+\beta}{2} \cdot \log n, \frac{\alpha-\beta}{\alpha+\beta}, x)$ *for some balanced* x *and then for each vertex removing at most* $\Delta \leqslant O(\log^2 n)$ *adjacent edges arbitrarily. Then on input* $\mathbf{G}$ *and a balanced rough estimate* $\tilde{x}$ *satisfying* $\mathrm{Ham}(\tilde{x}, x) \leqslant n/16$, *Algorithm 7.37 efficiently outputs* $\hat{x}(\mathbf{G})$ *such that for each vertex* v,

$$\mathbb{P}(\hat{x}(\mathbf{G})_v \neq x_v) \leqslant \exp\left(-\frac{1}{64} \cdot \varepsilon(\alpha-\beta) \cdot \log n\right) + 2 \cdot \exp\left(-\frac{1}{16^2} \cdot \frac{(\alpha-\beta)^2}{\alpha+\beta} \cdot \log n\right).$$

Proof. Let us fix an arbitrary vertex v and analyze the probability $\mathbb{P}(\hat{x}(\mathbf{G})_v \neq x_v)$. Let $r := \mathrm{Ham}(\tilde{x}, x)/n$. Then it is not hard to see

$$\mathbb{P}(\hat{x}(\mathbf{G})_v \neq x_v) \leqslant \mathbb{P}(\mathbf{B} + \mathbf{A}' - \mathbf{A} - \mathbf{B}' + \mathbf{W} > 0) \tag{7.4.8}$$

where

- $\mathbf{A} \sim \mathrm{Binomial}((1/2 - r)n - \Delta, \alpha\frac{\log n}{n})$, corresponding to the number of neighbors that are from the same community and correctly labeled by $\tilde{x}$,
- $\mathbf{B}' \sim \mathrm{Binomial}(rn - \Delta, \beta\frac{\log n}{n})$, corresponding to the number of neighbors that are from the different community but incorrectly labeled by $\tilde{x}$,
- $\mathbf{B} \sim \mathrm{Binomial}((1/2 - r)n, \beta\frac{\log n}{n})$, corresponding to the number of neighbors that are from the different community and correctly labeled by $\tilde{x}$,
- $\mathbf{A}' \sim \mathrm{Binomial}(rn, \alpha\frac{\log n}{n})$, corresponding to the number of neighbors that are from the same community but incorrectly labeled by $\tilde{x}$,
- $\mathbf{W} \sim \mathrm{Lap}(0, 2/\varepsilon)$, independently.

The Δ term appearing in both $\mathbf{A}$ and $\mathbf{B}'$ corresponds to the worst case where Δ "favorable" edges are removed. If $r \geqslant \Omega(1)$, then $\Delta = O(\log^2 n)$ is negligible to $rn = \Theta(n)$ and we can safely ignore the effect of removing Δ edges. If $r = o(1)$, then we can safely assume $\tilde{x}$ is correct on all vertices and ignore the effect of removing Δ edges as well. Thus, we will assume $\Delta = 0$ in the following analysis.

For any t, t', we have

$$\begin{aligned}\mathbb{P}(\mathbf{A}' + \mathbf{B} - \mathbf{A} - \mathbf{B}' + \mathbf{W} > 0) &\leqslant \mathbb{P}(\mathbf{A}' + \mathbf{B} + \mathbf{W} > t) + \mathbb{P}(\mathbf{A} + \mathbf{B}' \leqslant t) \\ &\leqslant \mathbb{P}(\mathbf{A}' + \mathbf{B} \geqslant t - t') + \mathbb{P}(\mathbf{W} \geqslant t') + \mathbb{P}(\mathbf{A} + \mathbf{B}' \leqslant t).\end{aligned}$$

We choose t, t' by first picking two constants $a, b > 0$ satisfying $a + b < 1$ and then solving

- $\mathbb{E}[\mathbf{A}' + \mathbf{B}] - t = a \cdot (\mathbb{E}[\mathbf{A} + \mathbf{B}'] - \mathbb{E}[\mathbf{A}' + \mathbf{B}])$ and

- $t' = (1-a-b)\cdot(\mathbb{E}[\mathbf{A}+\mathbf{B}'] - \mathbb{E}[\mathbf{A}'+\mathbf{B}])$.

By Fact 7.14,

$$\mathbb{P}(\mathbf{W} > t') \leqslant \exp\left(-\frac{t'\varepsilon}{2}\right) \leqslant \exp\left(-\frac{(1/4-r)(1-a-b)}{2}\cdot\varepsilon(\alpha-\beta)\cdot\log n\right).$$

By Fact E.8 and the assumption $r \leqslant 1/16$, we have

$$\mathbb{P}(\mathbf{A}+\mathbf{B}' \leqslant t) \leqslant \exp\left(-\frac{(\mathbb{E}[\mathbf{A}+\mathbf{B}']-t)^2}{2\,\mathbb{E}[\mathbf{A}+\mathbf{B}']}\right) \leqslant \exp\left(-(1/4-r)^2a^2\cdot\frac{(\alpha-\beta)^2}{\alpha+\beta}\cdot\log n\right).$$

Setting $b = 1/2$, by Fact E.8 and the assumption $r \leqslant 1/16$, we have

$$\mathbb{P}(\mathbf{A}'+\mathbf{B} \geqslant t-t') \leqslant \exp\left(-\frac{(t-t'-\mathbb{E}[\mathbf{A}'+\mathbf{B}])^2}{t-t'+\mathbb{E}[\mathbf{A}'+\mathbf{B}]}\right) \leqslant \exp\left(-\frac{2(1/4-r)^2}{7}\cdot\frac{(\alpha-\beta)^2}{\alpha+\beta}\cdot\log n\right).$$

Further setting $a = 1/3$, we have

$$\mathbb{P}(\hat{x}(\mathbf{G})_v \neq x_v) \leqslant \exp\left(-\frac{1/4-r}{12}\cdot\varepsilon(\alpha-\beta)\cdot\log n\right) + 2\cdot\exp\left(-\frac{(1/4-r)^2}{9}\cdot\frac{(\alpha-\beta)^2}{\alpha+\beta}\cdot\log n\right).$$

Finally, plugging the assumption $r \leqslant 1/16$ to conclude. □

Then it is not difficult to show the utility guarantee of our priavte exact recovery algorithm.

Lemma 7.41 (Utility). *Suppose α, β are fixed constants satisfying*

$$\sqrt{\alpha} - \sqrt{\beta} \geqslant 16 \quad \textit{and} \quad \alpha - \beta \geqslant \Omega\left(\frac{1}{\varepsilon^2}\cdot\frac{\log(2/\delta)}{\log n} + \frac{1}{\varepsilon}\right).$$

Then for any balanced $x \in \{\pm1\}^n$, on input $\mathbf{G} \sim \mathsf{SBM}_n(\frac{\alpha+\beta}{2}\cdot\log n, \frac{\alpha-\beta}{\alpha+\beta}, x)$, Algorithm 7.36 efficiently outputs $\hat{x}(\mathbf{G})$ satisfying $\hat{x}(\mathbf{G}) \in \{x, -x\}$ with probability $1 - o(1)$.

Proof. We will show the probability of a fixed vertex being misclassified is at most $o(1/n)$. Then by union bound, exact recovery can be achieved with probability $1 - o(1)$.

As the graph-splitting probability is $1/2$, $\mathbf{G}_1$ follows $\mathsf{SBM}_n(\frac{\alpha}{2}\cdot\frac{\log n}{n}, \frac{\beta}{2}\cdot\frac{\log n}{n}, x)$. By Theorem 7.28, the rough estimate $\tilde{x}(\mathbf{G}_1)$ satisfies[23]

$$\operatorname{err}(\tilde{x}(\mathbf{G}_1), x) \leqslant r := o(1) + \frac{14000}{(\alpha-\beta)\varepsilon^2}\cdot\frac{\log(2/\delta)}{\log n}. \tag{7.4.9}$$

with probability at least $1 - \exp(-\Omega(n))$. Without loss of generality, we can assume $\operatorname{Ham}(\tilde{x}(\mathbf{G}_1), x) \leqslant rn$, since we consider $-x$ otherwise. By Fact E.6, the maximum degree of

[23] It is easy to make the output of Algorithm 7.29 balanced at the cost of increasing the error rate by a factor of at most 2.

$\mathbf{G}_1$ is at most $\Delta := 2\log^2 n$ with probability at least $1 - n\exp(-(\log n)^2/3)$. In the following, we condition our analysis on the above two events regarding $\tilde{x}(\mathbf{G}_1)$ and $\mathbf{G}_1$.

Now, let us fix a vertex and analyze the probability p_e that it is misclassified after majority voting. With G_1 being fixed, $\mathbf{G_2}$ can be thought of as being generated by first sampling $\mathbf{G}$ and then removing G_1 from $\mathbf{G}$. To make $r \leqslant 1/16$, it suffices to ensure $\alpha - \beta > \frac{500^2}{\varepsilon^2} \cdot \frac{\log(2/\delta)}{\log n}$ by Eq. (7.4.9).Then by Lemma 7.40, we have

$$p_e \leqslant \exp\left(-\frac{1}{64} \cdot \varepsilon(\alpha-\beta) \cdot \log n\right) + 2 \cdot \exp\left(-\frac{1}{16^2} \cdot \frac{(\alpha-\beta)^2}{\alpha+\beta} \cdot \log n\right).$$

To make p_e at most $o(1/n)$, it suffices to ensure

$$\frac{1}{64} \cdot \varepsilon(\alpha-\beta) > 1 \quad \text{and} \quad \frac{1}{16^2} \cdot \frac{(\alpha-\beta)^2}{\alpha+\beta} > 1.$$

Note $(\alpha-\beta)^2/(\alpha+\beta) > (\sqrt{\alpha}-\sqrt{\beta})^2$ for $\alpha > \beta$. Therefore, as long as

$$\sqrt{\alpha} - \sqrt{\beta} \geqslant 16 \quad \text{and} \quad \alpha - \beta \geqslant \frac{500^2}{\varepsilon^2} \cdot \frac{\log(2/\delta)}{\log n} + \frac{64}{\varepsilon},$$

Algorithm 7.36 recovers the hidden communities exactly with probability $1 - o(1)$. □

Proof of Theorem 7.34. By Lemma 7.39 and Lemma 7.41. □

7.4.3 Inefficient recovery using the exponential mechanism

In this section, we will present an inefficient algorithm satisfying pure privacy which succeeds for all ranges of parameters - ranging from weak to exact recovery. The algorithm is based on the exponential mechanism [MT07] combined with the majority voting scheme introduced in section Section 7.4.2. In particular, we will show

Theorem 7.42 (Full version of Theorem 7.3)**.** *Let $\gamma\sqrt{d} \geqslant 12800$ and $x \in \{\pm 1\}^n$ be balanced. Let $\zeta \geqslant 2\exp\left(-\frac{\gamma^2 d}{512}\right)$. For any $\varepsilon \geqslant \frac{64\log(2/\zeta)}{\gamma d}$, there exists an algorithm, Algorithm 7.43, which on input $\mathbf{G} \sim \mathsf{SBM}_n(\gamma, d, x^*)$ outputs an estimate $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ satisfying*

$$\operatorname{err}(\hat{x}(\mathbf{G}), x^*) \leqslant \zeta$$

with probability at least $1 - \zeta$. In addition, the algorithm is ε-private. Further, by slightly modifying the algorithm, we can achieve error $20/\sqrt{\log(1/\zeta)}$ with probability $1 - e^{-n}$.[24]

[24]The first, smaller, error guarantee additionally needs the requirement that $\zeta \leqslant \exp(-640)$. The second one does not.

A couple of remarks are in order. First, our algorithm works across all degree-regimes in the literature and matches known non-private thresholds and rates up to constants. We remark that for ease of exposition we did not try to optimize these constants. In particular, for $\gamma^2 d$ a constant we achieve weak recovery. We reiterate, that $\gamma^2 d > 1$ is the optimal non-private threshold. For the regime, where $\gamma^2 d = \omega(1)$, it is known that the optimal error rate is $\exp\left(-(1-o(1))\gamma^2 d\right)$ even non-privately [ZZ16], where $o(1)$ goes to zero as $\gamma^2 d$ tends to infinity. We match this up to constants. Moreover, our algorithm achieves exact recovery as soon as $\gamma^2 d \geqslant 512 \log n$ since then $\zeta < \frac{1}{n}$. This also matches known non-private threshholds up to constants [ABH15, MNS15a]. Also, our dependence on the privacy parameter ε is also optimal as shown by the information-theoretic lower bounds in Section 7.4.4.

We also emphasize, that if we only aim to achieve error on the order of

$$\frac{1}{\gamma\sqrt{d}} = \Theta\left(\frac{1}{\sqrt{\log(1/\zeta)}}\right),$$

we can achieve exponentially small failure probability in n, while keeping the privacy parameter ε the same. This can be achieved, by ommitting the boosting step in our algorithm and will be clear from the proof of Theorem 7.42. We remark that in this case, we can also handle non-balanced communities.

Again, for an input graph G, consider the matrix $Y(G) = \frac{1}{\gamma d}\left(A(G) - \frac{d}{n}J\right)$. For $x \in \{\pm 1\}^n$ we define the score function

$$s_G(x) = \langle x, Y(G)x \rangle .$$

Since the entries of $A(G)$ are in $[0,1]$ and adjacent graphs differ in at most one edge, it follows immediately, that this score function has sensitivity at most

$$\Delta = \max_{\substack{G \sim G', \\ x \in \{\pm 1\}^n}} |s_G(x) - s_{G'}(x)| = \frac{2}{\gamma d} \cdot \max_{\substack{G \sim G', \\ x \in \{\pm 1\}^n}} |\langle x, (A(G) - A(G'))x \rangle| \leqslant \frac{2}{\gamma d} .$$

Algorithm 7.43 (Inefficient algorithm for SBM).
Input: Graph G, privacy parameter $\varepsilon > 0$
Operations:

1. Graph-splitting: Initialize $\mathbf{G}_1$ to be an empty graph on vertex set $V(G)$. Independently assign each edge of G to $\mathbf{G}_1$ with probability $1/2$. Let $\mathbf{G}_2 = G \setminus \mathbf{G}_1$.
2. Rough estimation on $\mathbf{G}_1$: Sample $\tilde{x}$ from the distribution with density
$$p(x) \propto \exp\left(\frac{\varepsilon}{2\Delta}\langle x, Y(\mathbf{G}_1)x\rangle\right),$$
where $\Delta = \frac{2}{\gamma d}$.
3. Majority voting on $\mathbf{G}_2$: Run the ε-DP majority voting algorithm (Algorithm 7.37) with input $(\mathbf{G}_2, \tilde{x}(\mathbf{G}_1))$. Denote its output by $\hat{\mathbf{x}}$.
4. Return $\hat{\mathbf{x}}$.

We first analyze the privacy guarantees of the above algorithm.

Lemma 7.44. *Algorithm 7.43 is ε-DP.*

Proof. For simplicity and clarity of notation, we will show that the algorithm satisfies 2ε-DP. Clearly, the graph splitting step is 0-DP. Step 2 corresponds to the exponential mechanism. Since the sensitivity of the score function is at most $\Delta = \frac{2}{\gamma d}$ it follows by the standard analysis of the mechanism that this step is ε-DP [MT07]. By Lemma 7.38, the majority voting step is also ε-DP. Hence, the result follows by composition (cf. Lemma 7.11). □

Next, we will analyze its utility.

Lemma 7.45. *Let $\gamma\sqrt{d} \geqslant 12800$ and $x \in \{\pm 1\}^n$ be balanced. Let $\exp(-640) \geqslant \zeta \geqslant 2\exp\left(-\frac{\gamma^2 d}{512}\right)$, $\varepsilon \geqslant \frac{64\log(2/\zeta)}{\gamma d}$, and $\mathbf{G} \sim \mathsf{SBM}_n(\gamma, d, x^*)$, the output $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ of Algorithm 7.43 satisfies*
$$\operatorname{err}(\hat{x}(\mathbf{G}), x^*) \leqslant \zeta$$
with probability at least $1 - \zeta$.

Proof. We will first show that the rough estimate $\tilde{x}$ obtained in step 2 achieves
$$\operatorname{err}(\tilde{x}, x^*) \leqslant \frac{20}{\sqrt{\log(1/\zeta)}}$$
with probability e^{-n}. This will prove the second part of the theorem - for this we don't need that $\zeta \leqslant \exp(-640)$. In fact, arbitrary ζ works. The final error guarantee will then follow by

Lemma 7.40. First, notice that similar to the proof of [GV16, Lemma 4.1], using Bernstein's inequality and a union bound, we can show that (cf. Fact E.2 for a full proof)

$$\max_{x\in\{\pm1\}^n}\left|\langle x,\left[Y(\mathbf{G})-\frac{1}{n}x^*(x^*)^\top\right]x\rangle\right|\leqslant\frac{100n}{\gamma\sqrt{d}}\leqslant\frac{5}{\sqrt{\log(1/\zeta)}}$$

with probability at least $1-\exp^{-10n}$. Recall that $s_{\mathbf{G}}(x)=\langle x,Y(\mathbf{G})x\rangle$. Let $\alpha=\frac{5}{\sqrt{\log(1/\zeta)}}$. We call $x\in\{\pm1\}^n$ *good* if $s_{\mathbf{G}}(x)\geqslant(1-3\alpha)n$. It follows that for good x it holds that

$$\frac{1}{n}\cdot\langle x,x^*\rangle^2\geqslant\langle x,Y(\mathbf{G})x\rangle-\left|\left\langle x,\left[Y(\mathbf{G})-\frac{1}{n}x^*(x^*)^\top\right]x\right\rangle\right|\geqslant(1-4\alpha)n\,.$$

Which implies that

$$2\operatorname{err}(x,x^*)\leqslant1-\sqrt{1-4\alpha}=1-\frac{1-4\alpha}{\sqrt{1-4\alpha}}\leqslant1-\frac{1-4\alpha}{1-2\alpha}=\frac{2\alpha}{1-2\alpha}\leqslant4\alpha\,,$$

where we used that $\alpha\leqslant1/4$ and that $\sqrt{1-4x}\leqslant1-2x$ for $x\geqslant0$. Hence, we have for good x that

$$\operatorname{err}(x,x^*)\leqslant\frac{20}{\sqrt{\log(1/\zeta)}}\,.$$

Since $s_{\mathbf{G}}(x^*)\geqslant(1-\alpha)n$,there is at least one good candidate. Hence, we can bound the probability that we do not output a good x as

$$\frac{\exp\left(\frac{\varepsilon}{2\Delta}(1-3\alpha)n\right)\cdot e^n}{\exp\left(\frac{\varepsilon}{2\Delta}(1-\alpha)n\right)\cdot1}=\exp\left(\left(1-\frac{2\varepsilon\alpha}{\Delta}\right)n\right)\leqslant e^{-n}\,,$$

where we used that

$$\frac{2\varepsilon\alpha}{\Delta}\geqslant\frac{64\log(2/\zeta)}{\gamma d}\cdot\frac{5\gamma d}{\sqrt{\log(1/\zeta)}}\geqslant320\sqrt{\log(1/\zeta)}\geqslant2\,.$$

We will use Lemma 7.40 to proof the final conclusion of the theorem. In what follows, assume without loss of generality that $\operatorname{Ham}(x,x^*)<\operatorname{Ham}(x,-x^*)$. The above discussion implies that

$$\operatorname{Ham}(x,x^*)\leqslant8\alpha n\leqslant\frac{40n}{\sqrt{\log(1/\zeta)}}\leqslant\frac{n}{16}\,,$$

where the last inequality uses $\zeta\leqslant e^{-640}$. Further, by Fact E.6 it also follows that the maximum degree of $\mathbf{G}_2$ is at most $O\left(\log^2 n\right)$ (by some margin). Recall that $\mathbf{G}_2\sim\mathsf{SBM}(d,\gamma,x^*)$. In the parametrization of Lemma 7.40 this means that

$$\alpha=\frac{(1+\gamma)d}{\log n}\,,\qquad\beta=\frac{(1-\gamma)d}{\log n}\,,$$

$$\alpha - \beta = \frac{2\gamma d}{\log n}, \qquad \alpha + \beta = \frac{2d}{\log n}.$$

Thus, it follows that the output $\hat{x}$ of the majority voting step satisfies for every vertex v

$$\begin{aligned}\mathbb{P}(\hat{x}(\mathbf{G})_v \neq x_v) &\leqslant \exp\left(-\frac{1}{64}\cdot\varepsilon(\alpha-\beta)\cdot\log n\right) + 2\cdot\exp\left(-\frac{1}{16^2}\cdot\frac{(\alpha-\beta)^2}{\alpha+\beta}\cdot\log n\right)\\ &\leqslant \exp\left(-\frac{1}{32}\cdot\varepsilon\gamma d\right) + \exp\left(-\frac{1}{16^2}\cdot\gamma^2 d\right)\\ &\leqslant \zeta^2/4 + \zeta^2/4 \leqslant \zeta^2.\end{aligned}$$

By Markov's Inequality it now follows that

$$\mathbb{P}(\mathrm{err}(\hat{x}(\mathbf{G}), x^*) \geqslant \zeta) \leqslant \zeta.$$

□

7.4.4 Lower bound on the parameters for private recovery

In this section, we prove a tight lower bound for private recovery for stochastic block models. Recall the definition of error rate, $\mathrm{err}(u,v) := \frac{1}{n}\cdot\min\{\mathrm{Ham}(u,v), \mathrm{Ham}(u,-v)\}$ for $u, v \in \{\pm 1\}^n$. Our main result is the following theorem.

Theorem 7.46 (Full version of Theorem 7.4)**.** *Suppose there exists an ε-differentially private algorithm such that for any balanced $x \in \{\pm 1\}^n$, on input $\mathbf{G} \sim \mathsf{SBM}_n(d,\gamma,x)$, outputs $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ satisfying*

$$\mathbb{P}(\mathrm{err}(\hat{x}(\mathbf{G}), x) < \zeta) \geqslant 1 - \eta,$$

where[25] *$1/n \leqslant \zeta \leqslant 0.04$ and the randomness is over both the algorithm and stochastic block models. Then,*

$$e^{2\varepsilon} - 1 \geqslant \Omega\left(\frac{\log(1/\zeta)}{\gamma d} + \frac{\log(1/\eta)}{\zeta n \gamma d}\right). \tag{7.4.10}$$

Remark 7.47. Both terms in lower bound Eq. (7.4.10) are tight up to constants by the following argument. Considering typical privacy parameters $\varepsilon \leqslant 1$, then $e^{2\varepsilon} - 1 \approx 2\varepsilon$. For exponentially small failure probability, i.e. $\eta = 2^{-\Omega(n)}$, the lower bound reads $\varepsilon \geqslant \Omega(\frac{1}{\gamma d}\cdot\frac{1}{\zeta})$, which is achieved by Algorithm 7.43 without the boosting step - see the discussion after Theorem 7.42. For polynomially small failure probability, i.e.$\eta = 1/\mathrm{poly}(n)$, the lower bound Eq. (7.4.10) reads $\varepsilon \geqslant \Omega(\frac{1}{\gamma d}\cdot\log\frac{1}{\zeta})$, which is achieved by Theorem 7.42.

By setting $\zeta = 1/n$ in Theorem 7.46, we directly obtain a tight lower bound for private exact recovery as a corollary.

[25] Error rate less than $1/n$ already means exact recovery. Thus it does not make sense to set ζ to any value strictly smaller than $1/n$. The upper bound $\zeta \leqslant 0.04$ is just a technical condition our proof needs for Eq. (7.4.12).

Corollary 7.48. *Suppose there exists an ε-differentially private algorithm such that for any balanced $x \in \{\pm 1\}^n$, on input $\mathbf{G} \sim \mathsf{SBM}_n(d, \gamma, x)$, outputs $\hat{x}(\mathbf{G}) \in \{\pm 1\}^n$ satisfying*

$$\mathbb{P}(\hat{x}(\mathbf{G}) \in \{x, -x\}) \geqslant 1 - \eta,$$

where the randomness is over both the algorithm and stochastic block models. Then,

$$e^{2\varepsilon} - 1 \geqslant \Omega\left(\frac{\log(n) + \log\frac{1}{\eta}}{\gamma d}\right). \tag{7.4.11}$$

Remark 7.49. The lower bound Eq. (7.4.11) for priavte exact recovery is tight up to constants, since there exists an (inefficient) ε-differentially priavte exact recovery algorithm with $\varepsilon \leqslant O(\frac{\log n}{\gamma d})$ and $\eta = 1/\mathrm{poly}(n)$ by Theorem 7.42 and [MNVT22, Theorem 3.7].

In rest of this section, we will prove Theorem 7.46. The proof applies the packing lower bound argument similar to [HKM22, Theorem 7.1]. To this end, we first show $\mathrm{err}(\cdot, \cdot)$ is a semimetric over $\{\pm 1\}^n$.

Lemma 7.50. $\mathrm{err}(\cdot, \cdot)$ *is a semimetric over* $\{\pm 1\}^n$.

Proof. Symmetry and non-negativity are obvious from the definition. We will show $\mathrm{err}(\cdot, \cdot)$ satisfies triangle inequality via case analysis. Let $u, v, w \in \{\pm 1\}^n$ be three arbitrary sign vectors. By symmetry, we only need to consider the following four cases.

Case 1: $\mathrm{Ham}(u, v), \mathrm{Ham}(u, w), \mathrm{Ham}(v, w) \leqslant n/2$. This case is reduced to showing Hamming distance satisfies triangle inequality, which is obvious.

Case 2: $\mathrm{Ham}(u, v), \mathrm{Ham}(u, w) \leqslant n/2$ and $\mathrm{Ham}(v, w) \geqslant n/2$. We need to check two subcases. First,

$$\begin{aligned}\mathrm{err}(u, v) \leqslant \mathrm{err}(u, w) + \mathrm{err}(v, w) &\Leftrightarrow \mathrm{Ham}(u, v) + \mathrm{Ham}(v, w) \leqslant \mathrm{Ham}(u, w) + n \\ &\Leftarrow \mathrm{Ham}(u, v) + H(u, v) + H(u, w) \leqslant \mathrm{Ham}(u, w) + n \\ &\Leftrightarrow \mathrm{Ham}(u, v) \leqslant n/2.\end{aligned}$$

Second,

$$\begin{aligned}\mathrm{err}(v, w) \leqslant \mathrm{err}(u, v) + \mathrm{err}(u, w) &\Leftrightarrow n \leqslant \mathrm{Ham}(v, w) + \mathrm{Ham}(u, v) + \mathrm{Ham}(u, w) \\ &\Leftarrow n \leqslant 2\,\mathrm{Ham}(v, w).\end{aligned}$$

Case 3: $\mathrm{Ham}(u, v) \leqslant n/2$ and $\mathrm{Ham}(u, w), \mathrm{Ham}(v, w) \geqslant n/2$. This case can be reduced to case 1 by considering $u, v, -w$.

Case 4: $\mathrm{Ham}(u, v), \mathrm{Ham}(u, w), \mathrm{Ham}(v, w) \geqslant n/2$. This case can be reduced to case 2 by considering $-u, v, w$. □

Proof of Theorem 7.46. Suppose there exists an ε-differentially private algorithm satisfying the theorem's assumption.

We first make the following notation. Given a semimetric ρ over $\{\pm 1\}^n$, a center $v \in \{\pm 1\}^n$, and a radius $r \geqslant 0$, define $B_\rho(v, r) := \{w \in \{\pm 1\}^n : \mathbf{1}^\top w = 0, \rho(w, v) \leqslant r\}$.

Pick an arbitrary balanced $x \in \{\pm 1\}^n$. Let $M = \{x^1, x^2, \ldots, x^m\}$ be a maximal 2ζ-packing of $B_{\mathrm{err}}(x, 4\zeta)$ in semimetric $\mathrm{err}(\cdot, \cdot)$. By maximality of M, we have $B_{\mathrm{err}}(x, 4\zeta) \subseteq \cup_{i=1}^m B_{\mathrm{err}}(x^i, 2\zeta)$, which implies

$$
\begin{aligned}
&|B_{\mathrm{err}}(x, 4\zeta)| \leqslant \sum_{i=1}^m \left|B_{\mathrm{err}}(x^i, 2\zeta)\right| \\
\Longrightarrow\ &|B_{\mathrm{Ham}}(x, 4\zeta)| \leqslant \sum_{i=1}^m 2 \cdot \left|B_{\mathrm{Ham}}(x^i, 2\zeta)\right| = 2m \cdot |B_{\mathrm{Ham}}(x, 2\zeta)| \\
\Longrightarrow\ &2m \geqslant \frac{|B_{\mathrm{Ham}}(x, 4\zeta n)|}{|B_{\mathrm{Ham}}(x, 2\zeta n)|} = \frac{\binom{n/2}{2\zeta n}^2}{\binom{n/2}{\zeta n}^2} \geqslant \frac{\left(\frac{1}{4\zeta}\right)^{4\zeta n}}{\left(\frac{e}{2\zeta}\right)^{2\zeta n}} = \left(\frac{1}{8e\zeta}\right)^{2\zeta n}
\end{aligned} \tag{7.4.12}
$$

For each $i \in [m]$, define $Y_i := \{w \in \{\pm 1\}^n : \mathrm{err}(w, x^i) \leqslant \zeta\}$. Then Y_i's are pairwise disjoint. For each $i \in [m]$, let P_i be the distribution over n-vertex graphs generated by $\mathsf{SBM}_n(d, \gamma, x^i)$. By our assumption on the algorithm, we have for any $i \in [m]$ that

$$
\mathop{\mathbb{P}}_{\mathbf{G} \sim P_i}(\hat{x}(\mathbf{G}) \in Y_i) \geqslant 1 - \eta.
$$

Combining the fact that Y_i's are pairwise disjoint, we have

$$
\sum_{i=1}^m \mathop{\mathbb{P}}_{\mathbf{G} \sim P_1}(\hat{x}(\mathbf{G}) \in Y_i) = \mathop{\mathbb{P}}_{\mathbf{G} \sim P_1}\left(\hat{x}(\mathbf{G}) \in \cup_{i=1}^m Y_i\right) \leqslant 1 \implies \sum_{i=2}^m \mathop{\mathbb{P}}_{\mathbf{G} \sim P_1}(\hat{x}(\mathbf{G}) \in Y_i) \leqslant \eta. \tag{7.4.13}
$$

In the following, we will lower bound $\mathbb{P}_{\mathbf{G} \sim P_1}(\hat{x}(\mathbf{G}) \in Y_i)$ for each $i \in [m] \setminus \{1\}$ using group privacy.

Note each P_i is a product of $\binom{n}{2}$ independent Bernoulli distributions. Thus for any $i, j \in [m]$, there exists a coupling ω_{ij} of P_i and P_j such that, if $(\mathbf{G}, \mathbf{H}) \sim \omega$, then

$$
\mathrm{Ham}(\mathbf{G}, \mathbf{H}) \sim \mathrm{Binomial}(N_{ij}, p),
$$

where $p = 2\gamma d/n$ and $N_{ij} = \mathrm{Ham}(x^i, x^j) \cdot (n - \mathrm{Ham}(x^i, x^j))$. Applying group privacy, we have for any two graphs G, H and for any $S \subseteq \{\pm 1\}^n$ that[26]

$$
\mathbb{P}(\hat{x}(G) \in S) \leqslant \exp(\varepsilon \cdot \mathrm{Ham}(G, H)) \cdot \mathbb{P}(\hat{x}(H) \in S). \tag{7.4.14}
$$

For each $i \in [m]$, taking expectations on both sides of Eq. (7.4.14) with respect to coupling ω_{i1} and setting $S = Y_i$, we have

$$
\mathop{\mathbb{E}}_{(\mathbf{G}, \mathbf{H}) \sim \omega_{i1}} \mathbb{P}(\hat{x}(\mathbf{G}) \in Y_i) \leqslant \mathop{\mathbb{E}}_{(\mathbf{G}, \mathbf{H}) \sim \omega_{i1}} \exp(\varepsilon \cdot \mathrm{Ham}(\mathbf{G}, \mathbf{H})) \cdot \mathbb{P}(\hat{x}(\mathbf{H}) \in Y_i). \tag{7.4.15}
$$

[26] In Eq. (7.4.14), the randomness only comes from the algorithm.

The left side of Eq. (7.4.15) is equal to

$$\mathop{\mathbb{E}}_{(\mathbf{G},\mathbf{H})\sim\omega_{i1}} \mathbb{P}(\hat{x}(\mathbf{G}) \in Y_i) = \mathop{\mathbb{P}}_{\mathbf{G}\sim P_i}(\hat{x}(\mathbf{G}) \in Y_i) \geqslant 1-\eta.$$

Upper bounding the right side of Eq. (7.4.15) by Cauchy-Schwartz inequality, we have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{(\mathbf{G},\mathbf{H})\sim\omega_{i1}} \exp(\varepsilon \cdot \mathrm{Ham}(\mathbf{G},\mathbf{H})) \cdot \mathbb{P}(\hat{x}(\mathbf{H}) \in Y_i) \\
\leqslant &\left(\mathop{\mathbb{E}}_{(\mathbf{G},\mathbf{H})\sim\omega_{i1}} \exp(2\varepsilon \cdot \mathrm{Ham}(\mathbf{G},\mathbf{H}))\right)^{1/2} \cdot \left(\mathop{\mathbb{E}}_{(\mathbf{G},\mathbf{H})\sim\omega_{i1}} \mathbb{P}(\hat{x}(\mathbf{H}) \in Y_i)^2\right)^{1/2} \\
= &\left(\mathop{\mathbb{E}}_{\mathbf{X}\sim\mathrm{Binomial}(N_{i1},p)} \exp(2\varepsilon \cdot \mathbf{X})\right)^{1/2} \cdot \left(\mathop{\mathbb{E}}_{\mathbf{H}\sim P_1} \mathbb{P}(\hat{x}(\mathbf{H}) \in Y_i)^2\right)^{1/2}.
\end{aligned}$$

Using the formula for the moment generating function of binomial distributions, we have

$$\mathop{\mathbb{E}}_{\mathbf{X}\sim\mathrm{Binomial}(N_{i1},p)} \exp(2\varepsilon \cdot \mathbf{X}) = (1-p+p\cdot e^{2\varepsilon})^{N_{i1}},$$

and it is easy to see

$$\mathop{\mathbb{E}}_{\mathbf{H}\sim P_1} \mathbb{P}(\hat{x}(\mathbf{H}) \in Y_i)^2 = \mathop{\mathbb{E}}_{\mathbf{H}\sim P_1} \left(\mathbb{E}\, \mathbb{1}_{[\hat{x}(\mathbf{H})\in Y_i]}\right)^2 \leqslant \mathop{\mathbb{P}}_{\mathbf{H}\sim P_1}(\hat{x}(\mathbf{H}) \in Y_i).$$

Putting things together, Eq. (7.4.15) implies for each $i \in [m]$ that

$$\mathop{\mathbb{P}}_{\mathbf{H}\sim P_1}(\hat{x}(\mathbf{H}) \in Y_i) \geqslant \frac{(1-\eta)^2}{(1-p+p\cdot e^{2\varepsilon})^{N_{i1}}}. \tag{7.4.16}$$

Since $x^i \in B_{\mathrm{err}}(x, 4\zeta)$ for $i \in [m]$, by assuming $\zeta \leqslant 1/16$, we have

$$N_{i1} = \mathrm{Ham}(x^i, x^1) \cdot (n - \mathrm{Ham}(x^1, x^i)) \leqslant 8\zeta n(n - 8\zeta n). \tag{7.4.17}$$

Recalling $p = 2\gamma d/n$ and combining Eq. (7.4.12), Eq. (7.4.13), Eq. (7.4.16) and Eq. (7.4.17), we have

$$(m-1) \cdot \frac{(1-\eta)^2}{(1-p+p\cdot e^{2\varepsilon})^{8\zeta n(n-8\zeta n)}} \leqslant \eta.$$

By taking logarithm on both sides, using $t \geqslant \log(1+t)$ for any $t > -1$, and assuming $\zeta \leqslant 1/(8e)$, we have

$$e^{2\varepsilon} - 1 \gtrsim \frac{\log\frac{1}{8e\zeta}}{\gamma d} + \frac{\log\frac{1}{\eta}}{\zeta n \gamma d}.$$

□

7.5 Private algorithms for learning mixtures of spherical Gaussians

In this section we present a private algorithm for recovering the centers of a mixtures of k Gaussians (cf. Model 7.5). Let $\mathcal{Y} \subseteq \left(\mathbb{R}^d\right)^{\otimes n}$ be the collection of sets of n points in $\mathbb{R}^d$. We consider the following notion of adjacency.

Definition 7.51 (Adjacent databases)**.** We say that $Y, Y' \in \mathcal{Y}$ are adjacent if $|Y \cap Y'| \geqslant n-1$.

Remark 7.52 (Problem parameters as public information)*.* We consider the parameters n, k, Δ to be *public information* given as input to the algorithm.

Next we present the main theorem of the section.

Theorem 7.53 (Privately learning spherical mixtures of Gaussians)**.** *Consider an instance of Model 7.5. Let $t \in \mathbb{N}$ be such that $\Delta \geqslant O\left(\sqrt{t}k^{1/t}\right)$. For $n \geqslant \Omega\left(k^{O(1)} \cdot d^{O(t)}\right)$, $k \geqslant (\log n)^{1/5}$, there exists an algorithm, running in time $(nd)^{O(t)}$, that outputs vectors $\hat{\mu}_1, \ldots, \hat{\mu}_\ell$ satisfying*

$$\max_{\ell \in [k]} \left\| \hat{\mu}_\ell - \mu_{\pi(\ell)} \right\|_2 \leqslant O(k^{-12}),$$

with high probability, for some permutation $\pi : [k] \to [k]$.[27] *Moreover, for $\varepsilon \geqslant k^{-10}, \delta \geqslant n^{-10}$, the algorithm is (ε, δ)-differentially private for any input Y.*

We remark that our algorithm not only works for mixtures of Gaussians but for all mixtures of $2t$-explicitly bounded distributions (cf. Definition 7.21).

Our algorithm is based on the sum-of-squares hierarchy and at the heart lies the following sum-of-squares program. The indeterminates $z_{11}, \ldots, z_{1k}, \ldots, z_{nk}$ and vector-valued indeterminates $\mu'_1, \ldots, \mu'_k$, will be central to the proof of Theorem 7.53. Let n, k, t be fixed parameters.

$$\left\{ \begin{aligned}
& z_{i\ell}^2 = z_{i\ell} && \forall i \in [n], \ell \in [k] && \text{(indicators)} \\
& \sum_{\ell \in [k]} z_{i\ell} \leqslant 1 && \forall i \in [n] && \text{(cluster mem.)} \\
& z_{i\ell} \cdot z_{i\ell'} = 0 && \forall i \in [n], \ell \in [k] && \text{(uniq. mem.)} \\
& \sum_i z_{i\ell} \leqslant n/k && \forall \ell \in [k] && \text{(size of clusters)} \\
& \mu'_\ell = \frac{k}{n} \sum_i z_{i\ell} \cdot y_i && \forall \ell \in [k] && \text{(means of clusters)} \\
& \forall v \in \mathbb{R}^d : \frac{k}{n} \sum_{i=1}^n z_{i\ell} \langle y_i - \mu'_\ell, v \rangle^{2s} + \left\| Q v^{\otimes s} \right\|^2 = (2s)^s \cdot \|v\|_2^{2s} && \forall s \leqslant t, \ell \in [k] && (t \text{ moment})
\end{aligned} \right\} \quad (\mathcal{P}_{n,k,t}(Y))$$

[27] We remark that we chose constants to optimize readibility and not the smallest possible ones.

We remark that the moment constraint encodes the $2t$-explicit 2-boundedness constraint introduced in Definition 7.21. Note that in the form stated above there are infinitely many constraints, one for each vector v. This is just for notational convenience. This constraint postulates equality of two polynomials in v. Formally, this can also be encoded by requiring there coefficients to agree and hence eliminating the variable v. It is not hard to see that this can be done adding only polynomially many constraints. Further, the matrix variable Q represents the SoS proof of the $2t$-explicit 2-boundedness constraint and we can hence deduce that for all $0 \leqslant s \leqslant t$

$$\mathcal{P} \vdash_{2s}^{v} \left\{ \frac{k}{n} \sum_{i=1}^{n} z_{i\ell} \langle y_i - \mu'_\ell, v \rangle^{2s} \leqslant (2s)^s \|s\|_2^{2s} \right\}.$$

Before presenting the algorithm we will introduce some additional notation which will be convenient. We assume t, n, k to be *fixed* throughout the section and drop the corresponding subscripts. For $Y \in \mathcal{Y}$, let $\mathcal{Z}(Y)$ be the set of degree-$10t$ pseudo-distributions satisfying $\mathcal{P}(Y)$. For each $\zeta \in \mathcal{Z}(Y)$ define $W(\zeta)$ as the n-by-n matrix satisfying

$$W(\zeta)_{ij} = \tilde{\mathbb{E}}_\zeta \left[\sum_{\ell \in [k]} z_{i\ell} \cdot z_{j\ell} \right].$$

We let $\mathcal{W}(Y) := \{W(\zeta) \mid \zeta \in \mathcal{Z}(Y)\}$.

Recall that J denotes the all-ones matrix. We define the function $g : \mathbb{R}^{n \times n} \to \mathbb{R}$ as

$$g(W) = \|W\|_{\mathrm{F}}^2 - (10)^{10} k^{300} \langle J, W \rangle$$

and let

$$W(\hat{\zeta}(Y)) := \operatorname{argmin}_{W \in \mathcal{W}(Y)} g(W).$$

We also consider the following function

Definition 7.54 (Soft thresholding function)**.** We denote by $\phi : [0,1] \to [0,1]$ the function

$$\phi(x) = \begin{cases} 0 & \text{if } x \leqslant 0.8, \\ 1 & \text{if } x \geqslant 0.9, \\ \frac{x-0.8}{0.9-0.8} & \text{otherwise}. \end{cases}$$

Notice that $\phi(\cdot)$ is $\frac{1}{0.9-0.8} = 10$ Lipschitz. Next we introduce our algorithm. Notice the algorithm relies on certain private subroutines. We describe them later in the section to improve the presentation.

Algorithm 7.55 (Private algorithm for learning mixtures of Gaussians)**.**
Input: Set of n points $Y \subseteq \mathbb{R}^d$, $\varepsilon, \delta > 0$, $k, t \in \mathbb{N}$, $d^* = 100 \log n$, $b = k^{-15}$.

1. Compute $W = W(\hat{\zeta}(Y))$.
2. Pick $\tau \sim \mathrm{tLap}\left(-n^{1.6}\left(1 + \frac{\log(1/\delta)}{\varepsilon}\right), \frac{n^{1.6}}{\varepsilon}\right)$.
3. If $|\tau| \geqslant n^{1.7}$ or $\left\|\phi(W)\right\|_1 \leqslant \frac{n^2}{k} \cdot \left(1 - \frac{1}{n^{0.1}} - \frac{1}{k^{100}}\right) + \tau$ reject.
4. For all $i \in [n]$, compute the n-dimensional vector
$$v^{(i)} = \begin{cases} \mathbf{0} & \text{if } \left\|\phi(W_i)\right\|_1 = 0 \\ \left\|\phi(W_i)\right\|_1^{-1} \sum_j \phi(W_{ij}) \cdot y_j & \text{otherwise.} \end{cases}$$
5. Pick a set $\mathcal{S}$ of $n^{0.01}$ indices $i \in [n]$ uniformly at random.
6. For each $i \in \mathcal{S}$ let $\bar{\nu}^{(i)} = v^{(i)} + \mathbf{w}$ where $\mathbf{w} \sim N\left(0, n^{-0.18} \cdot \frac{\log(2/\delta)}{\varepsilon^2} \cdot \mathrm{Id}\right)$.
7. Pick $\Phi \sim N\left(0, \frac{1}{d^*}\right)^{d^* \times d}$, $\mathbf{q} \overset{u.a.r.}{\sim} [0, b]$ and run the histogram learner of Lemma 7.20 with input $\Phi\bar{\nu}^{(1)}, \ldots, \Phi\bar{\nu}^{(n^{0.01})}$ and parameters
$$\mathbf{q}, b, \alpha = k^{-10}, \beta = n^{-10}, \delta^* = \frac{\delta}{n}, \varepsilon^* = \varepsilon \cdot \frac{10k^{50}}{n^{0.01}}.$$
Let $\mathbf{B}_1, \ldots, \mathbf{B}_k$ be the resulting d^*-dimensional bins with highest counts. Break ties randomly.
8. Reject if $\min_{i \in [k]} \left|\left\{j \mid \Phi\bar{\nu}^{(j)} \in \mathbf{B}_i\right\}\right| < \frac{n^{0.01}}{2k}$.
9. For each $l \in [k]$ output
$$\hat{\mu}_l := \frac{1}{\left|\left\{j \mid \Phi\bar{\nu}^{(j)} \in \mathbf{B}_i\right\}\right|} \cdot \left(\sum_{\Phi\bar{\nu}^{(j)} \in \mathbf{B}_l} \bar{\nu}^{(j)}\right) + \mathbf{w}',$$
where $\mathbf{w}' \sim N\left(0, N\left(0, 32 \cdot k^{-120} \cdot \frac{\log(2kn/\delta)}{\varepsilon^2} \cdot \mathrm{Id}\right)\right)$.

For convenience, we introduce some preliminary facts.

Definition 7.56 (Good Y)**.** Let $\mathbf{Y}$ be sampled according to Model 7.5. We say that $\mathbf{Y}$ is *good* if:

1. for each $\ell \in [k]$, there are at least $\frac{n}{k} - n^{0.6}$ and most $\frac{n}{k} + n^{0.6}$ points sampled from D_ℓ in $\mathbf{Y}$. Let $\mathbf{Y}_\ell \subseteq \mathbf{Y}$ be such set of points.

2. Each $\mathbf{Y}_\ell$ is $2t$-explicitly 2-bounded.

It turns out that typical instances $\mathbf{Y}$ are indeed good.

Lemma 7.57 ([HL18, KSS18])**.** *Consider the settings of Theorem 7.53. Then $\mathbf{Y}$ is good with high probability. Further, in this case the sets $\mathcal{Z}(Y)$ and $\mathcal{W}(Y)$ are non-empty.*

7.5.1 Privacy analysis

In this section we show that our clustering algorithm is private.

Lemma 7.58 (Differential privacy of the algorithm)**.** *Consider the settings of Theorem 7.53. Then Algorithm 7.55 is (ε, δ)-differentially private.*

We split our analysis in multiple steps and combine them at the end. On a high level, we will argue that on adjacent inputs Y, Y' many of the vectors $v^{(i)}$ by the algorithm are close to each other and a small part can be very far. We can then show that we can mask this small difference using the Gaussian mechanism and afterwards treat this subset of the vectors as privatized (cf. Lemma E.4). Then we can combine this with known histogram learners to deal with the small set of $v^{(i)}$'s that is far from each other on adjacent inputs.

7.5.1.1 Sensitivity of the matrix W

Here we use Lemma 7.25 to reason about the sensitivity of $\phi(W(\hat{\zeta}(Y)))$. For adjacent datasets $Y, Y' \in \mathcal{Y}$ we let $\hat{\zeta}, \hat{\zeta}'$ be the pseudo-distribution corresponding to $W(\hat{\zeta}(Y))$ and $W(\hat{\zeta}(Y'))$ computed in step 1 of the algorithm, respectively. We prove the following result.

Lemma 7.59 (ℓ_1-sensitivity of $\phi(W)$)**.** *Consider the settings of Theorem 7.53. Let W, W' be respectively be the matrices computed in step 1 by Algorithm 7.55 on adjacent inputs $Y, Y' \in \mathcal{Y}$. Then*

$$\left\|\phi(W) - \phi(W')\right\|_1 \leqslant n^{1.6}\,.$$

For all but $n^{0.8}$ rows i of $\phi(W), \phi(W')$, it holds

$$\left\|\phi(W)_i - \phi(W')_i\right\|_1 \leqslant n^{0.8}\,.$$

Proof. The second inequality is an immediate consequence of the first via Markov's inequality. Thus it suffices to prove the first. Since $\phi(\cdot)$ is 10-Lipschitz, we immediately obtain the result if

$$\left\|W(\hat{\zeta}(Y)) - W(\hat{\zeta}(Y'))\right\|_1 \leqslant n^{1.55}\,.$$

Thus we focus on this inequality. To prove it, we verify the two conditions of Lemma 7.25. First notice that g is 2-strongly convex with respect to its input W. Indeed for $W, W' \in \mathcal{W}(Y)$, since $\forall i, j \in [n], W_{ij} \geqslant 0$ it holds that

$$\begin{aligned}\|W'\|_F^2 &= \|W\|_F^2 + \|W - W'\|_F^2 + 2\langle W' - W, W\rangle \\ &= \|W\|_F^2 + \|W - W'\|_F^2 + 2\langle W' - W, W\rangle + \langle W' - W, (10)^{10}k^{300}(J - J)\rangle \\ &= g(W) + \|W - W'\|_F^2 + \langle W' - W, \nabla g(W)\rangle + \langle W', (10)^{10}k^{300}J\rangle ,\end{aligned}$$

where we used that $\nabla g(W) = 2W - (10)^{10}k^{300}J$. Thus it remain to prove *(i)* of Lemma 7.25.

Let $\hat{\zeta} \in \mathcal{Z}(Y), \hat{\zeta}' \in \mathcal{Z}(Y')$ be the pseudo-distributions such that $W_Y(\hat{\zeta}) = W$ and $W_Y(\hat{\zeta}') = W'$. We claim that there always exists $\zeta_{\text{adj}} \in \mathcal{Z}(Y) \cap \mathcal{Z}(Y')$ such that

1. $|g(W(\zeta)) - g(W(\zeta_{\text{adj}})| \leqslant \frac{2n}{k} \cdot \left((10)^{10}k^{300} + 1\right) \leqslant 3 \cdot (10)^{10}k^{300}n ,$
2. $|g_{Y'}(W(\zeta_{\text{adj}})) - g(W(\zeta_{\text{adj}})| = 0 .$

Note that in this case the second point is always true since g doesn't depend on Y. Together with Lemma 7.25 these two inequalities will imply that

$$\left\|W(\hat{\zeta}(Y)) - W(\hat{\zeta}(Y'))\right\|_F^2 \leqslant 18 \cdot (10)^{10}k^{300}n .$$

By assumption on n, an application of Cauchy-Schwarz will give us the desired result.

So, let i be the index at which Y, Y' differ. We construct ζ_{adj} as follows: for all polynomials p of degree at most $10t$ we let

$$\tilde{\mathbb{E}}_{\zeta_{\text{adj}}}[p] = \begin{cases} \tilde{\mathbb{E}}_{\zeta}[p] & \text{if } p \text{ does not contain variables } z_{i\ell} \text{ for any } \ell \in [k] \\ 0 & \text{otherwise.} \end{cases}$$

By construction $\zeta_{\text{adj}} \in \mathcal{Z}(Y) \cap \mathcal{Z}(Y')$. Moreover, $W(\zeta), W(\zeta_{\text{adj}})$ differ in at most $2n/k$ entries. Since all entries of the two matrices are in $[0, 1]$, the first inequality follows by definition of the objective function. □

7.5.1.2 Sensitivity of the resulting vectors

In this section we argue that if the algorithm does not reject in step 3 then the vectors $v^{(i)}$ are stable on adjacent inputs. Concretely our statement goes as follows:

Lemma 7.60 (Stability of the $v^{(i)}$'s)**.** *Consider the settings of Theorem 7.53. Suppose Algorithm 7.55 does not reject in step 3, on adjacent inputs $Y, Y' \in \mathcal{Y}$. Then for all but $\frac{6n}{k^{50}}$ indices $i \in [n]$, it holds:*

$$\left\|v_Y^{(i)} - v_{Y'}^{(i)}\right\|_2 \leqslant O\left(n^{-0.1}\right) .$$

The proof of Lemma 7.60 crucially relies on the next statement.

Lemma 7.61 (Covariance bound)**.** *Consider the settings of Theorem 7.53. Let W be the matrix computed by Algorithm 7.55 on input $Y \in \mathcal{Y}$. For $i \in [n]$, if $\|\phi(W_i)\|_1 \geqslant \frac{n}{k} \cdot \left(1 - \frac{10}{k^{50}}\right)$ then $\nu^{(i)}$ induces a 2-explicitly 40-bounded distribution over Y.*

Proof. First, by assumption notice that there must be at least $\frac{n}{k} \cdot \left(1 - \frac{10}{k^{50}}\right)$ entries of $\phi(W_i)$ larger than 0.8. We denote the set of $j \in [n]$ such that $W_{ij} \geqslant 0.8$ by $\mathcal{G}$. Let $\zeta \in \mathcal{Z}(Y)$ be the degree $10t$ pseudo-distribution so that $W = W(\zeta(Y))$. Since ζ satisfies $\mathcal{P}(Y)$, for $\ell \in [k]$ it follows from the moment bound constraint for $s = 1$ that for all unit vectors u it holds that

$$\mathcal{P} \left|\frac{}{4}\right. \left\{ 0 \leqslant \frac{k}{n} \sum_{j=1}^{n} z_{j\ell} \langle \mathbf{y}_j - \mu'_l, u \rangle^2 \leqslant 2 \right\},$$

Using the SoS triangle inequality (cf. Fact E.14) $\left|\frac{a,b}{2}\right. (a+b)^2 \leqslant 2(a^2+b^2)$ it now follows that

$$\mathbf{0} \preceq \tilde{\mathbb{E}}_\zeta \left[\frac{k^2}{n^2} \sum_{j,j' \in [n]} z_{j\ell} z_{j'\ell} \cdot \left(y_j - y_{j'}\right)^{\otimes 2} \right] \preceq 8\mathrm{Id}$$

and thus

$$\mathbf{0} \preceq \tilde{\mathbb{E}}_\zeta \left[\frac{k^2}{n^2} \sum_{\ell \in [k]} \sum_{j,j' \in [n]} z_{i\ell} z_{j\ell} z_{j'\ell} \cdot \left(y_j - y_{j'}\right)^{\otimes 2} \right] \preceq 8\mathrm{Id}\,.$$

Furthermore using $\mathcal{P}(Y) \left|\frac{}{2}\right. \{z_{i\ell} z_{i\ell'} = 0\}$ for $\ell \neq \ell'$ we have

$$\tilde{\mathbb{E}}_\zeta \left[\sum_{\ell \in [k]} \sum_{j,j' \in [n]} z_{i\ell} z_{j\ell} z_{j'\ell} \right] = \tilde{\mathbb{E}}_\zeta \left[\left(\sum_{\ell \in [k], j \in [n]} z_{i\ell} z_{j\ell} \right) \cdot \left(\sum_{\ell' \in [k], j' \in [n]} z_{i\ell'} z_{j'\ell'} \right) \right].$$

Now, for fixed $j, j' \in [n]$, using

$$\left\{ a^2 = a\,, b^2 = b \right\} \left|\frac{}{O(1)}\right. \left\{ 1 + ab - a - b = 1 - ab - (a-b)^2 \geqslant 0 \right\}$$

with $a = \sum_{\ell \in [k]} z_{i\ell} z_{j\ell}$ and $b = \sum_{\ell' \in [k]} z_{i\ell'} z_{j'\ell'}$ we get

$$\begin{aligned} \tilde{\mathbb{E}}_\zeta \left[\left(\sum_{\ell \in [k]} z_{i\ell} z_{j\ell} \right) \left(\sum_{\ell' \in [k]} z_{i\ell'} z_{j'\ell'} \right) \right] &\geqslant \tilde{\mathbb{E}}_\zeta \left[\sum_{\ell \in [k]} z_{i\ell} z_{j\ell} + \sum_{\ell' \in [k]} z_{i\ell'} z_{j'\ell'} \right] - 1 \\ &= W_{ij} + W_{ij'} - 1\,. \end{aligned}$$

Now if $j, j' \in \mathcal{G}$ we must have

$$\sum_{\ell \in [k]} \tilde{\mathbb{E}}_\zeta \left[z_{i\ell} z_{j\ell} z_{j'\ell} \right] = \tilde{\mathbb{E}}_\zeta \left[\left(\sum_{\ell \in [k]} z_{i\ell} z_{j\ell} \right) \left(\sum_{\ell' \in [k]} z_{i\ell'} z_{j'\ell'} \right) \right] \geqslant 0.6\,.$$

Since $\phi(W_{ij}) \leqslant 1$ by definition and $\left\|\phi(W_i)\right\|_1 \geqslant \frac{n}{k} \cdot \left(1 - \frac{10}{k^{50}}\right)$, we conclude

$$\begin{aligned}
&\left\|\phi(W_i)\right\|_1^{-2} \left[\sum_{j,j' \in [n]} \phi(W_{ij})\phi(W_{ij'})\left(y_j - y_{j'}\right)^{\otimes 2}\right] \\
&\preceq 5 \cdot \frac{k^2}{n^2} \sum_{j,j' \in [n], \ell \in [k]} \tilde{\mathbb{E}}_\zeta\left[z_{i\ell} z_{j\ell} z_{j'\ell}\right] \cdot \left(y_j - y_{j'}\right)^{\otimes 2} \\
&\preceq 40\mathrm{Id}\,.
\end{aligned}$$

as desired. □

We can now prove Lemma 7.60.

Proof of Lemma 7.60. Let W, W' be the matrices computed by Algorithm 7.55 in step 1 on input Y, Y', respectively. Let $\mathcal{G} \subseteq [n]$ be the set of indices i such that

$$\left\|\phi(W)_i - \phi(W')_i\right\|_1 \leqslant n^{0.8}\,.$$

Notice that $|\mathcal{G}| \geqslant n - n^{0.8}$ by Lemma 7.59. Since on input Y the algorithm did not reject in step 3 we must have

$$\left\|\phi(W)\right\|_1 \geqslant \frac{n^2}{k} \cdot \left(1 - \frac{1}{n^{0.1}} - \frac{1}{k^{100}}\right) - n^{1.7} \geqslant \frac{n^2}{k} \cdot \left(1 - \frac{2}{k^{100}}\right).$$

Let g_W be the number of indices $i \in \mathcal{G}$ such that $\left\|\phi(W)_i\right\|_1 \geqslant \frac{n}{k} \cdot \left(1 - \frac{1}{k^{50}}\right)$. It holds that

$$\begin{aligned}
\frac{n^2}{k} \cdot \left(1 - \frac{2}{k^{100}}\right) &\leqslant g_W \cdot \frac{n}{k} + (n - |\mathcal{G}|) \cdot \frac{n}{k} + \left(|G| - g_w\right)\frac{n}{k} \cdot \left(1 - \frac{1}{k^{50}}\right) \\
&\leqslant g_W \cdot \frac{n}{k} \cdot \frac{1}{k^{50}} + \frac{n^{1.8}}{k} + \frac{n^2}{k} \cdot \left(1 - \frac{1}{k^{50}}\right) \\
&\leqslant g_W \cdot \frac{n}{k} \cdot \frac{1}{k^{50}} + \frac{n^2}{k} \cdot \left(1 + \frac{1}{k^{100}} - \frac{1}{k^{50}}\right).
\end{aligned}$$

Rearring now yields

$$g_W \geqslant n \cdot \left(1 - \frac{3}{k^{50}}\right).$$

Similarly, let $g_{W'}$ be the number of indices $i \in \mathcal{G}$ such that $\left\|\phi(W')_i\right\|_1 \geqslant \frac{n}{k} \cdot \left(1 - \frac{1}{k^{50}}\right)$. By an analogous argument it follows that $g_{W'} \geqslant n \cdot \left(1 - \frac{3}{k^{50}}\right)$. Thus, by the pigeonhole principle there are at least $g_W \geqslant n \cdot \left(1 - \frac{6}{k^{50}}\right)$ indices i such that

1. $\left\|\phi(W)_i\right\|_1 \geqslant \frac{n}{k}\left(1 - \frac{1}{k^{50}}\right)$,

2. $\left\|\phi(W')_i\right\|_1 \geqslant \frac{n}{k}\left(1-\frac{1}{k^{50}}\right)$,

3. $\left\|\phi(W)_i - \phi(W')_i\right\|_1 \leqslant n^{0.8}$.

Combining these with Lemma 7.61 we may also add

4. the distribution induced by $\left\|\phi(W_i)\right\|_1^{-1}\phi(W_i)$ is 2-explicitly 40-bounded,

5. the distribution induced by $\left\|\phi(W'_i)\right\|_1^{-1}\phi(W'_i)$ is 2-explicitly 40-bounded.

Using that for non-zero vectors x, y it holds that $\left\|\frac{x}{\|x\|} - \frac{y}{\|y\|}\right\| \leqslant \frac{2}{\|x\|}\|x-y\|$ points 1 to 3 above imply that

$$\left\|\left\|\phi(W_i)\right\|_1^{-1}\phi(W_i) - \left\|\phi(W'_i)\right\|_1^{-1}\phi(W'_i)\right\|_1 \leqslant \frac{2n^{0.8}}{\frac{n}{k}\cdot\left(1-\frac{1}{k^{50}}\right)} = O\left(n^{-0.2}\right).$$

Hence, applying Theorem 7.23 with $t = 1$ it follows that

$$\left\|\nu_Y^{(i)} - \nu_{Y'}^{(i)}\right\|_2 \leqslant O\left(n^{-0.1}\right).$$

□

7.5.1.3 From low sensitivity to privacy

In this section we argue privacy of the whole algorithm, proving Lemma 7.58. Before doing that we observe that low-sensitivity is preserved with high probability under subsampling.

Fact 7.62 (Stability of $\mathcal{S}$)**.** *Consider the settings of Theorem 7.53. Suppose Algorithm 7.55 does not reject in step 3, on adjacent inputs $Y, Y' \in \mathcal{Y}$. With probability at least $1 - e^{-n^{\Omega(1)}}$ over the random choices of $\mathcal{S}$, for all but $\frac{10n^{0.01}}{k^{50}}$ indices $i \in \mathcal{S}$, it holds:*

$$\left\|\nu_Y^{(i)} - \nu_{Y'}^{(i)}\right\|_2 \leqslant O\left(n^{-0.1}\right).$$

Proof. There are at most $\frac{6n}{k^{50}}$ such indices in $[n]$ by Lemma 7.60. By Chernoff's bound, cf. Fact E.8, the claim follows. □

Finally, we prove our main privacy lemma.

Proof of Lemma 7.58. For simplicity, we will prove that the algorithm is $(5\varepsilon, 5\delta)$-private. Let $Y, Y' \in \mathcal{Y}$ be adjacent inputs. By Lemma 7.17 and Lemma 7.59 the test in step 3 of Algorithm 7.55 is (ε, δ)-private.

Thus suppose now the algorithm did not reject in step 3 on inputs Y, Y'. By composition (cf. Lemma 7.11) it is enough to show that the rest of the algorithm is (ε, δ)-private with respect to Y, Y' under this condition. Next, let $\nu_Y^{(1)}, \ldots, \nu_Y^{(n)}$ and $\nu_{Y'}^{(1)}, \ldots, \nu_{Y'}^{(n)}$ be the vectors

computed in step 4 of the algorithm and $\mathcal{S}$ be the random set of indices computed in step 5.[28] By Lemma 7.60 and Fact 7.62 with probability $1 - e^{-n^{\Omega(1)}}$ over the random choices of $\mathcal{S}$ we get that for all but $\frac{10n^{0.01}}{k^{50}}$ indices $i \in \mathcal{S}$, it holds that

$$\left\| \nu_Y^{(i)} - \nu_{Y'}^{(i)} \right\|_2 \leqslant O\left(n^{-0.1}\right).$$

Denote this set of indices by $\mathcal{G}$. Note, that we may incorporate the failure probability $e^{-n^{\Omega(1)}} \leqslant \min\{\varepsilon/2, \delta/2\}$ into the final privacy parameters using Fact E.15.

Denote by $\mathbf{V}, \mathbf{V}'$ the $|\mathcal{S}|$-by-d matrices respectively with rows $\nu_Y^{(i_1)}, \ldots, \nu_Y^{(i_{|\mathcal{S}|})}$ and $\nu_{Y'}^{(i_1)}, \ldots, \nu_{Y'}^{(i_{|\mathcal{S}|})}$, where $i_1, \ldots, i_{|\mathcal{S}|}$ are the indices in $\mathcal{S}$. Recall, that $|\mathcal{G}|$ rows of $\mathbf{V}$ and $\mathbf{V}'$ differ by at most $O\left(n^{-0.1}\right)$ in ℓ_2-norm. Thus, by the Gaussian mechanism used in step 6 (cf. Lemma 7.19) and Lemma E.4 it is enough to show that step 7 to step 9 of the algorithm are private with respect to pairs of inputs V and V' differing in at most 1 row.[29] In particular, suppose these steps are $(\varepsilon_1, \delta_1)$-private. Then, for $m = n^{0.01} - |\mathcal{G}| \leqslant \frac{10n^{0.01}}{k^{50}}$, by Lemma E.4 it follows that step 6 to step 9 are (ε', δ')-differentially private with

$$\begin{aligned} \varepsilon' &:= \varepsilon + m\varepsilon_1, \\ \delta' &:= e^{\varepsilon} m e^{(m-1)\varepsilon_1} \delta_1 + \delta. \end{aligned}$$

Consider steps 7 and 8. Recall, that in step 7 we invoke the histogram learner with parameters

$$b = k^{-15}, \mathbf{q} \overset{u.a.r.}{\sim} [0, b], \alpha = k^{-10}, \beta = n^{-10}, \delta^* = \frac{\delta}{n}, \varepsilon^* = \varepsilon \cdot \frac{10k^{50}}{n^{0.01}}.$$

Hence, by Lemma 7.20 this step is $(\varepsilon^*, \delta^*)$-private since

$$\frac{8}{\varepsilon^*\alpha} \cdot \log\left(\frac{2}{\delta^*\beta}\right) \leqslant \frac{200 \cdot k^{10} \cdot n^{0.01}}{10 \cdot k^{50} \cdot \varepsilon} \cdot \log n = \frac{20 \cdot n^{0.01}}{k^{40} \cdot \varepsilon} \cdot \log n \leqslant n,$$

for $\varepsilon \geqslant k^{-10}$. Step 8 is private by post-processing.

Next, we argue that step 9 is private by showing that the average over the bins has small ℓ_2-sensitivity. By Lemma 7.11 we can consider the bins $\mathbf{B}_1, \ldots, \mathbf{B}_k$ computed in the previous step as fixed. Further, we can assume that the algorithm did not reject in step 8, i.e., that each bin contains at least $\frac{n^{0.01}}{2k}$ points of V and V' respectively. As a consequence, every bin contains at least two (projections of) points of the input V or V' respectively. In particular, it contains at least one (projection of a) point which is present in both V and V'. Fix a bin $\mathbf{B}_l$ and let $\bar{v}^*$ be such that it is both in V and V' and $\mathbf{\Phi}\bar{v}^* \in \mathbf{B}_l$. Also, define

$$S_l := \left| \left\{ j \;\middle|\; \mathbf{\Phi}\bar{v}_Y^{(j)} \in \mathbf{B}_i \right\} \right|,$$

[28] Note that since this does not depend on Y or Y', respectively, we can assume this to be the same in both cases. Formally, this can be shown, e.g., via a direct calculation or using Lemma 7.11.

[29] Note that for the remainder of the analysis, these do *not* correspond to $\mathbf{V}$ and $\mathbf{V}'$, since those differ in m rows. Lemma E.4 handles this difference.

$$S'_l := \left|\left\{ j \;\middle|\; \mathbf{\Phi}\bar{v}^{(j)}_{Y'} \in \mathbf{B}_i \right\}\right|.$$

Assume V and V' differ on index j. We consider two cases. First, assume that $\mathbf{\Phi}\bar{v}^{(j)}_{Y}$ and $\mathbf{\Phi}\bar{v}^{(j)}_{Y'}$ both lie in $\mathbf{B}_l$. In this case, $S_l = S'_l$ and using Lemma E.9 it follows that with probability $n^{-100} \leqslant \min\{\varepsilon/2, \delta/2\}$ it holds that

$$\begin{aligned}\left\|\bar{v}^{(j)}_{Y} - \bar{v}^{(j)}_{Y'}\right\|_2 &\leqslant \left\|\bar{v}^{(j)}_{Y} - \bar{v}^{*}\right\|_2 + \left\|\bar{v}^{*} - \bar{v}^{(j)}_{Y'}\right\| \leqslant 10 \cdot \left(\left\|\mathbf{\Phi}\bar{v}^{(j)}_{Y} - \mathbf{\Phi}\bar{v}^{*}\right\|_2 + \left\|\mathbf{\Phi}\bar{v}^{(j)}_{Y'} - \mathbf{\Phi}\bar{v}^{*}\right\|_2\right) \\ &\leqslant 20 \cdot \sqrt{d^*} \cdot b \leqslant 200 \cdot k^{-12}.\end{aligned}$$

And hence we can bound

$$\left\|\frac{1}{S_l} \cdot \left(\sum_{\mathbf{\Phi}\bar{v}^{(j)}_{Y} \in \mathbf{B}_l} \bar{v}^{(j)}_{Y}\right) - \frac{1}{S'_l} \cdot \left(\sum_{\mathbf{\Phi}\bar{v}^{(j)}_{Y'} \in \mathbf{B}_l} \bar{v}^{(j)}_{Y'}\right)\right\|_2 \leqslant \frac{\left\|\bar{v}^{(j)}_{Y} - \bar{v}^{(j)}_{Y'}\right\|_2}{S_l} \leqslant \frac{400 \cdot k^{-11}}{n^{0.01}}.$$

Next, assume that $\mathbf{\Phi}\bar{v}^{(j)}_{Y} \notin \mathbf{B}_l$ and $\mathbf{\Phi}\bar{v}^{(j)}_{Y'} \in \mathbf{B}_l$ (the other case works symetrically). It follows that $S_l = S'_l - 1$ and we can bound

$$\begin{aligned}\left\|\frac{1}{S_l} \cdot \left(\sum_{\mathbf{\Phi}\bar{v}^{(j)}_{Y} \in \mathbf{B}_l} \bar{v}^{(j)}_{Y}\right) - \frac{1}{S'_l} \cdot \left(\sum_{\mathbf{\Phi}\bar{v}^{(j)}_{Y'} \in \mathbf{B}_l} \bar{v}^{(j)}_{Y'}\right)\right\|_2 &= \frac{1}{S_l \cdot S'_l} \cdot \left\|S'_l\left(\sum_{\mathbf{\Phi}\bar{v}^{(j)}_{Y} \in \mathbf{B}_l} \bar{v}^{(j)}_{Y}\right) - (S'_l - 1)\left(\sum_{\mathbf{\Phi}\bar{v}^{(j)}_{Y'} \in \mathbf{B}_l} \bar{v}^{(j)}_{Y'}\right)\right\|_2 \\ &= \frac{1}{S_l \cdot S'_l} \cdot \left\|S'_l \cdot \bar{v}^{(j)}_{Y'} + \left(\sum_{\mathbf{\Phi}\bar{v}^{(j)}_{Y'} \in \mathbf{B}_l} \bar{v}^{(j)}_{Y'}\right)\right\|_2 \\ &= \frac{1}{S_l} \cdot \left\|\bar{v}^{(j)}_{Y'} - \frac{1}{S'_l}\left(\sum_{\mathbf{\Phi}\bar{v}^{(j)}_{Y'} \in \mathbf{B}_l} \bar{v}^{(j)}_{Y'}\right)\right\|_2 \\ &\leqslant \frac{\sqrt{d^*} \cdot b}{S_l} \leqslant \frac{20 \cdot k^{-11}}{n^{0.01}}.\end{aligned}$$

Hence, the ℓ_2-sensitivity is at most $\Delta := \frac{400 \cdot k^{-11}}{n^{0.01}}$. Since

$$2\Delta^2 \cdot \frac{\log(2/(\delta^*/k))}{(\varepsilon^*/k)^2} = 32 \cdot k^{-120} \cdot \frac{\log(2kn/\delta)}{\varepsilon^2}$$

and $\mathbf{w}' \sim N\left(0, 32 \cdot k^{-120} \cdot \frac{\log(2kn/\delta)}{\varepsilon^2} \cdot \mathrm{Id}\right)$ it follows that outputing $\hat{\boldsymbol{\mu}}_l$ is $(\varepsilon^*/k, \delta^*/k)$-DP by the Gaussian Mechanism that. By Lemma 7.11 it follows step 9 is $(\varepsilon^*, \delta^*)$-private.

Hence, by Lemma 7.11 it follows that step 7 to step 9 are $(2\varepsilon^*, 2\delta^*)$-differentially private. Using $m \leqslant \frac{10n^{0.01}}{k^{10}}$ it now follows by Lemma E.4 that step 6 to step 9 are (ε', δ')-private for

$$\varepsilon' = \varepsilon + 2m\varepsilon^* \leqslant 3\varepsilon,$$

$$\delta' = 2e^{\varepsilon} m e^{(m-1)2\varepsilon^*}\delta^* + \delta \leqslant 2me^{3\varepsilon}\cdot\frac{\delta}{n} + \delta \leqslant 3\delta\,.$$

Thus, combined with the private check and Fact E.15 in step 3 the whole algorithm is $(5\varepsilon, 5\delta)$-private.

□

7.5.2 Utility analysis

In this section we reason about the utility of Algorithm 7.55 and prove Theorem 7.53. We first introduce some notation.

Definition 7.63 (True solution)**.** Let $\mathbf{Y}$ be an input sampled from Model 7.5. Denote by $W^*(\mathbf{Y}) \in \mathcal{W}(\mathbf{Y})$ the matrix induced by the *true solution* (or *ground truth*). I.e., let

$$W^*(\mathbf{Y})_{ij} = \begin{cases} 1 & \text{if } i\,,j \text{ were both sampled from the same component of the mixture,} \\ 0 & \text{otherwise.} \end{cases}$$

Whenever the context is clear, we simply write $\mathbf{W}^*$ to ease the notation.

First, we show that in the utility case step 3 of Algorithm 7.55 rejects only with low probability.

Lemma 7.64 (Algorithm does not reject on good inputs)**.** *Consider the settings of Theorem 7.53. Suppose* $\mathbf{Y}$ *is a good set as per Definition 7.56. Then* $\left\|W(\hat{\zeta}(\mathbf{Y}))\right\|_1 \geqslant \frac{n^2}{k}\cdot\left(1 - n^{-0.4} - \frac{1}{(10)^{10}k^{300}}\right)$ *and Algorithm 7.55 rejects with probability at most* $\exp\left(-\Omega\left(n^{1.7}\right)\right)$.

Proof. Since $\mathbf{Y}$ is good, there exists $\mathbf{W}^* \in \mathcal{W}(\mathbf{Y})$, corresponding to the indicator matrix of the true solution, such that

$$\begin{aligned} g(\mathbf{W}^*) &= \|\mathbf{W}^*\|_{\mathrm{F}}^2 - 10^{10}k^{300}\langle J, \mathbf{W}^*\rangle \leqslant \frac{n^2}{k} + n^{1.6} - (10)^{10}k^{300}\left(\frac{n^2}{k} - n^{1.6}\right) \\ &= \frac{n^2}{k}\left(1 + \frac{k}{n^{0.4}} - (10)^{10}k^{300}\left(1 - \frac{k}{n^{0.4}}\right)\right). \end{aligned}$$

Since $g(W(\hat{\zeta}(\mathbf{Y}))) \leqslant g(\mathbf{W}^*)$ it follows that

$$(10)^{10}k^{300}\langle J, W(\hat{\zeta}(\mathbf{Y}))\rangle \geqslant |g(W(\hat{\zeta}(\mathbf{Y})))| \geqslant \frac{n^2}{k}\left((10)^{10}k^{300}\left(1 - \frac{k}{n^{0.4}}\right) - 1 - \frac{k}{n^{0.4}}\right).$$

Since, $\left\|W(\hat{\zeta}(\mathbf{Y}))\right\|_1 \geqslant \langle J, W(\hat{\zeta}(\mathbf{Y}))\rangle$ the first claim follows rearranging the terms. This means that the algorithm rejects only if $|\tau| \geqslant n^{1.7}$. Recall that $\tau \sim \mathrm{tLap}\left(-n^{1.6}\left(1 + \frac{\log(1/\delta)}{\varepsilon}\right), \frac{n^{1.6}}{\varepsilon}\right)$. Hence,by Lemma 7.18 it follows that

$$\mathbb{P}\left(|\tau| \geqslant n^{1.7}\right) \leqslant \frac{\exp\left(-n^{1.7} + \varepsilon + \log(1/\delta)\right)}{2 - \exp\left(-\varepsilon - \log(1/\delta)\right)} = \exp\left(-\Omega\left(n^{1.7}\right)\right).$$

□

The next step shows that on a good input $\mathbf{Y}$ the matrix $\phi(W(\hat{\zeta}(\mathbf{Y})))$ is close to the true solution.

Lemma 7.65 (Closeness to true solution on good inputs)**.** *Consider the settings of Theorem 7.53. Suppose $\mathbf{Y}$ is a good set as per Definition 7.56. Let $W(\mathbf{Y}) \in \mathcal{W}(\mathbf{Y})$ be the matrix computed by Algorithm 7.55. Suppose the algorithm does not reject. Then*

$$\left\|\phi(W(\mathbf{Y})) - \mathbf{W}^*\right\|_1 \leqslant \frac{n^2}{k} \cdot \frac{3}{k^{98}} .$$

The proof is similar to the classical utility analysis of the sum-of-squares program found, e.g., in [HL18, FKP+19]. We defer it to Appendix E.2.

Together, the above results imply that the vectors $v^{(i)}$ computed by the algorithm are close to the true centers of the mixture.

Lemma 7.66 (Closeness to true centers)**.** *Consider the settings of Theorem 7.53. Suppose $\mathbf{Y}$ is a good set as per Definition 7.56. Let $\mathbf{W} \in \mathcal{W}(\mathbf{Y})$ be the matrix computed by Algorithm 7.55. Suppose the algorithm does not reject in step 3. Then for each $\ell \in [k]$, there exists $\frac{n}{k} \cdot \left(1 - \frac{2}{k^{47}}\right)$ indices $i \in [n]$, such that*

$$\left\|v^{(i)}(\mathbf{W}) - \mu_\ell\right\|_2 \leqslant O\left(k^{-25}\right) .$$

Proof. We aim to show that for most indices $i \in [n]$ the vectors $\left\|\phi(\mathbf{W}_i)\right\|_1^{-1}\phi(\mathbf{W}_i)$ and $\left\|\mathbf{W}_i^*\right\|_1^{-1}\mathbf{W}_i^*$ induce a 2-explicitly 40-bounded distribution over $\mathbf{Y}$. If additionally the two vectors are close in ℓ_1-norm, the result will follow by Theorem 7.23.

Note that $\left\|\mathbf{W}_i^*\right\|_1^{-1}\mathbf{W}_i^*$ induces a 2-explicitly 40-bounded distribution by Lemma 7.57. By Markov's inequality and Lemma 7.65 there can be at most n/k^{48} indices $j \in [n]$ such that

$$\left\|\phi(\mathbf{W})_j - \mathbf{W}_j^*\right\|_1 \geqslant \frac{n}{k} \cdot \frac{3}{k^{50}} .$$

Consider all remaining indices i. It follows that

$$\left\|\phi(\mathbf{W}_i)\right\|_1 \geqslant \left\|\mathbf{W}_i^*\right\|_1 - \left\|\phi(\mathbf{W})_i - \mathbf{W}_i^*\right\|_1 \geqslant \frac{n}{k} \cdot \left(1 - \frac{k}{n^{0.4}} - \frac{3}{k^{50}}\right) \geqslant \frac{n}{k} \cdot \left(1 - \frac{10}{k^{50}}\right) .$$

Hence, by Lemma 7.61 the distribution induced by $\left\|\phi(\mathbf{W}_i)\right\|_1^{-1}\phi(\mathbf{W}_i)$ is 2-explicitly 40-bounded distribution. Further, using $\left\|\mathbf{W}_i^*\right\|_1 \geqslant \frac{n}{k}\left(1 - \frac{k}{n^{0.4}}\right)$ we can bound

$$\begin{aligned}
&\left\|\left\|\phi(\mathbf{W}_i)\right\|_1^{-1}\phi(\mathbf{W}_i) - \left\|\mathbf{W}_i^*\right\|_1^{-1}\mathbf{W}_i^*\right\|_1 = \left\|\phi(\mathbf{W}_i)\right\|_1^{-1}\left\|\mathbf{W}_i^*\right\|_1^{-1} \cdot \left\|\left\|\mathbf{W}_i^*\right\|_1\phi(\mathbf{W}_i) - \left\|\phi(\mathbf{W}_i)\right\|_1\mathbf{W}_i^*\right\|_1 \\
&\leqslant \left\|\phi(\mathbf{W}_i)\right\|_1^{-1}\left\|\mathbf{W}_i^*\right\|_1^{-1} \cdot \left(\left|\left\|\phi(\mathbf{W}_i)\right\|_1 - \left\|\mathbf{W}_i^*\right\|_1\right| \cdot \left\|\phi(\mathbf{W}_i)\right\|_1 + \left\|\phi(\mathbf{W}_i)\right\|_1 \cdot \left\|\phi(\mathbf{W}_i) - \mathbf{W}_i^*\right\|_1\right) \\
&\leqslant \left\|\mathbf{W}_i^*\right\|_1^{-1} \cdot 2\left\|\phi(\mathbf{W}_i) - \mathbf{W}_i^*\right\|_1 \leqslant \frac{6}{k^{50} \cdot \left(1 - \frac{k}{n^{0.4}}\right)} \leqslant \frac{7}{k^{50}} .
\end{aligned}$$

Hence, by Theorem 7.23 for each $l \in [k]$ there are at least $\frac{n}{k} - n^{0.6} - \frac{n}{k^{48}} \geqslant \frac{n}{k} \cdot \left(1 - \frac{2}{k^{47}}\right)$ indices i such that

$$\left\| v^{(i)}(\mathbf{W}) - \left\|\mathbf{W}_i^*\right\|_1^{-1} \sum_{j=1}^{n} \mathbf{W}_{i,j}^* \mathbf{y}_j \right\|_2 \leqslant O\left(k^{-25}\right).$$

The result now follows by standard concentration bounds applied to the distribution induced by $\left\|\mathbf{W}_i^*\right\|_1^{-1} \mathbf{W}_i^*$. □

An immediate consequence of Lemma 7.66 is that the vectors $\bar{\boldsymbol{\nu}}^{(i)}$ inherits the good properties of the vectors $v^{(i)}$ with high probability.

Corollary 7.67 (Closeness to true centers after sub-sampling)**.** *Consider the settings of Theorem 7.53. Suppose* $\mathbf{Y}$ *is a good set as per Definition 7.56. Let* $\mathbf{W} \in \mathcal{W}(\mathbf{Y})$ *be the matrix computed by Algorithm 7.55. Suppose the algorithm does not reject. Then with high probability for each* $\ell \in [k]$*, there exists* $\frac{n^{0.01}}{k} \cdot \left(1 - \frac{150}{k^{47}}\right)$ *indices* $i \in \boldsymbol{\mathcal{S}}$*, such that*

$$\left\|\bar{\boldsymbol{\nu}}^{(i)} - \mu_\ell\right\|_2 \leqslant O\left(k^{-25}\right).$$

Proof. For each $\ell \in [k]$, denote by $\mathcal{T}_\ell$ the set of indices in $[n]$ satisfying

$$\left\|v^{(i)}(\mathbf{W}) - \mu_\ell\right\|_2 \leqslant O\left(k^{-25}\right).$$

By Lemma 7.66 we know that $\mathcal{T}_\ell$ has size at least $\frac{n}{k} \cdot \left(1 - \frac{2}{k^{47}}\right)$. Further, let $\mathcal{S}$ be the set of indices selected by the algorithm. By Chernoff's bound Fact E.8 with probability $1 - e^{-n^{\Omega(1)}}$, we have $|\boldsymbol{\mathcal{S}} \cap \mathcal{T}_\ell| \geqslant \frac{n^{0.01}}{k} \cdot \left(1 - \frac{150}{k^{47}}\right)$. Taking a union bound over all $\ell \in [k]$ we get that with probability $1 - e^{-n^{\Omega(1)}}$, for each $\ell \in [k]$, there exists $\frac{n^{0.01}}{k} \cdot \left(1 - \frac{150}{k^{47}}\right)$ indices $i \in \boldsymbol{\mathcal{S}}$ such that

$$\left\|v^{(i)}(\mathbf{W}) - \mu_\ell\right\|_2 \leqslant O\left(k^{-25}\right).$$

Now, we obtain the corollary observing (cf. Fact E.5 with $m = 1$) that with probability at least $1 - e^{-n^{\Omega(1)}}$, for all $i \in \boldsymbol{\mathcal{S}}$

$$\left\|\bar{\boldsymbol{\nu}}^{(i)} - v^{(i)}(\mathbf{W})\right\|_2 = \|\mathbf{w}\|_2 \leqslant n^{-0.05} \cdot \frac{\sqrt{\log(2/\delta)}}{\varepsilon} \cdot \sqrt{d} \leqslant n^{-0.04} \leqslant O\left(k^{-25}\right).$$

□

For each ℓ, denote by $\boldsymbol{\mathcal{G}}_\ell \subseteq \boldsymbol{\mathcal{S}}$ the set of indices $i \in \boldsymbol{\mathcal{S}}$ satisfying

$$\left\|\bar{\boldsymbol{\nu}}^{(i)} - \mu_\ell\right\|_2 \leqslant O\left(k^{-25}\right).$$

Let $\boldsymbol{\mathcal{G}} := \bigcup_{\ell \in [k]} \boldsymbol{\mathcal{G}}_\ell$. We now have all the tools to prove utility of Algorithm 7.55. We achieve this by showing thst with high probability, each bin returned by the algorithm at step 7 satisfies $\boldsymbol{\mathcal{G}}_{\ell'} \subseteq \mathbf{B}_\ell$ for some $\ell, \ell' \in [k]$. Choosing the bins small enough will yield the desired result.

Lemma 7.68 (Closeness of estimates)**.** *Consider the settings of Theorem 7.53. Suppose* $\mathbf{Y}$ *is a good set as per Definition 7.56. Let* $\mathbf{W} \in \mathcal{W}(\mathbf{Y})$ *be the matrix computed by Algorithm 7.55. Suppose the algorithm does not reject. Then with high probability, there exists a permutation* $\pi : [k] \to [k]$ *such that*

$$\max_{\ell \in [k]} \left\| \mu_\ell - \hat{\boldsymbol{\mu}}_{\pi(\ell)} \right\|_2 \leqslant O\left(k^{-20}\right)$$

Proof. Consider distinct $\ell, \ell' \in [k]$. By Corollary 7.67 for each $\bar{\boldsymbol{\nu}}^{(i)}, \bar{\boldsymbol{\nu}}^{(j)} \in \boldsymbol{\mathcal{G}}_\ell$ it holds that

$$\left\| \bar{\boldsymbol{\nu}}^{(i)} - \bar{\boldsymbol{\nu}}^{(j)} \right\|_2 \leqslant C \cdot k^{-25} ,$$

for some universal constant $C > 0$. Moreover, by assumption on $\mu_\ell, \mu_{\ell'}$ for each $\bar{\boldsymbol{\nu}}^{(i)} \in \boldsymbol{\mathcal{G}}_\ell$ and $\bar{\boldsymbol{\nu}}^{(j)} \in \boldsymbol{\mathcal{G}}_{\ell'}$

$$\left\| \bar{\boldsymbol{\nu}}^{(i)} - \bar{\boldsymbol{\nu}}^{(j)} \right\|_2 \geqslant \Delta - O\left(k^{-25}\right) .$$

Thus, by Lemma E.9 with probability at least $1 - e^{\Omega(d^*)} \geqslant 1 - n^{-100}$ it holds that or each $\bar{\boldsymbol{\nu}}^{(i)}, \bar{\boldsymbol{\nu}}^{(j)} \in \boldsymbol{\mathcal{G}}_\ell$ and $\bar{\boldsymbol{\nu}}^{r} \in \boldsymbol{\mathcal{G}}_{\ell'}$ with $\ell' \neq \ell$,

$$\left\| \boldsymbol{\Phi} \bar{\boldsymbol{\nu}}^{(i)} - \boldsymbol{\Phi} \bar{\boldsymbol{\nu}}^{(j)} \right\|_2 \leqslant C^* \cdot k^{-25} \qquad \text{and} \qquad \left\| \boldsymbol{\Phi} \bar{\boldsymbol{\nu}}^{(i)} - \boldsymbol{\Phi} \bar{\boldsymbol{\nu}}^{(r)} \right\|_2 \geqslant \Delta - C^* \cdot k^{-25}$$

for some other universal constant $C^* > C$. Let $Q_{\boldsymbol{\Phi}}(\boldsymbol{\mathcal{G}}_\ell) \subseteq \mathbb{R}^{d^*}$ be a ball of radius $C^* \cdot \left(k^{-25}\right)$ such that $\forall i \in \boldsymbol{\mathcal{G}}_\ell$ it holds $\boldsymbol{\Phi} \bar{\boldsymbol{\nu}}^{(i)} \in Q_{\boldsymbol{\Phi}}(\boldsymbol{\mathcal{G}}_\ell)$. That is, $Q_{\boldsymbol{\Phi}}(\boldsymbol{\mathcal{G}}_\ell)$ contains the projection of all points in $\boldsymbol{\mathcal{G}}_\ell$.

Recall that $d^* = 100 \log(n) \leqslant 100k^5$ and $b = k^{-15}$. Let $\boldsymbol{\mathcal{B}} = \{\mathbf{B}_i\}_{i=1}^{\infty}$ be the sequence of bins computed by the histogram learner of Lemma 7.20 for $\mathbb{R}^{d^*}$ at step 7 of the algorithm. By choice of b, and since $\mathbf{q}$ is chosen uniformly at random in $[0, b]$, the probability that there exists a bin $\mathbf{B} \in \boldsymbol{\mathcal{B}}$ containing $Q_{\boldsymbol{\Phi}}(\boldsymbol{\mathcal{G}}_\ell)$ is at least

$$1 - d^* \cdot \frac{C^*}{b} \cdot \left(k^{-25}\right) \geqslant 1 - \frac{100 C^*}{b} \cdot k^{-20} \geqslant 1 - O\left(k^{-5}\right) ,$$

where we used that $d^* = 100 \log n \leqslant 100k^5$. A simple union bound over $\ell \in [k]$ yields that with high probability for all $\ell \in [k]$, there exists $\mathbf{B} \in \boldsymbol{\mathcal{B}}$ such that $Q_{\boldsymbol{\Phi}}(\boldsymbol{\mathcal{G}}_\ell) \subseteq \mathbf{B}$. For simplicity, denote such bin by $\mathbf{B}_\ell$.

We continue our analysis conditioning on the above events, happening with high probability. First, notice that for all $l \in [k]$

$$\max_{u, u' \in \mathbf{B}_\ell} \|u - u'\|_2^2 \leqslant d^* \cdot b^2 \leqslant 100 k^{-25} \leqslant \frac{\Delta - C^* k^{-25}}{k^{10}} ,$$

and thus there cannot be $\ell, \ell' \in [k]$ such that $Q_{\boldsymbol{\Phi}}(\boldsymbol{\mathcal{G}}_\ell) \subseteq \mathbf{B}_\ell$ and $Q_{\boldsymbol{\Phi}}(\boldsymbol{\mathcal{G}}'_\ell) \subseteq \mathbf{B}_\ell$. Moreover, by Corollary 7.67 and

$$\min_{\ell \in [k]} |\boldsymbol{\mathcal{G}}_\ell| \geqslant \frac{n^{0.01}}{k} \cdot \left(1 - \frac{150}{k^{47}}\right) ,$$

and hence

$$|\mathcal{S} \setminus \mathcal{G}| \leqslant n^{0.01} \cdot \frac{150}{k^{47}} = \frac{n^{0.01}}{k} \cdot \frac{150}{k^{46}}$$

it must be that step 7 returned bins $\mathbf{B}_1, \ldots, \mathbf{B}_k$. This also implies that the algorithm does not reject. Further, by Lemma E.9 for all $\bar{\boldsymbol{\nu}}^{(i)}, \bar{\boldsymbol{\nu}}^{(j)}$ such that $\boldsymbol{\Phi}\bar{\boldsymbol{\nu}}^{(i)}, \boldsymbol{\Phi}\bar{\boldsymbol{\nu}}^{(j)} \in \mathbf{B}_l$ it holds that

$$\left\|\bar{\boldsymbol{\nu}}^{(i)} - \bar{\boldsymbol{\nu}}^{(j)}\right\|_2 \leqslant C^* \cdot \left\|\boldsymbol{\Phi}\bar{\boldsymbol{\nu}}^{(i)} - \boldsymbol{\Phi}\bar{\boldsymbol{\nu}}^{(j)}\right\|_2 \leqslant C^* \cdot \sqrt{d^*} \cdot b \leqslant O\left(k^{-12}\right).$$

And hence, by triangle inequality, we get

$$\left\|\bar{\boldsymbol{\nu}}^{(i)} - \mu_l\right\|_2 \leqslant O\left(k^{-12}\right).$$

Finally, recall that for each $\ell \in [k]$,

$$\hat{\boldsymbol{\mu}}_l \coloneqq \frac{1}{\left|\left\{j \;\middle|\; \boldsymbol{\Phi}\bar{\boldsymbol{\nu}}^{(j)} \in \mathbf{B}_i\right\}\right|} \cdot \left(\sum_{\boldsymbol{\Phi}\bar{\boldsymbol{\nu}}^{(j)} \in \mathbf{B}_l} \bar{\boldsymbol{\nu}}^{(j)}\right) + \mathbf{w}',$$

where $\mathbf{w}' \sim N\left(0, N\left(0, 32 \cdot k^{-120} \cdot \frac{\log(2kn/\delta)}{\varepsilon^2} \cdot \mathrm{Id}\right)\right)$. Since by choice of n, k, ε it holds that

$$32 \cdot k^{-120} \cdot \frac{\log(2kn/\delta)}{\varepsilon^2} \leqslant O\left(k^{-90}\right),$$

we get with probability at least $1 - e^{-k^{\Omega(1)}}$ for each $\ell \in [k]$, by Fact E.5, with $m = 1$, and a union bound that

$$\|\mathbf{w}'\| \leqslant O\left(k^{-20}\right).$$

Since all $\bar{\boldsymbol{\nu}}^{(i)}$ such that $\boldsymbol{\Phi}\bar{\boldsymbol{\nu}}^{(i)} \in \mathbf{B}_l$ are at most $O\left(k^{-12}\right)$-far from μ_l, also their average is. We conclude that

$$\left\|\hat{\boldsymbol{\mu}}_\ell - \mu_l\right\|_2 \leqslant O(k^{-12}) + \|\mathbf{w}\|_2 \leqslant O(k^{-12}).$$

This completes the proof. □

Now Theorem 7.53 is a trivial consequence.

Proof of Theorem 7.53. The error guarantees and privacy guarantees immediately follows combining Lemma 7.58, Lemma 7.65, Lemma 7.64 and Lemma 7.68. The running time follows by Fact 2.9. □

Part III

Speeding up robust algorithms

Chapter 8

Fast and robust algorithm for graph partitioning problems

In this chapter, based on [CdM23], we continue the discussion started in Section 1.1.6, prove Theorem 1.13 and other related results. We start by considering the balanced cut problem:

Problem 8.1 (a-balanced cut)**.** Let $a \in [0\,, 1/2]$ and let G be a graph on n vertices. Find the partition (A, B) with maximum cut that also satisfies $\min\{|A|, |B|\} \geqslant an$.

We restate the semi-random model introduced in [MMV12], which subsumes many of the other semi-random models for balanced-cut considered in the literature [BCLS87, DF86, Bop87, FK01, McS01, JS93, DI98, BL12, MMV12, MMV14, Pen20, CPRT22].

Model 8.2 (Random cut with monotone perturbations)**.** We consider graphs over n vertices generated through the following process. Let $a \in (0, 1/2)$, $\eta(n) \in (0, 1)$:

(i) The adversary partition $[n]$ into sets A, B satisfying $|A|, |B| \geqslant an$.

(ii) Each edge between A and B is drawn randomly and independently with probability η.

(iii) The adversary arbitrarily adds edges within A and within B.

(iv) The adversary arbitrarily removes edges between A and B.

In the worst-case, combining [AK07, She09, CKL$^+$22] it is possible to achieve a $O(\sqrt{\log(n)}/\varepsilon)$-approximation algorithm for Problem 8.1 in time $\tilde{O}\big(n^{1+\varepsilon}\big)$. In contrast, [MMV12] introduced a polynomial time $O(1)$-approximation algorithm for Model 8.2, when a constant fraction of the edges in the cut are untouched. However, the algorithm runs in time n^C for a large constant $C \geqslant \Omega(1)$. Indeed it requires solving polylogarithmically many SDPs with n^3 constraints, each followed by a rounding procedure taking time $O(n^3)$.

For any $\rho > 1$, our algorithm achieves an $O(\sqrt{\rho})$-approximation with almost linear running time, namely $\tilde{O}\left(|V(G)|^{1+O(1/\rho^2)} + |E(G)|\right)$. [1] The lower bound on η is also needed in [MMV12]. Thus, for $\rho \leqslant O(1)$, the error guarantees match [MMV12].

Theorem 8.3 (Restatement fo Theorem 1.13)**.** *Let $\rho > 0$. Let G be a graph over n vertices generated through Model 8.2 with parameters $a > 0, \eta \geqslant \Omega(\frac{(\log n)^2 \cdot (\log\log n)^2}{n})$. There exists an algorithm that on input G, with probability $1 - o(1)$, outputs an $\Omega(a)$-balanced cut of value at most $O(n^2 \cdot \eta \cdot \rho)$, namely a cut where each side has size at least $\Omega(a \cdot n)$.*

Moreover, the algorithm runs in time $\tilde{O}\left(|V(G)|^{1+O(1/\rho^2)} + |E(G)|\right)$.

As in [MMV12], Theorem 8.3 achieves constant approximation guarantees when the true partition has a cut of size $\Omega(n^2 \cdot \eta)$. This can also be seen as assuming that the parameter η is known and "tight", in the sense that the adversary leaves a constant fraction of edges in the true cut unmodified. Furthermore, as we see next, the ideas in Theorem 8.3 can be naturally extended to more sophisticated graph problems.

Generalizations. In [CKMM19], the authors considered the celebrated objective function for hierarchical clustering introduced by Dasgupta [Das16] and investigate how well it can be approximated beyond-the-worst-case. Assuming the Small Set Expansion conjecture [Ste10b], the problem cannot be approximated within any constant factor. The authors thus introduced a generative model for hierarchical clustering inputs called the *hierarchical stochastic block model* that naturally generalizes the classic stochastic block model, and show that one can approximate Dasgupta's objective up to a constant factor in that model and under semi-random perturbation (the precise definition of the model can be found in Section 8.5). The framework introduced here can be used to improve the complexity of the algorithm of [CKMM19].

Theorem 8.4. *Let G be a graph generated from the HSM (Definition 8.40) with $p_{min} = \Omega(\log n / n^{2/3})$. Let $\rho > 0$ be a large constant. Then, there exists a randomized algorithm that runs in time $\tilde{O}\left(|V(G)|^{1+O(1/\rho^2)} + |E(G)|\right)$ with probability $1 - o(1)$ outputs a tree T such that*

$$cost(T; G) = O(OPT(\bar{G})), \tag{8.0.1}$$

where $OPT(\bar{G})$ denotes the value of the optimal tree for $\bar{G}$ and we note that $OPT(\bar{G}) = cost(\widetilde{T}; \bar{G})$, where $\widetilde{T}$ is the generating tree. Furthermore, the above holds even in the semi-random case, i.e., when an adversary is allowed to remove any subset of the edges from G.

We leave it as an open question to understand whether the techniques presented here could lead to further improvements for other related problems for which the beyond-worst-case analysis has been studied (e.g.: Bilu-Linial stability for multicut [BL12, AMM17]).

[1] We hide multiplicative terms polylogarithmic in $|V(G)|$ with the notation $\tilde{O}$.

8.1 Techniques

We present here the main ideas contained in the proof of Theorem 8.3. Throughout the section let $\mathbf{G}$ be a random graph sampled through steps (i) and (ii) of Model 8.2 and let G° be the resulting graph after steps (iii) and (iv). We let $\alpha \leqslant (1 + o(1)) \cdot n^2 \cdot \eta$ be the number of edges in $\mathbf{G}$ and thus an upper bound on the optimal balanced cut in G° with high probability.

Slow algorithms for balanced cut in the semi-random model. In order to see how to design a near linear time algorithm with constant approximation factor for Model 8.2, it is necessary to first understand how the known *slow* algorithm of [MMV12] works . Consider the random graph $\mathbf{G}$ and let $v_1, \ldots, v_n \in \mathbb{R}^n$ be the embedding corresponding to the returned SDP solution where v_i corresponds to the embedding of the ith vertex of $\mathbf{G}$.[2] Their algorithm is an iterative procedure that cycles over two subroutines: first, the algorithm solves the canonical balanced cut SDP as in [ARV09]; second, the algorithm *carefully* removes clusters of vertices that are particularly *close* to each other in the embedding given by the SDP solution.

Concretely, the latter step identifies so-called (δ, n)-*heavy* vertex, namely a vertex i such that its embedding v_i in the SDP solution is at distance at most δ from at least $10\delta^2 n$ embeddings of other vertices in the SDP solution[3]. Then, the algorithms *carves out* a ball of radius δ; It creates a cluster containing i and all the vertices j s.t. v_j is at distance at most δ from v_i. Here δ is a parameter chosen appropriately. At the end of this process, the algorithm has removed a set $H_\delta \subseteq V(\mathbf{G})$ of vertices from the instance.

The crucial observation here is that, in *every* feasible embedding of the random graph G, the random cut *restriced* to the non-heavy vertices satisfies a one-sided Chebyshev-like inequality of the form:

$$\mathbb{P}_{ij \overset{u.a.r.}{\sim} E(\mathbf{G} \setminus H_\delta)} \left(\|v_i - v_j\|^2 \leqslant \delta \right) \leqslant 1/\delta^2 .$$

That is, with high probability only a $O(\delta^2)$-fraction of the edges between non-heavy vertices is shorter than δ in the embedding. This property is called *geometric expansion*.

Now the crux of the argument is that, given a feasible embedding $v_1, \ldots, v_n \in \mathbb{R}^n$ of G, if by removing heavy vertices we don't cut more than $O(\alpha)$ edges, then the geometric expansion property guarantees that the minimum balanced cut in the *remaining* graph has cardinality at most $O(\delta^2 \cdot \alpha)$. Thus after several iterations of the algorithm we have decreased the value of the minimum balanced cut by at least a $O(1/\sqrt{\log n})$ factor and so a simple application of the SDP rounding of [ARV09] returns now a cut of optimal value α. Importantly, the geometric expansion property is robust to monotone changes – namely to

[2] See Section 8.3 for a definition of the program.

[3] Notice that a ball of radius 2δ centered at any heavy vertex v contains at least $\delta^2 n$ vertices.

the changes that the adversary can make to the graph at the last two steps of the generative model (Model 8.2) and thus, the exact same reasoning applies for the graph G° as well.

Speeding up the approach through the matrix multiplicative weights framework. As previously discussed, the matrix multiplicative weight method (e.g. see [AK07, She09, Ste10a]) can be used in place of exact semidefinite programs solvers. Recall the framework aims at obtaining a "feasible enough" solution to the SDP so that the desired rounding argument works out (these are not approximately feasible solutions, in fact they may appear to be *far* from satisfying the constraints of the program). The technique requires an ORACLE algorithm that efficiently answers **yes** if the candidate solution is feasible enough or otherwise answers **no** and exhibits constraints that are violated by the current solution. These will then be used to pick the direction of movement in the mirror descend algorithm. The running time of the algorithm depends on the so called "width" of the oracle and the challenge is usually to design oracles of bounded width.

We have seen that in the context of balanced cut for arbitrary graphs, this approach has been extremely successful, leading to an $O(\sqrt{\log n}/\varepsilon)$ approximation algorithm running in time $O(n^{1+\varepsilon}\log n + m)$ (combining [She09] with [CKL+22]). However, at each iteration of [MMV12], to apply the rounding framework and obtain matching guarantees, it seems needed to obtain first a feasible optimal solution. Moreover, the rounding framework is not efficient since it requires to find out all the heavy vertices of the graph – or in other words, all the particularly dense balls in the SDP solution. To overcome these obstacles we will need to deviate from the canonical matrix multiplicative weights paradigm.

Approximate heavy vertices removal. Our first step is thus to speed up the procedure that identifies all the dense balls of the SDP solution. Identifying dense balls in high-dimensional Euclidean spaces (say of dimension $\Omega(\varepsilon^{-2}\log n)$) is a well studied problem. We thus make use of locally-sensitive hashing (LSH) schemes to *approximately* solve this problem. Namely, our procedure will recover all the heavy vertices but may yield false positives, namely balls that are almost dense – up to a constant factor away from the target density. Concretely, we ask for balls that contain at least $10\delta^2 n$ vertices but are of radius $\sqrt{\rho}\delta$ instead of δ, where ρ is a parameter. While we can achieve this in time $|V(\mathbf{G})|^{1+1/\rho+o(1)}$, we now have to modify the next steps of the rounding to take the false positive into account in the rounding.

Rounding and the matrix multiplicative weights framework. The challenge here is more important. We can show that the above approximate heavy vertex removal procedure guarantees with constant probability that only a few edges will be cut *if* applied to a feasible solution of small objective value. Note here that the procedure may cut few edges while still being applied to a infeasible solution of high objective value, but we would still be satisfied since we are making good progress in the graph partitioning at low cost. That is,

even if a candidate solution is overall far from being feasible, it turns out to be sufficiently good for us if it is close to being feasible on the heavy vertices and their neighborhoods. We thus mainly have to deal with the case where the procedure cuts too many edges. The key idea here is that if the above procedure cuts too many edges, then we can show that with reasonable probability there exists a hyperplane of *small width* separating our solution from the set of feasible solution of small objective value, and moreover identify it efficiently. Then, we can make progress and obtain a better solution through applying a step of the matrix multiplicative weights framework.

In this case, the intuition is that *on average* the probability that an edge with exactly one endpoint in these heavy balls is cut throughout the removal procedure must be larger than for feasible embeddings with small objective value. Indeed otherwise we could have expected our procedure to cut fewer edges. Thus we can conclude that, either the current solution has significantly larger objective value, or several triangle inequalities must be violated at the same time and thus we can provide a feedback matrix of small width. To ensure that our heavy vertices removal procedure would have indeed cut fewer edges if given a feasible solution, we repeat this process poly-logarithmically times.

Remark 8.5 (On potential practical implementations). It is important to remark that we believe the hidden constant behind the algorithm of Theorem 8.3 to be small. In particular, we do not try to optimize the constants appearing in our analysis and therefore they should not be seen as an inherent barrier to practical implementations.

8.2 Preliminaries

Contrarily to Chapter 2, we hide multiplicative factors *poly-logarithmic* in n using the notation $\tilde{O}(\cdot), \tilde{\Omega}(\cdot)$. Similarly, we hide absolute constant multiplicative factors using the standard notation $O(\cdot), \Omega(\cdot), \Theta(\cdot)$. Often times we use the letter C to denote universal constants independent of the parameters at play.

Vectors and matrices. For a matrix M, we denote its eigenvalues by $\lambda_1(M), \ldots, \lambda_n(M)$; we simply write λ_i when the context is clear. Let $\mathcal{S}_n \subset \mathbb{R}^n$ be the set of real symmetric n-by-n matrices and let $\Delta_n(r) := \{X \in \mathcal{S}_n \mid \operatorname{Tr} X = r, X \succeq 0\}$. For $X \in \mathcal{S}_n$, the matrix exponential is $\exp(X) = \sum_{i=0}^{\infty} \frac{X^i}{i!}$. We remark that $\exp(X)$ is positive semidefinite for all symmetric X as $\exp(X) = \left(\exp(\frac{1}{2}X)\right)^\intercal \exp(\frac{1}{2}X)$. For a vector $v \in \mathbb{R}^n$, we write $v \geqslant \mathbf{0}$ if all entries of v are non-negative. We use $\mathbb{S}^n \subseteq \mathbb{R}^n$ to denote the unit sphere.

Graphs. For a graph G we write L_G for the associated combinatorial Laplacian, which is a matrix with rows and columns indexed by the nodes of G such that $(L_G)_{ii} = \sum_{ij \in E(G)} 1$, i.e. the degree of node i, and for $i \neq j$ $(L_G)_{ij}$ is -1 if $ij \in E(G)$ and 0 otherwise. When the context is clear we drop the specification of G. Unless specified otherwise, we use n to denote $|V(G)|$. For a partition (A, B) of the vertices of G, we write $E(A, B) \subseteq E$ for the set of

edges in the A-B cut. We say that a partition (A, B) is a-balanced if $|A|/|B| \geqslant a$ assuming $|A| \leqslant |B|$.

Maximum flow. Let $G(V, E)$ be a graph. For a flow which assigns value f_p to path p define f_e to be the flow on edge $e \in E(G)$, i.e. $f_e := \sum_{p \ni e} f_p$. Define define f_{ij} to be the total flow between nodes i , j, i.e. $f_{ij} = \sum_{p \in P_{ij}} f_p$, where P_{ij} is the set of paths from i to j. A valid d-regular flow is one that satisfies the capacity constraints: $\forall e \in E : f_e \leqslant 1$ and $\forall i \in V : f_i \leqslant d$. For a partition (A, B) of G, the maximum d-regular flow between A and B is the maximum d-regular flow between vertices s and t in the graph obtain from G as follows: (1) connect all vertices in A to a new vertex s by edges of capacity d, (2) connect all vertices in B to a new vertex t by edges of capacity d.

Through the chapter we always assume the capacities d to be integral and bounded by $O(\text{poly}(n))$. We assume the algorithm used to compute the maximum flow is the near linear time algorithm in [CKL+22], captured by the result below:

Theorem 8.6 (Maximum flow in almost linear time [CKL+22])**.** *Let G be a graph on n vertices and let d be integral of value at most $O(\text{poly}(n))$. There exists an algorithm computing the maximum d-regular flow between two vertices in time at most $O(|E(G)|^{1+o(1)})$.*

8.2.1 The matrix multiplicative weights method for SDPs

We recall here how the matrix multiplicative method can be used to approximately solve semidefinite programs. For a more in-depth discussion we redirect the reader to [AK07, Ste10a]. We focus on minimization problems although the same framework applies to maximization problems. A primal semidefinite program over n^2 variables (i.e. the n-by-n matrix variable X) and m constraints can be written in its canonical form as

$$\begin{Bmatrix} \min & \langle L, X \rangle \\ \forall j \in [m], & \langle A_j, X \rangle \geqslant b_j \\ & X \succeq \mathbf{0} \end{Bmatrix} \tag{8.2.1}$$

Here $A_1, \ldots, A_m, L$ are symmetric matrices. We denote the feasible set of solutions by $\mathcal{X}$ and the optimal objective value by α. For simplicity we assume that $A_1 = -\text{Id}_n$ and $b_1 = -r$. This serves to bound the trace of the solution so that $\mathcal{X} \subseteq \Delta_n(r)$. The associated dual, with variables $y_1, \ldots, y_m$, is the following program

$$\begin{Bmatrix} \max & \langle b, y \rangle \\ & \sum_{j \in [m]} A_j y_j \preceq L \\ & y \geqslant \mathbf{0} \end{Bmatrix} \tag{8.2.2}$$

where b is the m-dimensional vector with entries $b_1, \ldots, b_m$.

For a convex set $\mathcal{X}^* \subseteq \Delta_n(r)$ (think of $\mathcal{X}^*$ as the set of feasible solution to a program of the form Eq. (8.2.1) with objective value close to the optimum) a *γ-separation* ORACLE is an algorithm that , given a candidate matrix X, outputs one of the following:

yes the ORACLE determines X is "close" (the precise notion of closeness is problem dependent) to $\mathcal{X}^*$.

no the ORACLE finds a hyperplane that separates X from $\mathcal{X}^*$ by a γ-margin. That is, it outputs a symmetric matrix M such that for all $X' \in \mathcal{X}^*$ we have $\langle M, X' \rangle \geqslant 0$ while $\langle M, X \rangle < \gamma\alpha$.

A γ-separation ORACLE is said to be ζ-bounded if $\|M\| \leqslant \zeta$ for any hyperplane M found by the ORACLE. The boundedness of the ORACLE will be relevant for the running time of our algorithms. It is important to notice that the parameters ζ, γ are not independent, in particular one may increase γ by scaling up the corresponding matrix M. We keep them distinct for convenience .

Concretely, given a program of the form Eq. (8.2.1) and a candidate solution X, we will consider ORACLE algorithms that, in the **no** case, find a pair (y, F) where F is a matrix in $\mathcal{S}_n$ satisfying $F \preceq L$ and y is a candidate solution[4] for the dual program Eq. (8.2.2) such that $y \in \left\{ y \;\middle|\; \langle b, y \rangle \geqslant \alpha \,, y \geqslant 0 \right\}$ and

$$\langle \sum_{j \in [m]} A_j y_j - F, X \rangle \leqslant -\gamma \cdot \alpha \,.$$

It is easy to see that this is indeed a separating hyperplane as for any feasible solution X' with objective value less than $\alpha(1+\gamma)$

$$\langle \sum_{j \in [m]} A_j y_j - F, X' \rangle \geqslant \sum_{j \in [m]} b_j y_j - \langle L, X' \rangle > \alpha - (1+\gamma)\alpha = -\gamma \cdot \alpha$$

, .

We will use our oracle algorithms in the following framework.

[4] Not necessarily feasible.

Algorithm 8.7 (Matrix multiplicative weights algorithm for SDPs)**.**
Input: A program of the form Eq. (8.2.1) with optimal value α, a ζ-bounded γ-separation ORACLE, parameters T, ε, r.
Set $X^{(1)} = \frac{r}{n}\mathrm{Id}_n$. For $t = 1, \ldots, T$:

1. Run the ORACLE with candidate solution $X^{(t)}$.
2. If the ORACLE outputs **yes**, return $X^{(t)}$.
3. Else, let $(y^{(t)}, F)$ be the pair generated by ORACLE. Set $Y^{(t)} = \left(\sum_{j\in[m]} A_j y_j^{(t)} - F + \zeta \mathrm{Id}_n\right)/2\zeta$.
4. Compute $X^{(t+1)} = r \cdot \exp\left(\varepsilon \sum_{t'\leqslant t} Y^{(t')}\right)/\mathrm{Tr}\exp\left(\varepsilon \sum_{t'\leqslant t} Y^{(t')}\right)$ and continue.

The choice of the iterative updates in step 4 is based on the matrix multiplicative weights method. In particular, this allows one to obtain the following crucial statement.

Theorem 8.8 ([AK07])**.** *Consider Algorithm 8.7. Let $\varepsilon \leqslant \gamma\alpha/(2\zeta \cdot r)$ and $T \geqslant 2\varepsilon^{-2}\log n$. If there exists a feasible solution with value at most $\alpha(1+\gamma)$, then* ORACLE *will output* ***yes*** *within T iterations.*

8.2.1.1 Approximate matrix exponentiation, robust and reliable oracles

There are two issues with Theorem 8.8 if one aims for near linear running time: first, already writing down $X^{(t)}$ requires time quadratic in n; second, algorithms known to compute the matrix exponentiation are slow . One can circumvent these obstacles computing the exponentiation only approximately while also keeping only an approximate representation of $X^{(t)}$. To formalize this we introduce additional notation. For a positive semidefinite n-by-n matrix M, we let $P_{\leqslant p}(M)$ be the degree-p approximation of the matrix exponential $\exp(X)$:

$$P_{\leqslant p}(M) := \sum_{i\leqslant p} \frac{1}{i!} M^i \,.$$

Recall that for a matrix M, the Gram decomposition of the exponential $\exp(M)$ is $\exp(M) = \exp(\frac{1}{2}M)^{\mathsf{T}}\exp(\frac{1}{2}M)$ thus we may see $P_{\leqslant p}(\frac{\varepsilon}{2}\sum_{t'\leqslant t} Y^{(t')})$ as a a matrix having as columns low-degree approximations of the Gram vectors of $\exp(M)$. One can then embed these vectors in a low dimensional space, without distorting their pair-wise distance by projecting them onto a random d-dimensional subspace:

Lemma 8.9 ([Joh84])**.** *Let Φ be a d-by-n Gaussian matrix, with each entry independently chosen from $N(0, 1/d)$. Then, for every vector $u \in \mathbb{R}^n$ and every $\varepsilon \in (0,1)$*

$$\mathbb{P}(\|\Phi u\| = (1 \pm \varepsilon)\|u\|) \geqslant 1 - e^{-\Omega(\varepsilon^2 d)} \,.$$

We will follow this strategy to speed up Algorithm 8.7.

Algorithm 8.10 (Approximate matrix multiplicative weights algorithm for SDPs)**.**
Input: A program of the form Eq. (8.2.1) with optimal value α, a ζ-bounded γ-separation ORACLE, parameters T, ε, r, d, p, a d-by-n random matrix Φ with i.i.d entries from $N(0, 1/d)$.
Set $W^{(1)} = \frac{r}{n}(\Phi \mathrm{Id}_n)/\mathrm{Tr}(\Phi \mathrm{Id}_n)$. For $t = 1, \ldots, T$:

1. Run the ORACLE with candidate solution $W^{(t)}$.
2. If the ORACLE outputs **yes**, return $W^{(t)}$.
3. Else, let $(y^{(t)}, F)$ be the pair generated by ORACLE. Set $Y^{(t)} = \left(\sum_{j\in[m]} A_j y_j^{(t)} - F + \zeta \mathrm{Id}_n\right)/2\zeta$.
4. Sample a d-by-n random matrix Φ with i.i.d entries from $N(0, 1/d)$.
5. Compute $W^{(t)} = r \cdot \Phi P_{\leqslant p}(\frac{\varepsilon}{2}\sum_{t'\leqslant t} Y^{(t')})/\mathrm{Tr}\big(\Phi P_{\leqslant p}(\frac{\varepsilon}{2}\sum_{t'\leqslant t} Y^{(t')})\big)$ and continue.

Observe that $P_{\leqslant p}(\frac{\varepsilon}{2}\sum_{t'\leqslant t} Y^{(t')})$ corresponds to a low degree approximation of the Gram vectors of the matrix $X^{(t)}$ in step 4 of Algorithm 8.7. We then compute $W^{(t)}$ by embedding these vectors in a random d-dimensional space.

The statement below shows that in many cases we can compute such matrices $W^{(t)}$ very efficiently.

Lemma 8.11 ([Ste10a])**.** *Suppose we can perform matrix-vector multiplication with the matrices $Y^{(t)}$ in time $\mathcal{T}$. Then, for every t, we can compute $W^{(t)}$ in time $O(t \cdot p \cdot d \cdot \mathcal{T})$.*

A priori it is not clear whether Algorithm 8.10 can provide the same guarantees of Algorithm 8.7. However, the next result show this is the case under reasonable circumstances.

Definition 8.12 (d-robust oracle, extension of [Ste10a])**.** We say that a ζ-bounded γ-separation oracle is d-robust if for every matrix $X \in \Delta(r)$ with $X = W^{\mathsf{T}}W$

$$\mathbb{P}_{\Phi\sim N(0,1/d)^{d\times n}}\left(\text{ORACLE outputs } \mathbf{no} \text{ on input } (\Phi W)^{\mathsf{T}}\Phi W \text{ and } \langle Y^{(t)}, X\rangle \geqslant -\frac{3}{4}\gamma\alpha\right) \leqslant \frac{(\gamma\alpha/\zeta r)^2}{(\log n)^{10}}\,.$$

Lemma 8.13 ([Ste10a])**.** *Consider Algorithm 8.10. Let $\varepsilon \leqslant \gamma\alpha/(2\zeta \cdot r)$, $T \geqslant 2\varepsilon^{-2}\log n$ and $p \geqslant 10\varepsilon^{-1}\log n$. Suppose we have a d-robust ζ-bounded γ-separation ORACLE. If there exists a feasible solution with value at most $\alpha(1 + 2\gamma)$, then ORACLE will output **yes** within T iterations with probability at least $1 - O(\log n)^{-10}$.*

We can combine Lemma 8.13, Lemma 8.11 and Fact 8.15 to obtain a user-friendly statement concerning the running time of Algorithm 8.10. We introduce two additional definitions.

Definition 8.14 ($\mathcal{T}$-lean oracle)**.** We say that a ζ-bounded γ-separation d-robust is $\mathcal{T}$-lean if:

- the oracle compute its outputs in time at most $O(\mathcal{T})$.
- If the oracle outputs **no**, the matrix-vector multiplication between an arbitrary vector and the feedback matrix $\left(\sum_{j\in[m]} A_j y_j - F + \zeta \mathrm{Id}_n\right)/2\zeta$ can be computed in time $O(\mathcal{T})$.

We remark that one can upper bound the time needed for matrix-vector multiplication by the number of non-zero entries in the matrix of interest.

Fact 8.15. *Let $M \in \mathbb{R}^{n\times n}$ be a matrix with m non-zero entries and let $v \in \mathbb{R}^n$. There exists an algorithm that computes Mv in time $O(m+n)$.*

The next definition formalizes the idea of oracles that may find a separating hyperplane only with certain probability.

Definition 8.16 (q-reliable)**.** We say that a ζ-bounded, γ-separation, d-robust, $\mathcal{T}$-lean oracle is q-reliable if the probability (over random bits) that it outputs **no** for any feasible solution with objective value at most $1(+2\gamma)\alpha$ is at most $1-q$.

For oracles that are 1-reliable we omit mentioning their reliability. We are ready to present a user-friendly running time statement, which we will use as a black box.

Corollary 8.17 (Running time of Algorithm 8.10)**.** *Let* ORACLE *be a ζ-bounded, γ-separation, d-robust, $\mathcal{T}$-lean q-reliable oracle. Then, for $\varepsilon \leqslant \gamma\alpha/(2\zeta\cdot r)$, $T \geqslant 2\varepsilon^{-2}\log n$ and $p \geqslant 10\varepsilon^{-1}\log n$, with probability at least $1 - O(\log n)^{-10} - (1-q)T$ over random bits, Algorithm 8.10 terminates in time*

$$O\left(T^2\cdot\mathcal{T}\cdot d\cdot p\right).$$

Proof. The Corollary follows immediately from Lemma 8.11, Lemma 8.13, Definition 8.14 and Definition 8.16. □

8.3 A fast algorithm for semi-random balanced cut

We present here our main theorem which implies Theorem 8.3. To solve the balanced cut problem we consider its basic SDP relaxation. Given a graph G, the relaxation for the

a-balanced cut problem is:

$$\left\{\begin{array}{lll} \min \sum_{ij\in E} c_{ij}\left\|v_i - v_j\right\|^2 & & \\ \|v_i\|^2 = 1 & \forall i \in [n] & \text{(unit norm)} \\ \left\|v_i - v_j\right\|^2 + \left\|v_j - v_k\right\|^2 \geqslant \|v_i - v_k\|^2 & \forall i,j,k, \in [n] & \text{(triangle inequality)} \\ \sum_{i,j\in[n]} \left\|v_i - v_j\right\|^2 \geqslant 4an^2 & & \text{(balance)} \end{array}\right\} \tag{8.3.1}$$

which may be rewritten in its canonical form

$$\left\{\begin{array}{lll} \min\langle L, X\rangle & & \\ X_{ii} = 1 & \forall i \in [n] & \text{(unit norm)} \\ \langle T_p, X\rangle \geqslant 0 & \forall \text{ paths } p \text{ of length 2} & \text{(triangle inequality)} \\ \langle K_V, X\rangle \geqslant 4an^2 & & \text{(balance)} \end{array}\right\} \tag{8.3.2}$$

where L is the combinatorial Laplacian of the graph, K_S is the Laplacian of the full graph over vertex set S and, for a path p, T_p is the difference between the Laplacian of p and the Laplacian of the single edge connecting its endpoints. Notice that $v_1, \ldots, v_n$ are the Gram vectors of X. To ease the reading we will sometimes use the vectors representation and others the matrix representation. As in Section 8.2, we denote by α the optimal value for a given instance of Eq. (8.3.2). In particular, we say a graph G has optimal cut α if minimum solutions to Eq. (8.3.1) have objective value α. Notice that for graphs generated as in Model 8.2, with high probability we have $\alpha \leqslant (1 + o(1))n^2 \cdot \eta$.

Before stating the main theorem we require a couple of definitions concerning the embedding of graphs. These are based on [MMV12].

Definition 8.18 (Heavy vertex)**.** Let $\delta, n, n' > 0$. Let $G(V, E)$ be a graph on n vertices and let X be the Gram matrix of an embedding of G onto $\mathbb{R}^n$. A vertex $i \in V$ is said (δ, n')-heavy if

$$\left|\left\{j \in V(G) \;\middle|\; \left\|v_i - v_j\right\|^2 \leqslant \delta\right\}\right| \geqslant \delta^2 \cdot n' .$$

We denote the set of (δ, n')-heavy vertices in the embedding X by $H_{\delta,n'}(X, V)$. For a subset of vertices V' we let $H_{\delta,n'}(X, V')$ be the set of vertices that are (δ, n') heavy in the subgraph induced by V'.

In other words, a vertex is (δ, n')-heavy if it is close to $\delta^2 n'$ other vertices in the given embedding. The next structural property of graphs is what will separate semirandom instances from worst-case instances.

Definition 8.19 (Geometric expansion)**.** A graph $G(V, E)$ on n vertices satisfies the geometric expansion property at scale (δ, n', α) if, for every feasible solution X to Eq. (8.3.1) on input G and every subset $V' \subseteq V$ such that $H_{\delta,n'}(X, V') = \emptyset$, it holds

$$\left|\left\{ij \in E \cap (V' \times V') \;\middle|\; \left\|v_i - v_j\right\|^2 \leqslant \delta\right\}\right| \leqslant 10 \cdot \delta^2 \cdot \alpha .$$

That is, a graph is geometrically expanding if the uniform distribution over the edges of non-heavy vertices satisfies a one-sided Chebyshev's inequality. For simplicity, we say that a graph G is geometrically expanding up to scale $(100^{-z}, n, \alpha)$ if it is geometrically expanding at scale $(100^{-i}, n, \alpha)$ for all $1 \leqslant i \leqslant z$. Sometimes we say that a graph is

We remark that Definition 8.19 is equivalent to the geometric expansion property defined in [MMV12].

We are now ready to present the main theorem of the section.

Theorem 8.20 (Main theorem)**.** *There exists a randomized algorithm that on input $a, \alpha > \Omega(1)$, κ , $\delta > 1/\log n$, $\Omega(1) \leqslant \rho < 1/\delta$ and a graph G such that:*

1. *there exists an a-balanced partition (A, B) with $|E(A, B)| \leqslant \alpha$,*
2. *$G(V, E(A, B))$ is a geometric expander up to scale $(100\delta, n, \alpha)$,*

returns a $\Omega(a)$-balanced partition (S, T) with cut $|E(S, T)| \leqslant O(\alpha \cdot \rho)(1 + \delta \cdot \kappa \cdot \sqrt{\log n})$ with probability $1 - o(1)$. Moreover, the algorithm runs in time $\tilde{O}\left(|V(G)|^{1+O(1/\rho^2)+O(1/\kappa^2)} + |E(G)|^{1+O(1/\kappa^2)}\right)$.

Theorem Theorem 8.3 essentially follows from Theorem 8.20 observing that graphs generated through Model 8.2 are good geometric expanders. To show this first observe that random bipartite graphs are good geometric expanders.

Theorem 8.21 (Geometric expansion of random graphs, [MMV12])**.** *Let $\mathbf{G}$ be a graph over n vertices generated through the first two steps (i), (ii) of Model 8.2 with parameters $a, \eta > 0$. Then, with probability $1 - n^{-\Omega(1)}$, $\mathbf{G}$ is geometrically expanding up to scale $\left(100^{-t}, n, \Theta(n^2 \cdot \eta + 100^t \cdot n \cdot t^2)\right)$.*

Second, observe that geometric expansion in bipartite graphs is a property that is to some extend robust to changes that are monotone with respect to the bipartition.

Fact 8.22 (Robustness of geometric expansion, [MMV12])**.** *Let $\mathbf{G}$ be a graph over n vertices generated through the first two steps (i), (ii) of Model 8.2 and let G° be a graph obtained after applying steps (iii), (iv). If $\mathbf{G}$ is geometrically expanding up to scale (δ, n, τ), then so is $G^\circ(V, E(A, B))$.*

This statement above implies that for $\eta \geqslant \Omega\left(\frac{(\log n)^2 \cdot (\log\log n)^2}{n}\right)$, with high probaility $G^\circ(V, E(A, B))$ is a good geometric expander. Now Theorem 8.3 immediately follows combining Theorem 8.20 Theorem 8.21 and Fact 8.22.

8.3.1 The algorithm

We present here the algorithm behind Theorem 8.20. Since we will work using the matrix multiplicative frame work presented in Section 8.2 our main challenge is that of designing an appropriate oracle. For simplicity, we split ORACLE in three parts the first two are due to [AK07, She09] the third part is our crucial addition. Recall we denote by α the minimum objective of the program at hand.

Lemma 8.23 ([AK07]). *Let $a > \Omega(1)$. There exists a $\tilde{O}(\alpha/n)$-bounded, $\Theta(\log n)^2$-robust, $\Theta(1)$-separation oracle that, given a candidate solution to Eq. (8.3.1) with input graph G on n vertices and a-balanced cut of value at most α, outputs **no** if one of the following conditions is violated:*

(flatness): $W := \left\{ i \in [n] \;\middle|\; \|v_i\|^2 > 2 \right\} \subseteq [n]$ *satisfies* $|W| < \frac{n}{(\log n)^{100}}$.

(balance): $S := [n] \setminus W$ *satisfies* $\sum_{i,j \in S} \left\|v_i - v_j\right\|^2 \geqslant 2an$.

Moreover, the oracle is $\mathcal{T}$-lean for some $\mathcal{T} \leqslant O((|V(G)| + |E(G)|)$

We omit the proof of Lemma 8.23 as it can be found in [AK07]. If both the flatness and the balance condition are satisfied, then we apply the following oracle, due to [She09].

Lemma 8.24 ([She09]). *Let $\kappa, a > 0, 0 < \delta < 1/(100\rho), a > \Omega(1)$. There exists a $\tilde{O}(\alpha/n)$-bounded, $O(\log n)^2$-robust, $\Theta(1)$-separation oracle that, given a candidate solution to Eq. (8.4.2) with input graph G on n vertices and a-balanced cut of value at most α, outputs **yes** only if it finds an $\Omega(a)$-balanced partition (P, P') of $V(G)$ satisfying*

$$\sum_{i \in P, j \in P', ij \in E(G)} \left\|v_i - v_j\right\|^2 \leqslant O(\alpha)$$

$$|E(P, P')| \leqslant O(\alpha \cdot \kappa) \cdot \sqrt{\log n}\,.$$

Moreover, the oracle is $\mathcal{T}$-lean for some $\mathcal{T} \leqslant \tilde{O}\left(|V(G)|^{1+O(1/\kappa^2)} + |E(G)|^{1+O(1/\kappa^2)}\right)$.

The proof of Lemma 8.24 can be found in [She09].The improvement on the time complexity simply follows using Theorem 8.6. The next result is the *crucial* addition we need to the oracle of [AK07, She09]. We prove it in Section 8.4.1.

Lemma 8.25. *Let $0 < \ell \leqslant 1, \rho, \alpha, a > 0$ and $0 < \delta \leqslant 1/100\rho$. Let G be a graph on $\ell \cdot n$ vertices such that*

- *it has a a-balanced partition (A, B) with $|E(A, B)| \leqslant \alpha$,*
- *$G(V, E(A, B))$ is geometrically expanding up to scale (δ, n, α).*

*There exists a $\tilde{O}(\alpha/\ell \cdot n)$-bounded, $O(\log n)^{100}$-robust, $\Theta(1/\log n)$-separation oracle that, given a candidate solution to Eq. (8.4.2) with input graph G, either outputs **no**, or outputs a set of edges $E^* \subseteq E(G)$ of cardinality $O(\alpha/\delta)$ and partition (P_1, P_2, V') of $V(G)$ such that*

(1) $|E(P_1, P_2, V') \setminus E^*| \leqslant O\left(\frac{\alpha}{\rho\delta}\left(1 + \frac{\ell}{\rho\delta}\right)\right)$.

(2) $||P_1| - |P_2|| \leqslant a \cdot n/2$.

(3) $\forall ij \in E(G) \setminus E^*$ *with* $i, j \in V'$ *it holds* $\left\|v_i - v_j\right\|^2 \leqslant \delta$.

(4) $H_{\delta,n}(X, V') = \emptyset$.

Moreover the oracle is $\mathcal{T}$-lean for some $\mathcal{T} \leqslant \tilde{O}\left(|V(G)|^{1+O(1/\rho^2)} + |E(G)|\right)$ and $1 - O(\log n)^{-50}$-reliable.

Before presenting the algorithm that uses the oracle above, let's briefly discuss its meaning. Notice the *heavy vertices* condition (4). This ensures that in the subgraph $G(V', E \setminus E^*)$ any feasible embedding has weight at most $10\alpha \cdot \delta^2$ on the edges in $(E(A, B) \setminus E^*) \cap (V' \times V')$. In other words, after paying the edges in the cut of the partition (P_1, P_2, V'), geometric expansion of the undelrying graph guarantees that the minimum objective value of Eq. (8.3.1) has now decrease by a $10\delta^2$ factor. As in Lemma 8.24, we point out the trade-off between running time and cardinality of the set of edges E^* found by the oracle which may ultimately be part of the cut. This is capture by the requirement $\rho < 1/\delta$. We are ready to present the algorithm we denote by ORACLE a combination of the oracles in Lemma 8.23, Lemma 8.24 and Lemma 8.25 obtained applying them sequentially (in this order).

Algorithm 8.26 (Fast and robust algorithm for balanced cut)**.**
Input: A graph G with minimum a-balanced cut of value at most $\alpha, T, \rho, \kappa, d, \delta > 0$.

0. Set $\delta^{(0)} = 1/(100\rho)$.
1. Repeat T times:
   (a) Let i be the current iteration and $G^{(i)}$ be the current remaining graph with optimal cut value $\alpha^{(i)}$ and $\left|V(G^{(i)})\right| =: n^{(i)}$.
   (b) Run Algorithm 8.10 for program 8.3.2 using ORACLE (with parameter $\delta^{(i)}$). Let $W^* \in \mathbb{R}^{d,\times n}$ be the returned embedding, $E^{(i)}$ the set of edges found and $(P^{(i)}, P'^{(i)})$, $(P_1^{(i)}, P_2^{(i)}, V'^{(i)})$ the partitions found by ORACLE.
   (c) Remove the edges in $E^{(i)}$.
   (d) If $\left|E(P^{(i)}, P'^{(i)})\right| \leqslant O(\alpha \cdot \rho)(1 + \delta \cdot \kappa \cdot \sqrt{\log n})$ break the cycle. Else remove vertices in $P_1^{(i)}$ and in $P_2^{(i)}$. Set $\delta^{(i+1)}$ to $\delta^{(i)}/100$.
2. Let (P, P') be the bipartition found by ORACLE in its last iteration.
3. Arbitrarily assign sets $P_1^{(1)}, P_2^{(1)}, \ldots, P_1^{(i)}, P_2^{(i)}$ removed in previous iterations to P or P', keeping the two sides a-balanced.
4. Return the resulting bipartition.

Remark 8.27 (On practical implementations of the algorithm). The following precautions should be taken into account upon implementing the algorithm. First, replace the subroutine of [She09] with a simple maximum-flow computation as in [AK07]. Note that this different

subroutine provides worse approximation guarantees by a $O(\sqrt{\log n})$ factor. Nevertheless, the algorithms still provides the same error guarantees if one increases proportionally the number of iterations of the main procedure. Second, replace the maximum flow algorithm of Theorem 8.6 with a more practical –even if asymptotically slower– maximum flow algorithm. We remark that with these modifications, the algorithm amounts to a sequence of basic computations such as: matrix-vector multiplications, hash-functions and max-flow computations.

With Algorithm 8.26 we can now prove Theorem 8.20.

Proof of Theorem 8.20. We set T such that $\delta^{(T)} = 100\delta$ where $\delta^{(0)} = 1/(100\rho)$. By construction of ORACLE and Corollary 8.17 the algorithm runs in time $\tilde{O}\left(|V(G)|^{1+O(1/\kappa^2)} + |E(G)|^{1+O(1/\kappa^2)}\right)$.

Now consider a fixed iteration i, we assume $\alpha^{(0)} = \alpha$, $\ell^{(0)} = 1$. Let $\alpha^{(i)}$ be the cost of the minimum feasible solution on the remaining graph $G^{(i)}$ on $\ell^{(i)} \cdot n$ vertices. Let $E^{(i)}$ be the set of edges removed at iteration i and $(P_1^{(i)}, P_2^{(i)}, V'^{(i)})$ the partition at iteration i. Notice that if at some point $\alpha^{(i)} \leqslant O(\alpha \cdot \rho)(1 + \delta \cdot \kappa \cdot \sqrt{\log n})$ then the algorithm breaks the cycle and returns a balanced partition.

So we may assume that at the current iteration i, $\alpha^{(i)} \geqslant \omega(\alpha \cdot \rho)(1 + \delta \cdot \kappa \cdot \sqrt{\log n})$. Now the result follows by showing that, at each step, it holds

$$\alpha^{(i)} \leqslant 10 \cdot \alpha^{(i-1)} \cdot \left(\delta^{(i-1)}\right)^2 . \tag{8.3.3}$$

Indeed suppose the claim holds. By construction all the edges in the final cut are in

$$\left(\bigcup_{i<T} E^{(i)}\right) \cup \left(\bigcup_{i<T} E(P_1^{(i)}, P_2^{(i)}, V'^{(i)}) \setminus E^{(i)}\right) \cup E(P^{(T)}, P'^{(T)}) .$$

By Eq. (8.3.3) we can bound the first term as

$$\left|\bigcup_{i\leqslant T} E^{(i)}\right| \leqslant O(\alpha \cdot \rho) .$$

For the second term we have

$$\begin{aligned}
\left|\bigcup_{i\leqslant T} E(P_1^{(i)}, P_2^{(i)}, V'^{(i)})\right| &\leqslant \sum_{i\leqslant T} O\left(\frac{\alpha^{(i)}}{\delta^{(i)}}\left(1 + \frac{\ell^{(i)}}{\delta^{(i)}}\right)\right) \\
&\leqslant O\left(\sum_{i\leqslant T} \frac{\alpha^{(i)}}{\delta^{(i)}} + \frac{\alpha^{(i)} \cdot \ell^{(i)}}{\left(\delta^{(i)}\right)^2}\right) \\
&\leqslant O\left(\sum_{i\leqslant T} \alpha^{(i)} \cdot \left(\delta^{(i)} + \ell^{(i+1)}\right)\right)
\end{aligned}$$

$$\begin{aligned}
&\leqslant O\left(\sum_{i\leqslant T} \alpha^{(i)} \cdot \left(\delta^{(i)} + \ell^{(i)}\right)\right) \\
&\leqslant \alpha \cdot O\left(\sum_{i\leqslant T} \delta^{(i)} + \ell^{(i)}\right) \\
&\leqslant O(\alpha)\,,
\end{aligned}$$

where in the second step we used the inequalities

$$\begin{aligned}
\frac{\alpha^{(i)}}{\delta^{(i)}} &\leqslant 10^3 \cdot \alpha^{(i-1)} \cdot \delta^{(i-1)}\,, \\
\frac{\alpha^{(i)} \cdot \ell^{(i)}}{\left(\delta^{(i)}\right)^2} &\leqslant 10^5 \cdot \alpha^{(i-1)} \cdot \ell^{(i)}\,,
\end{aligned}$$

both following from Eq. (8.3.3). For the third term we have $\alpha^{(T)} \leqslant O(\alpha \cdot \delta)$ by construction. Thus by Lemma 8.24 we get

$$\left|E(P^{(T)}, P'^{(T)})\right| \leqslant O\left(\alpha \cdot \kappa \cdot \delta\sqrt{\log n}\right).$$

Thus it remains to prove Eq. (8.3.3). At each iteration i, the set $V(G^{(i)})$ does not contain edges of length more than $\delta^{(i)}$ in the embedding as well as $(\delta^{(i)}, n)$ heavy vertices. Thus by Definition 8.19, the set $V(G^{(i)})$ has a $\Omega(a)$-balanced cut of value $O(\alpha^{(i)})$. Then Eq. (8.3.3) follows as desired. □

8.4 The heavy vertices removal oracle

We prove here Lemma 8.25. In Section 8.4.1 we introduce a procedure that the oracle uses to find either the partition or a separating hyperplane. Then in Section 8.4.2 we prove the Lemma. Throughout the section we consider the following parameters range:

$$n, \alpha > 0\,, \Omega(1) \leqslant a \leqslant 1\,, 0 < \ell \leqslant 1\,, \Omega(1/\sqrt{\log n}) \leqslant \delta \leqslant 1/100\rho\,, \rho > \Omega(1)\,. \tag{8.4.1}$$

8.4.1 The fast heavy vertices removal procedure

We introduce the main procedure used by ORACLE. The central tool of the section is the following statement.

Lemma 8.28. *Consider the parameter settings of Eq. (8.4.1). Let G be a graph on $\ell \cdot n$ vertices with a-balanced cut of value at most α that is geometrically expanding up to scale (δ, n, α).*

Let X be a feasible solution for Eq. (8.3.1) on input G, with objective value $O(\alpha)$. There exists a randomized procedure (Algorithm 8.31) that outputs a set of vertices $E^* \subseteq E(G)$ of cardinality $O(\alpha/\delta)$ and a partition (P_1, P_2, V') of $V(G)$ satisfying (2) , (3) , (4) in Lemma 8.25 and such that

$$\mathbb{E}[|E(P_1, P_2, V') \setminus E^*|] \leqslant C \cdot \frac{\alpha}{\rho\delta}\left(1 + \frac{\ell}{\rho\delta}\right),$$

where $C > 0$ is a universal constant. Moreover, if the solution is given in the form of $v_1, \ldots, v_n \in \mathbb{R}^{O(\text{polylog}\, n)}$, the procedure runs in time $\tilde{O}(|V(G)|^{1+O(1/\rho^2)} + |E(G)|)$.

The first building block towards a proof of Lemma 8.28 is the result below, which introduces a subroutine to identify heavy vertices.

Lemma 8.29. *Consider the settings of Lemma 8.28. Let $\rho > 1$. There exists a randomized procedure that outputs a set of vertices V^* and a mapping $f : V \to V^* \cup \{(*)\}$ such that*

1. *Each vertex i of V^* satisfies $\left|\left\{j \in V \;\middle|\; \left\|v_i - v_j\right\|^2 \leqslant \rho\delta\right\}\right| \geqslant 10\delta^2 n$; and*
2. *The set $W := \{i \mid f(i) = (*)\}$ does not contain a vertex i such that $\left|\left\{j \in W \;\middle|\; \left\|v_i - v_j\right\|^2 \leqslant \delta\right\}\right| \geqslant 10\delta^2 n$.*

and

1. *$f(i) = j$ if there exists some $j \in V^*$ with $\left\|v_i - v_j\right\|^2 \leqslant \rho\delta$,*
2. *$f(i) = (*)$ otherwise.*

Moreover, if the solution is given in the form of $v_1, \ldots, v_n \in \mathbb{R}^{O(\text{polylog}\, n)}$, the procedure runs in time $O\left(\frac{1}{\delta^2} \cdot |V(G)|^{1+O(1/\rho^2)}\right)$.

Proof. We will make use of locality sensitive hash functions to solve the *spherical range counting problem*, namely the problem of, given a set of point $v_1, \ldots, v_n \in \mathbb{R}^{O(\text{polylog}\, n)}$ and a distance δ, estimating for each element v_i, the number of elements of $v_1, \ldots, v_n$ at distance at most δ from v_i. For each point v_i and distance x, let $w(v_i, x)$ be the number of points at distance at most x from v_i.

We provide a procedure with running time $\frac{1}{10\delta^2} \cdot n^{1+1/\rho^2+o(1)}$ that constructs the desired V^* and f by iteratively finding "heavy balls" and removing them from the instance. For each point v_i, the procedure computes an estimate $\hat{w}_i$ such that $w(v_i, x) \leqslant \hat{w}_i \leqslant w(v_i, \rho x)$ and take an arbitrary vertex i such that $\hat{w}_i \geqslant 10\delta^2 n$ if there is one, or otherwise stops and set $f(j) = (*)$ for all the remaining vertices. If there exists such a vertex i, then by a linear scan of the vertices, the procedure identifies the set S_i of all the vertices j such that $\left\|v_i - v_j\right\|^2 \leqslant \rho\delta$, sets $f(j) = i$ for each of them, and removes S_i from the instance. Then, the algorithm repeats the above procedure on the remaining vertices. Since the procedure removes at least $10\delta^2 n$ vertices at each step, the procedure stops after $\frac{1}{10\delta^2}$ steps.

We now provide more details on how to perform each step. We show that each step takes time $n^{1+1/\rho^2+o(1)}$. Defining $f(j)$ for all the vertices in S_i takes linear time. The bulk of the computation comes from computing the estimate $\hat{w}_i$ for each vertex i. We now provide more details on the spherical range counting problem. Let H be a family of hash functions mapping $\mathbb{R}^d$ to some universe U. We say that H is $(\eta, \rho\eta, p_1, p_2)$-sensitive if for any $x, y \in \mathbb{R}^d$ it satisfies the following properties:

1. If $\left\|x - y\right\|^2 \leqslant \eta$ then $\Pr_{h\in H}[h(x) = h(y)] \geqslant p_1$.
2. If $\left\|x - y\right\|^2 \geqslant \rho\eta$ then $\Pr_{h\in H}[h(x) = h(y)] \leqslant p_2$.

We will use the following lemma.

Lemma 8.30 ([AI06])**.** *For any "scale" $\eta > 0$, dimension $d > 0$, and $\rho > 1$, there exists a $(\eta, O(\rho\eta), 1/n^{1/c^2}, 1/n^3)$-sensitive family of hash functions for $\mathbb{R}^d$.*

Following the reduction from the spherical range counting problem to (ρ, r)-approximate nearest neighbors provided by Indyk [Ind01] (see also [AAP17]). The reduction makes $O(\log^2 n/(\rho - 1)^3)$ queries, which, using the above family take time $O(n^{1/\rho^2})$, resulting in a total running time of $n^{1+1/\rho^2+o(1)}$. □

We use the procedure in Lemma 8.29 as a subroutine of the one presented next, which for feasible embeddings finds a paritition satisfying (2) , (3) , (4) in Lemma 8.25 and (1) in expectation.

Algorithm 8.31 (Fast heavy vertex removal procedure)**.**
Input: A graph G on $\ell \cdot n$ vertices, a candidate solution X to Eq. (8.3.1) on input G, parameters $a, \rho, \delta > 0, C > 100$.

1. Remove all edges of length at least δ in the embedding. Let E^* be the set of such edges.
2. Find the set V^* via the subroutine in Lemma 8.29.
3. Pick a maximal set U of vertices in V^* at pairwise distance at least $10 \cdot C\rho\delta$ in the embedding.
4. If $|U| \geqslant \frac{a}{C \cdot \rho\delta}$:
   (a) Pick $r \overset{u.a.r.}{\sim} [1, 2]$.
   (b) For each $i \in U$, remove i and all vertices at distance $\leqslant r \cdot \rho\delta$ in the embedding. Let U_i be the set of removed vertices via i.
   (c) *Repeat* from step 1 on the remaining graph.
5. Else run the subroutine Algorithm 8.33 on the remaining graph and obtain additional sets U_i's.
6. Distribute evenly the vertices in the U_i's among two sets P_1, P_2 so that if $j, k \in U_i$ then j, k are in the same set. Let $V' = V \setminus (P_1 \cup P_2)$.
7. Return the partition (P_1, P_2, V').

Fact 8.32. *Algorithm 8.31 runs in time* $\tilde{O}\left(|V(G)|^{1+O(1/\rho^2)} + |E(G)|\right)$.

Proof. Step 1 requires $O(E)$ time. The steps 2-4 can be repeated at most $C \cdot \rho \cdot \ell/(a \cdot \delta)$ times. Indeed no vertex can be in both U_i and $U_{i'}$ at the same time (even if X does not satisfy the triangle inequality constraints) and since by definition each U_i contains at least $10\delta^2 \cdot n$ vertices, in $C \cdot \rho \cdot \ell/(a \cdot \delta)$ iterations we will have removed all vertices form the graph. For each of these iterations, step 2 requires time $O\left(\frac{1}{\delta^2}|V(G)|^{1+O(1/\rho^2)}\right)$ and step 3 requires time $O(|V(G)| \cdot \text{poly}(1/(a\rho\delta)))$. Step 4 runs in time at most $O(|V(G)|/(a\rho\delta))$.

As we show in Fact 8.34, Algorithm 8.33 also runs in time $\tilde{O}(|V(G)| + |E(G)|)$. Step 6 can be done in time $\tilde{O}(|V(G)|)$ after ordering the sets U_i's. Thus the statement follows. □

The subroutine of step 5 in Algorithm 8.31 is presented below.

Algorithm 8.33 (Subroutine of fast heavy vertex removal procedure)**.**
Input: A graph G on $\ell \cdot n$ vertices, a candidate solution X to Eq. (8.3.1) on input G, the list of vertices V^*, parameters $a, \rho, \delta > 0$.

0. Consider the graph $G^*(V^*, \emptyset)$.
1. For each $ij \in E(G)$ if $\|v_i - v_j\|^2 \leqslant 100\rho\delta$ and $f(i) \neq (*), f(i) \neq (*)$ connect $f(i)$ to $f(j)$ in G^* (excluding self-loops).
2. Pick $r \overset{u.a.r.}{\sim} [1,2]$ and for each connected component U in G^*, remove all vertices at distance $\leqslant r \cdot \rho\delta$ to some vertex $i \in U$. Index the resulting sets by arbitrary representative vertices in each component.
3. Return the resulting sets U_i's.

Fact 8.34. *Algorithm 8.33 runs in time* $\tilde{O}(|V(G)| + |E(G)|)$.

Proof. We use the mapping of Lemma 8.29. We can then construct the graph G^* in time $O(|E(G)|)$. Moreover, notice that $|E(G^*)| \leqslant |E(G)|$. We can find the connected components in G^* in time $O(|E(G^*)| + |V(G^*)|)$ and partition the vertices in G according to such connected components in time $\tilde{O}(|V(G)|)$. The result follows. □

Next we bound the probability that an edge gets cut in Algorithm 8.31.

Lemma 8.35. *Consider the settings of Lemma 8.28.*

At each iteration of steps 2-4 in Algorithm 8.31 as well as the one using Algorithm 8.33 the following holds:

$$\forall i \ s.t. \ f_i \neq (*) \quad \mathbb{P}\left(\exists U_k \in U \ s. .t \ i \in U_k, j \notin U_k\right) \leqslant \frac{\|v_i - v_j\|^2}{\rho \cdot \delta}.$$

Proof. Consider first an iteration of steps 1-3 in Algorithm 8.31. By construction each vertex i can be in at most one set U_k. Since r is chosen uniformly at random in the interval $[1,2]$ the claim follows. So consider Algorithm 8.33. Again, by construction each vertex i can be in at most one set U_k so by choice of r the inequality holds. □

Now Lemma 8.28 follows as a simple corollary.

Proof of Lemma 8.28. By Fact 8.32, steps 1-3 in Algorithm 8.31 are repeated at most $C \cdot \ell/(\rho \cdot \delta)$ times while Algorithm 8.33 runs only once. By Lemma 8.35 we then have $\forall ij \in E(G) \setminus E^*$

$$\mathbb{P}\left(ij \in E(P_1, P_2, V') \setminus E^*\right) \leqslant C^* \frac{\|v_i - v_j\|^2}{\rho \cdot \delta} \cdot \left(1 + C \cdot \ell/(\rho \cdot \delta)\right).$$

Then the bound on $\mathbb{E}[|E(P_1, P_2, V') \setminus E^*|]$ follows by linearity of expectation. By definition of V^*, the set V' does not contain (δ, n)-heavy vertices as well as no edges of length at least δ. so it satisfies conditions (3), (4) in Lemma 8.25 . Finally condition (2) follows by the spreadness condition of feasible solutions. □

8.4.2 The oracle

We prove here Lemma 8.25. To simplify the description of the oracle, as in [AK07], we consider the following modification of Eq. (8.3.2), which contains additional constraints. The two programs are equivalent as these constraints are *implied* by the ones in Eq. (8.3.2).

$$\left\{\begin{array}{lll} \min\langle L, X\rangle & & \\ \quad X_{ii} = 1 & \forall i \in [n] & \text{(unit norm)} \\ \quad \langle T_p, X\rangle \geqslant 0 & \forall \text{paths } p & \text{(triangle inequality)} \\ \quad \langle K, X\rangle \geqslant 4an^2 & & \text{(balance)} \end{array}\right\} \tag{8.4.2}$$

Remark 8.36. We remark that, as we need not to explicitly write down the program, but only efficiently find separating hyperplanes, the use of Eq. (8.4.2) does not imply an increase in the running time of the overall algorithm.

We consider the dual program of Eq. (8.4.2), which has variables $x_1, \ldots, x_n$ for each vertex, f_p for every path p and an additional variable z for the set $[n]$ considered in the primal.

$$\left\{\begin{array}{ll} \max \sum_{i\in[n]} x_i + an^2 z & \\ \quad \text{diag}(x) + \sum_p f_p T_p + zK \preceq L & \\ \quad f_p, z, \geqslant 0 & \forall \text{paths } p, \end{array}\right\} \tag{8.4.3}$$

Now, given a candidate solution X, our starting point is the procedure of Lemma 8.28, which we use to remove vertices that are heavy in the current embedding.

Proof of Lemma 8.25. Throughout the proof, whenever we write a feedback matrix, all the variables that are not specified are set to 0. Let X be the matrix denoting the current embedding. We may assume without loss of generality that both the oracles in Lemma 8.23 and Lemma 8.24 outputted **yes** on X. We claim there are at most $\bar{C} \cdot \alpha/\delta$ edges of length at least δ in the embedding for some large enough constant $\bar{C} > 0$. Suppose this is not the case, consider the following procedure:

- Randomly partition the vertices into two a balanced sets A, B
- Compute the max d-regular A-B flow with $d = O(\alpha)/n$.

In expectation the flow is larger than $\bar{C}\alpha$ (if it is smaller we have found a cut). If the flow is larger than $\bar{C}\alpha$ then let F be the Laplacian of the flow graph and let D be the Laplacian of the complete weighted graph where only edges ij with $i \in A$, $j \in B$ have weight f_{ij}, and the rest have 0 weight . Then by definition $\sum_p f_p T_p = F - D$. Thus we set $x_i = \alpha/|V(G)|$ for all $i \in V(G)$, f_p as in the computed flow for all p and all other variables to 0. The feedback matrix Y becomes $\frac{\alpha}{|V(G)|}\mathrm{Id} + F - D - F = \frac{\alpha}{|V(G)|}\mathrm{Id} - D$ and we have

$$\langle \frac{\alpha}{|V(G)|}\mathrm{Id} - D, X\rangle \leqslant \alpha - (\bar{C} - 1)\cdot\alpha < -\alpha < 0\,.$$

Notice also that $\left\|\frac{\alpha}{|V(G)|}\mathrm{Id} - D\right\| \leqslant O(\alpha/|V(G)|)$. We repeat this procedure $O(\log n)^{100}$ times, by Markov's inequality the claim follows with probability at least $1 - O(\log n)^{-99}$. Let $E^* \subseteq E(G)$ be the set of edges of length at least $O(\alpha/\delta)$ in the embedding.

Now we run the heavy vertex removal procedure Algorithm 8.31. Let (P_1, P_2, V') be the resulting partition. If the partition satisfies (1), (2), (3), (4) in Lemma 8.25 the result follows. Else, since X approximately satisfies the spreadness constraints, there are no edges longer than δ in $E(G) \setminus E^*$ and by construction V' does not contain (δ, n) heavy vertices, it must be that

$$\left|E(P_1, P_2, V') \cap \left\{ij \in E(G) \;\middle|\; \|v_i - v_j\|^2 \leqslant \delta\right\}\right| > C^* \cdot \frac{\alpha}{\rho\delta} \cdot \left(1 + \frac{\ell}{\rho\delta}\right), \tag{8.4.4}$$

for some large enough constant $C^* > 10^{10}C$, where $C > 0$ is the universal constant of Lemma 8.28.

Let $d = 100 \cdot C^* \cdot \left(\frac{\alpha}{|V(G)|\cdot\delta\rho}\left(1 + \frac{\ell}{\delta\rho}\right)\right)$. We compute the maximum d-regular flows for each of the partitions $(P_1 \cup P_2\,, V')$, $(P_1\,, P_2 \cup V')$, $(P_2\,, P_1 \cup V')$ as described in Section 8.2 using the algorithm in Theorem 8.6. By Eq. (8.4.4) at least one of these cuts has flow $\frac{C^*}{3} \cdot \frac{\alpha}{\rho\delta} \cdot \left(1 + \frac{\ell}{\rho\delta}\right)$ as otherwise by duality we have found a $(a/2)$-balanced cut of value at most $C^* \cdot \frac{\alpha}{\rho\delta} \cdot \left(1 + \frac{\ell}{\rho\delta}\right) + |E^*| \leqslant O\left(\frac{\alpha}{\rho\delta}\left(1 + \frac{\ell}{\rho\delta}\right)\right)$ as desired. Without loss of generality we may always assume this is the partition $(P_1 \cup P_2\,, V')$. We distinguish two cases:

1. $\sum_{ij \in E(P_1 \cup P_2\,, V') \setminus E^*} f_{ij}\|v_i - v_j\|^2 \geqslant \frac{C^*}{10^9}\left(\alpha\left(1 + \frac{\ell}{\rho\delta}\right)\right)$,
2. $\sum_{ij \in E(P_1 \cup P_2\,, V') \setminus E^*} f_{ij}\|v_i - v_j\|^2 < \frac{C^*}{10^9}\left(\alpha\left(1 + \frac{\ell}{\rho\delta}\right)\right)$.

Suppose we are in case 1. Let F to be the Laplacian of the weighted graph corresponding to the flow and let D be the Laplacian of the complete weighted graph where only edges ij with $i \in P_1 \cup P_2$ and $j \in V'$ have weight f_{ij}, and the rest have 0 weight. Since we are in case 1 we have

$$\langle D, X\rangle \geqslant \frac{C}{10^9}\left(\alpha\left(1 + \frac{\ell}{\rho\delta}\right)\right)$$

Moreover, by definition $\sum_p f_p T_p = F - D$. Thus we set $x_i = \alpha/|V(G)|$ for all $i \in V(G)$, f_p as in the computed flow for all p and all other variables to 0. The feedback matrix Y becomes $\frac{\alpha}{|V(G)|}\mathrm{Id} + F - D - F = \frac{\alpha}{|V(G)|}\mathrm{Id} - D$ and we have

$$\langle \frac{\alpha}{|V(G)|}\mathrm{Id} - D, X\rangle \leqslant \alpha - \frac{C}{10^9}\cdot\alpha\left(1+\frac{\ell}{\rho\delta}\right) < -\alpha < 0\,.$$

Moreover notice that $\left\|\frac{\alpha}{|V(G)|}\mathrm{Id} - D\right\| \leqslant O\left(\frac{\alpha}{|V(G)|}\right) + d \leqslant \tilde{O}\left(\frac{\alpha}{|V(G)|}\right)$ where in the last step we used the inequalities $\rho \geqslant \Omega(1)\,, \delta \geqslant \Omega(1/\log n)$. In conclusion, in this case the ORACLE finds a separating hyperplane and outputs **no**. Notice also that by construction D has at most $O(m+n)$ non zero entries so the feedback matrix can be computed in time $O(m+n)$.

It remains to consider case 2. By Lemma 8.28 and Markov's inequality, we know that for any feasible solution X^* to Eq. (8.3.1) with objective value at most α, it holds with probability at least $1/2$:

$$\mathbb{E}\left[\left|E(P_1,P_2,V')\cap\left\{ij\in E(G)\;\middle|\;\left\|v_i-v_j\right\|^2\leqslant\rho\delta\right\}\right|\right] \leqslant C\cdot\frac{\alpha}{\delta}\cdot\left(1+\frac{\ell}{\rho\delta}\right),$$

where $C < C^*/10^{10}$. Thus repeating the procedure $(\log n)^{100}$ times, we get that, for any feasible solution X^* with objective value α, with probability at least $1 - O(\log n)^{-100}$, for at least one of the resulting partitions $(P_1, P_2\,, V')$

$$\left|E(P_1,P_2\,,V')\cap\left\{ij\in E(G)\;\middle|\;\left\|v_i-v_j\right\|^2\leqslant\rho\delta\right\}\right| \leqslant C\cdot\frac{\alpha}{\delta}\cdot\left(1+\frac{\ell}{\rho\delta}\right).$$

Now consider again our candidate solution X satisfying Eq. (8.4.4). If after $(\log n)^{100}$ trials we still satisfy Eq. (8.4.4) and are always in case 2, then with probability $1 - O(\log n)^{-100}$ there exist at least $\frac{C^*}{10}\cdot\frac{\alpha}{\rho\delta}\cdot\left(1+\frac{\alpha}{\rho\delta}\right)$ edges of length at most $\rho\delta/10^8$ crossing the cut $E(P_1\cup P_2\,,V')\setminus E^*$.

By design of Algorithm 8.31, then it must be the case that there exists a set of size $\Omega(n)$ of triplets $\{i,j,k\}\subseteq C$ with $ij\in E(G)$ and $k\in V^*$ such that $\left\|v_i-v_j\right\|^2 \leqslant \delta\rho/10^8$, $\left\|v_j-v_k\right\|^2 \leqslant \|v_i-v_k\|^2$ but

$$\mathbb{P}_{r\overset{u.a.r.}{\sim}[1,2]}\left(\left\|v_k-v_j\right\|^2\leqslant\rho\delta(1+r)\text{ and }\|v_k-v_i\|^2>\rho\delta(1+r)\right)\geqslant 10^4\cdot\frac{\left\|v_i-v_j\right\|^2}{\rho\delta}\,.$$

Indeed if this scenario does not apply then we would have seen a partition violating Eq. (8.4.4) with probability at least $1 - O(\log n)^{-100}$ by the argument used in the proof of Lemma 8.35.

So suppose this scenario applies, and consider such a triplet $\{i,j,k\}$. Then we must have

$$\|v_i-v_k\|^2 \geqslant \left\|v_j-v_k\right\|^2 + 10^4\cdot\left\|v_i-v_j\right\|^2 \tag{8.4.5}$$

so we are violating the triangle inequality. Furthermore, we know that the sum over each such triplets must satisfy

$$\sum_{\{i,j,k\} \text{ satisfying Eq. (8.4.5)}} \|v_i - v_j\|^2 \geqslant \Omega\left(\frac{\alpha}{\rho\delta}\left(1 + \frac{\ell}{\rho\delta}\right)\right).$$

as otherwise with probability $1 - O(\log n)^{-100}$ we would have found a cut violating Eq. (8.4.4). Notice now that we can find such triangle inequalities in linear time by looking at the edges being cut and the vertices being picked at each iteration of Algorithm 8.31.

Thus set $x_i = \alpha/|V(G)|$ for all $i \in V(G)$ and $f_p = \frac{C^*\alpha}{n}$ for $\Theta(n)$ such violated triangle inequalities and a large enough constant $C^* > 0$. We set $F = \mathbf{0}$ and

$$\langle \frac{\alpha}{|V(G)|}\mathrm{Id} + \sum f_p T_p, X\rangle \leqslant \alpha - O\left(\frac{\alpha}{\rho\delta}\right) \leqslant -\alpha < 0\,,$$

where in the last step we used the assumption $\delta\rho < 1$. The width of the feedback matrix is at most $O(\alpha/|V(G)|)$ and it has $O(m+n)$ entries, thus it can be computed in $O(m+n)$ time.

Finally we remark that choosing $d = O(\log n)^{100}$ the oracle is d-robust by Lemma 8.9.

□

8.5 The semi-random hierarchical stochastic model

In this section we consider the semi-random hierarchical stochastic model (HSM) from [CKMM19] and develop a nearly linear time algorithm that estimates the Dasgupta's cost of the underlying hierarchical clustering model upto constant factor. The main idea is to recursively compute an $O(1)$-approximation to Balanced Cut which produces a graph with $O(1)$-approximation to the Dasgupta's cost [Das16]. Essentially most of this section is directly cited from [CKMM19] and we only provide it for the completeness. However, note that using Theorem 8.3 we can improve the running time of the algorithm to nearly linear time. In the following subsection, we formally define the Dasgupta's cost of the graph and the hierarchical stochastic model.

8.5.1 Related notions

Let $G = (V, E, w)$ be an undirected weighted graph with weight function $w : E \to \mathbb{R}^+$, where $\mathbb{R}^+$ denotes non-negative real numbers. For simplicity we let $w(x, y) = w(y, x) = w(\{x, y\})$. For set $U \subseteq V$ we define $G[U]$ to be the subgraph induced by U. A hierarchical clustering T of graph G is a rooted binary tree with exactly $|V|$ leaves, such that each leaf is labeled by a unique vertex $x \in V$.

For $G = (V, E)$ and a hierarchical-clustering tree T we denote the lowest common ancestor of vertex x and y in T by LCAT(x, y). For any internal node N of T, we let T_N to be the subtree of T rooted at N and we define $V(N)$ to be the set of leaves of the subtree rooted at N. Finally, for a weighted graph $G = (V, E, w)$ and any subset of vertices $A \subseteq V$ we define $w(A) = \sum_{x,y \in A} w(x, y)$, and for any set of edges $E0$, we let $w(E0) = \sum_{e \in E0} w(e)$.For any sets of vertices $A, B \subseteq V$, we also define $w(A, B) = \sum_{x \in A, y \in B} w(x, y)$.

Equipped with these notation we define the Dasgupta's cost of a graph for a tree as follows:

Definition 8.37 ((Dasgupta's cost[Das16, CKMM19]))**.** Dasgupta's cost of the tree T for the graph $G = (V, E, w)$ is defined as

$$\text{cost}(T; G) = \sum_{(x,y) \in E} \text{leaves}(T[LCA(x; y)]) \cdot w(x, y).$$

Definition 8.38 (Ultrametric [CKMM19])**.** A metric space (X, d) is an ultrametric if for every $x, y, z \in X$, $d(x, y) \leqslant \max\{d(x, z), d(y, z)\}$. We say that a weighted graph $G = (V, E, w)$ is generated from an ultrametric if there exists an ultrametric (X, d), such that $V \subseteq X$, and for every $x, y \in V$, $x \neq y$, $e = \{x, y\}$ exists, and $w(e) = f(d(x, y))$, where $f : \mathbb{R}^+ \to \mathbb{R}^+$ is a non-increasing function. For a weighted undirected graph $G = (V, E, w)$ generated from an ultrametric, in general there may be several ultrametrics and corresponding functions f mapping distances in the ultrametric to weights on the edges, that generate the same graph. It is useful to introduce the notion of a minimal ultrametric that generates G. Let (X, d) be an ultrametric that generates $G = (V, E, w)$ and f the corresponding function mapping distances to similarities. Then we consider the ultrametric $(V, \tilde{d})$ as follows: (i) $\tilde{d}(u, u) = 0$ and (ii) for $u \neq v$

$$\tilde{d}(u, v) = \tilde{d}(v, u) = \max_{u', v'} d(u', v') | f(d(u', v')) = f(d(u, v))$$

Definition 8.39 (Generating Tree [CKMM19])**.** Let $G = (V, E, w)$ be a graph generated by a minimal ultrametric (V, d). Let T be a rooted binary tree with $|V|$ leaves and $|V| - 1$ internal nodes; let N denote the internal nodes and L the set of leaves of T and let $\sigma : L \to V$ denote a bijection between the leaves of T and nodes of V . We say that T is a generating tree for G, if there exists a weight function $W : \mathcal{N} \to R^+$, such that for $N_1, N_2 \in \mathcal{N}$, if N_1 appears on the path from N_2 to the root, $W(N1) \leqslant W(N2)$. Moreover for every $x, y \in V$, $w(x, y) = W(\text{LCAT}(\sigma^{-1}(x), \sigma^{-1}(y)))$.

We say that a graph G is a ground-truth input if it is a graph generated from an ultrametric. Equivalently, there exists a tree T that is generating for G.

Now we are ready to define Hierarchical Stochastic Model graphs as follows:

Definition 8.40 (Hierarchical Stochastic Model (HSM) [CKMM19])**.** . Let $\widetilde{T}$ be a generating tree for an n-vertex graph $\bar{G}$, called the expected graph, such that all weights are in

$[0,1]$. A hierarchical stochastic model is a random graph G such that for every two vertices u and v, the edge $\{u,v\}$ is present independently with probability $w(\{u,v\}) = W(\mathrm{LCA}_T(\sigma^{-1}(u),\sigma^{-1}(v)))$, where w and W are the weights functions associated with $\widetilde{T}$ as per Definition 8.39. In words, the probability of an edge being present is given by the weight of the lowest common ancestor of the corresponding vertices in $\widetilde{T}$.

8.5.2 The algorithm for the semi-random hierarchical stochastic model

We generate a random graph, $G = (V,E)$, according to HSM (Definition 8.40). The semi-random model considers a random HSM graph generated as above where an adversary is allowed to only remove edges from G. Note that the comparison is to the cost of the generating tree on the graph $\bar{G}$ (Definition 8.40). In this section we present the proof of Theorem 8.4.

Proof of Theorem 8.4 is a variant of Theorem 6.1 from [CKMM19] with nearly-linear running time that uses our fast algorithm for finding Balanced-Cut.

Let $\bar{G}_n = (\bar{V}_n, \bar{E}_n, w)$ be a graph generated according to an ultrametric, where for each $e \in \bar{E}_n$, $w(e) \in (0,1)$ (Definition 8.39). Let $G = (V,E)$ be an unweighted random graph with $|V| = \left|\bar{V}_n\right| = n$ generated from $\bar{G}$ as follows. For every $u,v \in \bar{V}_n$ the edge (u,v) is added to G with probability $w((u,v))$ (Definition 8.40).

We assume that $V = \bar{V}_n$ and let T be a generating tree for $\bar{G}$. Let $U \subseteq V$. Let $\widetilde{T}|_U$ denote the restriction of T to leaves in U Let $N(U)$ be the root of $\widetilde{T}|_U$. Consider the following procedure where the nodes appear as leaves in the left and right subtrees of the root of $\widetilde{T}|_U$. Suppose we follow the convention that the left subtree is never any smaller than the right subtree in $\widetilde{T}|_U$. We say that the canonical node of $\widetilde{T}|_U$ is the first left node N_L encountered in a top-down traversal starting from $N(U)$ such that $(1-b)\cdot|U| \geqslant V(N_L) \geqslant b\cdot|U|$, where, $0 < b < 1/2$ is a constant. We define $U_L = V(N_L)$, and $U_R = U \setminus U_L$. We say that (U_L, U_R) is the *canonical* cut of U. It is easy to see that such a cut always exists since the tree is binary and left subtrees are never smaller than right subtrees. Let $E_{rnd} = \{(u,v) \in E | u \in U_L, v \in U_R\}$.

Lemma 8.41 ([CKMM19]). *For a random graph G generated as described in Theorem 8.4, with probability at least $1 - o(1)$, for every subset U of size at least $n^{2/3}\sqrt{\log n}$, the subgraph (U, E_{rnd}) is geometrically expanding up to scale $(1/\sqrt{D}, n, \alpha)$ where*

$$\alpha = C.\max\{w(L,R), |U|\cdot D\cdot\log^2 D, |U|\cdot D\cdot\log n\}, \tag{8.5.1}$$

Furthermore, the result also applies in the semi-random setting where an adversary may remove any subset of edges from the random graph G.

Theorem 8.42. *For any graph $G = (V,E)$, and weight function $w : E \to \mathbb{R}^+$, the ϕ-sparsest-cut algorithm from [CKMM19] outputs a solution of cost at most $O(\phi \cdot OPT)$.*

Now we show the proof of Theorem 8.4 which is a variant of Theorem 6.1 from [CKMM19] using our nearly-linear time agorithm for Balanced-Cut.

Proof of Theorem 8.4. Let $b = 1/3$. By Theorem 8.42, the recursive sparsest cut algorithm approximates Dasgupta's cost upto factor $O(\phi)$ assuming that at every recusrsion step, the algorithm is provided with a ϕ-approximation to the b-Balanced Cut problem (i.e., minimize cut subject to the constraint that both sides have at least b fraction of vertices being cut)

Note that $cost(\widetilde{T}; \bar{G}) = \Omega(n^3 \cdot p_{min}) = \Omega\left(n^{7/3} \cdot \log n\right)$. Therfore, once we obtain sets U of size $(n_0 = n^{2/3} \cdot \log n)$, since there are at most n/n_0 of them, even if we use an arbitrary tree on any such U, together this can only add $O(\frac{n}{n_0} \cdot n_0^3) = O(n^{13/9} \cdot (\log n)^2) = O\left(n^{7/3} \cdot \log n\right)$ to the cost. Thus, we only need to obtain suitable approximations during the recursive procedure as long as $|U| \geqslant n^{2/3} \cdot \log n$. This is precisely given by using Lemma 8.41. Let $D = O(\log n)$, $\delta = \frac{1}{\sqrt{D}} = O\left(\frac{1}{\sqrt{\log n}}\right)$, let $\rho > 0$ be a large constant, $\kappa \geqslant \Omega(\sqrt{\log n})$ be large constants. Observe that in Equation 8.5.1, $w(L, R) = \Omega(|U|^2 \cdot p_{min}) = \Omega\left(n^{2/3}(\log n)^3\right)$, $|U|D \log^2 D = o(|U|D \log n)$, and $D|U| \log n = O\left(n^{2/3}(\log n)^3\right)$. Let $\alpha = O(w(L, R))$. Thus, by Theorem 8.20 there exists an algorithm that runs in time $\tilde{O}\left(|V(G)|^{1+O(1/\rho^2)} + |E(G)|\right)$ and returns a cut that is an approximation to the $\Omega(b)$-balanced partition (S, T) with cut of size $|E(S, T)| \leqslant O(\alpha \cdot \rho)(1 + \delta \cdot \kappa \cdot \sqrt{\log n}) = O(\alpha \cdot \rho) = O(\alpha)$ on the induced subgraph of $\bar{G}$ on the vertex set U. This observation together with the case where subgraphs have size less than $n^{2/3} \log n$ finishes the proof. □

Chapter 9

Practical algorithms robust against adversarial distributions

In this chapter we prove Theorem 1.15. This result first appeared in [dKNS20]. Given that the algorithm is easy to implement, we further show through *experiments* how the algorithm performs against such adversaril perturbations (and how other algorithms fail instead). We do so in Section 9.2.

Recall the single-spiked covariance model. We reuse the notation introduced in Chapter 3.

Problem 9.1 (Restatement of Problem 3.1)**.** Given a matrix of the form

$$Y = W + \sqrt{\beta} u_0 v_0^\mathsf{T} + E, \text{ where} \tag{9.0.1}$$

- $v_0 \in \mathbb{R}^d$ is a unit k-sparse vector,
- $u_0 \sim N(0, \mathrm{Id}_n)$ is a standard Gaussian vector,
- $W \sim N(0,1)^{n \times d}$ is a Gaussian matrix and W, u_0, v_0 are distributionally independent,
- $E \in \mathbb{R}^{n \times d}$ is an arbitrary perturbation matrix satisfying

$$\|E\|_\infty \lesssim \sqrt{\beta/k} \cdot \min\{\sqrt{\beta}, 1\}\,. \tag{9.0.2}$$

Return a unit vector $\hat{v}$ having non-vanishing correlation with v_0.

We restate the Theorem of interest, which provides a spectral algorithms with guarantees matching that of the lower bound in Theorem 3.3.

Theorem 9.2 (Restatement of Theorem 1.15)**.** *Given an n-by-d matrix Y of the form,*

$$Y = \sqrt{\beta} u_0 v_0^\mathsf{T} + W + E\,,$$

for $\beta > 0$, a unit k-sparse vector $v_0 \in \mathbb{R}^d$, a Gaussian matrix $W \sim N(0,1)^{n \times d}$, a Gaussian vector $u_0 \sim N(0, \mathrm{Id}_n)$ such that v_0, u_0, W are distributionally independent, and $E \in \mathbb{R}^{n \times d}$ is a matrix from Theorem 3.3 for $t = 3$.[1] *Suppose that $d \gtrsim n^3 \log d \log n$, $k \gtrsim n \log n$ and*

$$\beta \gtrsim \frac{k}{n}\left(\frac{d}{k}\right)^{1/3}.$$

Then there exits an algorithm that computes in time $O(nd \log n)$ a unit vector $\hat{v} \in \mathbb{R}^d$ such that

$$1 - \langle v_0, \hat{v} \rangle \leqslant 0.01$$

with probability at least 0.99.

The algorithm behind the theorem (which we call SVD-t, where t is the corresponding sum-of-squares degree) captures the behavior of degree $\leqslant 6$ sum-of-squares in Theorem 3.2, but runs in time *nearly linear* in the input size.

9.1 The algorithm

We present here the algorithm and formally prove Theorem 9.2. We borrow the notation from Chapter 3 and Chapter 2, hence do not restate it here.

[1] More precisely, to prove Theorem 3.3 we consider a specific distribution over matrices E (this distribution depends on v_0, u_0 and W), and here we mean that E is sampled from this distribution.

Algorithm 9.3 (SVD-t: Sparse Vector Recovery)**.**

Given: Sample matrix $Y \in \mathbb{R}^{n\times d}$, let $y_1, \dots, y_d \in \mathbb{R}^n$ be its columns. Degree $j \in \{2, 4, 6\}$.

Estimate: The sparse vector v_0.

Operation:

1. Compute the top eigenvector $\hat{u}$ of the matrix
$$A := \sum_{i\in[d]} c_j(y_i, n) \cdot y_i y_i^\mathsf{T}$$
where for $x \in \mathbb{R}^n, t \in \mathbb{R}$, $c_2(x,t) := 1$, $c_4(x,t) := \left(\|x\|^2 - (t-1)\right)$, $c_6(x,t) := \left(c_4(x)^2 - 2(t-1)\right)$.
2. Compute $\hat{v} = \hat{u}^\mathsf{T} Y$.
3. Threshold the vector $\hat{v}$ in the following way (for some fixed $\tau \geqslant 0$):
$$\forall i \in [d], \eta(\hat{v})_i = \begin{cases} \hat{v}_i, & \text{if } |\hat{v}_i| \geqslant \frac{\tau}{\sqrt{k}} \\ 0, & \text{otherwise} \end{cases}$$
4. Output the thresholded vector $\eta(\hat{v})$.

Remark 9.4 (Running Time of the Algorithm). For $j \in \{2, 4, 6\}$, the terms $m_j(y_1, n), \dots, m_j(y_d, n)$ are computable in time $O(nd)$. Correctness of the algorithm will be proved showing that A has at least constant spectral gap. This means that we can compute the top eigenvalue with power iteration using $O(\log n)$ matrix-vector multiplications . A matrix multiplication requires computing $m_i = c_6(y_i, n)\langle y_i, z\rangle$ for each i and then taking the sum $\sum_{i\in[d]} m_i a_i$. Both operations take time $O(nd)$. Then, $\hat{v}$ can be computed in time $O(nd)$ and $\eta(\hat{v})$ in time $O(d)$. In conclusion the algorithm runs in time $O(nd \log n)$.

To get an intuition on the algorithm, consider SVD-6 and the simpler adversarial model $Y = \left(\mathrm{Id} - \frac{1}{\|u\|^2} uu^\mathsf{T}\right) W + \sqrt{\beta} uv^\mathsf{T}$.That is, standard sparse PCA in the Wishart model with the noise projected into the space orthogonal to u.[2] Now for $i \in \mathrm{supp}\{v\}$,
$$\left\| \left[\left(\|y_i\|^2 - n \right)^2 - 2n \right] y_i y_i^\mathsf{T} \right\| \approx \frac{\beta^3 n^3}{k^3},$$
while for $i \in [d] \setminus \mathrm{supp}\{v\}$,
$$\left\| \mathbb{E}\left[\left(\|y_i\|^2 - n \right)^2 - 2n \right] y_i y_i^\mathsf{T} \right\| = O(1).$$

[2]Note that the estimate $\hat{u}$ obtained by SVD-2 is the same returned by the standard SVD.

Indeed, the coefficient $c_6(y_i, n)$ has the effect of "killing" the expectation for Gaussian vectors. Then for $d \gg n^3$, the sum $\sum_{i\in[d]\setminus\mathrm{supp}\{v\}} c_6(y_i, n) y_i y_i^\mathsf{T}$ will be concentrated around its expectation d, while on the other hand

$$\left\| \sum_{i\in\mathrm{supp}\{v\}} \left[\left(\|y_i\|^2 - n \right)^2 - 2n \right] y_i y_i^\mathsf{T} \right\| \approx \frac{\beta^3 n^3}{k^2} .$$

Hence, for $\beta \gtrsim \frac{k}{n}\left(\frac{d}{k}\right)^{1/3}$ the leading eigenvector of A will be highly correlated with u.

We remark that it is an open question how these ideas could be generalize to construct an algorithm that works for $\beta \gtrsim \frac{k}{n}\left(\frac{d}{k}\right)^{1/t}$ and $d \gg n^t$. From this perspective, the result of this section can be seen as a proof of concept.

In order to define the adversarial perturbations, we will use the notation introduced for Problem 3.41, we recall that with high probability $\lambda = (1 \pm o(1))\sqrt{\frac{\beta n}{k}}$. The rest of the section is devoted to prove the Theorem below, which implies Theorem 9.2.

Theorem 9.5. *Consider a matrix of the form,*

$$Y = W + \lambda u v^\mathsf{T} + u\left(v' - W^\mathsf{T} u\right)^\mathsf{T}$$

for a Gaussian matrix $W \sim N(0,1)^{n\times d}$, a random unit vector u, a k-sparse vector v with entries in $\{0, \pm 1\}$ and a vector v' as defined in 3.41. For $d \gtrsim n^3 \log d \log n$, $\lambda \gtrsim \frac{\sqrt{\log d}}{\tau}$ and $k \geqslant n \log n$, Algorithm 9.3 with degree 6 returns a vector $\eta(\hat{v})$ such that

$$\left\|\eta(\hat{v}) - v\right\| \leqslant O\left(\frac{d}{k\lambda^6} + \tau\right) \cdot \sqrt{k}$$

with probability at least 0.99. Furthermore, for $\frac{d}{k\lambda^6} + \tau \leqslant 1$ and $\beta = \frac{\lambda^2 k}{n}$,

$$1 - \frac{\langle \eta(\hat{v}), v\rangle^2}{\left\|\eta(\hat{v})\right\|^2 \cdot \|v\|^2} \leqslant \left(\frac{k}{n\beta}\left(\frac{d}{k}\right)^{1/3} + \tau^2 \right).$$

We remark that the second inequality of the theorem follows from the first by direct substitution and using the fact that $\eta(\hat{v})$ is close to a unit vector. Comparing this result with Theorem 3.27 we see that both SVD-6 9.3 and degree-6 SoS 3.29 need $\beta \gtrsim \frac{k}{n}\frac{d}{k}^{1/3}$ in order to achieve correlation 0.9 with the sparse vector.

To prove Theorem 9.5, we will first show that the vector $\hat{u}$ computed by the algorithm is close to the true vector u. Then, thresholding the vector $\hat{u}^\mathsf{T} Y$ we will obtain a vector close to v. Concretely, we will prove two results. First,

Lemma 9.6. *Consider a matrix of the form,*

$$Y = W + \lambda u v^\mathsf{T} + u\big(v' - W^\mathsf{T} u\big)^\mathsf{T}$$

for a Gaussian matrix $W \sim N(0,1)^{n\times d}$, a random unit vector u, a k-sparse vector v with entries in $\{0, \pm 1\}$ and a vector v' as defined in 3.41. Let $\hat{u} \in \mathbb{R}^n$ be the top eigenvector of the matrix

$$\sum_{i\in[d]} c_6(y_i, n) \cdot y_i y_i^\mathsf{T}.$$

Then for $d \geqslant C^* n^3 \log d \log n$, $n \geqslant 10 \log d$

$$\|u - \hat{u}\| \leqslant O\left(\frac{d}{k\lambda^6} + \frac{1}{\lambda} + \frac{\sqrt{n \log n}}{\lambda\sqrt{k}}\right)$$

with probability at least 0.999, where C^* is a universal constants.

Second,

Lemma 9.7. *Let $\hat{u}$ be a vector such that $\|\hat{u} - u\| \leqslant \varepsilon$ for some $0 \leqslant \varepsilon \leqslant \frac{1}{10}$ and let $\hat{v} = \frac{1}{\lambda\sqrt{k}} \hat{u}^\mathsf{T} Y$. If $\lambda \gtrsim \frac{\sqrt{\log d}}{\tau}$, then with probability at least $1 - \exp(-n)$*

$$\|\hat{v} - v\| \lesssim (\varepsilon + \tau)\sqrt{k}\,,$$

where $\eta(\hat{v}) \in \mathbb{R}^d$ is the vector with coordinates

$$\eta(\hat{v})_i = \begin{cases} \hat{v}_i, & \text{if } |\hat{v}_i| \geqslant \tau \\ 0, & \text{otherwise.} \end{cases} \tag{9.1.1}$$

It is easy to see how the two results immediately imply Theorem 9.5.

Lemma 9.6 is proved in Section 9.1.1, in Section 9.1.2 we prove Lemma 9.7.

9.1.1 Recovery of the random vector *u*

The goal of this Section is to prove Lemma 9.6.

By rotational symmetry of the Gaussian distribution, we may assume without loss of generality that $u = e_1$. Now, for vectors $v, z \in \mathbb{R}^n$, define $M(v,z) := \left[\left(\|v+z\|^2 - (n-1)\right)^2 - 2(n-1)\right] vv^\mathsf{T}$. Recall that the adversarial vector v' is, by construction, orthogonal to the sparse vector v. Hence our strategy will be the following, first we bound the contribution of terms of the form $M(w, \gamma e_1)$ and $M(v'(i)u, w)$. Note that the first type of terms arise due to the noise, the second ones due to the adversarial distribution. Then, lower bounding $M(\lambda u, w)$, we will be able to show that

$$\left\| \sum_{i \in \operatorname{supp}\{v\}} M(\lambda u, w) \right\| \gg \left\| \sum_{i \in [d]} M(w_i, \gamma_i e_1) \right\| + \left\| \sum_{i \in [d] \setminus \operatorname{supp}\{v\}} M(v'(i)u, w_i) \right\| \text{ with high probability.}$$

Cross-terms will play a minor role.

First we bound the contribution of the Gaussian part. We will use Bernstein Inequality, the next results act as building blocks for the bound, which is then shown in Lemma 9.12.

Fact 9.8. *Let $x \sim N(0, \mathrm{Id}_n)$,*

$$\mathbb{E}\|x\|^2 x_i^2 = n + 2$$
$$\mathbb{E}\|x\|^4 x_i^2 = n^2 + 6n + 8.$$

Proof.

$$\begin{aligned}
\mathbb{E}\|x\|^2 x_i^2 &= \sum_{j \in [n], j \neq i} \mathbb{E}\, x_i^2 x_j^2 + \mathbb{E}\, x_i^4 = n + 2 \\
\mathbb{E}\|x\|^4 x_i^2 &= \sum_{\substack{j,k \in [n] \\ j \neq i, k \neq i, j \neq k}} \mathbb{E}\, x_i^2 x_j^2 x_k^2 + \sum_{\substack{j,k \in [n] \\ j = k \neq i}} \mathbb{E}\, x_j^4 x_i^2 + 2 \sum_{\substack{j,k \in [n] \\ j \neq k = i}} \mathbb{E}\, x_i^4 x_j^2 + \mathbb{E}\, x_i^6 \\
&= (n-1)(n-2) + 3(n-1) + 6(n-1) + 15 \\
&= n^2 + 6n + 8.
\end{aligned}$$

□

We bound the spectral norm of the expectation of the terms $M(w, \gamma e_1)$.

Lemma 9.9. *Let $w \sim N(0, \mathrm{Id}_n - e_1 e_1^{\mathsf{T}})$, $\gamma \in \mathbb{R}$. Then*

$$\|\mathbb{E}\, M(w, \gamma e_1)\| = \gamma^4 + 8\gamma^2 + 8.$$

Proof. We need only to look into diagonal entries. By construction, $\mathbb{E}\|w\|^2 = n - 1 =: m$,

$$\begin{aligned}
\mathbb{E}\left[\left(\|w + \gamma e_1\|^2 - m \right)^2 - 2m \right] w_i^2 &= \mathbb{E}\left[\left(\|w\|^2 + \gamma^2 - m \right)^2 - 2m \right] w_i^2 \\
&= \mathbb{E}\left(\|w\|^4 + \gamma^4 + m^2 + 2\|w\|^2\gamma^2 - 2m\|w\|^2 - 2\gamma^2 m - 2m \right) w_i^2
\end{aligned}$$

Applying Fact 9.8,

$$\begin{aligned}
&\mathbb{E}\left[\left(\|w + \gamma e_1\|^2 - m \right)^2 - 2m \right] w_i^2 \\
&= m^2 + 6m + 8 + \gamma^4 + m^2 + 2\gamma^2 m + 4\gamma^2 - 2m^2 - 4m - 2\gamma^2 m - 2m \\
&= \gamma^4 + 8\gamma^2 + 8.
\end{aligned}$$

□

The second property we need is a high probability bound on the maximum value of $\|M(w, \gamma e_1)\|$.

Lemma 9.10. *Let $w \sim N(0, \mathrm{Id}_n - e_1 e_1^{\mathsf{T}})$, $\gamma \in \mathbb{R}$. Then for any $q \geqslant 1$, with probability at least $1 - 2e^{-q}$,*

$$\|M(w, \gamma e_1)\| \leqslant C\big(\gamma^4 n + n \max\{nq, q^2\}\big),$$

where C is a universal constant.

Proof. For simplicity of the notation let $m = n - 1$, and let $p = \max\{q, \sqrt{mq}\}$. By Fact A.18,

$$\mathbb{P}\Big(\|w\|^2 \notin \big[m - 10p, m + 10p\big]\Big) \leqslant 2e^{-q}.$$

Hence with probability at least $1 - 2e^{-q}$,

$$\begin{aligned}
\left|\left[\Big(\|w + \gamma e_1\|^2 - m\Big)^2 - 2m\right]\right| &= \left|\left[\Big(\|w + \gamma e_1\|^2 - m\Big)^2 - 2m\right]\right| \\
&= \left|\left[\Big(\|w\|^2 + \gamma^2 - m\Big)^2 - 2m\right]\right| \\
&\leqslant \left[\big(\gamma^2 + 10p\big)^2 - 2m\right] \\
&\leqslant C\big(\gamma^4 + p^2\big)
\end{aligned}$$

for some universal constant $C > 0$. The result follows. □

And finally, the last ingredient we need for our Bernstein inequality is a bound on the variance.

Lemma 9.11. *Let $w \sim N(0, \mathrm{Id}_n - e_1 e_1^{\mathsf{T}})$, $\gamma \in \mathbb{R}$. Then*

$$\big\|\mathbb{E}\, M(w, \gamma e_1)^2\big\| \leqslant C\Big(\gamma^8 n + n \max\{\log^4 n\gamma, n^2 \log n\gamma\}\Big),$$

for a universal constant $C > 0$.

Proof. For simplicity of the notation let $m = n - 1$. Fix $q = 50 \log m\gamma$ and $p = \max\{q, \sqrt{mq}\}$. Define the event $\mathcal{E} = \Big\{\|w\|^2 \notin \big[m - 10p, m + 10p\big]\Big\}$, which happens with probability at least $1 - 2e^{-q}$. Then,

$$\left[\Big(\|w + \gamma e_1\|^2 - m\Big)^2 - 2m\right]^2 \leqslant C\big(\gamma^8 + p^4\big).$$

By triangle inequality,

$$\big\|\mathbb{E}\, M(w, \gamma e_1)^2\big\| = \big\|\mathbb{P}(\mathcal{E})\, \mathbb{E}\big[M(w, \gamma e_1)^2 \,\big|\, \mathcal{E}\big] + \mathbb{P}(\bar{\mathcal{E}})\, \mathbb{E}\big[M(w, \gamma e_1)^2 \,\big|\, \bar{\mathcal{E}}\big]\big\|$$

$$\leqslant \left\|\mathbb{P}(\mathcal{E})\,\mathbb{E}\left[M(w,\gamma e_1)^2 \mid \mathcal{E}\right]\right\| + \left\|\mathbb{P}(\bar{\mathcal{E}})\,\mathbb{E}\left[M(w,\gamma e_1)^2 \mid \bar{\mathcal{E}}\right]\right\|.$$

We bound the first term,

$$\begin{aligned}\left\|\mathbb{P}(\mathcal{E})\,\mathbb{E}\left[M(w,\gamma e_1)^2 \mid \mathcal{E}\right]\right\| &\leqslant \left\|O\left((\gamma^8+p^4)m\right)\mathbb{E}\left[ww^\mathsf{T} \mid \mathcal{E}\right]\right\| \\ &\leqslant O\left(\gamma^8 m + mp^4\right)\left\|\mathbb{E}\left[ww^\mathsf{T} \mid \mathcal{E}\right]\right\| \\ &\leqslant O\left(\gamma^8 m + mp^4\right)\left\|\mathbb{E}\left[ww^\mathsf{T}\right]\right\| \\ &\leqslant O\left(\gamma^8 m + mp^4\right).\end{aligned}$$

To bound the second term, observe that $\left\|M(w,\gamma e_1)^2\right\| \leqslant O\left(m^{12}+\gamma^{12}+\|w\|^{12}\right)$ for any $\gamma, w, m \geqslant 1$. For $i \in \mathbb{N}$, define the event

$$\begin{aligned}\mathcal{E}_{qi} := \left\{\|w\|^2 \in \left[m - 2\sqrt{mq\cdot(i+1)} - 2q(i+1),\, m + 2\sqrt{mq\cdot(i+1)} + 2q(i+1)\right]\right\} \\ \cap \left\{\|w\|^2 \notin \left[m - 2\sqrt{mqi} - 2qi,\, m + 2\sqrt{mqi} + 2qi\right]\right\}.\end{aligned}$$

By construction $\mathbb{P}(\mathcal{E}_{q_i}) \leqslant 2\max\left\{e^{-\frac{q_i^2}{4m}}, e^{-q_i/4}\right\}$ and $\bar{\mathcal{E}} \subseteq \bigcup_{i\in\mathbb{N}} \mathcal{E}_{qi}$. By choice of q, it follows that

$$\begin{aligned}\left\|\mathbb{P}(\bar{\mathcal{E}})\,\mathbb{E}\left[M(w,\gamma e_1)^2 \mid \bar{\mathcal{E}}\right]\right\| &\leqslant \sum_{i\in\mathbb{N}} \left\|\mathbb{P}(\mathcal{E}_{qi})\,\mathbb{E}\left[M(w,\gamma e_1)^2 \mid \mathcal{E}_{qi}\right]\right\| \\ &\leqslant O(1),\end{aligned}$$

concluding the proof.

□

We can now apply Bernstein Inequality A.21:

Lemma 9.12. *Let $w_1,\dots,w_l \sim N(0, \mathrm{Id}_n - e_1e_1^\mathsf{T})$, let $|\gamma_1|,\dots,|\gamma_l| \leqslant \gamma \in \mathbb{R}$. Then for $l \geqslant C^*\cdot\max\left\{n^3\log(l+\gamma n), n\log^3(l+\gamma n)\right\}$,*

$$\left\|\sum_{i\in[l]} M(w_i,\gamma_i e_1)\right\| \leqslant l(\gamma^4+8) + C^*\gamma^4\sqrt{ln\log n}$$

with probability at least $1-2l^{-10}-n^{-10}$, where C^* is a universal constant.

Proof. By triangle inequality,

$$\left\|\sum_{i\in[l]} M(w_i,\gamma_i e_1)\right\| \leqslant \left\|\sum_{i\in[l]} \mathbb{E}\,M(w_i,\gamma_i e_1)\right\| + \left\|\sum_{i\in[l]} M(w_i,\gamma_i e_1) - \sum_{i\in[l]} \mathbb{E}\,M(w_i,\gamma_i e_1)\right\|.$$

By Lemma 9.9,

$$\left\| \sum_{i\in[l]} \mathbb{E}\, M(w_i, \gamma_i e_1) \right\| \leqslant 8l + 8l\gamma^2 + l\gamma^4.$$

Let $q = 100\log(l + m\gamma)$ and $p := \max\{q, \sqrt{mq}\}$. Define the event $\mathcal{E} = \left\{\|w\|^2 \notin \left[m - 10p, m + 10p\right]\right\}$, which happens with probability at least $1 - 2e^{-q}$. By Lemma 9.10, with probability at least $1 - 2l^{-10}$, for each $i \in [l]$,

$$\|M(w_i, \gamma_i e_1) - \mathbb{E}\, M(w_i, \gamma_i e_1)\| \leqslant C\Big(8 + 8\gamma^2 + \gamma^4 + \gamma^4 n + n\max\{n\log l, \log^2 l\}\Big),$$

for a constant $C > 0$. Hence, by Lemma 9.11, applying Bernstein Inequality A.21

$$\left\| \sum_{i\in[l]} (M(w_i, \gamma_i e_1) - \mathbb{E}\, M(w_i, \gamma_i e_1)) \right\| \leqslant C' \cdot t\sqrt{ln\log n} \cdot \gamma^4$$

with probability at least $1 - 2l^{-10} - e^{-(t-1)\log n}$, where C' is a universal constant. □

The next lemma will be used to bound the contribution of the adversarial vector v'.

Lemma 9.13. *Let $|a_1|, \dots, |a_l| \leqslant a \in \mathbb{R}$. Let $w \sim N(0, \mathrm{Id}_n - e_1 e_1^\mathsf{T})$. Then, with probability at least $1 - e^{-(t-1)\log n} - 2l^{-10}$*

$$\left\| \sum_{i\in[l]} M(a_i e_1, w) \right\| \leqslant C \cdot t \cdot \sqrt{l\log n}\, a^2 \left(a^4 + \max\left\{\log l, \sqrt{n\log l}\right\}\right)$$

where $t \geqslant 1$ and $C > 0$ is a universal constant.

Proof. For simplicity let $m = n - 1$ and $q = 10\log l$ and $p := \max\{q, \sqrt{mq}\}$. By Fact A.18,

$$\mathbb{P}\Big(\|w\|^2 \notin \left[m - 10p, m + 10p\right]\Big) \leqslant 2e^{-q}.$$

Hence, as in Lemma 9.10

$$\left\| \left[\left(\|w + a_i e_1\|^2 - m\right)^2 - 2m \right] \right\| \leqslant O\left(a_i^4 + p^2\right) \leqslant O\left(a^4 + p^2\right).$$

This implies,

$$\left\| M(a_i e_1, w)^2 \right\| \leqslant O\left(a^{12} + p^4 a^4\right).$$

We have everything we need to apply Hoeffding Inequality A.22

$$\mathbb{P}\left(\left\| \sum_{i\in[l]} M(a_i e_1, w) \right\| \geqslant C \cdot t \cdot \sqrt{l\log n}\left(a^6 + p^2 a^2\right) \right) \leqslant e^{-(t-1)\log n},$$

for $p \geqslant 1$ and a universal constant C. □

The last intermediate result, is a high probability lower bound on the spectral norm of the matrix $\sum_{i\in\mathrm{supp}\{v\}} M(\lambda e_1, w)$, that is, the matrix corresponding to the sum of the columns that contain the spike.

Lemma 9.14. *Let $\zeta_1,\dots,\zeta_l \in \{-1,+1\}$ and $w_1,\dots,w_l \sim N(0, \mathrm{Id}_n - e_1e_1^\mathsf{T})$. Let $\gamma \in \mathbb{R}$, for $t \geqslant 1$ and auniversal constant $C > 0$, suppose $l \geqslant C \cdot t \log n \cdot \max\{n^3 \log l, \log^2 l\}$. Then*

$$\left\| \sum_{i\in[l]} M(\zeta_i \gamma e_1, w_i) \right\| \geqslant \frac{l\gamma^6}{2},$$

with probability at least $1 - 2e^{-t\log n} - 2l^{-10}$.

Proof. Now,

$$\begin{aligned}\sum_{i\in[l]} M(\zeta_i\gamma e_1, w_i) &= \left(\sum_{i\in[l]} \left[\left(\|w_i + \zeta_i\gamma e_1\|^2 - m \right)^2 - 2m \right] \right) \gamma^2 e_1 e_1^\mathsf{T} \\ &= \left[\gamma^4 l + m^2 l - 2ml - 2\gamma^2 ml + \left(\sum_{i\in[l]} \|w_i\|^4 - 2m\|w_i\|^2 + 2\gamma^2 \|w_i\|^2 \right) \right] \gamma^2 e_1 e_1^\mathsf{T}.\end{aligned}$$

We bound the terms in the parenthesis. Recall that by construction

$$\mathbb{E}\left[m^2 l - 2ml - 2\gamma^2 ml + \left(\sum_{i\in[l]} \|w_i\|^4 - 2m\|w_i\|^2 + 2\gamma^2\|w_i\|^2 \right) \right] = 0.$$

For $q := 10 \log l$ and $p = \max\{\sqrt{mq}, q\}$, we can condition on the event,

$$\mathcal{E} := \left\{ \forall i \in [l] \;\middle|\; \|w\|^2 \in \left[m - 10p, m + 10p \right] \right\},$$

which happens with probability at least $1 - 2e^{-q}$. Then, by Hoeffding Inequality A.22,

$$\begin{aligned}\left| \sum_{i\in[l]} \left[m^2 l - 2ml - 2\gamma^2 ml + \left(\sum_{i\in[l]} \|w_i\|^4 - 2m\|w_i\|^2 + 2\gamma^2\|w_i\|^2 \right) \right] \right| \\ \geqslant C \cdot t \cdot \sqrt{l}(p^2 + mp + \gamma^2 p)\end{aligned}$$

with probability at most $2e^{-t}$, for $t \geqslant 1$ and a universal constant C. By assumption on γ, l and n, it follows that

$$\left\| \sum_{i\in[l]} M(\zeta_i\gamma e_1, w_i)\gamma^2 e_1 e_1^\mathsf{T} \right\| \geqslant \left\| \frac{l\gamma^6}{2} e_i e_1^\mathsf{T} \right\| = \frac{l\gamma^6}{2}.$$

□

We are now ready to prove the main result of the section. Combining Lemma 9.15 with Lemma 9.14 and an application of Lemma A.28 we immediately get Lemma 9.6.

Lemma 9.15. *Let Y be defined as in Theorem 9.5, let $d \geqslant C \cdot n^3 \log d \log n \geqslant 100$. For $k\lambda^6 \geqslant C^* d$, $n \geqslant \log d$ and large enough constants C, C^*, with probability at least 0.999,*

$$\sum_{i\in[d]} \left[\left(\|y_i\|^2 - m \right)^2 - 2m \right] y_i y_i^\mathsf{T} = \sum_{i\in \mathrm{supp}\{v\}} M(v(i)\lambda e_1, w) + M,$$

where M is a matrix such that

$$\|M\| \leqslant O\left(d + k\lambda^5 + \sqrt{kn \log n}\lambda^5 \right).$$

Proof. Let $m = n - 1$. Recall the notation used in the algorithm with $c_6(y_i, n) = \left[\left(\|y_i\|^2 - m \right)^2 - 2m \right]$. Then we can rewrite the matrix A computed by SVD-6 as,

$$\begin{aligned}
\sum_{i\in[d]} \left[\left(\|y_i\|^2 - m \right)^2 - 2m \right] y_i y_i^\mathsf{T} = &\sum_{i\in \mathrm{supp}\{v\}} M(v(i)\lambda e_1, w) + \sum_{i\in \mathrm{supp}\{v\}} M(w_i, v(i)\lambda e_1) \\
&+ \sum_{i\in[d]\setminus \mathrm{supp}\{v\}} M(w_i, v'(i)e_1) + \sum_{i\in[d]\setminus \mathrm{supp}\{v\}} M(v'(i)e_1, w) \\
&+ \sum_{i\in \mathrm{supp}\{v\}} c_6(w_i + v(i)\lambda e_1, n)\left(v(i) w_i e_1^\mathsf{T} + v(i) e_1 w_i^\mathsf{T} \right) \\
&+ \sum_{i\in[d]\setminus \mathrm{supp}\{v\}} c_6(w_i + v'(i)e_1)\left(v'(i) w_i e_1^\mathsf{T} + v'(i) e_1 w_i^\mathsf{T} \right).
\end{aligned}$$

We first bound the cross-terms,

$$\begin{aligned}
&\left\| \sum_{i\in[d]\setminus \mathrm{supp}\{v\}} c_6(w_i + v'(i)e_1) v'(i) w_i e_1^\mathsf{T} \right\| \\
&\leqslant \left\| \sum_{i\in[d]\setminus \mathrm{supp}\{v\}} c_6(w_i + v'(i)e_1) v'(i)^2 e_1 e_1^\mathsf{T} \right\|^{1/2} \left\| \sum_{i\in[d]\setminus \mathrm{supp}\{v\}} c_6(w_i + v'(i)e_1) v'(i)^2 w_i w_i^\mathsf{T} \right\|^{1/2} \\
&\leqslant \left\| \sum_{i\in[d]\setminus \mathrm{supp}\{v\}} c_6(w_i + v'(i)e_1) v'(i)^2 e_1 e_1^\mathsf{T} \right\| + \left\| \sum_{i\in[d]\setminus \mathrm{supp}\{v\}} c_6(w_i + v'(i)e_1) v'(i)^2 w_i w_i^\mathsf{T} \right\|
\end{aligned}$$

And

$$\left\| \sum_{i\in \mathrm{supp}\{v\}} c_6(w_i + v(i)\lambda e_1, n) v(i) w_i e_1^\mathsf{T} \right\|$$

$$\leqslant \left\| \sum_{i \in \mathrm{supp}\{v\}} c_6(w_i + v(i)\lambda e_1, n) v(i)^2 e_1 e_1^\mathsf{T} \right\|^{1/2} \left\| \sum_{i \in \mathrm{supp}\{v\}} c_6(w_i + v(i)\lambda e_1, n) w_i w_i^\mathsf{T} \right\|^{1/2}$$

Observe that, by construction of the vector v' in Model 3.41 and since $k\lambda^6 \geqslant C^* d$, for a large enough constant C^*, we get that for all $i \in [d]$, $|v'(i)| \leqslant 100$. Moreover we get that with probability at least 0.999, all the following inequalities hold.

By Lemma 9.14,

$$\left\| \sum_{i \in \mathrm{supp}\{v\}} M(v(i)\lambda e_1, w_i) \right\| \geqslant \frac{k\lambda^6}{2}.$$

By Lemma 9.12,

$$\left\| \sum_{i \in \mathrm{supp}\{v\}} M(w_i, v(i)\lambda e_1) \right\| \leqslant O\left(k\lambda^4 + \sqrt{kn \log n}\lambda^4\right)$$
$$\left\| \sum_{i \in [d] \setminus \mathrm{supp}\{v\}} M(w_i, v'(i) e_1) \right\| \leqslant O\left(d + \sqrt{dn \log n}\right) \leqslant O(d).$$

By Lemma 9.13,

$$\left\| \sum_{i \in [d] \setminus \mathrm{supp}\{v\}} M(v'(i)e_1, w) \right\| \leqslant O\left(\sqrt{d \log n} \max\left\{\log d, \sqrt{n \log d}\right\}\right).$$

All in all we get,

$$\left\| \sum_{i \in [d]} \left[\left(\|y_i\|^2 - m \right)^2 - 2m \right] y_i y_i^\mathsf{T} - \sum_{i \in \mathrm{supp}\{v\}} M(v(i)\lambda e_1, w_i) \right\|$$
$$\leqslant O\left(d + \sqrt{dn \log n} + \sqrt{dn} \log d + \sqrt{d} \log d + k\lambda^5 + \sqrt{kn \log n}\lambda^5\right).$$

The result follows. □

9.1.2 Recovery of the sparse direction v

We now show how to obtain a good estimate of the sparse vector v from $\hat{u}$. Since several algorithms try to recover first u and then the sparse vector (e.g. SVD with thresholding) we turn back to the model 3.1. A corollary for model 3.41 is presented at the end of the section. So, for the rest of the section, let $Y = \sqrt{\beta} u v^\mathsf{T} + W + E$, where $v \in \mathbb{R}^d$ is a k-sparse

unit vector, $u \in \mathbb{R}^n$ is a vector such that $\|u\| \geqslant 0.9\sqrt{n}$, and $W \sim N(0,1)^{n\times d}$. We also assume that $n \leqslant k \leqslant d$.

The first observation is that on one hand the vector Yv is close to $\sqrt{\beta}u$ with high probability. On the other hand, the vector $Y^\mathsf{T}u$ may be far from the sparse vector; that is, even knowing exactly u, the thresholding step is required to recover v. The next theorem provides guarantees on the achievable correlation with the sparse vector given a vector close to u. Theorem 3.42 shows in which sense these guarantees are information theoretically tight.

Theorem 9.16. *Let $\hat{u}$ be a vector such that $\|\hat{u} - u\| \leqslant \varepsilon\sqrt{n}$ for some $0 \leqslant \varepsilon \leqslant \frac{1}{10}$, and let*

$$\hat{v} = \frac{1}{\sqrt{\beta}\cdot\|\hat{u}\|^2}\hat{u}^\mathsf{T}Y\,.$$

If $\beta \gtrsim \frac{k}{\tau^2 n}\left(\log d + \|E\|_{1\to 2}^2\right)$ for some $0 < \tau \leqslant 1$, then with probability at least $1 - 10\exp(-n)$,

$$\|\eta(\hat{v}) - v\| \lesssim \varepsilon + \tau\,,$$

where $\eta(\hat{v}) \in \mathbb{R}^d$ is the vector with coordinates

$$\eta(\hat{v})_i = \begin{cases} \hat{v}_i, & \text{if } |\hat{v}_i| \geqslant \tau/\sqrt{k} \\ 0, & \text{otherwise} \end{cases}$$

Proof. Assume that $\beta \geqslant 10^4 \cdot \frac{k}{\tau^2 n}\left(\log d + \|E\|_{1\to 2}^2\right)$. Let's rewrite $Y = \sqrt{\beta}\hat{u}v^\mathsf{T} + W + Z + E$ for a matrix $Z = \sqrt{\beta}(u - \hat{u})v^\mathsf{T} \in \mathbb{R}^{n\times d}$. Then for $i \in [d]$:

$$\left|\left(\hat{u}^\mathsf{T}Z\right)_i\right| = \left|\sqrt{\beta}\langle\hat{u}, u - \hat{u}\rangle v_i\right| \leqslant \varepsilon\sqrt{\beta n}\cdot\|\hat{u}\|\cdot|v_i|\,.$$

Let $S = \{i \mid v_i = 0\}$, $T = \{i \mid \hat{v}_i > \tau/\sqrt{k}\}$, $A = \{i \mid |v_i| \leqslant 2\tau/\sqrt{k}\}$ and $B = \{i \mid (\hat{u}W)_i \geqslant 10\|\hat{u}\|\sqrt{\log d}\}$. By Lemma A.25, with probability at least $1 - 2\exp(-n)$, $|B| \leqslant n$. Consider some $i \in S \cap T$. Since $v_i = 0$,

$$(\hat{u}W)_i = \sqrt{\beta}\cdot\|\hat{u}\|^2\cdot\hat{v}_i - (\hat{u}E)_i \geqslant 100\cdot\frac{\|\hat{u}\|}{\sqrt{n}}\cdot\|\hat{u}\|\left(\sqrt{\log d} + \|E\|_{1\to 2}\right) - \|\hat{u}\|\cdot\|E\|_{1\to 2} \geqslant 10\|\hat{u}\|\sqrt{\log d}\,,$$

which means that $S\cap T \subseteq B$. Hence $|T\setminus B| \leqslant |\overline{S}| = k$. Note that since $\varepsilon \leqslant \frac{1}{10}$ and $\|u\| \geqslant 0.9\sqrt{n}$, $\|\hat{u}\| \geqslant 0.8\sqrt{n}$. Hence

$$\begin{aligned}
\sum_{i\in T\setminus B}(\eta(\hat{v})_i - v_i)^2 &= \sum_{i\in T\setminus B}(\hat{v}_i - v_i)^2 \\
&\leqslant 2\sum_{i\in T\setminus B}\frac{n}{\|\hat{u}\|^2}\varepsilon^2 v_i^2 + 2\sum_{i\in T\setminus B}\frac{1}{\beta\|\hat{u}\|^4}\left((\hat{u}^\mathsf{T}W)_i^2 + (\hat{u}^\mathsf{T}E)_i^2\right) \\
&\leqslant 4\varepsilon^2 + 4\sum_{i\in T\setminus B}\frac{1}{\beta n}\left(100\log d + \|E\|_{1\to 2}^2\right)
\end{aligned}$$

$$\leqslant 4\varepsilon^2 + \tau^2 .$$

Note that since $|B| \leqslant n$, By Theorem A.23, with probability at least $1 - \exp(-n)$,

$$\sum_{i\in B} (\hat{u}^\mathsf{T} W)_i^2 \leqslant 100\|\hat{u}\|^2 \cdot n \log d .$$

Hence

$$\begin{aligned}
\sum_{i\in T\cap B} (\eta(\hat{v})_i - v_i)^2 &= \sum_{i\in T\cap B} (\hat{v}_i - v_i)^2 \\
&\leqslant 2\sum_{i\in B} \frac{n}{\|\hat{u}\|^2}\varepsilon^2 v_i^2 + 2\sum_{i\in B} \frac{1}{\beta\|\hat{u}\|^4}(\hat{u}^\mathsf{T} W)_i^2 + 2\sum_{i\in B} \frac{1}{\beta\|\hat{u}\|^4}(\hat{u}^\mathsf{T} E)_i^2 \\
&\leqslant 4\varepsilon^2 + 400\frac{\log d}{\beta} + 4\frac{\|E\|_{1\to 2}^2}{\beta} \\
&\leqslant 4\varepsilon^2 + \tau^2 .
\end{aligned}$$

If $i \in S \setminus T$, then $\eta(\hat{v})_i = v_i = 0$. If $i \in \overline{S} \cap \overline{T} \cap \overline{A}$, then

$$\frac{\tau}{\sqrt{k}} \geqslant |\hat{v}_i| \geqslant \left(1 - \frac{\sqrt{n}}{\|\hat{u}\|}\varepsilon\right)|v_i| - \left|\frac{(\hat{u}^\mathsf{T} W)_i}{\sqrt{\beta}\|\hat{u}\|^2}\right| - \left|\frac{(\hat{u}^\mathsf{T} E)_i}{\sqrt{\beta}\|\hat{u}\|^2}\right| \geqslant 1.8\frac{\tau}{\sqrt{k}} - \left|\frac{(\hat{u}^\mathsf{T} W)_i}{\sqrt{\beta}\|\hat{u}\|^2}\right| - 0.1\frac{\tau}{\sqrt{k}} ,$$

hence in this case $|(\hat{u}W)_i| > 0.7 \cdot 0.8 \cdot 100\sqrt{\log d} \geqslant 10\sqrt{\log d}$, so $i \in B$. Moreover,

$$|v_i| \leqslant \frac{1.1\tau}{0.9\sqrt{k}} + 2\left|\frac{1}{\sqrt{\beta}n}(\hat{u}^\mathsf{T} W)_i\right| .$$

Therefore

$$\sum_{i\in\overline{S}\cap\overline{T}\cap\overline{A}} (\eta(\hat{v})_i - v_i)^2 = \sum_{i\in\overline{S}\cap\overline{T}\cap\overline{A}} v_i^2 \leqslant 2\sum_{i\in B} \frac{2\tau^2}{k} + 4\sum_{i\in B} \frac{1}{\beta n^2}(\hat{u}W)_i^2 \leqslant 4\tau^2 + \tau^2 = 5\tau^2 .$$

It follows that

$$\|\eta(\hat{v}) - v\|^2 \leqslant \sum_{i\in T}(\eta(\hat{v})_i - v_i)^2 + \sum_{i\in\overline{S}\cap\overline{T}\cap A} v_i^2 + \sum_{i\in\overline{S}\cap\overline{T}\cap\overline{A}} v_i^2 \leqslant 8\varepsilon^2 + 2\tau^2 + 4\tau^2 + 5\tau^2 = 8\varepsilon^2 + 11\tau^2 .$$

Hence with probability at least $1 - 3\exp(-n)$,

$$\|\eta(\hat{v}) - v\| \lesssim \varepsilon + \tau .$$

□

An immediate consequence is the following corollary.

Corollary 9.17. *Consider a matrix of the form,*

$$Y = W + \lambda u v^\mathsf{T} + u\big(v' - W^\mathsf{T} u\big)^\mathsf{T}$$

for a Gaussian matrix $W \sim N(0,1)^{n\times d}$, a random unit vector u, a k-sparse vector v with entries in $\{0, \pm 1\}$ and a vector v' as defined in 3.41. Let $\hat{u}$ be a vector such that $\|\hat{u} - u\| \leqslant \varepsilon$ for some $0 \leqslant \varepsilon \leqslant \frac{1}{10}$. If $\lambda \gtrsim \frac{\sqrt{\log d}}{\tau}$, then we can compute in time $O(nd)$ an estimator $\hat{v}$ such that with probability at least $1 - \exp(-n)$

$$\|\hat{v} - v\| \lesssim (\varepsilon + \tau)\sqrt{k}\,.$$

9.2 Experimental results

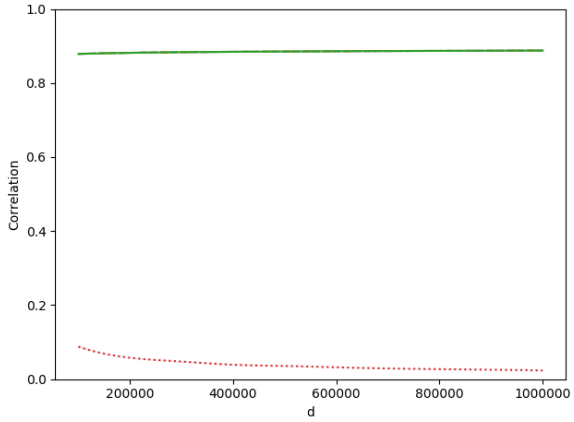
In this section we compare the performance of Diagonal Thresholding and SVD of degree 2, 4, 6 as in 9.3 on practical instances. The table below explains the regimes of the figures presented. We refer to Robust Sparse PCA as model 3.1 where the adversarial matrix E follows the distribution shown in 3.41. 9.2.1 contains a detailed report of the experimental setup.

	Standard Sparse PCA	**Robust Sparse PCA**
$k \geqslant \sqrt{d}$	Figure 9.1a for $\beta \geqslant \sqrt{\frac{d}{n}}$	Figure 9.1b for $\beta \geqslant \frac{k}{n}\left(\frac{d}{k}\right)^{1/2}$ Figure 9.1c for $\beta \geqslant \frac{k}{n}\left(\frac{d}{k}\right)^{1/3}$
$k \leqslant \sqrt{d}$	Figure 9.2 for $\beta \geqslant \frac{k}{\sqrt{n}}\sqrt{\log \frac{d}{k}}$	

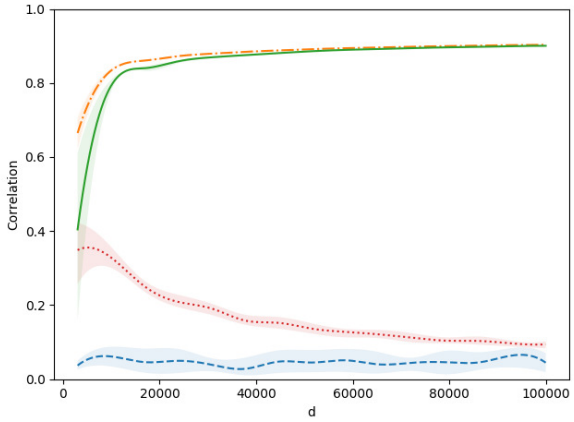
Table 9.1: Plots

9.2.1 Experimental Setup

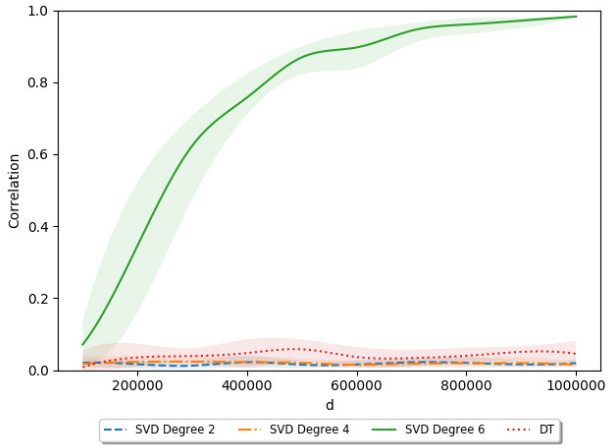
In the experiments, the instances were sampled from the planted distributions of 3.41 with the difference that $u \sim N(0, \mathrm{Id}_n)$ and v is a k-sparse unit vector obtained sampling a random k-subset $S \subseteq [d]$ and then a unit vector with support S. All the algorithms returned the top k coordinates of their estimation vector. Figure 9.1, 9.1 plot the absolute correlation between v and its estimate. Each plot was obtained averaging multiple independent runs on the same parameters, for each algorithm the shadowed part corresponds to the interval containing 50% of the results, the line corresponds to the mean of the results in such interval.
In Figure 9.1b, the adversarial matrix E is sampled according to model 3.41 for $s = 2$, that is the first 2 moments of Y are Gaussian. Similarly, in Figure 9.1c, E is sampled according to model 3.41 for $s = 4$, so the first 4 moments of Y are Gaussian.



(a) Standard Sparse PCA, with $k \geqslant \sqrt{d}$, $\beta \geqslant \sqrt{\frac{d}{n}}$



(b) Robust Sparse PCA with $k \geqslant \sqrt{d}$, $\beta \geqslant \frac{k}{n}\left(\frac{d}{n}\right)^{1/2}$



(c) Robust Sparse PCA with $k \geqslant \sqrt{d}$, $\beta \geqslant \frac{k}{n}\left(\frac{d}{n}\right)^{1/3}$

Figure 9.1: Forthe single spiked covariance model with $k > \sqrt{d}$, Figure 9.1a shows how the SVD algorithms works (with information theoretically optimal guarantees) and Diagonal Thresholding fails. Figures 9.1b, 9.1c show however how adversarial noise immediately breaks SVD with thresholding. In Figure 9.1b $\beta \gtrsim \frac{k}{n}\left(\frac{d}{k}\right)^{1/2}$, hence as d increases and becomes larger than n^2, SVD-4 returns a good estimate. We point out how how SVD-6 performs well even for $d \ll n^3$ when the signal is much larger than $\frac{k}{n}\left(\frac{d}{k}\right)^{1/3}$. Finally, Figure 9.1c shows how DT, SVD-4 and SVD-2 fails for $\beta = \Theta\left(\frac{k}{n}\left(\frac{d}{k}\right)^{1/3}\right)$, but as d grows towards n^3, SVD-6 approaches correlation 1.

Experiments were done on a laptop computer with a 3.5 GHz Intel Core i7 CPU and 16 GB of RAM, random instances were obtained using `Numpy` pseudo-random generator.

In non-robust settings, as well as in the adversarial model 3.41, the algorithm achieves high correlation under conditions similar (up to logarithmic terms) to those of the Sum-of-

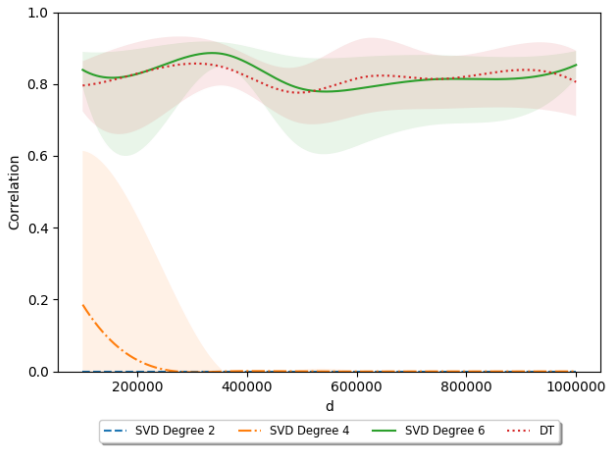


Figure 9.2: The figure shows settings in which $k \leqslant \sqrt{d}$. In this regime, among the algorithms considered, Diagonal Thresholding achieves asymptotically the most correlation. In practical settings however it is often the case that $\frac{k}{\sqrt{n}}\sqrt{\log d} \geqslant \Omega\left(\frac{k}{n}\left(\frac{d}{k}\right)^{1/3}\right)$ and hence also SVD-6 can accurately recover the signal.

Squares algorithm 3.29 (of degree 2, 4 and 6).

Bibliography

[20221] *Disclosure avoidance for the 2020 census: An introduction*, https://www2.census.gov/library/publications/decennial/2020/2020-census-disclosure-avoidance-handbook.pdf, 2021. 15

[AAP17] Thomas D Ahle, Martin Aumüller, and Rasmus Pagh, *Parameter-free locality sensitive hashing for spherical range reporting*, Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2017, pp. 239–256. 294

[AB09] Sanjeev Arora and Boaz Barak, *Computational complexity: a modern approach*, Cambridge University Press, 2009. 2

[ABARS20] Emmanuel Abbe, Enric Boix-Adsera, Peter Ralli, and Colin Sandon, *Graph powering and spectral robustness*, SIAM Journal on Mathematics of Data Science **2** (2020), no. 1, 132–157. 10

[Abb17] Emmanuel Abbe, *Community detection and stochastic block models: recent developments*, The Journal of Machine Learning Research **18** (2017), no. 1, 6446–6531. 9, 18

[ABH15] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall, *Exact recovery in the stochastic block model*, IEEE Transactions on information theory **62** (2015), no. 1, 471–487. 18, 230, 254

[ABS15] Sanjeev Arora, Boaz Barak, and David Steurer, *Subexponential algorithms for unique games and related problems*, Journal of the ACM (JACM) **62** (2015), no. 5, 1–25. 19

[ABW10] Benny Applebaum, Boaz Barak, and Avi Wigderson, *Public-key cryptography from different assumptions*, Proceedings of the forty-second ACM symposium on Theory of computing, 2010, pp. 171–180. 13, 14

[Ach09] Dimitris Achlioptas, *Random satisfiability.*, Handbook of Satisfiability **185** (2009), 245–270. 14

[AGK76] Rudolf Ahlswede, Peter Gács, and János Körner, *Bounds on conditional probabilities with applications in multi-user communication*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **34** (1976), no. 2, 157–177. 86, 107

[AI06] Alexandr Andoni and Piotr Indyk, *Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions*, 2006 47th annual IEEE symposium on foundations of computer science (FOCS'06), IEEE, 2006, pp. 459–468. 294

[AJT19] Vedat Levi Alev, Fernando Granha Jeronimo, and Madhur Tulsiani, *Approximating constraint satisfaction problems on high-dimensional expanders*, 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2019, pp. 180–201. 14, 178, 216

[AK07] Sanjeev Arora and Satyen Kale, *A combinatorial, primal-dual approach to semidefinite programs*, Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, 2007, pp. 227–236. 19, 20, 21, 237, 277, 280, 282, 284, 288, 289, 290, 297

[AL22] Hassan Ashtiani and Christopher Liaw, *Private and polynomial time algorithms for learning gaussians and beyond*, Conference on Learning Theory, PMLR, 2022, pp. 1075–1076. 15, 231

[AMM17] Haris Angelidakis, Konstantin Makarychev, and Yury Makarychev, *Algorithms for stable and perturbation-resilient problems*, Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, 2017, pp. 438–451. 278

[AN04] Noga Alon and Assaf Naor, *Approximating the cut-norm via grothendieck's inequality*, Proceedings of the thirty-sixth annual ACM symposium on Theory of computing, 2004, pp. 72–80. 31, 79, 163

[AOW15] Sarah R Allen, Ryan O'Donnell, and David Witmer, *How to refute a random csp*, 2015 IEEE 56th Annual Symposium on Foundations of Computer Science, IEEE, 2015, pp. 689–708. 14, 176, 178, 186, 187, 199, 205, 213, 214

[app17] *Learning with privacy at scale*, `https://docs-assets.developer.apple.com/ml-research/papers/learning-with-privacy-at-scale.pdf`, 2017, Accessed: 2022-11-06. 15

[ARV09] Sanjeev Arora, Satish Rao, and Umesh Vazirani, *Expander flows, geometric embeddings and graph partitioning*, Journal of the ACM (JACM) **56** (2009), no. 2, 1–37. 19, 20, 21, 279

[AS16] Emmanuel Abbe and Colin Sandon, *Achieving the ks threshold in the general stochastic block model with linearized acyclic belief propagation*, Advances in Neural Information Processing Systems **29** (2016). 79

[AW08] Arash A Amini and Martin J Wainwright, *High-dimensional analysis of semidefinite relaxations for sparse principal components*, 2008 IEEE international symposium on information theory, IEEE, 2008, pp. 2454–2458. 3, 4, 38, 47

[AWH13] George B Arfken, Hans J Weber, and Frank E Harris, *Chapter 15-legendre functions*, Mathematical Methods for Physicists (2013), 715–772. 350

[Bas92] Hyman Bass, *The ihara-selberg zeta function of a tree lattice*, International Journal of Mathematics **3** (1992), no. 06, 717–797. 189

[BBAP05] Jinho Baik, Gérard Ben Arous, and Sandrine Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Annals of Probability (2005), 1643–1697. 4, 36, 158

[BBH+12] Boaz Barak, Fernando GSL Brandao, Aram W Harrow, Jonathan Kelner, David Steurer, and Yuan Zhou, *Hypercontractivity, sum-of-squares proofs, and their applications*, Proceedings of the forty-fourth annual ACM symposium on Theory of computing, 2012, pp. 307–326. 31

[BCLS87] Thang Nguyen Bui, Soma Chaudhuri, Frank Thomson Leighton, and Michael Sipser, *Graph bisection algorithms with good average case behavior*, Combinatorica **7** (1987), no. 2, 171–191. 20, 277

[BDH+20] Ainesh Bakshi, Ilias Diakonikolas, Samuel B Hopkins, Daniel Kane, Sushrut Karmalkar, and Pravesh K Kothari, *Outlier-robust clustering of gaussians and other non-spherical mixtures*, 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2020, pp. 149–159. 12

[BDJ+22] Ainesh Bakshi, Ilias Diakonikolas, He Jia, Daniel M Kane, Pravesh K Kothari, and Santosh S Vempala, *Robustly learning mixtures of k arbitrary gaussians*, Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, 2022, pp. 1234–1247. 12

[BGG+16] Vijay VSP Bhattiprolu, Mrinalkanti Ghosh, Venkatesan Guruswami, Euiwoong Lee, and Madhur Tulsiani, *Multiplicative approximations for polynomial optimization over the unit sphere.*, Electron. Colloquium Comput. Complex., vol. 23, 2016, p. 185. 22

[BHK+19] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin, *A nearly tight sum-of-squares lower bound for the planted clique problem*, SIAM Journal on Computing **48** (2019), no. 2, 687–735. 37, 47, 62, 63

[BJK05] Andrei Bulatov, Peter Jeavons, and Andrei Krokhin, *Classifying the complexity of constraints using finite algebras*, SIAM journal on computing **34** (2005), no. 3, 720–742. 14

[BKS14] Boaz Barak, Jonathan A Kelner, and David Steurer, *Rounding sum-of-squares relaxations*, Proceedings of the forty-sixth annual ACM symposium on Theory of computing, 2014, pp. 31–40. 22

[BKS15] _____, *Dictionary learning and tensor decomposition via the sum-of-squares method*, Proceedings of the forty-seventh annual ACM symposium on Theory of computing, 2015, pp. 143–151. 14, 22

[BKW20] Afonso S Bandeira, Dmitriy Kunisky, and Alexander S Wein, *Computational hardness of certifying bounds on constrained pca problems*, 11th Innovations in Theoretical Computer Science Conference (ITCS 2020), vol. 151, 2020, p. 78. 62, 340

[BL12] Yonatan Bilu and Nathan Linial, *Are stable instances easy?*, Combinatorics, Probability and Computing **21** (2012), no. 5, 643–660. 20, 277, 278

[BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié, *Non-backtracking spectrum of random graphs: community detection and non-regular ramanujan graphs*, 2015 IEEE 56th Annual Symposium on Foundations of Computer Science, IEEE, 2015, pp. 1347–1357. 79, 83, 181, 182, 193, 194, 504

[BMR21] Jess Banks, Sidhanth Mohanty, and Prasad Raghavendra, *Local statistics, semidefinite programming, and community detection*, Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2021, pp. 1298–1316. 10, 81, 83

[Bop87] Ravi B Boppana, *Eigenvalues and graph bisection: An average-case analysis*, 28th Annual Symposium on Foundations of Computer Science (sfcs 1987), IEEE, 1987, pp. 280–285. 20, 277

[BR13] Quentin Berthet and Philippe Rigollet, *Computational lower bounds for sparse pca*, arXiv preprint arXiv:1304.0828 (2013). 3, 4, 38

[BRS11] Boaz Barak, Prasad Raghavendra, and David Steurer, *Rounding semidefinite programming hierarchies via global correlation*, 2011 ieee 52nd annual symposium on foundations of computer science, IEEE, 2011, pp. 472–481. 178, 183, 217, 219, 226

[BS16] Boaz Barak and David Steurer, *Proofs, beliefs, and algorithms through the lens of sum-of-squares*, Course notes: http://www. sumofsquares. org/public/index. html **1** (2016). 28

[BSB02] Eli Ben-Sasson and Yonatan Bilu, *A gap in average proof complexity*, Electronic Colloquium on Computational Complexity (ECCC), vol. 9, Citeseer, 2002. 13

[BST14] Raef Bassily, Adam Smith, and Abhradeep Thakurta, *Private empirical risk minimization: Efficient algorithms and tight error bounds*, 2014 IEEE 55th annual symposium on foundations of computer science, IEEE, 2014, pp. 464–473. 237

[CCAd+23] Hongjie Chen, Vincent Cohen-Addad, Tommaso d'Orsi, Alessandro Epasto, Jacob Imola, David Steurer, and Stefan Tiegel, *Private estimation algorithms for stochastic block models and mixture models*, arXiv preprint arXiv:2301.04822 (2023). 228

[Cd21] Davin Choo and Tommaso d'Orsi, *The complexity of sparse tensor pca*, Advances in Neural Information Processing Systems **34** (2021), 7993–8005. 4, 25, 39, 44

[Cd22] Hongjie Chen and Tommaso d'Orsi, *On the well-spread property and its relation to linear regression*, Conference on Learning Theory, PMLR, 2022, pp. 3905–3935. 25

[CdM23] Vincent Cohen-Addad, Tommaso d'Orsi, and Aida Mousavifar, *A near-linear time approximation algorithm for beyond-worst-case graph clustering*, to appear (2023). 24, 277

[Cha16] Siu On Chan, *Approximation resistance from pairwise-independent subgroups*, Journal of the ACM (JACM) **63** (2016), no. 3, 1–32. 14

[CKL+22] Li Chen, Rasmus Kyng, Yang P Liu, Richard Peng, Maximilian Probst Gutenberg, and Sushant Sachdeva, *Maximum flow and minimum-cost flow in almost-linear time*, 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2022, pp. 612–623. 20, 277, 280, 282

[CKM+21] Edith Cohen, Haim Kaplan, Yishay Mansour, Uri Stemmer, and Eliad Tsfadia, *Differentially-private clustering of easy instances*, International Conference on Machine Learning, PMLR, 2021, pp. 2049–2059. 15, 17, 24, 231

[CKMM19] Vincent Cohen-Addad, Varun Kanade, Frederik Mallmann-Trenn, and Claire Mathieu, *Hierarchical clustering: Objective functions and algorithms*, Journal of the ACM (JACM) **66** (2019), no. 4, 1–42. 21, 278, 300, 301, 302, 303

[CLP02] Andrea Crisanti, Luca Leuzzi, and Giorgio Parisi, *The 3-sat problem with large number of clauses in the infinity-replica symmetry breaking scheme*, Journal of Physics A: Mathematical and General **35** (2002), no. 3, 481. 13

[CMS11] Kamalika Chaudhuri, Claire Monteleoni, and Anand D Sarwate, *Differentially private empirical risk minimization.*, Journal of Machine Learning Research **12** (2011), no. 3. 237

[CMW13] T Tony Cai, Zongming Ma, and Yihong Wu, *Sparse pca: Optimal rates and adaptive estimation*, The Annals of Statistics **41** (2013), no. 6, 3074. 4

[COGL07] Amin Coja-Oghlan, Andreas Goerdt, and André Lanka, *Strong refutation heuristics for random k-sat*, Combinatorics, Probability and Computing **16** (2007), no. 1, 5–28. 14

[CPRT22] Flavio Chierichetti, Alessandro Panconesi, Giuseppe Re, and Luca Trevisan, *Spectral robustness for correlation clustering reconstruction in semi-adversarial models*, International Conference on Artificial Intelligence and Statistics, PMLR, 2022, pp. 10852–10880. 20, 22, 277

[CRV15] Peter Chin, Anup Rao, and Van Vu, *Stochastic block model and community detection in sparse graphs: A spectral algorithm with optimal rate of recovery*, Conference on Learning Theory, PMLR, 2015, pp. 391–423. 164, 498

[Dan16] Amit Daniely, *Complexity theoretic limitations on learning halfspaces*, Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, 2016, pp. 105–117. 14

[Das16] Sanjoy Dasgupta, *A cost function for similarity-based hierarchical clustering*, Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, 2016, pp. 118–127. 278, 300, 301

[DdH23] Jingqiu Ding, Tommaso d'Orsi, and Yiding Hua, *Node robust recovery for stochastic block models*. 24, 157

[DdL+22] Jingqiu Ding, Tommaso d'Orsi, Chih-Hung Liu, David Steurer, and Stefan Tiegel, *Fast algorithm for overcomplete order-3 tensor decomposition*, Conference on Learning Theory, PMLR, 2022, pp. 3741–3799. 22, 25

[DdNS22] Jingqiu Ding, Tommaso d'Orsi, Rajai Nasser, and David Steurer, *Robust recovery for stochastic block models*, 2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2022, pp. 387–394. 24, 83, 167

[DF86] Martin E Dyer and Alan M Frieze, *Fast solution of some random np-hard problems*, 27th Annual Symposium on Foundations of Computer Science (sfcs 1986), IEEE Computer Society, 1986, pp. 331–336. 20, 277

[dGJL04] Alexandre d'Aspremont, Laurent Ghaoui, Michael Jordan, and Gert Lanckriet, *A direct formulation for sparse pca using semidefinite programming*, Advances in neural information processing systems **17** (2004). 38, 47

[DI98] Tassos Dimitriou and Russell Impagliazzo, *Go with the winners for graph bisection.*, SODA, vol. 98, 1998, pp. 510–520. 20, 277

[DK19] Ilias Diakonikolas and Daniel M Kane, *Recent advances in algorithmic high-dimensional robust statistics*, arXiv preprint arXiv:1911.05911 (2019). 2, 12

[DKMZ11] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová, *Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications*, Physical Review E **84** (2011), no. 6, 066106. 9, 18

[dKNS20] Tommaso d'Orsi, Pravesh K Kothari, Gleb Novikov, and David Steurer, *Sparse pca: algorithms, adversarial perturbations and certificates*, 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2020, pp. 553–564. 2, 4, 8, 22, 24, 34, 35, 36, 38, 46, 304

[DKWB23] Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira, *Subexponential-time algorithms for sparse pca*, Foundations of Computational Mathematics (2023), 1–50. 4, 39, 44, 45, 64

[dLN+21] Tommaso d'Orsi, Chih-Hung Liu, Rajai Nasser, Gleb Novikov, David Steurer, and Stefan Tiegel, *Consistent estimation for pca and sparse regression with oblivious outliers*, Advances in Neural Information Processing Systems **34** (2021), 25427–25438. 25

[DLSS14] Amit Daniely, Nati Linial, and Shai Shalev-Shwartz, *From average case complexity to improper learning complexity*, Proceedings of the forty-sixth annual ACM symposium on Theory of computing, 2014, pp. 441–448. 14

[DM14] Yash Deshpande and Andrea Montanari, *Sparse pca via covariance thresholding*, Advances in Neural Information Processing Systems **27** (2014). 4, 6, 38, 43, 344, 355

[DMNS06] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith, *Calibrating noise to sensitivity in private data analysis*, Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006. Proceedings 3, Springer, 2006, pp. 265–284. 16, 240

[DMS17] Amir Dembo, Andrea Montanari, and Subhabrata Sen, *Extremal cuts of sparse random graphs*. 78

[dNNS23] Tommaso d'Orsi, Rajai Nasser, Gleb Novikov, and David Steurer, *Higher degree sum-of-squares relaxations robust against oblivious outliers*, Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2023, pp. 3513–3550. 22, 25

[dNS21] Tommaso d'Orsi, Gleb Novikov, and David Steurer, *Consistent regression when oblivious outliers overwhelm*, International Conference on Machine Learning, PMLR, 2021, pp. 2297–2306. 25

[DNT15] Cynthia Dwork, Aleksandar Nikolov, and Kunal Talwar, *Efficient algorithms for privately releasing marginals via convex relaxations*, Discrete & Computational Geometry **53** (2015), 650–673. 15

[DSS15] Jian Ding, Allan Sly, and Nike Sun, *Proof of the satisfiability conjecture for large k*, Proceedings of the forty-seventh annual ACM symposium on Theory of computing, 2015, pp. 59–68. 14

[dT23] Tommaso d'Orsi and Luca Trevisan, *A ihara-bass formula for non-boolean matrices and strong refutations of random csps*, Computational Complexity Conference (2023). 22, 24

[EKZ22] Ronen Eldan, Frederic Koehler, and Ofer Zeitouni, *A spectral condition for spectral gap: fast mixing in high-temperature ising models*, Probability Theory and Related Fields **182** (2022), no. 3, 1035–1051. 238

[Ela15] Andrew Eland, *Tackling urban mobility with technology*, Google Europe Blog, November **18** (2015). 15

[FC20] Yingjie Fei and Yudong Chen, *Achieving the bayes error rate in synchronization and block models by sdp, robustly*, IEEE Transactions on Information Theory **66** (2020), no. 6, 3929–3953. 239

[Fei02] Uriel Feige, *Relations between average case complexity and approximation complexity*, Proceedings of the thiry-fourth annual ACM symposium on Theory of computing, 2002, pp. 534–543. 13, 14

[Fei07] ———, *Refuting smoothed 3cnf formulas*, 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), IEEE, 2007, pp. 407–417. 13, 178

[FGK05] Joel Friedman, Andreas Goerdt, and Michael Krivelevich, *Recognizing more unsatisfiable random k-sat instances efficiently*, SIAM Journal on Computing **35** (2005), no. 2, 408–430. 14, 176

[FK01] Uriel Feige and Joe Kilian, *Heuristics for semirandom graph problems*, Journal of Computer and System Sciences **63** (2001), no. 4, 639–671. 9, 20, 79, 277

[FKP+19] Noah Fleming, Pravesh Kothari, Toniann Pitassi, et al., *Semialgebraic proofs and efficient algorithm design*, Foundations and Trends® in Theoretical Computer Science **14** (2019), no. 1-2, 1–221. 272, 511

[FLM20] Andreas Emil Feldmann, Euiwoong Lee, and Pasin Manurangsi, *A survey on approximation in parameterized complexity: Hardness and algorithms*, Algorithms **13** (2020), no. 6, 146. 2

[FLP15] Dimitris Fotakis, Michael Lampis, and Vangelis Th Paschos, *Sub-exponential approximation schemes for csps: From dense to almost sparse*, arXiv preprint arXiv:1507.04391 (2015). 14

[FM17] Zhou Fan and Andrea Montanari, *How well do local algorithms solve semidefinite programs?*, Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, 2017, pp. 604–614. 181, 187

[FO05] Uriel Feige and Eran Ofek, *Spectral techniques applied to sparse random graphs*, Random Structures & Algorithms **27** (2005), no. 2, 251–275. 161, 164, 180, 498

[GKM22] Venkatesan Guruswami, Pravesh K Kothari, and Peter Manohar, *Algorithms and certificates for boolean csp refutation: smoothed is no harder than random*, Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, 2022, pp. 678–689. 15, 178, 183

[GLM16] Lennart Gulikers, Marc Lelarge, and Laurent Massoulié, *Non-backtracking spectrum of degree-corrected stochastic block models*, arXiv preprint arXiv:1609.02487 (2016). 83

[GLS81] Martin Grötschel, László Lovász, and Alexander Schrijver, *The ellipsoid method and its consequences in combinatorial optimization*, Combinatorica **1** (1981), 169–197. 28, 29

[GV16] Olivier Guédon and Roman Vershynin, *Community detection in sparse networks via grothendieck's inequality*, Probability Theory and Related Fields **165** (2016), no. 3-4, 1025–1049. 9, 17, 77, 78, 234, 238, 239, 247, 256, 509

[GW95] Michel X Goemans and David P Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, Journal of the ACM (JACM) **42** (1995), no. 6, 1115–1145. 19, 78

[HKM22] Samuel B Hopkins, Gautam Kamath, and Mahbod Majid, *Efficient mean estimation with pure differential privacy via a sum-of-squares exponential mechanism*, Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, 2022, pp. 1406–1417. 238, 258

[HKP+17] Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer, *The power of sum-of-squares for detecting hidden structures*, 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2017, pp. 720–731. 4, 5, 22, 37, 62, 63, 64, 81

[HL18] Samuel B Hopkins and Jerry Li, *Mixture models, robustness, and sum of squares proofs*, Proceedings of the 50th Annual ACM SIGACT Symposium on Theory

of Computing, 2018, pp. 1021–1034. 12, 17, 22, 54, 231, 232, 235, 244, 264, 272, 511

[Hop18] Samuel Brink Klevit Hopkins, *Statistical inference and the sum of squares method*. 2, 5, 6, 37, 47, 62, 63

[HS17] Samuel B Hopkins and David Steurer, *Efficient bayesian estimation from few samples: community detection and related problems*, 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2017, pp. 379–390. 37, 47, 62, 63, 64, 79, 80, 81, 166

[HSS15] Samuel B Hopkins, Jonathan Shi, and David Steurer, *Tensor principal component analysis via sum-of-square proofs*, Conference on Learning Theory, PMLR, 2015, pp. 956–1006. 22

[HSS19] Samuel B Hopkins, Tselil Schramm, and Jonathan Shi, *A robust spectral algorithm for overcomplete tensor decomposition*, Conference on Learning Theory, PMLR, 2019, pp. 1683–1722. 22

[HSSS16] Samuel B Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer, *Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors*, Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, 2016, pp. 178–191. 22, 23

[HST06] Matthew D Horton, HM Stark, and Audrey A Terras, *What are zeta functions of graphs and what are they good for?*, Contemporary Mathematics **415** (2006), 173–190. 181, 187

[HY04] Jun He and Xin Yao, *A study of drift analysis for estimating computation time of evolutionary algorithms*, Natural Computing **3** (2004), 21–35. 169

[IKW02] Russell Impagliazzo, Valentine Kabanets, and Avi Wigderson, *In search of an easy witness: Exponential time vs. probabilistic polynomial time*, Journal of Computer and System Sciences **65** (2002), no. 4, 672–694. 14

[Ind01] Piotr Indyk, *High-dimensional computational geometry*, stanford university, 2001. 294

[JL09] Iain M Johnstone and Arthur Yu Lu, *On consistency and sparsity for principal components analysis in high dimensions*, Journal of the American Statistical Association **104** (2009), no. 486, 682–693. 4, 6, 38

[JMRT16] Adel Javanmard, Andrea Montanari, and Federico Ricci-Tersenghi, *Phase transitions in semidefinite relaxations*, Proceedings of the National Academy of Sciences **113** (2016), no. 16, E2218–E2223. 10

[Joh84] William B Johnson, *Extensions of lipschitz mappings into a hilbert space*, Contemp. Math. **26** (1984), 189–206. 284

[JS93] Mark Jerrum and Gregory B Sorkin, *Simulated annealing for graph bisection*, IEEE, 1993. 20, 277

[Kho10] Subhash Khot, *Inapproximability of np-complete problems, discrete fourier analysis, and geometry*, Proceedings of the International Congress of Mathematicians 2010 (ICM 2010) (In 4 Volumes) Vol. I: Plenary Lectures and Ceremonies Vols. II–IV: Invited Lectures, World Scientific, 2010, pp. 2676–2697. 2

[KKM18] Adam Klivans, Pravesh K Kothari, and Raghu Meka, *Efficient algorithms for outlier-robust regression*, Conference On Learning Theory, PMLR, 2018, pp. 1420–1430. 54

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell, *Optimal inapproximability results for max-cut and other 2-variable csps?*, SIAM Journal on Computing **37** (2007), no. 1, 319–357. 19

[KLR22] Frederic Koehler, Holden Lee, and Andrej Risteski, *Sampling approximately low-rank ising models: Mcmc meets variational methods*, Conference on Learning Theory, PMLR, 2022, pp. 4945–4988. 238

[KMV22] Pravesh Kothari, Pasin Manurangsi, and Ameya Velingker, *Private robust estimation by stabilizing convex relaxations*, Conference on Learning Theory, PMLR, 2022, pp. 723–777. 15, 231, 237, 241

[KN11] Subhash Khot and Assaf Naor, *Grothendieck-type inequalities in combinatorial optimization*, arXiv preprint arXiv:1108.2464 (2011). 79

[KNV13] Robert Krauthgamer, Boaz Nadler, and Dan Vilenchik, *Do semidefinite relaxations solve sparse pca up to the information limit?*, Annals of Statistics **43** (2013), no. 3, 1300–1322. 4, 6, 36, 47

[Kot22] Pravesh K. Kothari, Personal communication (2022). 179

[KS17] Pravesh K Kothari and Jacob Steinhardt, *Better agnostic clustering via relaxed tensor norms*, arXiv preprint arXiv:1711.07465 (2017). 54

[KSS18] Pravesh K Kothari, Jacob Steinhardt, and David Steurer, *Robust moment estimation and improved clustering via sum of squares*, Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, 2018, pp. 1035–1046. 12, 17, 22, 54, 55, 231, 232, 235, 243, 244, 264, 519

[KSSU19] Gautam Kamath, Or Sheffet, Vikrant Singhal, and Jonathan Ullman, *Differentially private algorithms for learning mixtures of separated gaussians*, Advances in Neural Information Processing Systems **32** (2019). 15, 17, 231

[KST12] Daniel Kifer, Adam Smith, and Abhradeep Thakurta, *Private convex empirical risk minimization and high-dimensional regression*, Conference on Learning Theory, JMLR Workshop and Conference Proceedings, 2012, pp. 25–1. 237

[KT13] Michael Kapralov and Kunal Talwar, *On differentially private low rank approximation*, Proceedings of the twenty-fourth annual ACM-SIAM symposium on Discrete algorithms, SIAM, 2013, pp. 1395–1414. 239

[KV15] Subhash A Khot and Nisheeth K Vishnoi, *The unique games conjecture, integrality gap for cut problems and embeddability of negative-type metrics into ℓ-1*, Journal of the ACM (JACM) **62** (2015), no. 1, 1–39. 19

[KV17] Vishesh Karwa and Salil Vadhan, *Finite sample differentially private confidence intervals*, arXiv preprint arXiv:1711.03908 (2017). 242

[KWB22] Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira, *Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio*, Mathematical Analysis, its Applications and Computation: ISAAC 2019, Aveiro, Portugal, July 29–August 2, Springer, 2022, pp. 1–50. 5

[Las01] Jean B Lasserre, *New positive semidefinite relaxations for nonconvex quadratic programs*, Advances in Convex Analysis and Global Optimization: Honoring the Memory of C. Caratheodory (1873–1950) (2001), 319–331. 28

[LCY90] Lucien Le Cam and Lo Yang, *Locally asymptotically normal families*, Asymptotics in Statistics: Some Basic Concepts (1990), 52–98. 6, 63

[LL22] Allen Liu and Jerry Li, *Clustering mixtures with almost optimal separation in polynomial time*, Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing, 2022, pp. 1248–1261. 17, 231, 232

[LM00] Beatrice Laurent and Pascal Massart, *Adaptive estimation of a quadratic functional by model selection*, Annals of Statistics (2000), 1302–1338. 352

[LM22] Allen Liu and Ankur Moitra, *Minimax rates for robust community detection*, 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2022, pp. 823–831. 2, 12, 13, 157, 164, 498

[Mar86] Katalin Marton, *A simple proof of the blowing-up lemma (corresp.)*, IEEE Transactions on Information Theory **32** (1986), no. 3, 445–446. 86, 107

[Mar96] ———, *Bounding d-distance by informational divergence: a method to prove measure concentration*, The Annals of Probability **24** (1996), no. 2, 857–866. 86, 107

[Mas14] Laurent Massoulié, *Community detection thresholds and the weak ramanujan property*, Proceedings of the forty-sixth annual ACM symposium on Theory of computing, 2014, pp. 694–703. 9, 18, 79

[McS01] Frank McSherry, *Spectral partitioning of random graphs*, Proceedings 42nd IEEE Symposium on Foundations of Computer Science, IEEE, 2001, pp. 529–537. 20, 277

[MMV12] Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan, *Approximation algorithms for semi-random partitioning problems*, Proceedings of the forty-fourth annual ACM symposium on Theory of computing, 2012, pp. 367–384. 20, 21, 277, 278, 279, 280, 287, 288

[MMV14] ———, *Constant factor approximation for balanced cut in the pie model*, Proceedings of the forty-sixth annual ACM symposium on Theory of computing, 2014, pp. 41–49. 20, 277

[MNS15a] Elchanan Mossel, Joe Neeman, and Allan Sly, *Consistency thresholds for the planted bisection model*, Proceedings of the forty-seventh annual ACM symposium on Theory of computing, 2015, pp. 69–75. 230, 254

[MNS15b] ———, *Reconstruction and estimation in the planted partition model*, Probability Theory and Related Fields **162** (2015), 431–461. 9, 18, 172

[MNS18] ———, *A proof of the block model threshold conjecture*, Combinatorica **38** (2018), no. 3, 665–708. 9, 18, 79, 80

[MNVT22] Mohamed S Mohamed, Dung Nguyen, Anil Vullikanti, and Ravi Tandon, *Differentially private community detection for stochastic block models*, International Conference on Machine Learning, PMLR, 2022, pp. 15858–15894. 15, 18, 229, 230, 258

[MPW16] Ankur Moitra, William Perry, and Alexander S Wein, *How robust are reconstruction thresholds for community detection?*, Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, 2016, pp. 828–841. 10, 20, 79

[MR08] Dana Moshkovitz and Ran Raz, *Two-query pcp with subconstant error*, Journal of the ACM (JACM) **57** (2008), no. 5, 1–29. 14

[MR16] Pasin Manurangsi and Prasad Raghavendra, *A birthday repetition theorem and complexity of approximating dense csps*, arXiv preprint arXiv:1607.02986 (2016). 219

[MS16] Andrea Montanari and Subhabrata Sen, *Semidefinite programs on sparse random graphs and their application to community detection*, Proceedings of the forty-eighth annual ACM symposium on Theory of Computing, 2016, pp. 814–827. 9, 10, 11, 17, 77, 158, 164, 172, 234, 498

[MSS16] Tengyu Ma, Jonathan Shi, and David Steurer, *Polynomial-time tensor decompositions with sum-of-squares*, 2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2016, pp. 438–446. 22, 28

[MSVV21] Andres Munoz, Umar Syed, Sergei Vassilvtiskii, and Ellen Vitercik, *Private optimization without constraint violations*, International Conference on Artificial Intelligence and Statistics, PMLR, 2021, pp. 2557–2565. 237, 238

[MT07] Frank McSherry and Kunal Talwar, *Mechanism design via differential privacy*, 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), IEEE, 2007, pp. 94–103. 229, 238, 253, 255

[MWA06] Baback Moghaddam, Yair Weiss, and Shai Avidan, *Generalized spectral bounds for sparse lda*, Proceedings of the 23rd international conference on Machine learning, 2006, pp. 641–648. 47

[Nat95] Balas Kausik Natarajan, *Sparse approximate solutions to linear systems*, SIAM journal on computing **24** (1995), no. 2, 227–234. 47

[Nes00] Yurii Nesterov, *Squared functional systems and optimization problems*, High performance optimization (2000), 405–440. 28

[NP33] Jerzy Neyman and Egon Sharpe Pearson, *Ix. on the problem of the most efficient tests of statistical hypotheses*, Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character **231** (1933), no. 694-706, 289–337. 6, 62

[NRS07] Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith, *Smooth sensitivity and sampling in private data analysis*, Proceedings of the thirty-ninth annual ACM symposium on Theory of computing, 2007, pp. 75–84. 231

[O'D17] Ryan O'Donnell, *Sos is not obviously automatizable, even approximately*, 8th Innovations in Theoretical Computer Science Conference (ITCS 2017), Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2017. 28

[Par00] Pablo A Parrilo, *Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization*, California Institute of Technology, 2000. 28

[Pen20] Pan Peng, *Robust clustering oracle and local reconstructor of cluster structure of graphs*, Proceedings of the Fourteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, 2020, pp. 2953–2972. 20, 277

[PR22] Aaron Potechin and Goutham Rajendran, *Sub-exponential time sum-of-squares lower bounds for principal components analysis*, Advances in Neural Information Processing Systems, 2022. 4

[PS17] Aaron Potechin and David Steurer, *Exact tensor completion with sum-of-squares*, Conference on Learning Theory, PMLR, 2017, pp. 1619–1673. 22

[PWBM18] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra, *Optimality and sub-optimality of pca i: Spiked random matrix models*, The Annals of Statistics **46** (2018), no. 5, 2416–2451. 158

[Rag08] Prasad Raghavendra, *Optimal algorithms and inapproximability results for every csp?*, Proceedings of the fortieth annual ACM symposium on Theory of computing, 2008, pp. 245–254. 19

[Rag09] ———, *Approximating np-hard problems efficient algorithms and their limits*, University of Washington, 2009. 14

[RRS17] Prasad Raghavendra, Satish Rao, and Tselil Schramm, *Strongly refuting random csps below the spectral threshold*, Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, 2017, pp. 121–131. 14, 22, 183

[RS+13] Maxim Raginsky, Igal Sason, et al., *Concentration of measure inequalities in information theory, communications, and coding*, Foundations and Trends® in Communications and Information Theory **10** (2013), no. 1-2, 1–246. 86, 107

[RSS18] Prasad Raghavendra, Tselil Schramm, and David Steurer, *High dimensional estimation via sum-of-squares proofs*, Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018, World Scientific, 2018, pp. 3389–3423. 2, 81

[RST12] Prasad Raghavendra, David Steurer, and Madhur Tulsiani, *Reductions between expansion problems*, 2012 IEEE 27th Conference on Computational Complexity, IEEE, 2012, pp. 64–73. 19

[RT12] Prasad Raghavendra and Ning Tan, *Approximating csps with global cardinality constraints using sdp hierarchies*, Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms, SIAM, 2012, pp. 373–387. 219

[Rud99] Mark Rudelson, *Random vectors in the isotropic position*, Journal of Functional Analysis **164** (1999), no. 1, 60–72. 351

[RV17] Oded Regev and Aravindan Vijayaraghavan, *On learning mixtures of well-separated gaussians*, 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2017, pp. 85–96. 231, 232

[S+17] Konrad Schmüdgen et al., *The moment problem*, vol. 9, Springer, 2017. 349

[Sch22] Tselil Schramm, *The sum-of-squares algorithmic paradigm in statistics*, Lecture notes (2022). 32, 219

[SCS13] Shuang Song, Kamalika Chaudhuri, and Anand D Sarwate, *Stochastic gradient descent with differentially private updates*, 2013 IEEE global conference on signal and information processing, IEEE, 2013, pp. 245–248. 237

[She09] Jonah Sherman, *Breaking the multicommodity flow barrier for o (sqrtlog n)-approximations to sparsest cut*, 2009 50th Annual IEEE Symposium on Foundations of Computer Science, IEEE, 2009, pp. 363–372. 19, 20, 277, 280, 288, 289, 290

[Sho87] Naum Z Shor, *Quadratic optimization problems*, Soviet Journal of Computer and Systems Sciences **25** (1987), 1–11. 28

[SM19] Ludovic Stephan and Laurent Massoulié, *Robustness of spectral methods for community detection*, Conference on Learning Theory, PMLR, 2019, pp. 2831–2860. 12

[SS17] Tselil Schramm and David Steurer, *Fast and robust tensor decomposition with applications to dictionary learning*, Conference on Learning Theory, PMLR, 2017, pp. 1760–1793. 22

[ST21] David Steurer and Stefan Tiegel, *Sos degree reduction with applications to clustering and robust moment estimation*, Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA), SIAM, 2021, pp. 374–393. 17, 231, 232

[Ste10a] David Steurer, *Fast sdp algorithms for constraint satisfaction problems*, Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms, SIAM, 2010, pp. 684–697. 19, 237, 280, 282, 285

[Ste10b] ———, *On the complexity of unique games and graph expansion*, Citeseer, 2010. 278

[TCK+22] Eliad Tsfadia, Edith Cohen, Haim Kaplan, Yishay Mansour, and Uri Stemmer, *Friendlycore: Practical differentially private aggregation*, International Conference on Machine Learning, PMLR, 2022, pp. 21828–21863. 15, 231

[Tro12] Joel A Tropp, *User-friendly tail bounds for sums of random matrices*, Foundations of computational mathematics **12** (2012), 389–434. 353, 508

[Wai19] Martin J Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*, vol. 48, Cambridge university press, 2019. 1, 2, 352, 353, 510

[WEAM19] Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore, *The kikuchi hierarchy and tensor pca*, 2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2019, pp. 1446–1468. 183

[WF09] Yusuke Watanabe and Kenji Fukumizu, *Graph zeta function in the bethe free energy and loopy belief propagation*, Advances in Neural Information Processing Systems **22** (2009). 181

[WYX17] Di Wang, Minwei Ye, and Jinhui Xu, *Differentially private empirical risk minimization revisited: Faster and more general*, Advances in Neural Information Processing Systems **30** (2017). 237

[ZSWB22] Ilias Zadik, Min Jae Song, Alexander S Wein, and Joan Bruna, *Lattice-based methods surpass sum-of-squares in clustering*, Conference on Learning Theory, PMLR, 2022, pp. 1247–1248. 5

[ZZ16] Anderson Y Zhang and Harrison H Zhou, *Minimax rates of community detection in stochastic block models*, The Annals of Statistics **44** (2016), no. 5, 2252–2280. 230, 254

Part IV

Appendices

Appendix A

Deferred proofs and addendum to Chapter 3

A.1 Thresholding algorithms are fragile

In this section we formalize the discussions of Chapter 3 on fragile algorithms and show that SVD with Thresholding, Diagonal Thresholding and Covariance Thresholding are indeed not resilient to adversarial perturbations.

A.1.1 SVD with thresholding is fragile

The polynomial-time algorithm presented in Chapter 3 for the strong-signal regime is highly sensitive to small adversarial perturbations. Concretely, this can be shown constructing E with entries bounded $\tilde{O}(1/\sqrt{n})$ so that eigenvectors of $Y^\mathsf{T}Y$ cannot be used to recover u_0.

Consider $E = -\gamma u_0 u_0^\mathsf{T} W$ for some $0 < \gamma < \|u_0\|^2$ that we will choose later. Then $Y = \sqrt{\beta} u_0 v_0^\mathsf{T} + (\mathrm{Id} - \gamma u_0 u_0^\mathsf{T}) W$ and

$$\begin{aligned} YY^\mathsf{T} =& \beta u_0 u_0^\mathsf{T} + (\mathrm{Id} - \gamma u u_0^\mathsf{T}) W W^\mathsf{T} (\mathrm{Id} - \gamma u_0 u_0^\mathsf{T}) \\ &+ \sqrt{\beta}\left(u_0 v_0^\mathsf{T} W (\mathrm{Id} - \gamma u_0 u_0^\mathsf{T}) + (\mathrm{Id} - \gamma u_0 u_0^\mathsf{T}) W^\mathsf{T} v_0 u_0^\mathsf{T}\right) \end{aligned}$$

Hence with high probability,

$$\frac{1}{\|u_0\|^2}\|u_0^\mathsf{T} Y\|^2 = z + \beta\|u_0\|^2 + \gamma^2\|u_0\|^2 d - 2\gamma\|u_0\| d + \tilde{O}\left(\sqrt{\beta/n}\right),$$

where z has a χ^2-distribution with d degrees of freedom. On the other hand notice that for a unit vector x orthogonal to u_0 and independent of W, we get $\|x^\mathsf{T} Y\|^2 = \|x^\mathsf{T} W\|^2$ which has the same distribution as z. So our claim follows choosing γ so that $2\gamma\|u_0\| - \gamma^2\|u_0\|^2 = \beta \cdot \frac{\|u_0\|^2}{d} + \tilde{O}\left(\frac{1}{d}\sqrt{\beta/n}\right)$. Indeed then $u_0 Y$ has the same distribution as z. Now, since with high probability $\|u_0\|^2 \leqslant 2n$, if $d/n \gtrsim \beta$, such a γ exists.

A.1.2 Diagonal thresholding is fragile

Recall that Diagonal thresholding finds the top k diagonal entries of the covariance matrix and output a top eigenvector of the corresponding $k \times k$ principal submatrix. We shows here that a simple adversary can make diagonal entries in $[d] \setminus \operatorname{supp}\{v_0\}$ larger than diagonal entries in $\operatorname{supp}\{v_0\}$, hence leading the algorithm to choose a submatrix which contain no information about the sparse vector.

Concretely, the algorithm can be written as follows:

Algorithm A.1 (Diagonal Thresholding)**.**

Given: Sample matrix Y of form 3.1 where v_0 is a flat vector.

Estimate: The sparse vector v_0.

Operation:

1. Let $S := \{i_1, \dots, i_k\} \subseteq [d]$ be the set of indexes denoting the k largest diagonal entries of $Y^\top Y$.
2. Output a top eigenvector of $Y^\top Y[S \times S]$.

We start by defining the adversarial matrix.

Definition A.2. Let $b \in R$ and denote with $W_1, \dots, W_d$ the columns of W. Define E to be the matrix with columns

$$E_i = \begin{cases} \frac{b}{\|W_i\|} W_i & \text{if} i \in [d] \setminus \operatorname{supp}\{v_0\} \\ 0 & \text{otherwise.} \end{cases}$$

The result is shown in the theorem below.

Theorem A.3. *Let $n \geqslant \omega(\log d)$, $\beta = o(k)$. Let Y be sampled according to Problem 3.1 where v_0 is a flat vector. Let E be as defined in Definition A.2 and $\|E\|_{1\to 2} \gtrsim \frac{\beta\sqrt{n}}{k} + \sqrt{\log d}$. Then for each $i \in [d] \operatorname{supp}\{v_0\}$ and $j \in \operatorname{supp}\{v_0\}$*

$$\|Ye_i\|^2 \geqslant \left\|Ye_j\right\|^2$$

with probability at least 0.99.

Notice how, for $\beta = \Theta\left(\frac{k}{\sqrt{n}}\sqrt{\log d}\right)$ the theorem implies that an adversary with $\|E\|_{1\to 2} \gtrsim \sqrt{\log d}$ suffices to fool Diagonal Thresholding. A perturbation resilient algorithm would succeed as long as $\|E\|_{1\to 2} \lesssim \min\left\{(n \log d)^{1/4}, \sqrt{k \log d}\right\}$.

Remark A.4. The same adversary also fools the limited exhaustive search algorithm from [BKW20] that runs in time $n^{O(t)}$ up to some very large t (say, up to some $t = n^{\Omega(1)}$).

Proof of Theorem A.3. Let $b = \|E\|_{1\to 2}$. We condition our analysis on the event that

$$\begin{aligned}
\forall i \in [d] \qquad & \|W_i\|^2 \in \left[n - 40\sqrt{n\log d}, n + 40\sqrt{n\log d}\right]. \\
& \|u_0\|^2 \leqslant n + 100\sqrt{n} \\
\forall i \in [d] \qquad & \langle u_0, W_i\rangle \leqslant 10\sqrt{n\log d}
\end{aligned}$$

which happen with probability at least 0.99 by Fact A.18. Denote with $e_1, \ldots, e_d$ the standard basis vectors in $\mathbb{R}^d$. Notice that, by construction of E, for $i \in [d] \setminus \mathrm{supp}\{v_0\}$, $\frac{Ee_i}{\|Ee_i\|} = \frac{We_i}{\|We_i\|}$. Thus,

$$\begin{aligned}
\|Ye_i\|^2 &= \left\|\left(W + E + \sqrt{\beta}u_0v_0^\mathsf{T}\right)e_i\right\|^2 \\
&= \left\|\left(1 + \frac{b}{\|W_i\|}\right)W_i\right\|^2 \\
&= \|W_i\|^2 + b^2 + 2b\|W_i\| \\
&\geqslant n + b^2 + b\sqrt{n} - O\left(\sqrt{n\log d}\right) \\
&\geqslant n + b\sqrt{n} - O\left(\sqrt{n\log d}\right).
\end{aligned}$$

On the other hand, for $j \in \mathrm{supp}\{v_0\}$,

$$\begin{aligned}
\left\|Ye_j\right\|^2 &= \left\|W_j\right\|^2 + \frac{\beta}{k}\|u_0\|^2 + \sqrt{\frac{\beta}{k}}\langle W_j, u_0\rangle \\
&\leqslant n + \frac{n\beta}{k} + O\left(\sqrt{n\log d} + \sqrt{\frac{\beta n\log d}{k}}\right) \\
&\leqslant n + \frac{n\beta}{k} + O\left(\sqrt{n\log d}\right)
\end{aligned}$$

where the last step follows as $\beta = o(k)$. Combining the two inequalities,

$$\|Ye_i\|^2 - \left\|Ye_j\right\|^2 \geqslant b\sqrt{n} - \frac{n\beta}{k} - O\left(\sqrt{n\log d}\right)$$

which is larger then zero whenever,

$$b \geqslant O\left(\frac{\beta\sqrt{n}}{k} + \sqrt{\log d}\right).$$

□

A.1.3 Covariance thresholding is fragile

In this section, we show how uder Model 3.1, the Covariance thresholding algorithm fails to output a good estimation of the vector v_0 in the presence of an adversarial distribution.

Specifically, we will show that the algorithm fails for $k \geqslant \frac{\sqrt{n \log \frac{d}{k^2}}}{\|E\|_{1\to 2}^2}$. This bound is significant in the sense that already for $\|E\|_{1\to 2} = d^{o(1)}\sqrt{\frac{\beta n}{d}}$, the algorithm breaks. We remark that a similar phenomenon can also be observed in the Wigner model, we omit this proof since it is simpler than in the Wishart model.

Recall that the central idea behind Covariance Thresholding is to threshold entries of the empirical covariance matrix. The thresholding operation should remove noise while leaving the submatrix $\beta\|u_0\|^2 v_0 v_0^\mathsf{T}$ untouched. The top eigenvector of $\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)$ will then be close to the sparse vector. The key observation behind the adversary is that it is possible to plant a matrix E with small norm $\|E\|_{1\to 2}$ such that the thresholded covariance matrix $\eta\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)$ has many large eigenvalues with eigenspace far from v_0.

Consider the Covariance Thresholding algorithm:

Algorithm A.5 (Standard Covariance Thresholding)**.**

Input: Threshold τ, sample matrix $Y = \sqrt{\beta}\cdot u_0 v_0^T + W + E \in \mathbb{R}^{n\times d}$ where v_0 is k-sparse, u_0 and W have i.i.d subgaussian entries of mean 0 and variance 1 and E has column norms bounded by b.

Estimate: The sparse vector v_0.

Operation:

1. Compute the thresholded matrix $\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)$.
2. Output a top eigenvector $\hat{v}$ of $\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)$.

The main result of the section is the Theorem below. Its significance is to be read under this perspective: it shows that there exists an adversary that can plant several (i.e. $\omega(\log d)$) large eigenvalues, as a consequence the top eigenvectors of $\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)$ will not be correlated with v_0.

Theorem A.6. *Suppose that $k \leqslant \sqrt{d}$ and $\log^{10} d \leqslant n \leqslant d$. Let Y be of the form Problem 3.1 for a flat vector v_0. Let $r \in [n]$ be such that $\omega\big(\log d\big) \leqslant r \leqslant d^{o(1)}$ and $\tau \in \mathbb{R}$ be such that $2\sqrt{n} \leqslant \tau \leqslant o\Big(\sqrt{n\log d}\Big)$ as $d \to \infty$.*

Then with probability at least $1 - o(1)$ (as $d \to \infty$) there exists an adversarial matrix E with maximal column norm $\|E\|_{1\to 2} \leqslant d^{o(1)}\sqrt{\frac{\beta n}{d}}$ and orthogonal vectors $z^1, \dots, z^r$ such that

$$\forall i \in [r], \qquad \frac{1}{\|z^i\|^2}\cdot \big(z^i\big)^\mathsf{T}\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)z^i \geqslant v_0{}^\mathsf{T}\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)v_0$$

and $\langle z^i, v_0\rangle = 0$.

The theorem shows that with these adversarial perturbations the first r eigenvectors of the thresholded covariance matrix are uncorrelated with the sparse vector v_0. Notice that for $\beta \geqslant 1$ a perturbation resilient algorithm should succeed with perturbations bounded by $\sqrt{\frac{\beta n}{k}}$, that is, much larger (in absolute value) than the ones used to fool Covariance Thresholding. In particular, for $\beta = \Theta\left(\frac{k}{\sqrt{n}}\sqrt{\log \frac{d}{k^2}}\right)$ Theorem A.6 implies that already with perturbations satisfying $\|E\|_{1\to 2} \leqslant d^{-1/4+o(1)}n^{1/4}$ the algorithm fails, while a perturbation resilient algorithm would succeed for $\|E\|_{1\to 2} \leqslant \tilde{O}\left(n^{1/4}\right)$.

Before showing the proof, we provide some intuition.

Algorithm intuition. Let's ignore cross-terms for a moment and consider the Wishart model with no adversarial distribution. Then the centered empirical Covariance Matrix looks like

$$Y^\mathsf{T}Y - n\mathrm{Id} \approx W^\mathsf{T}W - n\mathrm{Id} + \beta n v_0 v_0^\mathsf{T}.$$

If we set the threshold $\tau = C\sqrt{n \log \frac{d}{k^2}}$ for some large enough constant $C > 0$, then $d^2 \exp\left[-\Theta\left(\tau^2/n\right)\right] \approx k^4$ entries in $\left(W^\mathsf{T}W - n\mathrm{Id}\right)$ will be larger than τ. [1] On the other hand, for $\beta \gtrsim \frac{k}{\sqrt{n}}\sqrt{\log \frac{d}{k^2}}$ as $\left|\left(\beta n v_0 v_0^\mathsf{T}\right)_{ij}\right| \geqslant \tau$ whenever $i, j \in \mathrm{supp}\{v_0\}$, many entries of $\beta n v_0 v_0^\mathsf{T}$ will survive the thresholding. This means that,

$$\eta_\tau\left(Y^\mathsf{T}Y - n\mathrm{Id}\right) \approx \left(W^\mathsf{T}W - n\mathrm{Id}\right)[S] + \beta n v_0 v_0^\mathsf{T}$$

where $S \subseteq [d] \times [d]$ has cardinality approximately k^4. If the entries were independent, since the fourth moment of each entry is not much larger than the second moment, standard spectral matrix bounds suggest

$$\left\|\left(W^\mathsf{T}W - n\mathrm{Id}\right)[S]\right\| \leqslant O\left(\sigma\sqrt{d}\right),$$

where $\sigma \leqslant \tau \exp\left[-\frac{C\tau^2}{10n}\right] \leqslant \sqrt{n} \cdot \frac{k}{\sqrt{d}}$ is a standard deviation of each entry. Hence we get

$$\left\|\left(W^\mathsf{T}W - n\mathrm{Id}\right)[S]\right\| \leqslant O\left(k\sqrt{n}\right),$$

and

$$\left\|\beta n v_0 v_0^\mathsf{T}\right\| = \beta n.$$

In conclusion, for $\beta \gtrsim \frac{k}{\sqrt{n}}\sqrt{\log \frac{d}{k^2}}$ the top eigenvector of $\eta_\tau\left(Y^\mathsf{T}Y - n\mathrm{Id}\right)$ will be close to v_0.

[1] To see this, recall that in a $d \times d$ Gaussian matrix, with high probability there are at most k^4 entries larger than $\sqrt{\log \frac{d^2}{k^4}}$. While entries in $W^\mathsf{T}W - n\mathrm{Id}$ are dependent, a similar bound will hold.

The main technical difficulty here is that the entries of $W^\mathsf{T}W$ are not independent. In [DM14] the authors provide a method to bound the spectral norm of the thresholded matrix[2].

Adversarial strategy. Now we provide intuition on how to choose E such that with constant probability there exists a vector z orthogonal to v_0 for which

$$\frac{z^\mathsf{T}}{\|z\|}\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)\frac{z}{\|z\|} \gtrsim {v_0}^\mathsf{T}\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)v_0.$$

Let $x \in \mathbb{R}^n$ be a randomly chosen unit vector orthogonal to u_0, let z be a vector such that $\mathrm{supp}\{z\} = [d] \setminus \mathrm{supp}\{v_0\}$ and for $i \in \mathrm{supp}\{z\}$, $z_i = \sigma_i b$ for some $b \in \mathbb{R}_+$ to be set later and $\sigma_i \sim \{\pm 1\}$. We define the adversarial matrix as $E := xz^\mathsf{T}$, notice that $\|E\|_\infty \leqslant \tilde{O}\big(b/\sqrt{n}\big)$. For $i, j \in \mathrm{supp}\{z\}$, consider the entry ij of the centered empirical covariance matrix $\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)$,

$$\Big|\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)_{ij}\Big| = \big|\langle w_i, w_j\rangle + \langle w_i, x\rangle + \langle x, w_j\rangle + z_i z_j\big| \lesssim \big|\langle w_i, w_j\rangle + z_i z_j\big|,$$

by construction of z, the term $z_i z_j$ is symmetric and bounded by b^2. Hence for $b^2 = o(\sqrt{n})$, the thresholding of entry $\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)_{ij}$ will depend almost only on the Gaussian contribution $\big(W^\mathsf{T}W - n\mathrm{Id}\big)_{ij}$. Let $S \subseteq [d] \times [d]$ be the set of non-zero entries in $\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)$. By independence of z and W, and since S dependence of z is very limited, we expect, as in our previous discussion, $|S| \gtrsim k^4$. Now consider the quadratic form

$$\frac{z^\mathsf{T}}{\|z\|}\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)[S]\frac{z}{\|z\|} \approx \frac{z^\mathsf{T}}{\|z\|}\big(zz^\mathsf{T}\big)[S]\frac{z}{\|z\|} + \frac{z^\mathsf{T}}{\|z\|}\big(W^\mathsf{T}W - n\mathrm{Id}\big)[S]\frac{z}{\|z\|}.$$

As argued in the previous paragraph A.1.3, $\frac{z^\mathsf{T}}{\|z\|}\big(W^\mathsf{T}W - n\mathrm{Id}\big)[S]\frac{z}{\|z\|} \leqslant O\big(k\sqrt{n}\big)$. On the other hand,

$$\frac{z^\mathsf{T}}{\|z\|}\big(zz^\mathsf{T}\big)[S]\frac{z}{\|z\|} = \frac{1}{\|z^2\|^2}\sum_{(i,j)\in S} z_i^2 z_j^2 = \frac{|S|}{\|z\|^2}b^4 \gtrsim \frac{k^4}{d}b^2 \gtrsim d^{1-o(1)}b^2.$$

For the signal we instead have ${v_0}^\mathsf{T}\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)[S]v_0 \lesssim \beta n$. It follows that setting $b \gtrsim \sqrt{\frac{\beta n}{d^{1-o(1)}}}$ the top eigenvector of $\eta\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)$ will not achieve constant correlation with v_0. Recall now that $n \geqslant d^{1-o(1)}$ and that $\|E\|_\infty \lesssim \tilde{O}(b/\sqrt{n})$. Hence for $\beta \approx \frac{k}{\sqrt{n}}\sqrt{\log\frac{d}{k^2}} \leqslant n^{o(1)}$, adversarial perturbations are bounded by $n^{o(1)}/\sqrt{n}$ are enough to fool the algorithm.

Remark A.7. While this adversarial matrix is enough to break Covariance Thresholding it also allows an easy fix. Indeed, although the top eigenvector is now almost uncorrelated

[2]Formally, in [DM14] the authors provided a proof for a matrix obtained applying *soft-thresholding*. As we will see these can easily be extended to the hard-thresholded matrix $\eta\big(W^\mathsf{T}W - n\mathrm{Id}\big)$.

with v_0, the eigenspaces spanned by two largest eigenvectors contain a vector close to v_0 and a brute-force search over such space can be performed in polynomial time. The same approach however can be used to build an adversarial matrix E such that there exist vectors $z^1, \dots, z^r$ for which, with constant probability

$$i \in [r] \qquad \frac{z^{i\mathsf{T}}}{\|z^i\|} \eta_\tau\big(Y^\mathsf{T} Y - n\mathrm{Id}\big) \frac{z^i}{\|z^i\|} \gtrsim v_0{}^\mathsf{T} \eta_\tau\big(Y^\mathsf{T} Y - n\mathrm{Id}\big) v_0.$$

The idea is to chose $x^1, \dots, x^r$ to be orthonormal vectors orthogonal to u_0, and $z^1, \dots, z^r$ with non-intersecting supports and the same structure as before. This latter choice of E implies that the space containing eigenvectors associated with large eigenvalues has now dimension at least $\Omega(r)$. For $r \geqslant \omega\big(\log d\big)$, brute-force search of a vector close to v_0 in this space requires super-polynomial time.

A.1.3.1 Proving covariance thresholding fragile

Now we formally prove the theorem. First we define the adversarial matrix.

Definition A.8 (Adversarial matrix)**.** For $b \geqslant 1, r \in \mathbb{N}$, $W \sim N(0,1)^{n \times d}$, $u_0 \sim N(0, \mathrm{Id}_n)$ and v_0 k-sparse, the adversarial matrix is built as follows. Let $x^1, \dots, x^r \in \mathbb{R}^n$ be unit vectors that are independent of W such that for distinct $i, j \in [r]$, $\langle x^i, x^j \rangle = 0$. Partition the set $[d] \setminus \mathrm{supp}\{v_0\}$ in sets $Z_1, \dots, Z_r$ of cardinality $\frac{d - |\mathrm{supp}\{v_0\}|}{r}$. For each $i \in [r]$, let z^i be the vector with support Z_i such that:

$$\forall l \in Z_i, \qquad z^i_l = \begin{cases} b & \text{if } \langle w_l, x^i \rangle \geqslant 0 \\ -b & \text{otherwise.} \end{cases}$$

Then

$$E := \sum_{i \in [r]} x^i z^{i\mathsf{T}}.$$

Notice that $\|E\|_{1 \to 2} = b\sqrt{r}$.
Theorem A.6 follows immediately combining Theorem A.9, and Lemma A.10.

Theorem A.9. *Let Y be of the form 3.1 with E constructed as in definition A.8 with $\omega(\log d) \leqslant r \leqslant d^{o(1)}$ and $b \leqslant \sqrt[4]{n}$. Assume that $d \geqslant n \geqslant \log^{10} d$ and that $k \leqslant \sqrt{d}$. Let $2\sqrt{n} \leqslant \tau \leqslant o\Big(\sqrt{n \log d}\Big)$ as $d \to \infty$. Then with probability at least $1 - 2d^{-\Omega(1)}$ there exists a subset $R \subseteq [r]$ of size at least $\frac{r}{10}$ such that*

$$\forall i \in R, \qquad \frac{1}{\|z^i\|^2} \cdot \big(z^i\big)^\mathsf{T} \eta_\tau\big(Y^\mathsf{T} Y - n\mathrm{Id}\big) z^i \geqslant b^2 \cdot \frac{d^{1-o(1)}}{r}.$$

Lemma A.10. *Suppose the conditions of Theorem A.9 are satisfied and that the entries of v_0 are from $\{0, \pm 1/\sqrt{k}\}$ and $n \geqslant \omega(\log d)$ as $d \to \infty$. Then with probability $1 - O(d^{-10})$*

$$\left|v_0^{\mathsf{T}} \eta_\tau(Y^T Y - n\mathrm{Id}) v_0\right| \leqslant O\left(k\sqrt{n \log d} + \beta n\right).$$

of Lemma A.10. With probability $1 - O(d^{-10})$ the entries of $\eta_\tau(Y^T Y - n\mathrm{Id})$ are bounded by

$$O\left(\frac{\beta n}{k} + \sqrt{n \log d} + \sqrt{\frac{\beta n \log d}{k}}\right) \leqslant O\left(\frac{\beta n}{k} + \sqrt{n \log d}\right).$$

Since v_0 has at most k nonzero entries,

$$\left|v_0^{\mathsf{T}} \eta_\tau(Y^T Y - n\mathrm{Id}) v_0\right| \leqslant k \cdot \|\eta_\tau(Y^T Y - n\mathrm{Id})\|_\infty \leqslant O\left(\beta n + k\sqrt{n \log d}\right).$$

□

Tp prove Theorem A.9 we make use of intermediate steps A.11-A.12. Our plan is to show that many entries of $z^i z^{i\mathsf{T}}$ survive the thresholding due to the contribution of $W^{\mathsf{T}} W - n\mathrm{Id}$. So, we start our analysis lower bounding the number of entries of $W^{\mathsf{T}} W - n\mathrm{Id}$ that are above the threshold.

The following lemma shows that for each vector z^i, many entries in $\mathrm{supp}\{z^i\} \times \mathrm{supp}\{z^i\}$ will survive the thresholding.

Lemma A.11. *For any $b, r \in \mathbb{R}$ consider Y sampled from model 3.1 with E as in A.8. For some $10 \leqslant q \leqslant d^{o(1)}$ let $\tau = \sqrt{n \log q}$. For $i \in [r]$ define the set*

$$S_i := \left\{(j,l) \in \mathrm{supp}\{z^i\} \times \mathrm{supp}\{z^i\} \;\middle|\; j \neq l, \left(Y^{\mathsf{T}} Y - n\mathrm{Id}\right)_{jl} \geqslant \tau\right\}.$$

Then with probability at least $1 - \exp\left(d^{1-o(1)}\right)$,

$$|S_i| \geqslant \frac{d^2}{1000 r^2 q^{10}}.$$

Proof. Consider an off diagonal entry jl of $\eta_\tau\left(Y^{\mathsf{T}} Y - n\mathrm{Id}\right)$ such that $j, l \in \mathrm{supp}\{z^i\}$ for some $i \in [r]$. Since with probability at least $1 - 2\exp\left[-\Omega(n^{0.2})\right] \geqslant 1 - 2d^{\Omega(1)}$, $\langle w_j, x_l\rangle \leqslant n^{0.1}$, we get

$$\begin{aligned}\mathbb{P}\left(\left|\langle w_j, w_l\rangle + \langle w_j, x\rangle z_l^i + \langle w_l, x\rangle z_j^i + z_j^i z_l^i\right| \geqslant \tau\right) &\geqslant \mathbb{P}\left(\tfrac{1}{\sqrt{n}}\left|\langle w_j, w_l\rangle\right| \geqslant 2\sqrt{\log q}\right) - d^{-\Omega(1)} \\ &\geqslant \frac{1}{10 q^{10}}.\end{aligned}$$

For fixed z^i and fixed row $j \in [d]$, the $\langle w_j, w_l\rangle$ (for different $l \in \mathrm{supp}\{z^i\}$) are independent from each other. Since $r \leqslant d^{o(1)}$, with probability $1 - \exp\left[\frac{d}{100 r q^{10}}\right] = 1 - \exp\left(d^{1-o(1)}\right)$ number of different l such that $\left(Y^{\mathsf{T}} Y - n\mathrm{Id}\right)_{jl} \geqslant \tau$ is at least $\frac{d}{1000 r q^{10}}$. Hence if with probability at least $1 - \exp\left(d^{1-o(1)}\right)$, for each z^i, $S_i \geqslant \frac{d^2}{1000 r^2 q^{10}}$. □

The last ingredient needed for Theorem A.9 is a proof that the cross-terms in the quadratic form $z^{i\mathsf{T}}\eta_\tau\big(Y^\mathsf{T}Y - n\mathrm{Id}\big)z^i$ do not remove the contribution of the adversarial vector.

Lemma A.12. *Let Y be sampled from model 3.1 with E as in Definition A.8. Let S_i be as in Lemma A.11 Then with probability at least $\frac{1}{2}$,*

$$\big(z^i\big)^\mathsf{T}\big(W^\mathsf{T}W + W^\mathsf{T}x^iz^{i\mathsf{T}} + z^ix^{i\mathsf{T}}W\big)[S_i]\,z^i \geqslant 0.$$

Proof. For simplicity of the notation we will refer to x^i, z^i simply as x, z. Opening up the sum,

$$\begin{aligned} z^\mathsf{T}\big(W^\mathsf{T}W + W^\mathsf{T}xz^\mathsf{T} + zx^\mathsf{T}W\big)[S_i]\,z &= 2\sum_{(j,l)\in S_i}\langle w_j, w_l\rangle z_jz_l + \langle w_j, x\rangle b^2z_j + \langle w_l, x\rangle b^2z_l \\ &\geqslant 2\sum_{(j,l)\in S_i}\langle w_j, w_l\rangle z_jz_l \\ &\geqslant 2\sum_{(j,l)\in S_i}\langle w_j, \big(\mathrm{Id} - xx^\mathsf{T}\big)w_l\rangle z_jz_l, \end{aligned}$$

using the fact that by construction $\langle w_j, x\rangle b^2z_j \geqslant 0, \langle w_l, x\rangle b^2z_l \geqslant 0$. So it is enough to prove that

$$\mathbb{P}\left(\sum_{j=1}^{d'}\sum_{(j,l)\in S_i}\langle w_j, \big(\mathrm{Id} - xx^\mathsf{T}\big)w_l\rangle z_jz_l \geqslant 0\right) \geqslant \frac{1}{2}.$$

Let

$$a_{jl} = \langle w_j, \big(\mathrm{Id} - xx^\mathsf{T}\big)w_l\rangle z_jz_l = \big(\langle w_j, w_l\rangle - \langle w_j, x\rangle\langle w_l, x\rangle\big)\cdot z_jz_l$$

and

$$\begin{aligned} p_{jl} &= \big(\langle w_j, x\rangle z_l + \langle w_l, x\rangle z_j + \langle w_j, x\rangle\langle w_i, x\rangle + z_jz_l\big)\cdot z_jz_l \\ &= \langle w_j, x\rangle b^2z_j + \langle w_l, x\rangle b^2z_l + \langle w_j, x\rangle\langle w_i, x\rangle z_jz_l + b^4\,. \end{aligned}$$

Notice that $(j,l) \in S_i$ if and only if $j \neq l$ and $\big|a_{jl} + p_{jl}\big| \geqslant b^2\tau$. Also notice that $p_{jl} \geqslant 0$ and that with probability at least $1 - 2\exp\big[-\Omega(n^{0.2})\big]$, $p_{jl} < b^2\tau$.

Since $|\mathrm{supp}\{z\}| = d'$, without loss of generality assume $\mathrm{supp}\{z\} = [d']$. For $q \in [d'-1]$ define

$$T_q^* := \sum_{j=q}^{d'-1}\sum_{\substack{j<l\leqslant d' \text{ s.t.} \\ |a_{jl}|\geqslant b^2\tau}}\langle\big(\mathrm{Id} - xx^\mathsf{T}\big)w_j, \big(\mathrm{Id} - xx^\mathsf{T}\big)w_l\rangle z_jz_l = \sum_{j=q}^{d'-1}\sum_{\substack{j<l\leqslant d' \text{ s.t.} \\ |a_{jl}|\geqslant b^2\tau}} a_{jl}$$

and

$$T_q := \sum_{j=q}^{d'-1}\sum_{\substack{j<l\leqslant d' \text{ s.t.} \\ |a_{jl}+p_{jl}|\geqslant b^2\tau}}\langle\big(\mathrm{Id} - xx^\mathsf{T}\big)w_j, \big(\mathrm{Id} - xx^\mathsf{T}\big)w_l\rangle z_jz_l = \sum_{j=q}^{d'-1}\sum_{\substack{j<l\leqslant d' \text{ s.t.} \\ |a_{jl}+p_{jl}|\geqslant b^2\tau}} a_{jl}\,.$$

Let $T_{d'} = T_{d'} = 0$. For $j \in [d'-1]$ consider

$$T_j^* - T_{j+1}^* = \sum_{\substack{j<l\leqslant d' \text{ s.t.} \\ |a_{jl}|\geqslant b^2\tau}} \langle (\mathrm{Id} - xx^\mathsf{T})w_j, (\mathrm{Id} - xx^\mathsf{T})w_l\rangle z_j z_l \,.$$

$(\mathrm{Id} - xx^\mathsf{T})w_j$ is symmetric around zero and independent from all z_j and all w_l for $l > w_j$. Moreover, the sign of $(\mathrm{Id} - xx^\mathsf{T})w_j$ does not influence on the condition $|a_{jl}| \geqslant b^2\tau$. It follows that the conditional disribution of T_j^* given z_j, z_l, w_l for $l > j$ is symmetric around T_{j+1}^* and thus by induction T_1^* is symmetric around zero. It remains to show that $\mathbb{P}(T_1 \geqslant 0) \geqslant \mathbb{P}(T_1^* \geqslant 0)$, which is true since if $T_1^* \geqslant 0$, then $T_1^* \geqslant T_1$. Indeed, if $T_1^* \geqslant 0$, then any $a_{jl} \geqslant 0$ such that $|a_{jl}| \geqslant b^2\tau$ satisfies $|a_{jl} + p_{jl}| \geqslant b^2\tau$, and any $a_{jl} < 0$ such that $|a_{jl} + p_{jl}| \geqslant b^2\tau$ satisfies $|a_{jl}| \geqslant b^2\tau$. □

We are now ready to prove Theorem A.9.

Proof of Theorem A.9. Let $10 \leqslant q \leqslant d^{o(1)}$ so that $\tau = \sqrt{n \log q}$. By construction of z^i,

$$\begin{aligned} z^{i\mathsf{T}}(W^\mathsf{T} uv^\mathsf{T})[S_i]z^i &= 0 \\ z^{i\mathsf{T}}(E^\mathsf{T} uv^\mathsf{T})[S_i]z^i &= 0 \\ z^{i\mathsf{T}}(E^\mathsf{T} E)[S_i]z^i &= z^{i\mathsf{T}}(z^i z^{i\mathsf{T}})[S_i]z^i. \end{aligned}$$

With probability $1 - d^{\Omega(1)}$ sum over diagonal entries is bounded by:

$$\sum_j \left(z_j^i\right)^2 (\|w_j\|^2 - n) \leqslant O\left(b^2 d\sqrt{n \log d}\right) \leqslant b^2 d^{1.5+o(1)}\,.$$

Notice that for different $i, m \in [r]$ the events $(z^i)^\mathsf{T}(W^\mathsf{T}W + W^\mathsf{T}x^i z^{i\mathsf{T}} + z^i x^{i\mathsf{T}}W)[S_i]\, z^i \geqslant 0$ and $(z^m)^\mathsf{T}(W^\mathsf{T}W + W^\mathsf{T}x^m z^{m\mathsf{T}} + z^m x^{m\mathsf{T}}W)[S_m]\, z^m \geqslant 0$ are independend. Hence, by Lemma A.12 with probability at least $1 - 2^{-0.1r}$ for at least $r/10$ different $i \in [r]$,

$$\begin{aligned} &(z^i)^\mathsf{T}\eta_\tau(Y^\mathsf{T}Y - n\mathrm{Id})\, z^i \\ &\quad = (z^i)^\mathsf{T}(z^i z^{i\mathsf{T}})[S_i]\, z^i + z^{i\mathsf{T}}(W^\mathsf{T}W + W^\mathsf{T}xz^{i\mathsf{T}} + z^i x^{i\mathsf{T}}W)[S_i]\, z^i \\ &\quad \geqslant (z^i)^\mathsf{T}(z^i z^{i\mathsf{T}})[S_i]\, z^i \\ &\quad = b^4|S_i|. \end{aligned}$$

By Lemma A.11 $|S_i| \geqslant \frac{d^2}{1000r^2q^{10}}$. The theorem follows observing that $\left\|z^i\right\|^2 \leqslant \frac{db^2}{r}$.

□

A.2 Existence of the adversarial distribution of Model 3.41

Let $\mathbb{R}[x]_{\leqslant s}$ be the space of one variable polynomials of degree at most s. To construct the desired distribution we will need the following theorem.

Theorem A.13 (Theorem 1.26 in [S+17]). *Suppose that $m_1, \dots, m_s \in \mathbb{R}$ and $K \subseteq \mathbb{R}$ is compact. Consider a linear functional $\mathcal{L} : \mathbb{R}[x]_{\leqslant s} \to \mathbb{R}$ such that $\mathcal{L}(1) = 1$ and*

$$\mathcal{L}(x^r) = m_r, \quad 1 \leqslant r \leqslant s\,.$$

If $\mathcal{L}(p) \geqslant 0$ for every $p \in \mathbb{R}[x]_{\leqslant s}$ that is nonnegative on K, then there exists a finitely supported probability distribution η such that $\operatorname{supp}(\eta) \subseteq K$ and $\mathbb{E}_{x\sim\eta}\, x^r = m_r$ for $1 \leqslant r \leqslant s$.

Let's take the maximal even number s such that $\delta\lambda^s \leqslant 2^{-10s}$. We will show that there exists a distribution with compact support such that with probability δ it takes values $\pm\lambda$ and its first s moments coincide with the first s Gaussian moments. Such a distribution is a mixture $\eta = (1-\delta)\eta_0 + \delta\eta_1$, where η_1 takes values $\pm\lambda$ with probability $\frac{1}{2}$ each, and η_0 has particular moments up to s.

Proposition A.14. *Suppose that $s \geqslant 2$ is even, $0 < \delta < 1$, $\lambda \geqslant 2$ and $\delta\lambda^s \leqslant 2^{-10s}$. Then there exists a finitely supported probability distribution η_0 such that $\operatorname{supp}(\eta_0) \subseteq [-10\sqrt{s\ln s}, 10\sqrt{s\ln s}]$ and $\mathbb{E}_{x\sim\eta_0}\, x^r = M_r$, where*

$$M_r = \begin{cases} 0, & \text{if } r \text{ is odd,} \\ \frac{1}{1-\delta}\big((r-1)!! - \delta\lambda^r\big), & \text{if } 0 \leqslant r \leqslant s \text{ and } r \text{ is even.} \end{cases}$$

Proof. Consider a linear functional $\mathcal{L} : \mathbb{R}[x]_{\leqslant s} \to \mathbb{R}$ such that $\mathcal{L}(1) = 1$ and $\mathcal{L}(x^r) = M_r$ for $1 \leqslant r \leqslant s$. We need to show that $\mathcal{L}(p) \geqslant 0$ for every polynomial $p \in \mathbb{R}[x]_{\leqslant s}$ that is nonnegative on $[-10\sqrt{s\ln s}, 10\sqrt{s\ln s}]$. Notice that for any polynomial $p \in \mathbb{R}[x]_{\leqslant s}$

$$(1-\delta)\cdot\mathcal{L}(p) = \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} p(x) - \frac{\delta}{2}\big(p(\lambda) + p(-\lambda)\big)\,.$$

Consider an arbitrary polynomial $p(x) = \sum_{r=0}^{s} p_r x^r$ that is nonnegative on $[-10\sqrt{s\ln s}, 10\sqrt{s\ln s}]$. If $p = 0$, then obviously $\mathcal{L}(p) = 0$. So we can assume that $p \neq 0$ and without loss of generality $\max\limits_{0\leqslant r\leqslant s}\{|p_r|\} = 1$. Since p is nonnegative on $[-10\sqrt{s\ln s}, 10\sqrt{s\ln s}]$,

$$\mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} p(x) \geqslant \frac{1}{\sqrt{2\pi}}\int_{-1}^{1} p(x)e^{-x^2/2} + \frac{1}{\sqrt{2\pi}}\int_{|x|>10\sqrt{s\ln s}} p(x)e^{-x^2/2}dx\,.$$

The second integral can be bounded as follows

$$\left|\int_{|x|>10\sqrt{s\ln s}} p(x)e^{-x^2/2}\right| \leqslant \sum_{r=0}^{s}|p_r|\int_{|x|>10\sqrt{s\ln s}}|x|^r e^{-x^2/2}dx \leqslant (s+1)\int_{|x|>10\sqrt{s\ln s}} x^s e^{-x^2/2}dx\,.$$

Notice that since the function $s\ln x + 10s - 0.4x^2$ is monotone for $|x| \geqslant 10\sqrt{s\ln s}$,

$$x^s e^{-x^2/2} \leqslant e^{-10s-x^2/10}$$

for all x such that $|x| \geqslant 10\sqrt{s \ln s}$. Hence

$$\left| \int\limits_{|x|>10\sqrt{s\ln s}} p(x)e^{-x^2/2}dx \right| \leqslant (s+1)\cdot e^{-10s} \int\limits_{|x|>10\sqrt{s\ln s}} e^{-x^2/10}dx \leqslant \sqrt{10}(s+1)\cdot e^{-10s} \leqslant e^{-8s}\,.$$

Let's bound $\int_{-1}^{1} p(x)\exp\left(-\frac{x^2}{2}\right)dx$. Since $p(x)$ is nonnegative on $[-1,1]$,

$$\int\limits_{-1}^{1} p(x)e^{-x^2/2}dx \geqslant \int\limits_{-1}^{1} \frac{p^2(x)}{\max\limits_{|x|\leqslant 1} p(x)} e^{-x^2/2}dx \geqslant \frac{e^{-1/2}}{\sum_{r=0}^{s}|p_r|}\int\limits_{-1}^{1} p^2(x)dx \geqslant \frac{1}{2(s+1)}\int\limits_{-1}^{1} p^2(x)dx\,.$$

To bound $\int_{-1}^{1} p^2(x)dx$ we can use Legendre polynomials (see for example [AWH13]). The degree j Legendre polynomial is

$$L_j(x) = \sum_{r=0}^{j} L_{j,r}x^r = \sum_{r=0}^{j} \sqrt{\frac{2j+1}{2}}\cdot 2^j \binom{j}{r}\binom{\frac{j+r-1}{2}}{j} x^r\,.$$

They form an orthonormal system on $[-1,1]$ with respect to the unit weight. Hence there exist coefficients $c_0,\ldots,c_s$ such that $p(x) = \sum_{j=0}^{s} c_s L_j(x)$ and

$$\int\limits_{-1}^{1} p^2(x)dx = \sum_{j=0}^{s} c_j^2\,.$$

Recall that by assumption $\max\limits_{0\leqslant r\leqslant s}\{|p_r|\} = 1$, so there exists some r such that $|p_r| = 1$. Thus

$$1 = |p_r| = \left|\sum_{j=r}^{s} c_j L_{j,r}\right| \leqslant \sum_{j=r}^{s} |c_j||L_{j,r}| \leqslant \max_{r\leqslant j\leqslant s}|L_{j,r}|\sqrt{(s+1)}\sqrt{\textstyle\sum_{j=0}^{s} c_j^2}\,.$$

Notice that $|L_{j,r}| \leqslant \sqrt{s+1}\cdot 2^{2s}$ for $0 \leqslant r \leqslant j \leqslant s$. Hence we get a bound

$$\int\limits_{-1}^{1} p(x)e^{-x^2/2}dx \geqslant \frac{1}{3s}\sum_{j=0}^{s} c_j^2 \geqslant \frac{1}{2(s+1)^3}2^{-4s} \geqslant 2^{-7s}\,,$$

and

$$\mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} p(x) \geqslant \frac{1}{\sqrt{2\pi}}2^{-7s} - e^{-8s} \geqslant 2^{-8s}\,.$$

Notice that

$$\frac{\delta}{2}\big(p(\lambda)+p(-\lambda)\big) \leqslant \delta\sum_{r=0}^{s}|\lambda|^r \leqslant 2\delta\lambda^s \leqslant 2^{-9s}\,.$$

Hence finally we get

$$(1-\delta)\cdot\mathcal{L}(p) = \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} p(x) - \frac{\delta}{2}\big(p(\lambda)+p(-\lambda)\big) \geqslant 2^{-8s} - 2^{-9s} > 0\,.$$

Therefore by Theorem A.13 there exists a finitely supported probability distribution η_0 with moments $M_1,\dots,M_s$ such that $\operatorname{supp}(\eta_0)\subseteq[-10\sqrt{s\ln s}, 10\sqrt{s\ln s}]$. □

We can assume that η_0 is symmetric (since if $z\sim\eta_0$ and $w\sim N(0,1)$ are independend, $zw/|w|$ is symmetrically distributed and has the same first s moments as z). Thus the mixture distribution $\eta=(1-\delta)\eta_0+\delta\eta_1$ (where η_1 takes values $\pm\lambda$ with probability $\frac{1}{2}$ each) is symmetric and has Gaussian moments up to $s+1$:

$$\mathop{\mathbb{E}}_{x\sim\eta} x^r = \mathop{\mathbb{E}}_{x\sim\mathcal{N}(0,1)} x^r\,,\quad \text{if } 0\leqslant r\leqslant s+1\,,$$

and its higher moments satisfy

$$\delta\lambda^r \leqslant \mathop{\mathbb{E}}_{x\sim\eta} x^r \leqslant \delta\lambda^r + (10s)^r\,,\quad \text{if } r>s \text{ is even.}$$

A.3 Additional tools

This section contains additional tools used throughout the proofs of Chapter 3. The notation is consistent with the aforementioned chapter.

Matrix concentration of measure

We introduce here some standard matrix concentration inequalities.

The following general result by Rudelson showsconvergence of empirical covariances of random variables.

Fact A.15 (Theorem 1, [Rud99])**.** *Let Y be a random vector in the isotropic position. Let $Y_1, Y_2, \dots, Y_q$ be q independent copies of Y. Then, for some absolute constant $C>0$,*

$$\mathbb{E}\left\|\frac{1}{q}\sum_{i=1}^{q} Y_i Y_i^\top - I\right\| \leqslant C\frac{\sqrt{\log q}}{\sqrt{q}}\cdot\mathbb{E}(\|Y\|^{\log q})^{1/\log q}\,.$$

The next computation bounds the variances of low-degree polynomials of product subgaussian random vectors.

Lemma A.16 (Variance of Polynomials of Independent Subgaussians)**.** *Let Y be a product random variable on $\mathbb{R}^n$ with coordinates of mean 0, variance 1 satisfying $\mathbb{E}\langle Y,u\rangle^{2t}\leqslant C^t(2t)^t$ for every unit vector u for some absolute constant $C>0$. Let $p=\sum_{S:|S|\leqslant k} p_S y_S$ be a polynomial in $y\in\mathbb{R}^n$ of degree k where the sum ranges of multisets $S\subseteq[n]$ of size at most k. Then, $\sum_{S:|S|\leqslant k} p_S^2 \leqslant \mathbb{E}\,p^2(Y) \leqslant C^t(2t)^t\sum_{S:|S|\leqslant k} p_S^2$.*

Proof. For any polynomial p, we write $\|p\|_2^2$ to denote the sum of squares of its coefficients in the monomial basis. For any multilinear polynomial p, observe that $\mathbb{E}\, p^2 = \|p\|_2^2$. For a non-multilinear p, we write $p = \sum_{S:|S|\leqslant k/2} y_S^2 q_S$ such that q_S is a multilinear polynomial of degree at most $k - 2|S|$. Observe that $\|p\|_2^2 = \sum_S \|q_S\|_2^2$. Further, $\mathbb{E}\, y_S^2 y_{S'}^2 q_S q_{S'} = 0$ whenever $S \neq S'$. Now, $\mathbb{E}\, p^2 = \sum_{S:|S|\leqslant k/2} \mathbb{E}\, y_S^2 q_S^2$. Since $\mathbb{E}\, y_S^2 \geqslant 1$ for any S, $\mathbb{E}\, y_S^2 q_S^2 \geqslant \|q_S\|_2^2$. Thus, $\mathbb{E}\, p^2 \geqslant \sum_S \|q_S\|_2^2 = \|p\|_2^2$. On the other hand, $\mathbb{E}\, y_S^2 q_S^2 \leqslant \|q_S\|_2^2 \cdot \max_{|S|\leqslant k} \mathbb{E}\, y_S^2 \leqslant C^k (2k)^k$. □

Lemma A.17. *Let Y be a random vector in $\mathbb{R}^n$ with independent coordinates of mean 0 and variance 1 satisfying $\mathbb{E}\langle Y, u\rangle^{2t} \leqslant C^t (2t)^t$ for some absolute constant $C > 0$. Then, with probability at least 0.99 over the draw of $Y_1, Y_2, \ldots, Y_d$ i.i.d. copies of Y,*

$$\left\| \frac{1}{d} \sum_i (Y_i^{\otimes t})(Y_i^{\otimes t})^\top - \mathop{\mathbb{E}}_{Y\sim D} \left(Y^{\otimes t}\right)\left(Y^{\otimes t}\right)^\top \right\| \leqslant \frac{n^{t/2} \log^{(t+1)/2}(n) (C't)^t}{\sqrt{d}},$$

for some absolute constant $C' > 0$.

Proof. Let $M = \mathbb{E}(Y^{\otimes t})(Y^{\otimes t})^\top$. Then, quadratic forms $\langle u, Mu\rangle$ is the variance of polynomial $p = \langle u, Y^{\otimes t}\rangle$ of degree at most t. Thus, using Lemma A.16, we have that $\|u\|_2^2 \leqslant \langle u, Mu\rangle \leqslant \|u\|_2^2 C^t (2t)^t$. Thus, all eigenvalues of M are between 1 and $C^t(2t)^t$.

We will now apply Fact A.15 to the isotropic random vectors $M^{-1/2} Z_i$ for $Z_i = Y_i^{\otimes t}$ for $1 \leqslant i \leqslant d$. Then, we obtain:

$$\mathbb{E}\left\|M^{-1/2} Z M^{-1/2} Z^\top - I\right\| \leqslant C \frac{\sqrt{\log d}}{\sqrt{d}} \left(\mathbb{E}\left\|M^{-1/2} Z\right\|_2^{\log d}\right)^{1/\log d}.$$

To finish, we compute $\mathbb{E}\left\|M^{-1/2} Z\right\|^{\log d} \leqslant \left\|M^{-1/2}\right\|^{\log d} \mathbb{E}\|Z\|^{\log d}$. Next, $\mathbb{E}\|Z\|^{\log d} = \mathbb{E}\|Y\|^{t \log d} \leqslant n^{(t/2)\log d} C^{(t/2)\log d} ((t/2)\log d)^{(t/2)\log d}$. Using $\left\|M^{-1/2}\right\| \leqslant 1$, we obtain: $\left(\mathbb{E}\left\|M^{-1/2} Z\right\|^{\log d}\right)^{1/\log d} \leqslant n^{t/2} C^{t/2} (t \log d)^{t/2}$.

Thus, for using $n \geqslant \log d$ and Fact A.15, $\mathbb{E}\left\|M^{-1/2} Z M^{-1/2} Z^\top - I\right\| \leqslant \frac{n^{t/2} \log^{(t+1)/2}(n) (10Ct)^t}{\sqrt{d}}$. Applying Markov's inequality completes the proof.

□

We also state here some other concentration bounds used in the proofs.

Fact A.18. *[LM00]Let $X \sim \chi_m^2$, $x > 0$, then*

$$\mathbb{P}\left(X - m \geqslant 2x + 2\sqrt{mx}\right) \leqslant e^{-x}$$

$$\mathbb{P}(m - X \geqslant x) \leqslant e^{-\frac{x^2}{4m}}$$

Fact A.19. *[Wai19] Let $0 < \varepsilon < 1$. The $n-1$-dimensional Euclidean sphere has an ε-net of size $\left(\frac{3}{\varepsilon}\right)^n$. That is, there exists a set N_ε of unit vectors in $\mathbb{R}^n$ of size at most $\left(\frac{3}{\varepsilon}\right)^n$ such that for any unit vector $u \in \mathbb{R}^n$ there exists some $v \in N_\varepsilon$ such that $\|v - u\| \leqslant \varepsilon$.*

Theorem A.20. *[Wai19] Let $W \sim N(0,1)^{n\times d}$. Then with probability $1-\exp(-t/2)$,*

$$\|W\| \leqslant \sqrt{n}+\sqrt{d}+\sqrt{t}$$

and

$$\left\|W^{\mathsf{T}}W - n\mathrm{Id}\right\| \leqslant d + 2\sqrt{dn} + t + 4\sqrt{t(n+d)}\,.$$

Theorem A.21 (Matrix Bernstein [Tro12])**.** *Consider a finite sequence $\{Z_k\}$ of independent, random, self-adjoint matrices in $\mathbb{R}^{d_1\times d_2}$. Assume that each random matrix satisfies*

$$\mathbb{E}\,Z_k = 0 \text{ and } \|Z_k\| \leqslant R \text{ almost surely.}$$

Define

$$\sigma^2 := \max\left\{\left\|\sum_k \mathbb{E}\,Z_k Z_k^{\mathsf{T}}\right\|, \left\|\sum_k \mathbb{E}\,Z_k^{\mathsf{T}} Z_k\right\|\right\}.$$

Then, or all $t \geqslant 0$,

$$\mathbb{P}\left(\left\|\sum_k Z_k\right\| \geqslant t\right) \leqslant (d_1+d_2)\exp\left\{\frac{-t^2/2}{\sigma^2 + Rt/3}\right\}.$$

Theorem A.22 (Matrix Hoeffding [Tro12])**.** *Consider a finite sequence $\{Z_k\}$ of independent, random, self-adjoint matrices in $\mathbb{R}^{d\times d}$. Assume that each random matrix satisfies*

$$\mathbb{E}\,Z_k = 0 \text{ and } Z_k^2 \preceq A_k^2 \text{ almost surely.}$$

Then, for all $t \geqslant 0$,

$$\mathbb{P}\left(\left\|\sum_k Z_k\right\| \geqslant t\right) \leqslant d\exp\left\{-\frac{t^2}{8\sigma^2}\right\}$$

where $\sigma^2 := \left\|\sum_k A_k^2\right\|$.

Theorem A.23 (k-sparse norm of a Gaussian matrix)**.** *Let $W \sim N(0,1)^{n\times d}$ be a Gaussian matrix. Let $1 \leqslant k \leqslant d$. Then with probability at least $1-\left(\frac{k}{ed}\right)^k$*

$$\max_{\substack{u\in\mathbb{R}^n\\ \|u\|=1}} \max_{\substack{k\text{-sparse } v\in\mathbb{R}^d\\ \|v\|=1}} u^{\mathsf{T}}Wv \leqslant \sqrt{n} + 3\sqrt{k\ln\left(\frac{ed}{k}\right)}\,.$$

Proof. Let v be some k-sparse unit vector that maximizes the value, and let $S(v)$ be the set of nonzero coordinates of v. Consider some fixed (independend of W) unit k-sparse vector $x \in \mathbb{R}^d$ and the set $S(x)$ of nonzero coordinates of x. If we remove from W all the rows with indices not from $S(x)$, we get an $n\times k$ Gaussian matrix $W_{S(x)}$. By Theorem A.20 norm of

this matrix is bounded by $\sqrt{n}+\sqrt{k}+\sqrt{t}$ with probability at least $\exp(-t/2)$. Number of all subsets $S \subseteq [d]$ of size k is $\binom{d}{k}$. By the union bound, the probability that the norm of $W_{S(v)}$ is greater than $\sqrt{n}+\sqrt{k}+\sqrt{t}$ is at most

$$\binom{d}{k} \cdot \exp(-t/2) \leqslant \exp\left(k \log_2(ed/k) - t/2\right).$$

Taking $t = 4k \ln(ed/k)$, we get the desired bound. □

Lemma A.24. *Let $w \sim N(0,1)^d$ be a Gaussian vector and let $1 \leqslant k \leqslant d$. Let S_k be the set of k largest coordinates of w. Then with probability $1 - \left(\frac{k}{ed}\right)^k$, $\sum_{i \in S_k} w_i^2 \leqslant 10k \ln(ed/k)$.*

Proof. Let S be any fixed subset of $[d]$ of size k. Then w restricted on S is a k-dimensional Gaussian vector and by Fact A.18, $\mathbb{P}\left(\sum_{i \in S} w_i^2 \geqslant 2x + 2\sqrt{kx}\right) \leqslant e^{-x}$. By a union bound over all $\binom{d}{k}$ subsets of $[d]$ of size k, we get

$$\mathbb{P}\left(\sum_{i \in S} w_i^2 \geqslant k + 2x + 2\sqrt{kx}\right) \leqslant e^{-x + k \log_2(ed/k)}.$$

Taking $x = 4k \ln(ed/k)$ we get the desired bound. □

Lemma A.25. *For large enough n and d such that $n \leqslant d$, let $W \sim N(0,1)^{n \times d}$ be a Gaussian matrix and let $u \in \mathbb{R}^n$ be an arbitrary unit vector (which can possibly depend on W). For any $t \geqslant 0$ let $S_t = \{i \in [d] \mid |(u^\mathsf{T} W)_i| \geqslant t\}$. Also let $B \geqslant 1$.*

Then, for any $t \geqslant 3\sqrt{B \ln d}$, $|S_t| \leqslant n/B$ with probability at least $1 - 2\exp(-n)$.

Proof. Let $t \geqslant 3\sqrt{B \ln d}$. For any fixed (independend of W) unit vector $x \in \mathbb{R}^n$, $(x^\mathsf{T} W)_i$ are i.i.d. standard Gaussian variables. For large enough d,

$$\mathbb{P}\left[|(x^\mathsf{T} W)_i| \geqslant t - 1\right] \leqslant \exp\left(-t^2/3\right).$$

Hence the probability that there are $3 \leqslant k \leqslant d$ coordinates that are larger than $t-1$ is at most

$$\binom{d}{k} \exp\left(-k \cdot t^2/3\right) \leqslant \exp\left[k\left(1 + \ln \frac{d}{k} - t^2/3\right)\right] \leqslant \exp\left[k\left(\ln d - t^2/3\right)\right] \leqslant \exp\left(-\frac{2}{3} \cdot kt^2 \ln d\right).$$

If for unit vectors $x, y \in \mathbb{R}^n$, $\|x - y\| \leqslant \varepsilon$, then $\|Wx - Wy\| \leqslant \varepsilon \|W\|$. By Theorem A.20, with probability at least $1 - \exp(-n)$, $\|W\| \leqslant 10\sqrt{d}$. Hence if $\varepsilon \leqslant \frac{1}{10\sqrt{d}}$, $\left|(x^\mathsf{T} W)_i - (y^\mathsf{T} W)_i\right| \leqslant 1$. By Fact A.19, for any $0 < \varepsilon < 1$, for $\varepsilon = \frac{1}{10\sqrt{d}}$ there exists an ε-net in $n-1$-dimensional sphere of size $\exp\left(\frac{n}{2} \log d + n \log 100\right) \leqslant \exp\left(n \log d\right)$ (for large enough d). By the union bound, the probability that $|S_t| > n/B$ is at most

$$\exp\left(n \ln d - \frac{2}{3}|S_t| t^2\right) \leqslant \exp(-n).$$

□

The next lemma is the main technical challenge of Section 3.2.

Theorem A.26. *[DM14] Let $W \sim N(0,1)^{n\times d}$, where $n \geqslant \omega(\log d)$ as $d \to \infty$. Let $0 \leqslant \tau \leqslant o\left(\sqrt{n \log d}\right)$ and let N be the matrix whose diagonal entries are zeros and each non-diagonal entry N_{ij} is*

$$N_{ij} = \begin{cases} \left(W^{\mathsf{T}}W\right)_{ij} - \operatorname{sign}\left(W^{\mathsf{T}}W\right)_{ij} \cdot \tau & \text{if } \left|\left(W^{\mathsf{T}}W\right)_{ij}\right| \geqslant \tau \\ 0 & \text{otherwise} \end{cases}$$

Then there exists an absolute constant $C \geqslant 1$ such that with probability $1 - o(1)$

$$\|N\| \leqslant C\left(d + \sqrt{dn}\right) \exp\left[-\frac{\tau^2}{Cn}\right].$$

Linear algebra

Lemma A.27. *Let v and u be unit vectors such that $\|vv^{\mathsf{T}} - uu^{\mathsf{T}}\| \leqslant \varepsilon$. Then $\langle v, u\rangle^2 \geqslant 1 - 2\varepsilon^2$.*

Proof. Let w be a unit vector orthogonal to u such that $v = \rho u + \sqrt{1-\rho^2}w$ for some positive $\rho \leqslant 1$. Then

$$vv^{\mathsf{T}} - uu^{\mathsf{T}} = \left(\rho^2 - 1\right)uu^{\mathsf{T}} + \rho\sqrt{1-\rho^2}uw^{\mathsf{T}} + \rho\sqrt{1-\rho^2}wu^{\mathsf{T}} + \left(1-\rho^2\right)ww^{\mathsf{T}}.$$

Since $vv^{\mathsf{T}} - uu^{\mathsf{T}}$ has rank 2, its Frobenius norm is bounded by 2ε, hence

$$4\varepsilon^2 \geqslant \|vv^{\mathsf{T}} - uu^{\mathsf{T}}\|_F^2 = 2\left(1-\rho^2\right)^2 + 2\rho^2\left(1-\rho^2\right) = 2\left(1-\rho^2\right).$$

It follows that

$$\langle v, u\rangle^2 = \rho^2 \geqslant 1 - 2\varepsilon^2.$$

□

Lemma A.28. *Let M be a symmetric matrix such that $\|M - uu^{\mathsf{T}}\| \leqslant \varepsilon < \frac{1}{2}$ for some unit vector u. Then the top eigenvalue λ_1 of M satisfies $|\lambda_1 - 1| \leqslant \varepsilon$ and the top eigenvector v_1 of M satisfies $\langle v_1, u\rangle^2 \geqslant 1 - 100\varepsilon^2$.*

Proof. Consider an eigenvalue decomposition of M:

$$M = \sum_{j=1}^{d} \lambda_j v_j v_j^{\mathsf{T}},$$

where $|\lambda_1| \geqslant |\lambda_2| \geqslant \ldots \geqslant |\lambda_d|$ and $\{v_j\}_{j=1}^d$ is an orthonormal basis in $\mathbb{R}^d$. By triangle inequality

$$\|v_1 v_1^{\mathsf{T}} - uu^{\mathsf{T}}\| \leqslant \|M - v_1 v_1^{\mathsf{T}}\| + \|M - uu^{\mathsf{T}}\| \leqslant \|M - v_1 v_1^{\mathsf{T}}\| + \varepsilon.$$

Let's bound $\|M - v_1 v_1^\mathsf{T}\|$:

$$\|M - v_1 v_1^\mathsf{T}\| \leqslant \max\{|1 - \lambda_1|, |\lambda_2|\}.$$

Since $\|M - uu^\mathsf{T}\| \leqslant \varepsilon$, $u^\mathsf{T} M u \geqslant 1 - \varepsilon$, hence $|\lambda_1| \geqslant 1 - \varepsilon > \frac{1}{2}$. Notice that

$$|\lambda_1 - \langle v_1, u\rangle^2| = |v_1^\mathsf{T} M v_1 - \langle v_1, u\rangle^2| \leqslant \varepsilon,$$

hence $\lambda_1 \geqslant -\varepsilon > -\frac{1}{2}$, so $\lambda_1 > 0$, and $\lambda_1 \leqslant \langle v_1, u\rangle^2 + \varepsilon \leqslant 1 + \varepsilon$, so $|\lambda_1 - 1| \leqslant \varepsilon$.

By triangle inequality $|\langle v_1, u\rangle^2 - 1| \leqslant 2\varepsilon$. By Pythagorean theorem $\sum_{j=1}^d \langle v_j, u\rangle^2 = 1$, hence

$$\sum_{j=2}^d \langle v_j, u\rangle^2 \leqslant 2\varepsilon,$$

so $\langle v_2, u\rangle^2 \leqslant 2\varepsilon$. Now let's bound $|\lambda_2|$:

$$|\lambda_2 - \langle v_2, u\rangle^2| = |v_2^\mathsf{T} M v_2 - \langle v_2, u\rangle| \leqslant \varepsilon,$$

hence $|\lambda_2| \leqslant 3\varepsilon$. Therefore

$$\|v_1 v_1^\mathsf{T} - uu^\mathsf{T}\| \leqslant 4\varepsilon.$$

By lemma A.27, $\langle v, u\rangle^2 \geqslant 1 - 32\varepsilon^2$. □

Lemma A.29. *Let $M \in \mathbb{R}^{d\times d}$, $M \succeq 0$, $\operatorname{Tr} M = 1$ and let $z \in \mathbb{R}^d$ be a unit vector such that $z^\mathsf{T} M z \geqslant 1 - \varepsilon$. Then the top eigenvector v_1 of M satisfies $\langle v_1, z\rangle^2 \geqslant 1 - O(\varepsilon)$.*

Proof. Write $z = \alpha v_1 + \sqrt{1-\alpha^2} v_\perp$ where $v_\perp$ is a unit vector orthogonal to v_1.

$$\begin{aligned} z^\mathsf{T} M z &= \alpha^2 v_1^\mathsf{T} M v_1 + (1-\alpha^2) v_\perp^\mathsf{T} M v_\perp \\ &= \alpha^2 (\lambda_1 - v_\perp^\mathsf{T} M v_\perp) + v_\perp^\mathsf{T} M v_\perp \\ &\geqslant 1 - \varepsilon \end{aligned}$$

As $v_1^\mathsf{T} M v_1 \geqslant z^\mathsf{T} M z$ and $v_\perp^\mathsf{T} M v_\perp \leqslant \varepsilon$, rearranging

$$\alpha^2 \geqslant \frac{1 - \varepsilon - v_\perp^\mathsf{T} M v_\perp}{\lambda_1 - v_\perp^\mathsf{T} M v_\perp} \geqslant 1 - 2\varepsilon.$$

□

Fact A.30. *Let $A, B \in \mathbb{R}^{d\times d}$, $A, B \succeq 0$. Then $\langle A, B\rangle \geqslant 0$.*

Lemma A.31. *Let $X \in \mathbb{R}^{d\times d}$ be a positive semidefinite matrix. Then for any $A \in \mathbb{R}^{d\times d}$,*

$$|\langle A, X\rangle| \leqslant \|A\| \cdot \operatorname{Tr} X\,.$$

Proof. Since X is positive semidefinite, $X = \sum_{i=1}^{d} \lambda_i z_i z_i^\mathsf{T}$ for unit vectors z_i such that $\lambda_i \geqslant 0$ and $\sum_{i=1}^{d} \lambda_i = \operatorname{Tr} X$. Hence

$$|\langle A, X\rangle| = \left|\operatorname{Tr} X^\mathsf{T} A\right| = \left|\sum_{i=1}^{d} \lambda_i \operatorname{Tr} z_i z_i^\mathsf{T} A\right| = \left|\sum_{i=1}^{d} \lambda_i \operatorname{Tr} z_i^\mathsf{T} A z_i\right| \leqslant \sum_{i=1}^{d} \lambda_i \|A\| = \|A\| \cdot \operatorname{Tr} X\,.$$

□

Lemma A.32. *Let $X \in \mathbb{R}^{d\times d}$ be a positive semidefinite matrix. Then for any $a, b \in \mathbb{R}^d$,*

$$\langle ab^\mathsf{T}, X\rangle^2 \leqslant \langle aa^\mathsf{T}, X\rangle \cdot \langle bb^\mathsf{T}, X\rangle\,.$$

Proof. By Fact A.30, $\langle aa^\mathsf{T}, X\rangle \geqslant 0$ and $\langle bb^\mathsf{T}, X\rangle \geqslant 0$. Notice that if the inequality is true for some $a, b \in \mathbb{R}^d$, it is also true for $c_1 a, c_2 b$ for all positive numbers c_1, c_2. So we can assume without loss of generality that $\langle aa^\mathsf{T}, X\rangle = \langle bb^\mathsf{T}, X\rangle = 1$. Consider $\langle (a+b)(a+b)^\mathsf{T}, X\rangle \geqslant 0$ and $\langle (a-b)(a-b)^\mathsf{T}, X\rangle \geqslant 0$. We get

$$2\langle ab^\mathsf{T}, X\rangle \leqslant \langle aa^\mathsf{T}, X\rangle + \langle bb^\mathsf{T}, X\rangle \leqslant 2$$

and

$$-2\langle ab^\mathsf{T}, X\rangle \leqslant \langle aa^\mathsf{T}, X\rangle + \langle bb^\mathsf{T}, X\rangle \leqslant 2\,,$$

hence $\langle ab^\mathsf{T}, X\rangle^2 \leqslant 1$. □

Appendix B

Deferred proofs and addendum to Chapter 4

Organization of appendices

The appendices of Chapter 4 are organized as follows. In Appendix B.1 we introduce our techniques for computing the truncated expectation of block self-avoiding walks. and prove Lemma 4.56, Lemma 4.67, Lemma 4.68, Lemma 4.96 and Lemma 4.98.

In Appendix B.2 we study the expectation of polynomials arising in the computation of the centered Schatten norm and prove Lemma 4.75. Most of the technical details of Appendix B.1 and Appendix B.2 can be found in Appendix B.3. In Appendix B.4.1 we prove Fact 4.58, Fact 4.65 and Fact 4.94. The counting arguments required in Section 4.5 are proved in Appendix B.4.2. Specifically, we prove here Lemma 4.64, Lemma 4.70, Lemma 4.71, Lemma 4.93, Lemma 4.99, Lemma 4.82 and Lemma 4.83. Finally, Appendix B.5 contains additional basic tools that are used throughout the chapter.

To help the reader navigate the sections we provide here a table where the various proofs of each lemma can be found.

Statement A	is implied by	Statement B
Lemma 4.56		Lemma B.17
Fact 4.58		Fact B.87
Fact 4.65		Fact B.85
Lemma 4.64		Lemma B.89
Lemma 4.67		Lemma B.26
Lemma 4.68		Lemma B.26
Lemma 4.70		Lemma B.96
Lemma 4.71		Lemma B.97
Lemma 4.75		Lemma B.49
Lemma 4.82		Lemma B.94
Lemma 4.83		Lemma B.95
Lemma 4.93		Lemma B.99
Fact 4.94		Fact B.86
Lemma 4.96		Lemma B.35
Lemma 4.98		Lemma B.36
Lemma 4.99		Lemma B.100

B.1 Bounds for the non-centered matrix

B.1.1 Useful notation

Let Δ be the constant truncation threshold (which we will specify later), and let $\overline{\mathbf{G}}$ be the graph obtained from $\mathbf{G} \sim \mathsf{SBM}_n(d, \varepsilon)$ by deleting every vertex that has a degree strictly greater than Δ in $\mathbf{G}$.

For every $v \in \mathbf{G}$, let $\mathcal{E}_v$ be the event that v is truncated, i.e.,

$$\mathcal{E}_v := \left\{v \notin \overline{\mathbf{G}}\right\} = \{d_{\mathbf{G}}(v) > \Delta\}.$$

For every $V \subseteq [n] = V(\mathbf{G})$, we say that V is truncated if there exists at least one vertex in V that is truncated. Let $\mathcal{E}_V$ be the event that V is truncated, i.e.,

$$\mathcal{E}_V := \bigcup_{v \in V} \mathcal{E}_v = \left\{V \subsetneq V\left(\overline{\mathbf{G}}\right)\right\}.$$

Similarly, for every multigraph H with $V(H) \subseteq [n]$, we say that H is truncated if there exists at least one vertex in H that is truncated. Let $\mathcal{E}_H$ be the event that H is truncated, i.e.,

$$\mathcal{E}_H := \mathcal{E}_{V(H)} = \left\{V(H) \subsetneq V\left(\overline{\mathbf{G}}\right)\right\}.$$

Recall that for every $i, j \in [n]$, we have

$$\mathbf{Y}_{ij} = \begin{cases} 1 - \frac{d}{n} & \text{if } ij \in \mathbf{G}, \\ -\frac{d}{n} & \text{otherwise.} \end{cases}$$

We define $\overline{\mathbf{Y}}_{ij} \in \left\{-\frac{d}{n}, 0, 1-\frac{d}{n}\right\}$ as follows:

$$\overline{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij} \cdot \mathbb{1}_{\mathcal{E}_i^c} \cdot \mathbb{1}_{\mathcal{E}_j^c} = \mathbf{Y}_{ij} \cdot \mathbb{1}_{\mathcal{E}_i^c \cap \mathcal{E}_j^c} = \mathbf{Y}_{ij} \cdot \mathbb{1}_{(\mathcal{E}_i \cup \mathcal{E}_j)^c}.$$

For every multigraph H, we write $m_H(ij)$ to denote the multiplicity in H of an edge $ij \in E(H)$. Define

$$\begin{aligned}
\overline{\mathbf{Y}}_H &= \prod_{ij \in E(H)} \overline{\mathbf{Y}}_{ij}^{m_H(ij)} = \prod_{ij \in E(H)} \left(\mathbf{Y}_{ij}^{m_H(ij)} \cdot \mathbb{1}_{(\mathcal{E}_i \cup \mathcal{E}_j)^c} \right) \\
&= \left(\prod_{ij \in E(H)} \mathbf{Y}_{ij}^{m_H(ij)} \right) \cdot \mathbb{1}_{\bigcap_{ij \in E(H)} (\mathcal{E}_i \cup \mathcal{E}_j)^c} = \mathbf{Y}_H \cdot \mathbb{1}_{\left(\bigcup_{ij \in E(H)} (\mathcal{E}_i \cup \mathcal{E}_j)\right)^c} \\
&= \mathbf{Y}_H \cdot \mathbb{1}_{\mathcal{E}_{V(H)}^c} = \mathbf{Y}_H \cdot \mathbb{1}_{\mathcal{E}_H^c} = \begin{cases} \mathbf{Y}_H & \text{if } H \text{ is truncated,} \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

B.1.2 An upper bound for every multigraph

In this section, we derive an upper bound $\overline{U}_H(\mathbf{x})$ on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$ for every multigraph H with at most st vertices and at most st multi-edges, where $t = K \cdot \log n$.

B.1.2.1 Informal discussion and useful definitions

Before delving into the details of the upper bound $\overline{U}_H(\mathbf{x})$ and the lower bound $\overline{L}_H$ that we mentioned in Section 4.5.1.2, let us first explain a few key observations that provide some intuition, and which clarifies our proof strategy. This section provides a road map for the proof of the bounds, and introduces notations and concepts that will be useful later.

Let H be an arbitrary multigraph. Observe that in the non-truncated case, the computation of $\mathbb{E}[\mathbf{Y}_H|\mathbf{x}]$ is easy because the edges are conditionally mutually independent given $\mathbf{x}$. More precisely,

$$\mathbb{E}[\mathbf{Y}_H|\mathbf{x}] = \prod_{ij \in E(H)} \mathbb{E}\left[\mathbf{Y}_{ij}^{m_H(ij)} \middle| \mathbf{x}\right].$$

Unfortunately, this is not the case for the truncated case: Since the presence of one edge can lead to the truncation of an incident edge, the random variables $\left(\overline{\mathbf{Y}}_{ij}\right)_{ij \in E(H)}$ are not conditionally mutually independent given $\mathbf{x}$. The dependencies between $\left(\overline{\mathbf{Y}}_{ij}\right)_{ij \in E(H)}$ makes the exact computation of $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]$ very complicated.

Another level of complication is that, unlike the non-truncated case, $\left(\overline{\mathbf{Y}}_{ij}\right)_{ij \in E(H)}$ are not conditionally independent of $\mathbf{G} - G(H)$ given $\mathbf{x}$. More precisely, $\left(\overline{\mathbf{Y}}_{ij}\right)_{ij \in E(H)}$ depend on $\mathbf{G} - G(H)$ through $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v \in V(H)}$, where

$$d_{\mathbf{G}-G(H)}(v) = \left|\left\{e \in \mathbf{G} : \ e \notin E(H) \text{ and } e \text{ is incident to } v\right\}\right|.$$

Furthermore, $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$ are not conditionally mutually independent given $\mathbf{x}$ because of edges in $\{ij : i,j \in V(H) \text{ and } ij \notin E(H)\}$. We call such edges *H-cross-edges*. Fortunately, if $V(H)$ contains at most st vertices, we have very few H-cross-edges and their effect will be negligible. More precisely, with high probability, no H-cross-edge will be present in $\mathbf{G}$, and if we condition on this event and on $\mathbf{x}$, the random variables $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$ become conditionally mutually independent. This will be made precise later.

Another phenomenon that we should be aware of, is the effect of truncation on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}, \left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}\right]$. In order to explain this, we will divide the edges of $G(H)$ into two categories:

- Edges of multiplicity 1 in H.
- Edges of multiplicity at least 2 in H.

Before discussing the effect of truncation on these two categories of edges, let us quickly mention a remark regarding some tempting approaches to upper and lower bound $\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]$, and which we did not find very successful.

Remark B.1. Since $\overline{\mathbf{Y}}_H = \mathbf{Y}_H \cdot \mathbb{1}_{\mathcal{E}_H^c}$, we have $\left|\overline{\mathbf{Y}}_H\right| = |\mathbf{Y}_H| \cdot \mathbb{1}_{\mathcal{E}_H^c} \leqslant |\mathbf{Y}_H|$. Therefore,

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right| \leqslant \mathbb{E}\left[|\overline{\mathbf{Y}}_H|\middle|\mathbf{x}\right] \leqslant \mathbb{E}\left[|\mathbf{Y}_H|\middle|\mathbf{x}\right].$$

We might hope to get a good upper bound for $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right|$ using the above inequality since $\mathbb{E}\left[|\mathbf{Y}_H|\middle|\mathbf{x}\right] = \prod_{ij\in E(H)} \mathbb{E}\left[|\mathbf{Y}_{ij}^{m_H(ij)}|\middle|\mathbf{x}\right]$ is easy to compute. However, $\mathbb{E}\left[|\mathbf{Y}_H|\middle|\mathbf{x}\right]$ can be too large compared to $\mathbb{E}\left[\mathbf{Y}_H\middle|\mathbf{x}\right]$. This is mainly because of edges of multiplicity 1 in H: If $ij \in E(H)$ is of multiplicity 1 in H, then

$$|\mathbb{E}[\mathbf{Y}_{ij}|\mathbf{x}]| = \frac{\varepsilon d}{2n},$$

while

$$\mathbb{E}\left[|\mathbf{Y}_{ij}|\middle|\mathbf{x}\right] = (1+o(1))\left(2 + \frac{\varepsilon \mathbf{x}_i \mathbf{x}_j}{2}\right)\frac{d}{n}.$$

Combining this observation with the fact that $E(H)$ can contain $\Theta(\log n)$ edges of multiplicity 1, it becomes evident that in general, $\mathbb{E}\left[|\mathbf{Y}_H|\middle|\mathbf{x}\right]$ is not a good upper bound on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right|$.

Other in-the-same-spirit approaches attempt to separately analyze the cases when $\overline{\mathbf{Y}}_H$ is positive or negative, i.e., write $\overline{\mathbf{Y}}_H = \overline{\mathbf{Y}}_H^+ - \overline{\mathbf{Y}}_H^-$, where $\overline{\mathbf{Y}}_H^+ = \max\{0, \overline{\mathbf{Y}}_H\}$ and $\overline{\mathbf{Y}}_H^- = \max\{0, -\overline{\mathbf{Y}}_H\}$. One might hope to obtain good bounds on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right] = \mathbb{E}\left[\overline{\mathbf{Y}}_H^+\middle|\mathbf{x}\right] - \mathbb{E}\left[\overline{\mathbf{Y}}_H^-\middle|\mathbf{x}\right]$ by obtaining good upper and lower bounds on $\mathbb{E}\left[\overline{\mathbf{Y}}_H^+\middle|\mathbf{x}\right]$ and $\mathbb{E}\left[\overline{\mathbf{Y}}_H^-\middle|\mathbf{x}\right]$, respectively. However, it turns out that $\mathbb{E}\left[\overline{\mathbf{Y}}_H^+\middle|\mathbf{x}\right]$ and $\mathbb{E}\left[\overline{\mathbf{Y}}_H^-\middle|\mathbf{x}\right]$ can be too large compared to $\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]$. This is essentially for the same reason why $\mathbb{E}\left[|\mathbf{Y}_H|\middle|\mathbf{x}\right]$ can be too large compared to $\left|\mathbb{E}\left[\mathbf{Y}_H\middle|\mathbf{x}\right]\right|$. Therefore, in order for this approach to succeed in deriving good bounds for $\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]$, the

bounds on $\mathbb{E}\left[\overline{\mathbf{Y}}_H^+ \middle| \mathbf{x}\right]$ and $\mathbb{E}\left[\overline{\mathbf{Y}}_H^- \middle| \mathbf{x}\right]$ need to be extremely tight, and this is very hard to obtain for a general H.

Effect of truncation on edges of multiplicity 1. For edges of multiplicity 1, we observe that truncation can make the absolute value of $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}\right]$ grow too much. To illustrate this point, take the example where H is a cycle with st edges of multiplicity 1. For the non-truncated case, since $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$ is conditionally independent from $\mathbf{Y}_H$ given $\mathbf{x}$, if we condition on the event that $d_{\mathbf{G}-G(H)}(v) = \Delta$ for every $v \in V(H)$, we get

$$\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}, \left\{d_{\mathbf{G}-G(H)}(v) = \Delta, \forall v \in V(H)\right\}\right] = \prod_{uv\in E(H)} \frac{\varepsilon d \mathbf{x}_u \mathbf{x}_v}{2n} = \left(\frac{\varepsilon d}{2n}\right)^{st}.$$

On the other hand, for the truncated case, if we condition on the event that $d_{\mathbf{G}-G(H)}(v) = \Delta$ for every $v \in V(H)$, then $\overline{\mathbf{Y}}_H$ is non-zero if and only if all the edges in $E(H)$ are not present in $\mathbf{G}$, hence

$$\begin{aligned}\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \left\{d_{\mathbf{G}-G(H)}(v) = \Delta, \forall v \in V(H)\right\}\right] &= \prod_{ij\in E(H)} \left(1 - \left(1 + \frac{\varepsilon \mathbf{x}_i \mathbf{x}_j}{2}\right)\frac{d}{n}\right)\left(-\frac{d}{n}\right) \\ &= (1 \pm o(1))\left(-\frac{d}{n}\right)^{st}.\end{aligned}$$

Therefore, the absolute value was multiplied by a factor of approximately $\left(\frac{2}{\varepsilon}\right)^{st}$, which can be too large. Fortunately, this can be mitigated by choosing Δ to be large enough so that problematic events such as $\left\{d_{\mathbf{G}-G(H)}(v) = \Delta, \forall v \in V(H)\right\}$ will have a very small probability.

Roughly speaking, the contribution of an edge of multiplicity 1 in the non-truncated case is $|\mathbb{E}[\mathbf{Y}_{ij}|\mathbf{x}]| = \frac{\varepsilon d}{2n}$, and its contribution in the truncated case can be as large as $\frac{d}{n}$. Since we are trying to make our upper bound as tight as possible — i.e., as low as possible — we would like to find situations where edges of multiplicity 1 behave similarly to the non-truncated case, because their contribution will be relatively small.

This brings us to the following observation: Assume that $ij \in E(H)$ is an edge of multiplicity 1 such that $d_{\mathbf{G}-G(H)}(i) + d_{G(H)}(i) \leqslant \Delta$ and $d_{\mathbf{G}-G(H)}(j) + d_{G(H)}(j) \leqslant \Delta$, where $d_{G(H)}(i)$ denotes the degree of i in H without counting multiplicities, i.e., $d_{G(H)}(i)$ is the degree of i in the underlying graph $G(H)$. In this case, we have

$$d_{\mathbf{G}}(i) = d_{\mathbf{G}-G(H)}(i) + d_{\mathbf{G}\cap G(H)}(i) \leqslant d_{\mathbf{G}-G(H)}(i) + d_{G(H)}(i) \leqslant \Delta.$$

Similarly, we have $d_{\mathbf{G}}(j) \leqslant \Delta$. Therefore, no matter which edges of $E(H)$ are present in $\mathbf{G}$ and which are absent, we are sure that neither i nor j will be deleted, and so $\overline{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij}$. Furthermore, it is easy to see that given $\mathbf{x}$, and given that $d_{\mathbf{G}-G(H)}(i) + d_{G(H)}(i) \leqslant \Delta$ and $d_{\mathbf{G}-G(H)}(j) + d_{G(H)}(j) \leqslant \Delta$, the random variable $\overline{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij}$ is conditionally independent from $\overline{\mathbf{Y}}_{H-ij}$. This motivates the following definition:

Definition B.2. Let H be a multigraph. For every vertex $v \in V(H)$, if $d_{\mathbf{G}-G(H)}(v)+d_{G(H)}(v) \leqslant \Delta$, we say that v is $(\mathbf{G}, H)$-*safe*. We say that v is $(\mathbf{G}, H)$-*unsafe* if it is not $(\mathbf{G}, H)$-*safe*.

We say that a subset S of $V(H)$ is *completely* $(\mathbf{G}, H)$-*safe* if all the vertices in it are $(\mathbf{G}, H)$-*safe*. Similarly, we say that it is *completely* $(\mathbf{G}, H)$-*unsafe* if all the vertices in it are $(\mathbf{G}, H)$-*unsafe*.

An edge $ij \in E(H)$ is said to be $(\mathbf{G}, H)$-*safe* if both i and j are safe.

If $\mathbf{G}$ and H are clear from the context, we drop $(\mathbf{G}, H)$ and simply write safe, completely safe, unsafe, and completely unsafe.

In summary, we have the following observations:

- If a vertex $v \in V(H)$ is safe, we are sure that it will not be deleted. On the other hand, if v is unsafe, it may or may not be deleted.
- If $ij \in E(H)$ is safe, then $\overline{\mathbf{Y}}_{ij} = \mathbf{Y}_{ij}$ and $\overline{\mathbf{Y}}_{ij}$ is conditionally independent of $\overline{\mathbf{Y}}_{H-ij}$ given $\mathbf{x}$. On the other hand, if ij is unsafe, then $\overline{\mathbf{Y}}_{ij}$ may or may not be equal to $\mathbf{Y}_{ij}$.

If $ij \in E(H)$ is an edge of multiplicity 1 in H, then we have

$$\begin{aligned}
\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \left\{ij \text{ is safe}\right\}\right] &= \mathbb{E}\left[\overline{\mathbf{Y}}_{ij} \middle| \mathbf{x}, \left\{ij \text{ is safe}\right\}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H-ij} \middle| \mathbf{x}, \left\{ij \text{ is safe}\right\}\right] \\
&= \mathbb{E}\left[\mathbf{Y}_{ij} \middle| \mathbf{x}, \left\{ij \text{ is safe}\right\}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H-ij} \middle| \mathbf{x}, \left\{ij \text{ is safe}\right\}\right] \\
&= \mathbb{E}[\mathbf{Y}_{ij} | \mathbf{x}] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H-ij} \middle| \mathbf{x}, \left\{ij \text{ is safe}\right\}\right] \\
&= \frac{\varepsilon \mathbf{x}_i \mathbf{x}_j d}{2n} \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H-ij} \middle| \mathbf{x}, \left\{ij \text{ is safe}\right\}\right].
\end{aligned}$$

As we can see from the above discussion, edges of multiplicity 1 that are safe behave similarly to the non-truncated case. In order to benefit from this phenomenon as much as possible, we would like the probability that an edge ij is unsafe to be small. Equivalently, we would like the probability that i (respectively j) is unsafe to be small. By analyzing the event $\left\{d_{\mathbf{G}-G(H)}(i) + d_{G(H)}(i) > \Delta\right\}$, we notice the following:

- The distribution of the random variable $d_{\mathbf{G}-G(H)}(i)$ is Binomial$\left(n - O(st), \frac{d}{n}\right)$,[1] so we can approximate it by a Poisson(d) distribution.
- The probability of the event $\left\{d_{\mathbf{G}-G(H)}(i) + d_{G(H)}(i) > \Delta\right\}$ is lower bounded by the probability that $d_{\mathbf{G}-G(H)}(i) > \Delta$. Now since Δ must be a constant[2] that should not scale with n, it is easy to see that the probability that i is unsafe is lower bounded by a strictly positive constant[3] and cannot be made go to zero as n goes to infinity. Nevertheless, the larger the value of $\Delta - d_{G(H)}(i)$ is, the smaller is the probability that

[1] Note that here we are not conditioning on $\mathbf{x}$.

[2] Recall that the main motivation behind truncation is to get rid of the effect of large degree vertices, and so Δ must be a constant and cannot diverge with n. Nevertheless, we can tune Δ and make it as large as we wish, as long as it stays polynomial in s and d, and polylogarithmic in ε.

[3] This is because the probability that Poisson$(d) > \Delta$ is constant.

i is unsafe. In particular, if $\frac{\Delta}{2}$ is very large with respect to d and $d_{G(H)}(i) \leqslant \frac{\Delta}{2}$, then the probability that i is unsafe is at most the probability that $d_{\mathbf{G}-G(H)}(i) > \frac{\Delta}{2}$, which can be approximated by the probability that a Poisson(d) random variable is greater than $\frac{\Delta}{2}$, which in turn is small as long as $\frac{\Delta}{2}$ is large with respect to d.

We conclude that vertices satisfying $d_{G(H)}(i) \leqslant \frac{\Delta}{2}$ are "well-behaved", and edges whose end-vertices are both of this type are also "well-behaved" in the sense that with large probability they will be safe and will behave similarly to the non-truncated case.

For edges $ij \in E(H)$ of multiplicity 1 in H that are not well-behaved, i.e., $d_{G(H)}(i) > \frac{\Delta}{2}$ or $d_{G(H)}(j) > \frac{\Delta}{2}$, we will use a very loose[4] upper-bound to estimate the contribution of $\overline{\mathbf{Y}}_{ij}$. Nevertheless, we will mitigate the effect of such edges depending on what caused them to be not well-behaved. In principal, a vertex $v \in V(H)$ can have a large degree in H, i.e., $d_{G(H)}(v) > \frac{\Delta}{2}$, either because it has many edges of multiplicity 1 in H which are incident to it, or because it has many edges of multiplicity at least 2 in H which are incident to it. We will treat each case differently. This motivates the following definition:

Definition B.3. Let H be an arbitrary multigraph. For every $v \in V(H)$, we define the following:

- The 1-degree of v in H, denoted $d_1^H(v)$ is the number of edges in $E(H)$ that are incident to v, and which have multiplicity 1 in H.
- An edge in $E(H)$ is said to be of multiplicity $\geqslant 2$ in H if it has multiplicity at least 2 in H. The ($\geqslant 2$)-degree of v in H, denoted $d_{\geqslant 2}^H(v)$ is the number of edges in $E(H)$ that are incident to v, and which have multiplicity $\geqslant 2$ in H.

Clearly, $d_{G(H)}(v) = d_1^H(v) + d_{\geqslant 2}^H(v)$ is the degree of v in the underlying graph $G(H)$, i.e., the degree of v in H without counting multiplicities.

Roughly speaking, we will mitigate the effects of "not-well-behaved" edges using the following ideas:

- We will show that there are very few block self-avoiding-walks that have vertices with large 1-degree, and this will counter-act the loose upper bounds that we use for the edges of multiplicity 1 that are incident to such vertices.
- Assuming that H is a multigraph that does not contain any vertex with large 1-degree, we will show that there are very few vertices with large ($\geqslant 2$)-degree. Therefore, the number of edges of multiplicity 1 that are incident to such vertices is not large, and so the aggregate effect of the loose upper bounds that we use for such edges is not too severe.

[4] Essentially, we will upper bound $\left|\overline{\mathbf{Y}}_{ij}\right|$ by $|\mathbf{Y}_{ij}|$.

Effect of truncation on edges of multiplicity at least 2. Compared to edges of multiplicity 1 in H, edges of multiplicity $\geqslant 2$ in H are much easier to treat: Unlike edges of multiplicity 1 in H, for an edge ij of multiplicity $m_H(ij) \geqslant 2$, the conditional expectations $\mathbb{E}\left[\mathbf{Y}_{ij}^{m_H(ij)} \middle| \mathbf{x}\right]$ and $\mathbb{E}\left[|\mathbf{Y}_{ij}^{m_H(ij)}| \middle| \mathbf{x}\right]$ are approximately equal. Therefore, for many cases of interest, $\left|\mathbf{Y}_{ij}^{m_H(ij)}\right|$ is a good upper bound on $\overline{\mathbf{Y}}_{ij}^{m_H(ij)}$.

If $m_H(ij) \geqslant 2$, we have the following:

- If $ij \in \mathbf{G}$, then

$$\left|\mathbf{Y}_{ij}^{m_H(ij)} \cdot \mathbb{P}[ij \in \mathbf{G}]\right| = \left(1 - \frac{d}{n}\right)^{m_H(ij)} \cdot \left(1 + \frac{\varepsilon \mathbf{x}_i \mathbf{x}_j}{2}\right)\frac{d}{n} = \Omega\left(\frac{1}{n}\right).$$

- If $ij \notin \mathbf{G}$, then

$$\left|\mathbf{Y}_{ij}^{m_H(ij)} \cdot \mathbb{P}[ij \notin \mathbf{G}]\right| = \left|\left(-\frac{d}{n}\right)^{m_H(ij)} \cdot \left(1 - \left(1 + \frac{\varepsilon \mathbf{x}_i \mathbf{x}_j}{2}\right)\frac{d}{n}\right)\right| = O\left(\frac{1}{n^{m_H(ij)}}\right).$$

As we can see, the contribution of the case when $ij \in \mathbf{G}$ dominates the contribution of the case when $ij \notin \mathbf{G}$. Therefore, when we want to compute $\mathbb{E}\left[\mathbf{Y}_{ij}^{m_H(ij)} \middle| \mathbf{x}\right]$, or $\mathbb{E}\left[|\mathbf{Y}_{ij}^{m_H(ij)}| \middle| \mathbf{x}\right]$, we can discard the case for which $ij \notin \mathbf{G}$. This is essentially the main reason why $\mathbb{E}\left[\mathbf{Y}_{ij}^{m_H(ij)} \middle| \mathbf{x}\right]$ and $\mathbb{E}\left[|\mathbf{Y}_{ij}^{m_H(ij)}| \middle| \mathbf{x}\right]$ are approximately equal.

A similar phenomenon occurs for the truncated case, except that the contribution of the case for which $ij \in \mathbf{G}$ might be multiplied by zero due to truncation. Therefore, roughly speaking, when we want to compute $\mathbb{E}\left[\overline{\mathbf{Y}}_{ij}^{m_H(ij)} \middle| \mathbf{x}\right]$, it is sufficient to consider the cases where the number of edges of multiplicity $\geqslant 2$ in H that are present in $\mathbf{G}$ is as large as possible.[5]

Now notice that if there is a vertex v with $d^H_{\geqslant 2}(v) > \Delta$, it is impossible to include all the edges of multiplicity $\geqslant 2$ in H that are incident to v without causing truncation. Therefore, roughly speaking, $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$ will be at least $O\left(\frac{1}{n}\right)$ smaller compared to $|\mathbb{E}[\mathbf{Y}_H|\mathbf{x}]|$. This will allow us to show that the aggregate contribution of all multigraphs having at least one vertex v satisfying $d^H_{\geqslant 2}(v) > \Delta$ is negligible.[6]

[5] To be more precise, we should give priority to edges with larger multiplicities. For example, if we have a multigraph where the multiplicities are 2, 3 and 4, we first try to include as many edges of multiplicity 4 as possible without causing truncation, then, if there is still room to include edges of multiplicity 3, we include as many of them as possible before turning to edges of multiplicity 2.

[6] This is the main reason why the algorithm that is based on $Q^{(s)}(\overline{\mathbf{Y}})$ works whereas the one based on $Q^{(s)}(\mathbf{Y})$ does not work: When we write $\mathbb{E}[\mathrm{Tr}((Q^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T})^t)]$ as the sum of contributions of block self-avoiding-walks, we find that there are too many block self-avoiding-walks where $d^H_{\geqslant 2}(v) > \Delta$ for some vertex $v \in V(H)$. This makes the value of $\|Q^{(s)}(\mathbf{Y}) - \mathbf{x}\mathbf{x}^\mathsf{T}\|_t$ too large. On the other hand, for the truncated case, even though we have too many such block self-avoiding-walks, the contribution of each one of them is very small.

B.1.2.2 The truncation threshold

We choose the truncation threshold to be of the form

$$\Delta = \max\left\{128e^4d^4, 40Asd, 2\log(2As) + 12A\tau s^2 \cdot \log 2 + 8A^2\tau^2 s^2\left(\log\frac{6}{\varepsilon}\right)^2\right\}, \tag{B.1.1}$$

where

$$\tau = As\log\frac{6}{\varepsilon}, \tag{B.1.2}$$

and $A > \max\{1, 100K, \frac{100}{K}\}$ is some constant to be chosen later. This form of Δ was carefully chosen to allow for a proof of all the phenomena that were mentioned in Appendix B.1.2.1. Before starting to prove the upper bound $\overline{U}_H(\mathbf{x})$ on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right|$, we need to introduce a few definitions.

Definition B.4. Let H be a multigraph. We classify the vertices $v \in V(H)$ according to their degree-1 as follows:

- If $d_1^H(v) \leqslant \tau$, we say that v is 1-*small* in H. We denote the set of 1-small vertices in H as $\mathcal{S}_1(H)$.
- If $d_1^H(v) > \tau$, we say that v is 1-*large* in H. We denote the set of 1-large vertices in H as $\mathcal{L}_1(H)$.

Definition B.5. Let H be a multigraph. We classify the vertices $v \in V(H)$ according to their ($\geqslant$2)-degree as follows:

- If $d_{\geqslant 2}^H(v) \leqslant \frac{\Delta}{4}$, we say that v is ($\geqslant$2)-*small* in H. We denote the set of ($\geqslant$2)-small vertices in H as $\mathcal{S}_{\geqslant 2}(H)$.
- If $\frac{\Delta}{4} < d_{\geqslant 2}^H(v) \leqslant \Delta$, we say that v is ($\geqslant$2)-*intermediate* in H. We denote the set of ($\geqslant$2)-intermediate vertices in H as $\mathcal{I}_{\geqslant 2}(H)$.
- If $d_{\geqslant 2}^H(v) > \Delta$, we say that v is ($\geqslant$2)-*large* in H. We denote the set of ($\geqslant$2)-large vertices in H as $\mathcal{L}_{\geqslant 2}(H)$.

Definition B.6. Let H be a multigraph. We denote the set of edges in $G(H)$ of multiplicity 1 in H as $E_1(H)$, and denote the set of edges in $G(H)$ of multiplicity $\geqslant$2 in H as $E_{\geqslant 2}(H)$.

An edge in $E_1(H)$ is said to be *annoying* if it is incident to at least one vertex in $\mathcal{L}_1(H)$. We denote the set of annoying edges as $E_1^a(H)$.

We partition $E_{\geqslant 2}(H)$ into two sets:

$$E_{\geqslant 2}^a(H) = \left\{uv \in E_{\geqslant 2}(H) : u \notin \mathcal{L}_{\geqslant 2}(H) \text{ and } v \notin \mathcal{L}_{\geqslant 2}(H)\right\},$$

and

$$E_{\geqslant 2}^b(H) = \left\{uv \in E_{\geqslant 2}(H) : u \in \mathcal{L}_{\geqslant 2}(H) \text{ or } v \in \mathcal{L}_{\geqslant 2}(H)\right\}.$$

Definition B.7. For every multigraph H with at most $st = sK\log n$ vertices and at most st multi-edges, we define the quantity

$$\overline{U}_H(\mathbf{x}) = n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|}\frac{1}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d^H_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|\mathbb{E}_{\geqslant 2}(H)|}\prod_{uv\in E^a_{\geqslant 2}(H)}\left[1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}+\frac{3d}{\sqrt{n}}\right].$$

In the rest of Appendix B.1.2, we show that $\overline{U}_H(\mathbf{x})$ is an upper bound on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right|$ for n large enough, assuming that H is a multigraph with at most $st = sK\log n$ vertices and at most st multi-edges.

Note that with more refined calculations, it is possible to get a better upper bound. In any case, $\overline{U}_H(\mathbf{x})$ is good enough for our purposes.

B.1.2.3 Analyzing edges of multiplicity 1

We start by proving an upper bound on the probability that a subset of $\mathcal{S}_1(H)\cap\mathcal{S}_{\geqslant 2}(H)$ is completely unsafe. We first need to introduce some notation:

Definition B.8. An edge ij satisfying $i,j\in V(H)$ and $ij\notin E(H)$ is called an *H-cross-edge*.

For every $v\in V(H)$, define the following:

$$\begin{aligned} d_{\mathbf{G}-G(H)}(v) &= \left|\left\{e\in\mathbf{G}:\ e \text{ is incident to } v \text{ and } e\notin E(H)\right\}\right| \\ &= \left|\left\{uv\in\mathbf{G}:\ uv\notin E(H)\right\}\right|, \\ d^i_{\mathbf{G}-G(H)}(v) &= \left|\left\{uv\in\mathbf{G}:\ uv\notin E(H) \text{ and } u\in V(H)\right\}\right|, \end{aligned}$$

and

$$d^o_{\mathbf{G}-G(H)}(v) = \left|\left\{uv\in\mathbf{G}:\ uv\notin E(H) \text{ and } u\notin V(H)\right\}\right|.$$

Clearly, $d_{\mathbf{G}-G(H)}(v) = d^i_{\mathbf{G}-G(H)}(v) + d^o_{\mathbf{G}-G(H)}(v)$.

$d^o_{\mathbf{G}-G(H)}(v)$ can be thought of as the "$V(H)$-outside degree" of v in $\mathbf{G}-G(H)$, i.e., the number of edges in $\mathbf{G}-G(H)$ that go from v to the outside of $V(H)$. On the other hand, $d^i_{\mathbf{G}-G(H)}(v)$ can be thought of as the "$V(H)$-inside degree" of v in $\mathbf{G}-G(H)$, i.e., the number of edges in $\mathbf{G}-G(H)$ that are incident to v and are inside $V(H)$.

If a vertex $v\in V(H)$ is unsafe, then either $d^i_{\mathbf{G}-G(H)}(v)$ is large or $d^o_{\mathbf{G}-G(H)}(v)$ is large. It turns out that with high probability, we have[7] $d^i_{\mathbf{G}-G(H)}(v) = 0$. Therefore, the probability that v is unsafe is dominated by the probability that $d^o_{\mathbf{G}-G(H)}(v) > \Delta - d_{G(H)}(v)$.

Notice that if $v\in\mathcal{S}_1(H)\cap\mathcal{S}_{\geqslant 2}(H)$, then

$$d_{G(H)}(v) = d^H_1(v) + d^H_{\geqslant 2}(v) \leqslant \tau + \frac{\Delta}{4} \leqslant \frac{\Delta}{4}+\frac{\Delta}{4} = \frac{\Delta}{2}.$$

The following lemma derives an upper bound on the probability that the outside degree is larger than $\frac{\Delta}{2}$.

[7] This is essentially because we have at most $s^2t^2 = o(n)$ H-cross-edges, and the probability of any particular one of them being present in $\mathbf{G}$ is $O\left(\frac{1}{n}\right)$.

Lemma B.9. *Let H be a multigraph such that $|V(H)| \leqslant st$, where $t = K \log n$. For every $v \in \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$, we have*

$$\mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{4} \middle| \mathbf{x} \right] \leqslant \frac{\eta}{2},$$

where[8]

$$\eta := \frac{1}{As \cdot 2^{6A\tau s^2}} \left(\frac{\varepsilon}{6}\right)^{4s^2\tau^2} \leqslant \frac{1}{As} \left(\frac{\varepsilon}{6}\right)^{\tau}. \tag{B.1.3}$$

Proof. The proof is based on simple calculations based on the probability distribution of the sum of Bernoulli random variables. The detailed proof can be found in Appendix B.3.1.1. □

In the following, we will show that $d^i_{\mathbf{G}-G(H)}(v) = 0$ with high probability. The following definition will be useful.

Definition B.10. Let H be a multigraph. We say that a vertex $v \in V(H)$ is *H-cross-free in* $\mathbf{G}$ if there is no H-cross-edge that is present in $\mathbf{G}$, and which is incident to v. In other words, v is H-cross-free in $\mathbf{G}$ if $d^i_{\mathbf{G}-G(H)}(v) = 0$. We say that $v \in V(H)$ is *H-crossing in* $\mathbf{G}$ if it is not H-cross-free in $\mathbf{G}$.

A subset of $V(H)$ is said to be *completely H-cross-free in* $\mathbf{G}$ if all the vertices in it are H-cross-free in $\mathbf{G}$. We say that it is *completely H-crossing in* $\mathbf{G}$ if all the vertices in it are H-crossing in $\mathbf{G}$.

If $\mathbf{G}$ and H are clear from the context, we drop $\mathbf{G}$ and H and simply write cross-free, crossing, completely cross-free, and completely crossing.

The following lemma shows that the probability that a nonempty subset of $V(H)$ is completely crossing is small. Furthermore, the larger the set, the smaller is the probability. This essentially means that we have $d^i_{\mathbf{G}-G(H)}(v) = 0$ with high probability.

Lemma B.11. *Let H be a multigraph such that $|V(H)| \leqslant st$, where $t = K \cdot \log n$. If n is large enough, then for every $S \subseteq V(H)$, the conditional probability given $\mathbf{x}$ that S is completely H-crossing in $\mathbf{G}$ can be upper bounded by:*

$$\mathbb{P}\left[\{S \text{ is completely } H\text{-crossing in } \mathbf{G}\} \middle| \mathbf{x} \right] \leqslant \left(\frac{2ds^2t^2}{n} \right)^{|S|/2}.$$

Proof. The proof is based on a simple application of the union bound. The detailed proof can be found in Appendix B.3.1.1. □

By combining Lemma B.9 and Lemma B.11 and using the fact that an unsafe vertex $v \in V(H)$ is either crossing or satisfies $d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}$, we can show the following upper bound on the probability that a subset of $\mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$ is completely unsafe:

[8] In the calculations for $\mathbb{E}\left[\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)\right]$, we will need $\eta \leqslant \frac{1}{As}\left(\frac{\varepsilon}{6}\right)^{\tau}$. On the other hand, in the calculations for $\mathbb{E}\left[\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^{\mathsf{T}}\right)^t\right)\right]$, we will need $\eta \leqslant \frac{1}{s \cdot 2^{6A\tau s^2}}\left(\frac{\varepsilon}{6}\right)^{4s^2\tau^2}$.

Lemma B.12. *Recall the definition of safe vertices in Definition B.2, and let H be a multigraph such that $|V(H)| \leqslant st$, where $t = K \cdot \log n$. If n is large enough, then for every $V \subseteq \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$, the conditional probability that V is completely unsafe given $\mathbf{x}$ can be upper bounded by:*

$$\mathbb{P}\left[\left\{V \text{ is completely } (\mathbf{G}, H)\text{-unsafe}\right\}\middle|\mathbf{x}\right] \leqslant \eta^{|V|},$$

where η is as in Eq. (B.1.3).

Proof. The detailed proof can be found in Appendix B.3.1.1. □

Since η is very small for large A, Lemma B.12 implies that it is unlikely for unsafe edges to occur. On the other hand, the following lemma implies that safe edges behave similarly to the truncated case. These two observations are what ultimately makes it possible to derive a good upper bound on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right|$.

Lemma B.13. *Let H be a multigraph and let $uv \in E(H)$. Let $\mathcal{E}$ be an event satisfying:*

- *$\mathcal{E}$ implies that uv is $(\mathbf{G}, H)$-safe, i.e., $\forall (G, x) \in \mathcal{E}$, u and v are (G, H)-safe.*
- *The event $\mathcal{E}$ depends only on $\mathbf{x}$ and $\mathbf{G} - uv$, i.e., $\mathcal{E}$ is $\sigma(\mathbf{x}, \mathbf{G} - uv)$-measurable. In other words, if we condition on $\mathbf{x}$, then $\mathcal{E}$ depends only $(\mathbb{1}_{u'v' \in \mathbf{G}})_{u', v' \in [n]:\, u'v' \neq uv}$. This implies that given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent from $\mathbb{1}_{\{uv \in \mathbf{G}\}}$.*

Then,

$$\forall (G, x) \in \mathcal{E},\ \overline{\mathbf{Y}}_{uv}(G) = \mathbf{Y}_{uv}(G), \tag{B.1.4}$$

and

$$\mathbb{E}\left[\overline{\mathbf{Y}}_{uv}^{m_H(uv)}\middle|\mathbf{x}, \mathcal{E}\right] = \mathbb{E}\left[\mathbf{Y}_{uv}^{m_H(uv)}\middle|\mathbf{x}\right]. \tag{B.1.5}$$

Furthermore, if H' is a submultigraph of H containing uv with the same multiplicity as in H, then we have

$$\mathbb{E}\left[\overline{\mathbf{Y}}_{H'}\middle|\mathbf{x}, \mathcal{E}\right] = \mathbb{E}\left[\mathbf{Y}_{uv}^{m_H(uv)}\middle|\mathbf{x}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H'-uv}\middle|\mathbf{x}, \mathcal{E}\right]. \tag{B.1.6}$$

Proof. Equations Eq. (B.1.4) and Eq. (B.1.5) are trivial. In order to get Equation Eq. (B.1.6), observe that we have

$$\begin{aligned}
\mathbb{E}\left[\overline{\mathbf{Y}}_{H'}\middle|\mathbf{x}, \mathcal{E}\right] &= \mathbb{E}\left[\overline{\mathbf{Y}}_{uv}^{m_H(uv)} \cdot \overline{\mathbf{Y}}_{H'-e}\middle|\mathbf{x}, \mathcal{E}\right] \\
&= \mathbb{E}\left[\mathbf{Y}_{uv}^{m_H(uv)} \cdot \overline{\mathbf{Y}}_{H'-e}\middle|\mathbf{x}, \mathcal{E}\right] \\
&\overset{(*)}{=} \mathbb{E}\left[\mathbf{Y}_{uv}^{m_H(uv)}\middle|\mathbf{x}, \mathcal{E}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H'-uv}\middle|\mathbf{x}, \mathcal{E}\right] \\
&\overset{(\dagger)}{=} \mathbb{E}\left[\mathbf{Y}_{uv}^{m_H(uv)}\middle|\mathbf{x}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H'-uv}\middle|\mathbf{x}, \mathcal{E}\right],
\end{aligned}$$

where $(*)$ follows from the fact that $\mathcal{E}$ is $\sigma(\mathbf{x}, \mathbf{G} - uv)$-measurable, which implies that given $\mathbf{x}$ and $\mathcal{E}$, we have that $\mathbf{Y}_{uv} = \mathbb{1}_{\{uv \in \mathbf{G}\}} - \frac{d}{n}$ is conditionally independent from $\left(\overline{\mathbf{Y}}_{u'v'}\right)_{u'v' \in E(H'-uv)}$, and $(\dagger)$ follows from the fact that given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent of $\mathbb{1}_{\{uv \in \mathbf{G}\}}$. □

We will only use Lemma B.13 to upper bound the contribution of multiplicity-1 edges. For edges of multiplicity $\geqslant 2$, we do not need the safeness mechanism. In fact, for edges of multiplicity $\geqslant 2$, we will ignore the edges in $\mathbf{G} - G(H)$ and the edges in $\mathbf{G} \cap E_1(H)$, and focus on the truncation events that are caused by having too many edges in $\mathbf{G} \cap E_{\geqslant 2}(H)$. This motivates the following definition:

Definition B.14. For every multigraph H, define

$$\tilde{\mathbf{Y}}^H_{\geqslant 2} = \prod_{uv \in E_{\geqslant 2}(H)} \tilde{\mathbf{Y}}^{m_H(uv)}_{uv, E_{\geqslant 2}(H)},$$

where

$$\tilde{\mathbf{Y}}_{uv, E_{\geqslant 2}(H)} := \mathbf{Y}_{uv} \cdot \mathbb{1}_{\{d_{\mathbf{G} \cap E_{\geqslant 2}(H)}(u) \leqslant \Delta\}} \cdot \mathbb{1}_{\{d_{\mathbf{G} \cap E_{\geqslant 2}(H)}(v) \leqslant \Delta\}},$$

and

$$d_{\mathbf{G} \cap E_{\geqslant 2}(H)}(v) := \left|\left\{w \in V(H) : \ vw \in \mathbf{G} \cap E_{\geqslant 2}(H)\right\}\right|.$$

Lemma B.15. *For every multigraph H with at most $st = sK \log n$ vertices and at most st multi-edges, we have*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right| \leqslant n^{\frac{2K}{A}} \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[\left|\tilde{\mathbf{Y}}^H_{\geqslant 2}\right| \middle| \mathbf{x}\right],$$

where $E_1^a(H)$ is as in Definition B.6.

Proof. We only provide a proof sketch here. The detailed proof can be found in Appendix B.3.1.2.

The main idea of the proof is based on partitioning $E_1(H)$ into three sets:

$$E_1^a(H) = \left\{uv \in E_1(H) : \ u \in \mathcal{L}_1(H) \text{ or } v \in \mathcal{L}_1(H)\right\},$$

$$E_1^b(H) = \left\{uv \in E_1(H) \setminus E_1^a(H) : \ u \in \mathcal{S}_{\geqslant 2}(H) \text{ and } v \in \mathcal{S}_{\geqslant 2}(H)\right\}, \tag{B.1.7}$$

and

$$E_1^d(H) = \left\{uv \in E_1(H) \setminus E_1^a(H) : \ u \notin \mathcal{S}_{\geqslant 2}(H) \text{ or } v \notin \mathcal{S}_{\geqslant 2}(H)\right\}. \tag{B.1.8}$$

The end-vertices of edges in $E_1^b(H)$ belong to $\mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$. Lemma B.12 implies that it is unlikely for edges in $E_1^b(H)$ to be unsafe, and so Lemma B.13 implies that they will likely behave similarly to the truncated case. As we will see in the detailed proof in Appendix B.3.1.2, the total contribution of edges in $E_1^b(H)$ can be upper bounded by $n^{\frac{K}{A}} \left(\frac{\varepsilon d}{2n}\right)^{|E_1^b(H)|}$. Note that the term $\left(\frac{\varepsilon d}{2n}\right)^{|E_1^b(H)|}$ is the contribution of edges in $E_1^b(H)$ in the non-truncated case. The factor $n^{\frac{K}{A}}$ comes from the fact that the edges in $E_1^b(H)$ are not always safe: They are only likely to be so, and the small probability for the edges in $E_1^b(H)$

to be unsafe will ultimately cause a multiplication by a factor that can be upper bounded by $n^{\frac{K}{A}}$.

For the edges in $E_1^a(H) \cup E_1^d(H)$, we did not find an easy way to get a good upper bound on their contribution, so we used a potentially very loose upper bound. Roughly speaking, we used the fact that $\left|\overline{\mathbf{Y}}_{uv}\right| \leqslant |\mathbf{Y}_{uv}|$ for every $uv \in G(H)$ in order to upper bound the contribution of an edge $uv \in E_1^a(H) \cup E_1^d(H)$ by $\mathbb{E}\left[|\mathbf{Y}_{uv}|\middle|\mathbf{x}\right] \leqslant \frac{3d}{n}$. Therefore, the total contribution of edges in $E_1^a(H) \cup E_1^d(H)$ can be upper bounded by

$$\left(\frac{3d}{n}\right)^{|E_1^a(H)|+|E_1^d(H)|} = \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|}\left(\frac{6}{\varepsilon}\right)^{|E_1^d(H)|}\left(\frac{\varepsilon d}{2n}\right)^{|E_1^a(H)|+|E_1^d(H)|}.$$

Now since the vertices in $V(H) \setminus \mathcal{S}_{\geqslant 2}(H)$ have degrees of at least $\frac{\Delta}{4}$, we cannot have too many vertices in $V(H) \setminus \mathcal{S}_{\geqslant 2}(H)$. Now since every edge in $E_1^d(H)$ must be incident to a vertex in $\mathcal{S}_1(H) \cap \left(V(H) \setminus \mathcal{S}_{\geqslant 2}(H)\right)$ and since every vertex in $\mathcal{S}_1(H)$ is incident to at most τ vertices, we can deduce that we cannot have too many edges in $E_1^d(H)$. This observation be used to show that $\left(\frac{6}{\varepsilon}\right)^{|E_1^d(H)|} \leqslant n^{\frac{K}{A}}$.

For edges of multiplicity $\geqslant 2$, it is easy to see that

$$\prod_{uv \in E_{\geqslant 2}(H)} \left|\overline{\mathbf{Y}}_{uv}^{m_H(uv)}\right| \leqslant \left|\tilde{\mathbf{Y}}_{\geqslant 2}^H\right|.$$

By combining all these observations together, we get

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right| \leqslant n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H|\middle|\mathbf{x}\right].$$

□

B.1.2.4 Analyzing edges of multiplicity at least 2

Lemma B.16. *For every multigraph H with at most $st = sK\log n$ vertices and at most st multi-edges, and for n large enough, we have*

$$\mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H|\middle|\mathbf{x}\right] \leqslant \frac{1}{\prod_{v \in \mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d_{\geqslant 2}^H(v)-\Delta\right)}}\left(\frac{d}{n}\right)^{|\mathbb{E}_{\geqslant 2}(H)|}\prod_{uv \in E_{\geqslant 2}^a(H)}\left[1+\frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}+\frac{d}{n}\right],$$

where $\tilde{\mathbf{Y}}_{\geqslant 2}^H$ is as in Lemma B.15, and $E_{\geqslant 2}^a(H)$ and $E_{\geqslant 2}^b(H)$ are as in Definition B.6.

Proof. We only provide a proof sketch here. The detailed proof can be found in Appendix B.3.1.3.

For an edge $uv \in E^a_{\geqslant 2}(H)$, we use the fact that $\left|\tilde{\mathbf{Y}}_{uv,E_{\geqslant 2}(H)}\right| \leqslant |\mathbf{Y}_{uv}|$, which allows us to upper bound the contribution of uv by

$$\mathbb{E}\left[|\mathbf{Y}_{uv}|^{m_H(uv)}\middle|\mathbf{x}\right] \leqslant \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}.$$

For edges in $E^b_{\geqslant 2}(H)$, notice the following:

- If an edge $uv \in E^b_{\geqslant 2}(H)$ is present in $\mathbf{G}$, then its contribution to the expectation is at most:
$$\left(1 - \frac{d}{n}\right)^{m_H(uv)}\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} \leqslant \frac{2d}{n}. \tag{B.1.9}$$
- If an edge $uv \in E^b_{\geqslant 2}(H)$ is not present in $\mathbf{G}$, then its contribution to the expectation is at most:
$$\left(-\frac{d}{n}\right)^{m_H(uv)} \cdot \left[1 - \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right] = O\left(\frac{1}{n^2}\right) = \frac{2d}{n} \cdot O\left(\frac{1}{n}\right). \tag{B.1.10}$$

Now notice that every edge in $E^b_{\geqslant 2}(H)$ is incident to some vertex in $\mathcal{L}_{\geqslant 2}(H)$. On the other hand, every vertex $v \in \mathcal{L}_{\geqslant 2}(H)$ is incident to $d^H_{\geqslant 2}(v) > \Delta$ edges in $E^b_{\geqslant 2}(H)$. These edges cannot all be present in $\mathbf{G}$ without causing truncation. In fact, if more than Δ of these edges are present in $\mathbf{G}$, then $\tilde{\mathbf{Y}}^H_{\geqslant 2} = 0$, so we can consider only the cases where at most Δ of these edges are present in $\mathbf{G}$. This ultimately makes it possible to show that the total contribution of the edges in $E^b_{\geqslant 2}(H)$ that are incident to a vertex $v \in \mathcal{L}_{\geqslant 2}(H)$ is at most

$$\tilde{O}\left(\frac{1}{n}\right)^{d^H_{\geqslant 2}(v)-\Delta} \cdot \left(\frac{2d}{n}\right)^{d^H_{\geqslant 2}(v)} = \tilde{O}\left(\frac{1}{n}\right)^{d^H_{\geqslant 2}(v)-\Delta} \cdot \left(\frac{d}{n}\right)^{d^H_{\geqslant 2}(v)}. \tag{B.1.11}$$

If there is no edge in $E^b_{\geqslant 2}(H)$ which has both its end-vertices in $\mathcal{L}_{\geqslant 2}(H)$, then we can multiply the upper bounds Eq. (B.1.11) for every $v \in \mathcal{L}_{\geqslant 2}(H)$, and deduce that the total contribution of edges in $E^b_{\geqslant 2}(H)$ can be upper bounded by

$$\prod_{v \in \mathcal{L}_{\geqslant 2}(H)}\left[\tilde{O}\left(\frac{1}{n}\right)^{d^H_{\geqslant 2}(v)-\Delta} \cdot \left(\frac{d}{n}\right)^{d^H_{\geqslant 2}(v)}\right] = \frac{1}{\prod\limits_{v \in \mathcal{L}_{\geqslant 2}(H)} \tilde{\Omega}\left(n^{d^H_{\geqslant 2}(v)-\Delta}\right)} \cdot \left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H)|}.$$

However, since it is possible for an edge to have both its end-vertices in $\mathcal{L}_{\geqslant 2}(H)$, we cannot just multiply the upper bounds Eq. (B.1.11) for every $v \in \mathcal{L}_{\geqslant 2}(H)$ because some edges would be counted twice. Instead, it is possible to show that the total contribution of edges in $E^b_{\geqslant 2}(H)$ can be upper bounded by

$$\frac{1}{\prod\limits_{v \in \mathcal{L}_{\geqslant 2}(H)} \tilde{\Omega}\left(n^{\frac{1}{2}\left(d^H_{\geqslant 2}(v)-\Delta\right)}\right)} \cdot \left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H)|} \leqslant \frac{1}{\prod\limits_{v \in \mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d^H_{\geqslant 2}(v)-\Delta\right)}} \cdot \left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H)|}.$$

Combining the contribution of edges in $E^a_{\geqslant 2}(H)$ and edges in $E^b_{\geqslant 2}(H)$, we get

$$\mathbb{E}\left[\left|\tilde{\mathbf{Y}}^H_{\geqslant 2}\right|\middle|\mathbf{x}\right] \leqslant \frac{1}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d^H_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{d}{n}\right)^{|\mathbb{E}_{\geqslant 2}(H)|}\prod_{uv\in E^a_{\geqslant 2}(H)}\left[1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}+\frac{d}{n}\right],$$

See Appendix B.3.1.3 for the details. □

B.1.2.5 Proof of the upper bound for every multigraph

Now we are ready to prove the upper bound $\overline{U}_H(\mathbf{x})$ on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H|\mathbf{x}\right]\right|$:

Lemma B.17. *For every multigraph H with at most $st = sK\log n$ vertices and at most st multi-edges, if $\overline{U}_H(\mathbf{x})$ is as in Definition B.7 and n is large enough, then*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H|\mathbf{x}\right]\right| \leqslant \overline{U}_H(\mathbf{x}).$$

Proof. This is a direct corollary of Lemma B.15 and Lemma B.16, and the fact that[9] $\frac{d}{n} \leqslant \frac{3d}{\sqrt{n}}$. □

B.1.3 Bounds for nice multigraphs

In this section, we will prove a lower bound on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]$ for multigraghs H belonging to a nice family of block self-avoid-avoiding walks.

Definition B.18. A block self-avoiding-walk H is said to be *nice* if it satisfies the following:

- H contains at most one cycle, and it should be formed by edges of multiplicity 1 in H. In other words,
  - $E_1(H)$ is either empty or a cycle.[10]
  - The edges of multiplicity $\geqslant 2$ in H form a forest, i.e., $E_{\geqslant 2}(H)$ is a forest.
  - There is no path in $E_{\geqslant 2}(H)$ between any two vertices $u, v \in V(E_1(H))$.
- $\mathcal{L}_{\geqslant 2}(H) = \emptyset$, i.e., $E_{\geqslant 2}(H) = \mathcal{S}_{\geqslant 2}(H) \cup \mathcal{I}_{\geqslant 2}(H)$ and so $d^H_{\geqslant 2}(v) \leqslant \Delta$ for all $v \in V(H)$.
- $|E_1(H)| \geqslant \frac{t}{A}$.

We denote the set of nice (s,t)-block self-avoiding-walks as $\mathrm{NBSAW}^*_{s,t}$.

[9] Note that if we put $\frac{d^2}{n^2}$ instead of $\frac{3d^2}{n\sqrt{n}}$ in $\overline{U}_H(\mathbf{x})$, we still get a valid upper bound on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H|\mathbf{x}\right]\right|$. We used the term $\frac{3d^2}{n\sqrt{n}}$ because it will be convenient when we upper bound $\mathbb{E}\left[\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\intercal\right)^t\right)\right]$.

[10] Hence, if $E_1(H) \neq \emptyset$, then $d^H_1(v) = 2$ for all $v \in V(E_1(H))$.

Remark B.19. The careful reader may notice that the set $\mathrm{NBSAW}^*_{s,t}$ slightly differs from the definition in Lemma 4.57, due to the additional constraint $|E_1(H)| \geqslant \frac{t}{A}$. The set of nice block self-avoiding walks with $|E_1(H)| < \frac{t}{A}$ however is negligible as shown in Lemma 4.70.

In fact, we will provide tight bounds on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]$ for a family of multigraphs that is larger than $\mathrm{NBSAW}^*_{s,t}$. In the following two definitions, we introduce the family of (s,t)-pleasant multigraphs.

Definition B.20. A multigraph H is said to be *agreeable* if satisfies one of the following three conditions:

(1) The underlying graph of $G(H)$ is a cycle. In this case, we say that H is *type-1 agreeable*.

(2) There are two sets of edges $E'(H) \subset E(H)$, $E''(H) \subset E(H)$, and a vertex $u_H \in V(H)$ such that:

- $E(H) = E'(H) \cup E''(H)$.
- $E'(H)$ and $E''(H)$ are cycles.
- $V(E'(H)) \cap V(E''(H)) = \{u_H\}$.

In this case, we say that H is *type-2 agreeable*.

(3) There are two sets of edges $E'(H) \subset E(H)$ and $E''(H) \subset E(H)$ such that:

- $E(H) = E'(H) \cup E''(H)$.
- $E'(H)$ and $E''(H)$ are cycles.
- $E'(H) \cap E''(H) \neq \varnothing$ and $V(E'(H)) \cap V(E''(H)) = V\big(E'(H) \cap E''(H)\big)$.
- $E'(H) \cap E''(H)$ is a simple path.
- $E'(H) \cap E''(H) \subset E_{\geqslant 2}(H)$, i.e., all the edges in $E'(H) \cap E''(H)$ are of multiplicity at least 2 in H.

In this case, we say that H is *type-3 agreeable*.

Definition B.21. A multigraph H is said to be (s,t)-*pleasant* if it satisfies the following conditions:

- H contains at most st vertices and at most st multi-edges.
- $\mathcal{L}_{\geqslant 2}(H) = \varnothing$.
- There are r_H sub-multigraphs $H^{(1)}, \ldots, H^{(r_H)}$ of H such that:
  - For every $i \in [r_H]$, $H^{(i)}$ is an induced sub-multigraph of H, i.e., $H^{(i)} = H\big(V\big(H^{(i)}\big)\big)$.
  - $H^{(1)}, \ldots, H^{(r_H)}$ are vertex-disjoint, i.e., $V\big(H^{(1)}\big), \ldots, V\big(H^{(r_H)}\big)$ are mutually disjoint.

- $H^{(1)}, \dots, H^{(r_H)}$ are agreeable.
- $E_1(H) = \bigcup_{i \in [r_H]} E_1\left(H^{(i)}\right) = E_1\left(H^{(*)}\right)$, where $H^{(*)} = \bigcup_{i \in [r_H]} H^{(i)}$.[11]
- The only cycles in H are those inside $H^{(*)} = \bigcup_{i \in [r_H]} H^{(i)}$.[12]

- Every cycle in H contains at least $\frac{t}{A}$ multiplicity-1 edges.

The sub-multigraphs $H^{(1)}, \dots, H^{(r_H)}$ are said to be the *agreeable components* of H.

It is easy to see that every nice (s,t)-block self-avoiding-walk is (s,t)-pleasant.

B.1.3.1 Informal discussion and proof strategy

In order to be able to show tights bound on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]$ for an (s,t)-pleasant multigraph H, we need to be more precise in our calculations. The techniques that were developed in Appendix B.1.2 will be useful, but we need more ideas in order to get precise calculations that allow for a proof of tight bounds. In the following few paragraphs, we will informally describe how we will prove the tight bounds on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]$ for a pleasant multigraph H.

Roughly speaking, if we write the exact expression of $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]$, we will get a very complicated polynomial $g(\mathbf{x})$ of $\mathbf{x}$. We can decompose the complicated polynomial $g(\mathbf{x})$ into positive terms and negative terms, and we might hope to get a lower bound on $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]$ by showing that the negative terms are negligible. However, it does not seem that we can easily show that the negative monomials are negligible.[13]

Instead of proving a lower bound on $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right] = g(\mathbf{x})$, we will prove a lower bound on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] = \mathbb{E}[g(\mathbf{x})]$. The main reason why we considered lower bounding $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] = \mathbb{E}[g(\mathbf{x})]$ instead of $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right] = g(\mathbf{x})$ is that $\mathbb{E}[\mathbf{x}_u \mathbf{x}_v] = 0$ for every edge uv. This property implies that the expectation of the vast majority of the monomials that appear in $g(\mathbf{x})$ is actually zero. In fact, it is possible to show that for any set E of edges, we have $\mathbb{E}\left[\prod_{uv \in \mathsf{E}} \mathbf{x}_u \mathbf{x}_v\right] \neq 0$ if and only if E is the disjoint union of cycles, in which case we have $\mathbb{E}\left[\prod_{uv \in \mathsf{E}} \mathbf{x}_u \mathbf{x}_v\right] = 1$. So by computing $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] = \mathbb{E}[g(\mathbf{x})]$, we can get rid of the vast majority of the terms in $g(\mathbf{x})$. This is the main reason why analyzing $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] = \mathbb{E}[g(\mathbf{x})]$ is much simpler than analyzing $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right] = g(\mathbf{x})$.

[11] This means that all edges in $H - H^{(*)}$ are of multiplicity at least 2.

[12] This means that if we contract $H^{(1)}, \dots, H^{(r_H)}$ into r_H vertices, H becomes a forest.

[13] In fact, this might not even be possible. It might be the case that the negative terms are not negligible, but the total aggregate of the positive terms is more important than the total aggregate of the negative terms. For example, consider $n = 5n - 4n$: While the negative term $4n$ is not negligible with respect to the positive term $5n$, the positive term is more important. Something like this occurs in $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]$: In order to see this, consider the case of a single edge of multiplicity 1, and suppose that ε is very small.

The following lemma shows that the nice structure of pleasant multigraphs makes the behavior of the expectation of monomials in $(\mathbf{x}_u\mathbf{x}_v)_{uv\in E(H)}$ very simple:

Lemma B.22. *Let H be an arbitrary multigraph and let $\mathsf{E} \subset E(H)$. We have:*

- *If E is an edge-disjoint union of cycles, then* $\mathbb{E}\left[\prod_{uv\in\mathsf{E}} \mathbf{x}_u\mathbf{x}_v\right] = 1.$
- *If E is not an edge-disjoint union of cycles, then* $\mathbb{E}\left[\prod_{uv\in\mathsf{E}} \mathbf{x}_u\mathbf{x}_v\right] = 0.$

Proof. We have:

$$\prod_{uv\in\mathsf{E}} \mathbf{x}_u\mathbf{x}_v = \prod_{v\in V(\mathsf{E})} \mathbf{x}_v^{d_{\mathsf{E}}(v)},$$

where

$$d_{\mathsf{E}}(v) = \left|\left\{u \in V(H) : \ uv \in \mathsf{E}\right\}\right|.$$

Since $(\mathbf{x}_v)_{v\in V(H)}$ are i.i.d. Rademacher random variables, it follows that

$$\mathbb{E}\left[\prod_{uv\in\mathsf{E}} \mathbf{x}_u\mathbf{x}_v\right] = \begin{cases} 1 & \text{if } d_{\mathsf{E}}(v) \text{ is even for all } v \in V(\mathsf{E}), \\ 0 & \text{if there exists at least one vertex } v \in V(\mathsf{E}) \text{ such that } d_{\mathsf{E}}(v) \text{ is odd.} \end{cases}$$

Now notice the following:

- If E is an edge-disjoint union of cycles, then $d_{\mathsf{E}}(v)$ is even for every $v \in V(\mathsf{E})$, and so $\mathbb{E}\left[\prod_{uv\in\mathsf{E}} \mathbf{x}_u\mathbf{x}_v\right] = 1.$
- If E is not an edge-disjoint union of cycles, then there must exist one vertex $v \in V(\mathsf{E})$ such that $d_{\mathsf{E}}(v)$ is odd, and so $\mathbb{E}\left[\prod_{uv\in\mathsf{E}} \mathbf{x}_u\mathbf{x}_v\right] = 0.$

□

As can be seen from Lemma B.22, if H is a pleasant multigraph, the behavior of the expectation of monomials in $(\mathbf{x}_u\mathbf{x}_v)_{uv\in E(H)}$ is very simple: Only the monomials corresponding to $\bigcup_{i\in I} C_i(H)$ for some $I \subseteq [z]$ have nonzero expectation, where $C_1(H), \dots, C_z(H)$ are the cycles of $E_1(H)$. If we could write $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right] = g(\mathbf{x})$ as a polynomial that only depends on $(\mathbf{x}_u\mathbf{x}_v)_{uv\in G(H)}$, the expectation $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] = \mathbb{E}[g(\mathbf{x})]$ would be very simple. Unfortunately, this is not the case because of truncation: The presence or absence of edges outside $E(H)$ in $\mathbf{G}$ makes $g(\mathbf{x})$ also depend on $(\mathbf{x}_u\mathbf{x}_v)_{uv\notin E(H)}$. Therefore, even if we use the fact that for

every set E of edges[14], $\mathbb{E}\left[\prod_{uv\in\mathsf{E}} \mathbf{x}_u\mathbf{x}_v\right] \neq 0$ if and only if E is a disjoint union of cycles, the expression of $\mathbb{E}[g(\mathbf{x})]$ is still too complicated to analyze. In fact, we have two complications:

- H-cross-edges[15] makes it possible to have monomials in $g(\mathbf{x})$ corresponding to cycles in $V(H)$, which are different than $E_1(H)$.
- Edges from $V(H)$ to $[n] \setminus V(H)$ makes it possible to have monomials in $g(\mathbf{x})$ corresponding to cycles that include vertices from $V(H)$ and vertices from $[n] \setminus V(H)$.

The first complication is not too severe because we have very few H-cross-edges, namely $O(s^2t^2) = \tilde{O}(1)$ H-cross-edges, and the probability of any particular edge being present in $\mathbf{G}$ is $O\left(\frac{1}{n}\right)$. Therefore, the event that at least one H-cross-edge is present in $\mathbf{G}$ has negligible probability, and we can assume that no H-cross-edge is present in $\mathbf{G}$. This does not completely solve the first complication, because the probability that an H-cross-edge uv is not present in $\mathbf{G}$ is

$$\mathbb{P}[uv \notin \mathbf{G}] = 1 - \left(1 + \frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n},$$

which contains a term that depends on $\mathbf{x}_u\mathbf{x}_v$. Therefore, even if we focus on the event that no H-cross-edge is present in $\mathbf{G}$, we still have monomials that depend on $(\mathbf{x}_u\mathbf{x}_v)_{uv \text{ is an } H\text{-cross-edge}}$. Fortunately, the term in $\mathbb{P}[uv \notin \mathbf{G}]$ that depends on $\mathbf{x}_u\mathbf{x}_v$ is negligible with respect to the constant term that does not depend on $\mathbf{x}_u\mathbf{x}_v$. We can leverage this observation to show that the total contribution of monomials containing H-cross-edges is negligible.

The second complication is a bit trickier to overcome. We know that the truncation event depends only on $\left(d_{\mathbf{G}}(v)\right)_{v\in V(H)}$, and so depends only on $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$ and $(\mathbf{Y}_{uv})_{uv\in E(H)}$. Only $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$ is problematic for us because it will lead to terms that depend on $(\mathbf{x}_u\mathbf{x}_v)_{uv\notin E(H)}$. In order to analyze the conditional expectation of $\overline{\mathbf{Y}}_H$ given $\mathbf{x}$ and given that no H-cross-edge is present in $\mathbf{G}$, we will further condition over $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$. If no H-cross-edge is present in $\mathbf{G}$, then for every $v \in V(H)$, we have

$$d_{\mathbf{G}-G(H)}(v) = d^o_{\mathbf{G}-G(H)}(v) := \left|\left\{uv \in \mathbf{G} :\ u \notin V(H)\right\}\right|.$$

The nice thing about $\left(d^o_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$ is that, unlike $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$, the random variables $\left(d^o_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$ are conditionally mutually independent given $\mathbf{x}$.[16] Therefore, for every $(\mathsf{d}_v)_{v\in V(H)} \in \mathbb{N}^{V(H)}$, we have

$$\mathbb{P}\left[\left\{\forall v \in V(H), d^o_{\mathbf{G}-G(H)}(v) = \mathsf{d}_v\right\}\middle|\mathbf{x}\right] = \prod_{v\in V(H)} \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) = \mathsf{d}_v\middle|\mathbf{x}\right],$$

[14] Here E may or may not be a subset of $E(H)$.

[15] Recall that an H-cross-edge is an edge uv such that $u, v \in V(H)$ and $uv \notin E(H)$.

[16] Because of H-cross-edges, the random variables $\left(d_{\mathbf{G}-G(H)}(v)\right)_{v\in V(H)}$ are not conditionally mutually independent given $\mathbf{x}$.

and for each $v \in V(H)$, we have

$$\mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) = \mathsf{d}_v \middle| \mathbf{x}\right] = \sum_{\substack{U \subseteq [n]\setminus V(H): \\ |S| = \mathsf{d}_v}} \left[\prod_{u \in U} \mathbb{P}[uv \in \mathbf{G}|\mathbf{x}]\right] \cdot \left[\prod_{u \in [n]\setminus(U \cup V(H))} \mathbb{P}[uv \notin \mathbf{G}|\mathbf{x}]\right]$$

$$= \sum_{\substack{U \subseteq [n]\setminus V(H): \\ |S| = \mathsf{d}_v}} \left[\prod_{u \in U} \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right] \cdot \left[\prod_{u \in [n]\setminus(U \cup V(H))} \left[1 - \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right].$$

The second complication that we mentioned comes from the fact that for every $v \in V(H)$, the probability $\mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) = \mathsf{d}_v \middle| \mathbf{x}\right]$ depends on $(\mathbf{x}_u \mathbf{x}_v)_{u \in [n]\setminus V(H)}$, as can be seen from the above equation. Now here comes the crucial observation that allows us to overcome this complication: If $\mathbf{x}$ is balanced on $[n] \setminus V(H)$ in the sense that $[n] \setminus V(H)$ contains an equal number of vertices from each community, then $\mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) = \mathsf{d}_v \middle| \mathbf{x}\right]$ will not depend on $(\mathbf{x}_u \mathbf{x}_v)_{u \in [n] \notin V(H)}$, and it will only be a function of d_v.

Unfortunately, $[n] \setminus V(H)$ is not likely to be balanced. Nevertheless, from Hoeffding inequality it can be easily seen that $[n] \setminus V(H)$ is very likely to be approximately balanced. More precisely, if $\alpha > 0$ is a fixed (but small) constant, then with high probability we can find a subset $V_b(H, \mathbf{x})$ of $[n] \setminus V(H)$ that is exactly balanced, i.e., it contains an equal number of vertices from each community, and such that

$$\left|[n] \setminus \left(V(H) \cup V_b(H, \mathbf{x})\right)\right| \leqslant n^{\frac{1}{2}+\alpha}.$$

Notice that there are at most $st \cdot n^{\frac{1}{2}+\alpha} = \tilde{O}\left(n^{\frac{1}{2}+\alpha}\right) = o(n)$ edges from $V(H)$ to $[n] \setminus \left(V(H) \cup V_b(H, \mathbf{x})\right)$. On the other hand, the probability that any particular one of these edges is present in $\mathbf{G}$ is at most $O\left(\frac{1}{n}\right)$, hence, with high probability, none of these edges will be present in $\mathbf{G}$. Therefore, we can treat these edges exactly as we treated the H-cross-edges, i.e., we can assume that none of them will be present in $\mathbf{G}$. Furthermore, for any particular edge uv between $u \in [n] \setminus \left(V(H) \cup V_b(H, \mathbf{x})\right)$ and $v \in V(H)$, the term that contains $\mathbf{x}_u \mathbf{x}_v$ in $\mathbb{P}[uv \notin \mathbf{G}]$ is negligible with respect to the constant term that does not depend on $\mathbf{x}_u \mathbf{x}_v$. This means that we can completely discard the edges between $V(H)$ and $[n] \setminus \left(V(H) \cup V_b(H, \mathbf{x})\right)$ exactly as we did with H-cross-edges.

Now since $V_b(H, \mathbf{x})$ is exactly balanced, we can see that the remaining polynomial will contain monomials that depend only on $(\mathbf{x}_u \mathbf{x}_v)_{uv \in E(H)}$. By computing the expectation, many terms will disappear and we will get a simple expression that is very easy to analyze, and this will eventually yield tight bounds on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]$.

In the following, we will turn the above informal discussion into a formal proof.

B.1.3.2 The well-behaved event

In the following, we assume that there is a fixed ordering[17] of the vertices $\{1, \dots, n\}$ of $\mathbf{G}$. This ordering will be used to define some useful concepts. We emphasize that this ordering can be arbitrary, but it should not dependent on the (random) SBM sample $(\mathbf{G}, \mathbf{x})$.

Definition B.23. Let H be a multigraph with at most $st = sK \log n$ vertices. We say that $\mathbf{x}$ is *approximately balanced on* $[n] \setminus V(H)$ if there are at least $\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$ vertices from $[n] \setminus V(H)$ in the first community, and at least $\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$ vertices from $[n] \setminus V(H)$ in the second community, i.e.,

$$\left|\left\{v \in [n] \setminus V(H) : \ \mathbf{x}_v = +1\right\}\right| \geqslant \left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil,$$

and

$$\left|\left\{v \in [n] \setminus V(H) : \ \mathbf{x}_v = -1\right\}\right| \geqslant \left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil.$$

Now assume that $\mathbf{x}$ is approximately balanced on $[n] \setminus V(H)$. Let $V_b(H, \mathbf{x})$ be the set containing the first[18] $\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$ vertices in $[n] \setminus V(H)$ of the first community and the first $\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$ vertices in $[n] \setminus V(H)$ of the second community, i.e.,

$$V_b(H, \mathbf{x}) = \left\{v \in [n] \setminus V(H) : \ \left|\left\{u \in [n] \setminus V(H) : \ u \leqslant v, \ \mathbf{x}_u = \mathbf{x}_v\right\}\right| \leqslant \left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil\right\}. \quad \text{(B.1.12)}$$

If $\mathbf{x}$ is not approximately balanced, we still define $V_b(H, \mathbf{x})$ as in Eq. (B.1.12), but now $V_b(H, \mathbf{x})$ contains at most $2\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil - 1$ vertices, and $V_b(H, \mathbf{x})$ would not necessarily be exactly balanced.

Remark B.24. If $\mathbf{x}$ is approximately balanced on $[n] \setminus V(H)$ then:

- $|V_b(H, \mathbf{x})| = 2\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$.
- $V_b(H, \mathbf{x})$ contains exactly $\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$ vertices of the first community and $\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$ vertices of the second community.
- $\left|\left([n] \setminus V(H)\right) \setminus V_b(H, \mathbf{x})\right| \leqslant \left|[n] \setminus V_b(H, \mathbf{x})\right| \leqslant 2n^{\frac{3}{4}}$.

Definition B.25. Let H be a multigraph with at most $st = sK \log n$ vertices. We say that $(\mathbf{G}, \mathbf{x})$ is *H-well-behaved* if:

(1) $\mathbf{x}$ is approximately balanced on $[n] \setminus V(H)$.

[17] We can use the total order that is induced by the names of the vertices as integers between 1 and n, e.g., $3 \leqslant 5$ and $4 \leqslant 17$.

[18] Here the vertices are chosen according to the fixed ordering of $V(\mathbf{G}) = [n]$.

(2) In the sampled graph $\mathbf{G}$, no vertex in $V(H)$ is adjacent to any vertex in $\big([n] \setminus V(H)\big) \setminus V_b(H, \mathbf{x})$.

(3) There is no H-cross-edge that is present in $\mathbf{G}$.

(4) All edges in $E_{\geqslant 2}(H)$ are present[19] in $\mathbf{G}$.

We denote the event that $(\mathbf{G}, \mathbf{x})$ is H-well-behaved as $\mathcal{E}_{wb,H}$, i.e.,

$$\mathcal{E}_{wb,H} = \big\{(\mathbf{G}, \mathbf{x}) \text{ is } H\text{-well-behaved}\big\}.$$

We will decompose $\mathbb{E}\big[\overline{\mathbf{Y}}_H\big]$ into two parts by conditioning on the event $\mathcal{E}_{wb,H}$:

$$\mathbb{E}\big[\overline{\mathbf{Y}}_H\big] = \mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathcal{E}_{wb,H}\big] \cdot \mathbb{P}[\mathcal{E}_{wb,H}] + \mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathcal{E}^c_{wb,H}\big] \cdot \mathbb{P}[\mathcal{E}^c_{wb,H}].$$

We will prove tight bounds on $\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathcal{E}_{wb,H}\big] \cdot \mathbb{P}[\mathcal{E}_{wb,H}]$ and an upper bound on $\Big|\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathcal{E}^c_{wb,H}\big] \cdot \mathbb{P}[\mathcal{E}^c_{wb,H}]\Big|$. These bounds will imply that the contribution of the not-well-behaved event is negligible with respect to that of the well-behaved event.

B.1.3.3 Upper bound on the contribution of the not-well-behaved event

The following lemma provides an upper bound on the total contribution of the not-well-behaved event in the expectation of $\overline{\mathbf{Y}}_H$.

Lemma B.26. *Let H be a multigraph with at most $st = sK \log n$ vertices and at most st multi-edges. Assume that $\mathcal{L}_1(H) = \mathcal{L}_{\geqslant 2}(H) = \varnothing$, and that $E_{\geqslant 2}(H)$ forms a forest. If $A > \max\{100K, 1\}$ and n is large enough, then we have*

$$\Big|\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathcal{E}^c_{wb,H}\big] \cdot \mathbb{P}[\mathcal{E}^c_{wb,H}]\Big| \leqslant \frac{2}{n^{\frac{1}{6}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

Proof. We only provide a proof sketch here. The detailed proof can be found in Appendix B.3.1.4.

We start by further conditioning on more refined events corresponding to the reasons that caused $(\mathbf{G}, \mathbf{x})$ to be not-well-behaved. Define the following events:

$$\begin{aligned} \mathcal{E}_{H,b} &= \big\{\text{Condition (1) of Definition B.25 is satisfied}\big\} \\ &= \big\{\mathbf{x} \text{ is approximately balanced on } [n] \setminus V(H)\big\}, \end{aligned} \tag{B.1.13}$$

$$\mathcal{E}_{H,g} = \big\{\text{Conditions (2) and (3) of Definition B.25 are satisfied}\big\}, \tag{B.1.14}$$

[19] Note that it is possible to show tight bounds without adding Condition (4) to the definition of the well-behaved event. We only added this condition because it makes the proof of the bounds simpler and easier to describe.

and

$$\begin{aligned}\mathcal{E}_{H,d} &= \left\{\text{Condition (4) of Definition B.25 is satisfied}\right\} \\ &= \left\{\text{All edges in } E_{\geqslant 2}(H) \text{ are present in } \mathbf{G}\right\}.\end{aligned} \tag{B.1.15}$$

Clearly,

$$\mathcal{E}_{wb,H} = \mathcal{E}_{H,b} \cap \mathcal{E}_{H,g} \cap \mathcal{E}_{H,d}.$$

We also define the following:

$$\mathcal{E}_{H,bg} = \mathcal{E}_{H,b} \cap \mathcal{E}_{H,g}, \tag{B.1.16}$$

$$\mathcal{E}_{H,b\overline{g}} = \mathcal{E}_{H,b} \cap \mathcal{E}^c_{H,g}, \tag{B.1.17}$$

and

$$\mathcal{E}_{H,bg\overline{d}} = \mathcal{E}_{H,bg} \cap \mathcal{E}^c_{H,d}. \tag{B.1.18}$$

We will decompose $\mathcal{E}^c_{wb,H}$ into mutually exclusive events as follows:

$$\mathcal{E}^c_{wb,H} = \mathcal{E}^c_{H,b} \cup \mathcal{E}_{H,b\overline{g}} \cup \mathcal{E}_{H,bg\overline{d}}.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{wb,H}\right] \cdot \mathbb{P}[\mathcal{E}^c_{wb,H}] = \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{H,b}\right] \cdot \mathbb{P}[\mathcal{E}^c_{H,b}] + \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\overline{g}}] \\ + \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,bg\overline{d}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,bg\overline{d}}].\end{aligned}$$

We will separately upper bound the absolute value of each term in the right hand side of the above equation.

For $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{H,b}\right] \cdot \mathbb{P}[\mathcal{E}^c_{H,b}]$, we use Hoeffding's inequality to deduce that

$$\mathbb{P}[\mathcal{E}^c_{H,b}] \leqslant 2e^{-\frac{9}{8}\sqrt{n}}.$$

By combining this with Lemma B.17, it is possible to show that for n large enough, we have

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{H,b}\right] \cdot \mathbb{P}[\mathcal{E}^c_{H,b}]\right| \leqslant e^{-\sqrt{n}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}. \tag{B.1.19}$$

For $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\overline{g}}]$, we first write

$$\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\overline{g}}] = \mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathbf{x}|\mathcal{E}_{H,b\overline{g}}]\right].$$

It is possible to show that $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b\overline{g}}\right]\right|$ can be upper bounded using the same techniques that allowed us to upper bound $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$ in Lemma B.17. On the other hand, notice that there are very few H-cross-edges[20], and very few edges between[21] $V(H)$ and $\left([n] \setminus V(H)\right) \setminus$

[20] There are at most $(st)^2 = \tilde{O}(1)$ such edges.

[21] There are at most $st \cdot 2n^{\frac{3}{4}} = \tilde{O}\left(n^{\frac{3}{4}}\right)$ such edges.

$V_b(H, \mathbf{x})$. Now since each particular edge is present in $\mathbf{G}$ with probability $O\left(\frac{1}{n}\right)$, it can be easily seen that $\mathbb{P}[\mathcal{E}_{H,b\bar{g}}] = \tilde{O}\left(n^{-\frac{1}{4}}\right)$. By carefully combining these facts together, we can show that

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b\bar{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\bar{g}}]\right| &\leqslant \tilde{O}\left(\frac{n^{\frac{2K}{A}}}{n^{\frac{1}{4}}}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \\ &\leqslant \frac{1}{n^{\frac{1}{6}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},\end{aligned} \tag{B.1.20}$$

where the last inequality is true for n large enough.

For $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,bg\bar{d}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,bg\bar{d}}]$, notice that $\mathcal{E}_{H,bg\bar{d}}$ implies that at least one edge of multiplicity $\geqslant 2$ is absent from $\mathbf{G}$. On the other hand, Eq. (B.1.9) and Eq. (B.1.10) imply that the contribution of an edge $uv \in E_{\geqslant 2}(H)$ in case it is absent from $\mathbf{G}$ is at least $O\left(\frac{1}{n}\right)$ smaller than its contribution when it is present. By the same techniques that allowed us to upper bound $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$ in Lemma B.17, together with the fact that we have at most $st = \tilde{O}(1)$ edges in $E_{\geqslant 2}(H)$, and the fact that at least one of these edges is absent from $\mathbf{G}$, it is possible to show that

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,bg\bar{d}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,bg\bar{d}}]\right| &\leqslant \tilde{O}\left(\frac{n^{\frac{2K}{A}}}{n}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \\ &\leqslant \frac{1}{n^{\frac{1}{2}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},\end{aligned} \tag{B.1.21}$$

where the last inequality is true for n large enough.

By combining Eq. (B.1.19), Eq. (B.1.20) and Eq. (B.1.21), we get the lemma. See Appendix B.3.1.4 for the details. □

B.1.3.4 Upper bound on the negligible part of the contribution of the well-behaved event

Now in order to study $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{wb,H}\right] \cdot \mathbb{P}[\mathcal{E}_{wb,H}]$, we will divide this expression into the sum of two terms: one that is negligible, and one that is significant. In this section, we prove an upper bound on the negligible part. In the next section, we will prove tight bounds on the significant part.

If we denote the set of H-cross-edges as $E_c(H)$, then we have[22]

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{wb,H}|\mathbf{x}] &= \left[\prod_{e\in E_c(H)}\mathbb{P}[e\notin\mathbf{G}|\mathbf{x}]\right]\cdot\left[\prod_{\substack{v\in V(H),\\ u\in\left([n]\setminus V(H)\right)\setminus V_b(H,\mathbf{x})}}\mathbb{P}[uv\notin\mathbf{G}|\mathbf{x}]\right]\cdot\left[\prod_{e\in E_{\geqslant 2}(H)}\mathbb{P}[e\in\mathbf{G}|\mathbf{x}]\right]\cdot\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x})\\ &= \left[\prod_{e\in E_{\overline{bc}}(H,\mathbf{x})}\mathbb{P}[e\notin\mathbf{G}|\mathbf{x}]\right]\cdot\left[\prod_{e\in E_{\geqslant 2}(H)}\mathbb{P}[e\in\mathbf{G}|\mathbf{x}]\right]\cdot\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x}),\end{aligned}$$

where

$$E_{\overline{bc}}(H,\mathbf{x}) = E_c(H)\cup\left\{uv:\ u\in\left([n]\setminus V(H)\right)\setminus V_b(H,\mathbf{x}), v\in V(H)\right\}. \tag{B.1.22}$$

Note that we could write $\mathbb{1}_{\mathcal{E}_{H,b}} = \mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x})$ as a function of $\mathbf{x}$ because the event $\mathcal{E}_{H,b}$ depends only on $\mathbf{x}$, i.e., it is $\sigma(\mathbf{x})$-measurable. Hence,

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{wb,H}|\mathbf{x}] &= \left[\prod_{uv\in E_{\overline{bc}}(H,\mathbf{x})}\left[1-\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right]\cdot\left[\prod_{uv\in E_{\geqslant 2}(H)}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right]\cdot\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x})\\ &= \left[\sum_{S\subseteq E_{\overline{bc}}(H,\mathbf{x})}\prod_{uv\in S}\left[-\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right]\cdot\left[\prod_{uv\in E_{\geqslant 2}(H)}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right]\cdot\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x})\\ &= P^{wb}_{1,H}(\mathbf{x})+P^{wb}_{2,H}(\mathbf{x}),\end{aligned}$$

where

$$P^{wb}_{1,H}(\mathbf{x}) = \left[\prod_{uv\in E_{\geqslant 2}(H)}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right]\cdot\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x}), \tag{B.1.23}$$

and

$$P^{wb}_{2,H}(\mathbf{x}) = \left[\sum_{\substack{S\subseteq E_{\overline{bc}}(H,\mathbf{x}):\\ S\neq\varnothing}}\prod_{uv\in S}\left[-\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right]\cdot\left[\prod_{uv\in E_{\geqslant 2}(H)}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right]\cdot\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x}). \tag{B.1.24}$$

It is easy to see that $P^{wb}_{1,H}(\mathbf{x})$ contains the monomials in $\mathbb{P}[\mathcal{E}_{wb,H}|\mathbf{x}]$ that do not depend on $(\mathbf{x}_u\mathbf{x}_v)_{uv\in E_{\overline{bc}}(H,\mathbf{x})}$, whereas $P^{wb}_{2,H}(\mathbf{x})$ contains the monomials in $\mathbb{P}[\mathcal{E}_{wb,H}|\mathbf{x}]$ that depend on $(\mathbf{x}_u\mathbf{x}_v)_{uv\in E_{\overline{bc}}(H,\mathbf{x})}$. We will decompose $\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathcal{E}_{wb,H}\right]\cdot\mathbb{P}[\mathcal{E}_{wb,H}]$ as follows:

$$\begin{aligned}\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathcal{E}_{wb,H}\right]\cdot\mathbb{P}[\mathcal{E}_{wb,H}] &= \mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot\mathbb{P}[\mathcal{E}_{wb,H}|\mathbf{x}]\right]\\ &= \mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P^{wb}_{1,H}(\mathbf{x})\right]+\mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P^{wb}_{2,H}(\mathbf{x})\right].\end{aligned} \tag{B.1.25}$$

[22]Recall that $V_b(H,\mathbf{x})$ is always defined, i.e., it is defined even when $\mathbf{x}$ is not approximately balanced on $[n]\setminus V(H)$.

The following lemma provides an upper bound on $\left|\mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P^{wb}_{2,H}(\mathbf{x})\right]\right|$.

Lemma B.27. *Let H be a multigraph with at most $st = sK\log n$ vertices and at most st multi-edges. Assume that $\mathcal{L}_1(H) = \mathcal{L}_{\geqslant 2}(H) = \varnothing$, and that $E_{\geqslant 2}(H)$ forms a forest. If $A > \max\{100K, 1\}$ and n is large enough, then we have*

$$\mathbb{E}\left[\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P^{wb}_{2,H}(\mathbf{x})\right|\right] \leqslant \frac{1}{n^{\frac{1}{6}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},$$

where $P^{wb}_{2,H}(\mathbf{x})$ is defined in Eq. (B.1.24).

Proof. We only provide a proof sketch here. The detailed proof can be found in Appendix B.3.1.5.

Recall that $P^{wb}_{2,H}(\mathbf{x})$ contains only the monomials in $\mathbb{P}[\mathcal{E}_{wb,H}|\mathbf{x}]$ that depend on $(\mathbf{x}_u\mathbf{x}_v)_{uv\in E_{\overline{bc}}(H,\mathbf{x})}$, where $E_{\overline{bc}}(H,\mathbf{x})$ is as in Eq. (B.1.22). On the other hand, for every $uv \in E_{\overline{bc}}(H,\mathbf{x})$, the term in $\mathbb{P}[uv \notin \mathbf{G}|\mathbf{x}]$ that depends on $\mathbf{x}_u\mathbf{x}_v$ is equal to $O\left(\frac{1}{n}\right)$, whereas the term that does not depend on $\mathbf{x}_u\mathbf{x}_v$ is equal to $1 - O\left(\frac{1}{n}\right)$. Now since there are at most $s^2t^2 + 2st\cdot n^{\frac{3}{4}} = \tilde{O}\left(n^{\frac{3}{4}}\right)$ edges in $E_{\overline{bc}}(H,\mathbf{x})$, after carrying out a few careful calculations, it is possible to show that

$$|P^{wb}_{2,H}(\mathbf{x})| \leqslant \tilde{O}\left(\frac{1}{n^{\frac{1}{4}}}\right)\cdot\prod_{uv\in E_{\geqslant 2}(H)}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right].$$

By using the same techniques that allowed us to prove the upper bound in Lemma B.17, we can now show that

$$\begin{aligned}\mathbb{E}\left[\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P^{wb}_{2,H}(\mathbf{x})\right|\right] &\leqslant \tilde{O}\left(\frac{n^{\frac{K}{A}}}{n^{\frac{1}{4}}}\right)\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\\ &\leqslant \frac{1}{n^{\frac{1}{6}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.\end{aligned}$$

See Appendix B.3.1.5 for the details. □

B.1.3.5 Tight bounds on the significant part of the contribution of the well-behaved event

The following lemma provides tight bounds for $\mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P^{wb}_{1,H}(\mathbf{x})\right]$.

Definition B.28. Let H be an (s,t)-pleasant multigraph and let $H^{(1)},\dots,H^{(r_H)}$ be the agreeable components of H. For every $i\in[r_H]$, we define $E'''_{\geqslant 2}(H^{(i)})$ as follows:

- If $H^{(i)}$ is an agreeable component of type 1 or type 2, we define $E'''_{\geqslant 2}(H^{(i)}) = E_{\geqslant 2}\left(H^{(i)}\right)$.

- If $H^{(i)}$ is an agreeable component of type 3, let $E'\big(H^{(i)}\big)$ and $E''\big(H^{(i)}\big)$ be as in Definition B.20, i.e., $E'\big(H^{(i)}\big)$ and $E''\big(H^{(i)}\big)$ are cycles and $E'\big(H^{(i)}\big) \cap E''\big(H^{(i)}\big)$ is a simple path of edges of multiplicity at least 2. We define

$$E'''_{\geqslant 2}\big(H^{(i)}\big) = E_{\geqslant 2}\big(H^{(i)}\big) \setminus \Big(E'\big(H^{(i)}\big) \cap E''\big(H^{(i)}\big)\Big).$$

We also define $E'''_{\geqslant 2}(H) = \bigcup_{i \in [r_H]} E'''_{\geqslant 2}\big(H^{(i)}\big)$.

Lemma B.29. *Let H be an (s,t)-pleasant multigraph, where $t = K \log n$. If $A > \max\{1, 100K, \frac{100}{K}\}$ and n is large enough, then*

$$\begin{aligned}\left(P_s^H - \frac{2}{\sqrt{n}}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \\ \leqslant \mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathbf{x}, \mathcal{E}_{wb,H}\big] \cdot P^{wb}_{1,H}(\mathbf{x})\Big] \\ \leqslant \left(P_s^H + \frac{2}{\sqrt{n}}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|},\end{aligned}$$

where

$$P_s^H = \mathbb{P}\big[\big\{\forall v \in V(H),\ \mathbf{D}_v^H + d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta\big\}\big|\mathcal{E}_{H,b}\big],$$

and $\mathbf{D}_v^H = |\{u \in V_b(H,\mathbf{x}) :\ uv \in \mathbf{G}\}|$ is the number of edges from v to $V_b(H,\mathbf{x})$ which are present in $\mathbf{G}$.

Proof. Roughly speaking, since $\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x}, \mathcal{E}_{wb,H}\big] \cdot P^{wb}_{1,H}(\mathbf{x})$ contains monomials that depend only on $(\mathbf{x}_u\mathbf{x}_v)_{uv \in E(H)}$, Lemma B.22 implies that by taking the expectation, we can get rid of most of the terms, and obtain a simple expression that can be easily bounded. The detailed proof can be found in Appendix B.3.1.6. □

B.1.3.6 Tight bounds for pleasant multigraphs

The following two lemmas study the probability of safety P_s^H:

Lemma B.30. *There exists a non-decreasing function $P_s : \mathbb{Z} \to [0,1]$ such that:*

- *$P_s(\ell) = 0$ for every $\ell < 0$.*
- *$P_s(\ell) \geqslant e^{-4d}$ for every $\ell \geqslant 0$.*
- *$P_s(\ell) \geqslant 1 - \frac{\eta}{2}$ for every $\ell \geqslant \frac{\Delta}{4}$, where η is as in Lemma B.9.*
- *If n is large enough, then for every (s,t)-pleasant multigraph H with $t = K \log n$, we have*

$$P_s^H = \prod_{v \in V(H)} P_s\big(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\big),$$

where P_s^H is as in Lemma B.29.

Proof. The proof can be found in Appendix B.3.1.7. □

Lemma B.31. *Let H be an (s,t)-pleasant multigraph, where $t = K\log n$. Let P_s^H be as in Lemma B.29. Assume that $A > 1$. We have the following:*

- *If $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$, then for n large enough, we have $P_s^H \geqslant \frac{1}{n^{\frac{2K}{A}}}$.*
- *If there exists $v \in V(H)$ such that $d_1^H(v) + d_{\geqslant 2}^H(v) > \Delta$, then $P_s^H = 0$.*

Proof. The proof can be found in Appendix B.3.1.7. □

Now we are ready to prove tight lower and upper bounds on $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]$ for every pleasant multigraph.

Lemma B.32. *Let H be an (s,t)-pleasant multigraph, where $t = K\log n$. If $A > \max\left\{1, 100K, \frac{100}{K}\right\}$ and n is large enough, then*

- *If $|E_{\geqslant 2}'''(H)| \leqslant \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ and $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$, then for n large enough, we have*

$$P_s^H\left(1 - \frac{1}{n^{\frac{1}{16}}}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E_{\geqslant 2}'''(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E_{\geqslant 2}'''(H)|} \leqslant \mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant P_s^H\left(1 + \frac{1}{n^{\frac{1}{16}}}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E_{\geqslant 2}'''(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E_{\geqslant 2}'''(H)|}. \tag{B.1.26}$$

- *If $|E_{\geqslant 2}'''(H)| > \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ or there exists $v \in V(H)$ such that $d_1^H(v) + d_{\geqslant 2}^H(v) > \Delta$, then for n large enough, we have*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]\right| \leqslant \frac{4}{n^{\frac{1}{12}}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}, \tag{B.1.27}$$

where P_s^H is as in Lemma B.29.

Proof. We have:

$$\begin{aligned}\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] &= \mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathcal{E}_{wb,H}\right]\cdot\mathbb{P}[\mathcal{E}_{wb,H}] + \mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathcal{E}_{wb,H}^c\right]\cdot\mathbb{P}[\mathcal{E}_{wb,H}^c] \\ &= \mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P_{1,H}^{wb}(\mathbf{x})\right] + \mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P_{2,H}^{wb}(\mathbf{x})\right] \\ &\qquad + \mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathcal{E}_{wb,H}^c\right]\cdot\mathbb{P}[\mathcal{E}_{wb,H}^c],\end{aligned}$$

where the last equality follows from Eq. (B.1.25). By combining this with Lemma B.26, Lemma B.27 and Lemma B.29, we get

$$\begin{aligned}&\left(P_s^H - \frac{2}{\sqrt{n}} - \frac{3}{n^{\frac{1}{6}}}\cdot\left(\frac{2}{\varepsilon}\right)^{|E_{\geqslant 2}'''(H)|}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E_{\geqslant 2}'''(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E_{\geqslant 2}'''(H)|} \\ &\qquad \leqslant \mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant \left(P_s^H + \frac{2}{\sqrt{n}} + \frac{3}{n^{\frac{1}{6}}}\cdot\left(\frac{2}{\varepsilon}\right)^{|E_{\geqslant 2}'''(H)|}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E_{\geqslant 2}'''(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E_{\geqslant 2}'''(H)|},\end{aligned} \tag{B.1.28}$$

We distinguish between two cases:

(1) If $|E'''_{\geqslant 2}(H)| \leqslant \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$, we have

$$\frac{3}{n^{\frac{1}{6}}} \cdot \left(\frac{2}{\varepsilon}\right)^{|E'''_{\geqslant 2}(H)|} \leqslant \frac{3}{n^{\frac{1}{6}}} \cdot n^{\frac{1}{12}} = \frac{3}{n^{\frac{1}{12}}}.$$

In this case, we get

$$\begin{aligned}\left(P_s^H - \frac{4}{n^{\frac{1}{12}}}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \\ \leqslant \mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant \left(P_s^H + \frac{4}{n^{\frac{1}{12}}}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|},\end{aligned} \tag{B.1.29}$$

We will further split case (1) into two subcases:

(i) If $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$, then Lemma B.31 implies that for n large enough, we have $P_s^H \geqslant \frac{1}{n^{\frac{2K}{A}}}$. Since $A > 100K$, we have

$$\frac{4}{n^{\frac{1}{12}}P_s^H} \leqslant \frac{4n^{\frac{2K}{A}}}{n^{\frac{1}{12}}} \leqslant \frac{4n^{\frac{1}{50}}}{n^{\frac{1}{12}}} \leqslant \frac{1}{n^{\frac{1}{16}}},$$

where the last inequality is true for n large enough. By combining this with Eq. (B.1.29), we get Eq. (B.1.26).

(ii) If there exists $v \in V(H)$ such that $d_1^H(v) + d_{\geqslant 2}^H(v) > \Delta$, then Lemma B.31 implies that $P_s^H = 0$. By combining this with Eq. (B.1.29) and using the fact that $\frac{\varepsilon}{2} \leqslant 1$, we get Eq. (B.1.27).

(2) If $|E'''_{\geqslant 2}(H)| > \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$, we have

$$\left(\frac{\varepsilon}{2}\right)^{|E'''_{\geqslant 2}(H)|} = \frac{1}{\left(\frac{2}{\varepsilon}\right)^{|E'''_{\geqslant 2}(H)|}} \leqslant \frac{1}{n^{\frac{1}{12}}},$$

hence, Eq. (B.1.28) implies that

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]\right| &\leqslant \left(\left(P_s^H + \frac{2}{\sqrt{n}}\right)\left(\frac{\varepsilon}{2}\right)^{|E'''_{\geqslant 2}(H)|} + \frac{3}{n^{\frac{1}{6}}}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \\ &\leqslant \left(\left(P_s^H + \frac{2}{\sqrt{n}}\right)\frac{1}{n^{\frac{1}{12}}} + \frac{3}{n^{\frac{1}{6}}}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \\ &\leqslant \frac{4}{n^{\frac{1}{12}}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.\end{aligned}$$

□

The following lemma provides an upper bound on the expectation $\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]$ for a significant (s,t)-pleasant multigraph H in terms of the expectation $\mathbb{E}\left[\overline{\mathbf{Y}}_C\right]$ for a cycle C with st edges of multiplicity 1.

Lemma B.33. *Assume that $A > \max\{1, 100K, \frac{100}{K}\}$, and let $H \in NBSAW^*_{s,t}$.*

- *If $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$, then for n large enough, we have*

$$\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \geqslant \frac{1}{2n^{\frac{2K}{A}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

- *If there exists $v \in V(H)$ such that $d_1^H(v) + d_{\geqslant 2}^H(v) > \Delta$, then for n large enough, we have*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]\right| \leqslant \frac{4}{n^{\frac{1}{12}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

Proof. This is a direct corollary of Lemma B.32, Lemma B.31, and the fact that every $H \in \mathrm{NBSAW}^*_{s,t}$ is an (s,t)-pleasant multigraph that satisfies $E'''_{\geqslant 2}(H) = \varnothing$. □

B.1.4 Bounds for products of block self-avoiding-walks

Lemma B.34. *Let $H = H_{(1)} \oplus H_{(2)}$, where $H_{(1)}$ and $H_{(2)}$ are two (s,t)-pleasant multigraphs such that $V(H_{(1)}) \cap V(H_{(2)}) = \varnothing$, and $t = K \log n$. Assume that $|E'''_{\geqslant 2}(H)| \leqslant \frac{\log n}{12 \log\left(\frac{2}{\varepsilon}\right)}$ and $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$. If $A > \max\{1, 200K, \frac{100}{K}\}$, then for n large enough, we have*

$$\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant \left(1 + \frac{4}{n^{\frac{1}{16}}}\right) \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right].$$

Proof. From Lemma B.30, we have

$$\begin{aligned}
P_s^H &= \prod_{v \in V(H)} P_s\left(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\right) \\
&= \left(\prod_{v \in V(H_{(1)})} P_s\left(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\right)\right) \cdot \left(\prod_{v \in V(H_{(2)})} P_s\left(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\right)\right) \\
&= \left(\prod_{v \in V(H_{(1)})} P_s\left(\Delta - d_1^{H_{(1)}}(v) - d_{\geqslant 2}^{H_{(1)}}(v)\right)\right) \cdot \left(\prod_{v \in V(H_{(2)})} P_s\left(\Delta - d_1^{H_{(2)}}(v) - d_{\geqslant 2}^{H_{(2)}}(v)\right)\right) \\
&= P_s^{H_{(1)}} \cdot P_s^{H_{(2)}}.
\end{aligned} \tag{B.1.30}$$

Since $H_{(1)}$ and $H_{(2)}$ are (s,t)-pleasant and $V(H_{(1)}) \cap V(H_{(2)}) = \varnothing$, the multigraph $H = H_{(1)} \oplus H_{(2)}$ is $(s,2t)$-pleasant, i.e., H is $(s, 2K \log n)$-pleasant. Now since $V(H_1) \cap V(H_2) = \varnothing$, $|E'''_{\geqslant 2}(H)| \leqslant \frac{\log n}{12 \log\left(\frac{2}{\varepsilon}\right)}$ and $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$, we have:

- For every $v \in V(H_{(1)})$, we have $d_1^{H_{(1)}}(v) + d_{\geqslant 2}^{H_{(1)}}(v) = d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$.
- For every $v \in V(H_{(2)})$, we have $d_1^{H_{(2)}}(v) + d_{\geqslant 2}^{H_{(2)}}(v) = d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$.
- Since $|E_{\geqslant 2}'''(H)| = |E_{\geqslant 2}'''(H_{(1)})| + |E_{\geqslant 2}'''(H_{(2)})|$, we have $|E_{\geqslant 2}'''(H_{(1)})| \leqslant \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ and $|E_{\geqslant 2}'''(H_{(2)})| \leqslant \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$.

Now if we apply Lemma B.32 to $H, H_{(1)}$ and $H_{(2)}$ and use the fact that $|E_1(H)| = |E_1(H_{(1)})| + |E_1(H_{(2)})|$, $|E_{\geqslant 2}(H)| = |E_{\geqslant 2}(H_{(1)})| + |E_{\geqslant 2}(H_{(2)})|$ and $|E_{\geqslant 2}'''(H)| = |E_{\geqslant 2}'''(H_{(1)})| + |E_{\geqslant 2}'''(H_{(2)})|$, we get

$$\frac{\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]}{\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right]\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right]} \leqslant \frac{P_s^H\left(1+\frac{1}{n^{\frac{1}{16}}}\right)}{P_s^{H_{(1)}}\left(1-\frac{1}{n^{\frac{1}{16}}}\right)P_s^{H_{(2)}}\left(1-\frac{1}{n^{\frac{1}{16}}}\right)} \overset{(*)}{=} \frac{\left(1+\frac{1}{n^{\frac{1}{16}}}\right)}{\left(1-\frac{1}{n^{\frac{1}{16}}}\right)\left(1-\frac{1}{n^{\frac{1}{16}}}\right)} \overset{(\dagger)}{\leqslant} \left(1+\frac{4}{n^{\frac{1}{16}}}\right),$$

where $(*)$ follows from Equation (B.1.30) and $(\dagger)$ is true for n is large enough. □

Lemma B.35. *Let $H = H_{(1)} \oplus H_{(2)}$, where $H_{(1)}, H_{(2)} \in NBSAW_{s,t}^*$ are such that $V(H_{(1)}) \cap V(H_{(2)}) = \varnothing$, and $t = K\log n$. If $A > \max\left\{1, 200K, \frac{100}{K}\right\}$, then*

- *If $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$, then for n large enough, we have*

$$\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant \left(1+\frac{4}{n^{\frac{1}{16}}}\right)\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right]\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right].$$

- *If there exists $v \in V(H)$ such that $d_1^H(v) + d_{\geqslant 2}^H(v) > \Delta$, then for n large enough, we have*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]\right| \leqslant \frac{4}{n^{\frac{1}{12}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

Proof. Since $H_{(1)}, H_{(2)} \in \mathrm{NBSAW}_{s,t}^*$ and $V(H_{(1)}) \cap V(H_{(2)}) = \varnothing$, we have $|E_{\geqslant 2}'''(H)| = \varnothing$. The lemma now follows immediately from Lemma B.32 and Lemma B.34. □

Now we will study products of nice block self-avoiding-walks sharing a vertex.

Lemma B.36. *Let $H = H_{(1)} \oplus H_{(2)}$, where $H_{(1)}, H_{(2)} \in NBSAW_{s,t}^*$ are such that $V(H_{(1)}) \cap V(H_{(2)}) \neq \varnothing$, H is an $(s, 2t)$-pleasant multigraph, and $t = K\log n$. If $A > \max\left\{1, 200K, \frac{100}{K}\right\}$, then*

- *If $|E_{\geqslant 2}'''(H)| \leqslant \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ and $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$, then for n large enough, we have*

$$\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant \left(1+\frac{\eta}{2}\right)\cdot\left(\frac{4n\cdot e^{16d/\Delta}}{\varepsilon^2 d(1-\eta)}\right)^{|E(H_{(1)})\cap E(H_{(2)})|}\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right]\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right],$$

where η is as in Lemma B.9.

- *If* $|E'''_{\geqslant 2}(H)| > \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ *or there exists* $v \in V(H)$ *such that* $d_1^H(v) + d_{\geqslant 2}^H(v) > \Delta$*, then for* n *large enough, we have*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]\right| \leqslant \frac{4}{n^{\frac{1}{12}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

Proof. If $|E'''_{\geqslant 2}(H)| > \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ or there exists $v \in V(E_1(H))$ such that $d_1^H(v) + d_{\geqslant 2}^H(v) > \Delta$, then Lemma B.32 implies that

$$\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant \frac{4}{n^{\frac{1}{12}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

Now assume that $|E'''_{\geqslant 2}(H)| \leqslant \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ and $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$. If we apply Lemma B.32 to $H, H_{(1)}$ and $H_{(2)}$, and using the fact that $E'''_{\geqslant 2}(H_{(1)}) = E'''_{\geqslant 2}(H_{(2)}) = \varnothing$, we get

$$\frac{\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]}{\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right]} \leqslant \frac{P_s^H\left(1 + \frac{1}{n^{\frac{1}{16}}}\right)}{P_s^{H_{(1)}}\left(1 - \frac{1}{n^{\frac{1}{16}}}\right) P_s^{H_{(2)}}\left(1 - \frac{1}{n^{\frac{1}{16}}}\right)} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)| + |E'''_{\geqslant 2}(H)| - |E_1(H_{(1)})| - |E_1(H_{(2)})|} \times \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)| - |E'''_{\geqslant 2}(H)| - |E_{\geqslant 2}(H_{(1)})| - |E_{\geqslant 2}(H_{(2)})|}.$$

Now since $\varepsilon \leqslant 2$, if n is large enough, we get

$$\frac{\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]}{\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right]} \leqslant \left(1 + \frac{4}{n^{\frac{1}{16}}}\right) \cdot \frac{P_s^H}{P_s^{H_{(1)}} P_s^{H_{(2)}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)| - |E_1(H_{(1)})| - |E_1(H_{(2)})|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)| - |E_{\geqslant 2}(H_{(1)})| - |E_{\geqslant 2}(H_{(2)})|}. \tag{B.1.31}$$

Now since $H = H_{(1)} \oplus H_{(2)}$, we have

$$E_1(H) \subseteq E_1(H_{(1)}) \cup E_1(H_{(2)}) \quad \text{and} \quad E_{\geqslant 2}(H_{(1)}) \cup E_{\geqslant 2}(H_{(2)}) \subseteq E_{\geqslant 2}(H).$$

Furthermore,

$$\begin{aligned}&\left(E_1(H_{(1)}) \cup E_1(H_{(2)})\right) \setminus E_1(H)\\ &\qquad = \left(E_1(H_{(1)}) \cap E_1(H_{(2)})\right) \cup \left(E_1(H_{(1)}) \cap E_{\geqslant 2}(H_{(2)})\right) \cup \left(E_{\geqslant 2}(H_{(1)}) \cap E_1(H_{(2)})\right),\end{aligned}$$

which implies that,

$$\begin{aligned}&|E_1(H_{(1)})| + |E_1(H_{(2)})| - |E_1(H)|\\ &\qquad = |E_1(H_{(1)}) \cap E_1(H_{(2)})| + |E_1(H_{(1)}) \cap E_1(H_{(2)})| - |E_1(H)| \\ &\qquad = 2\left|E_1(H_{(1)}) \cap E_1(H_{(2)})\right| + \left|E_1(H_{(1)}) \cap E_{\geqslant 2}(H_{(2)})\right| + \left|E_{\geqslant 2}(H_{(1)}) \cap E_1(H_{(2)})\right|.\end{aligned} \tag{B.1.32}$$

On the other hand,

$$E_{\geqslant 2}(H) \setminus \big(E_{\geqslant 2}(H_{(1)}) \cup E_{\geqslant 2}(H_{(2)})\big) \;=\; \big(E_1(H_{(1)}) \cap E_1(H_{(2)})\big),$$

which implies that

$$\begin{aligned}
&|E_{\geqslant 2}(H)| - |E_{\geqslant 2}(H_{(1)})| - |E_{\geqslant 2}(H_{(2)})| \\
&\qquad = |E_{\geqslant 2}(H)| - |E_{\geqslant 2}(H_{(1)}) \cup E_{\geqslant 2}(H_{(2)})| - |E_{\geqslant 2}(H_{(1)}) \cap E_{\geqslant 2}(H_{(2)})| \\
&\qquad = |E_1(H_{(1)}) \cap E_1(H_{(2)})| - |E_{\geqslant 2}(H_{(1)}) \cap E_{\geqslant 2}(H_{(2)})|
\end{aligned} \tag{B.1.33}$$

By combining Eq. (B.1.32) and Eq. (B.1.33), we get

$$\begin{aligned}
&\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|-|E_1(H_{(1)})|-|E_1(H_{(2)})|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E_{\geqslant 2}(H_{(1)})|-|E_{\geqslant 2}(H_{(2)})|} \\
&\leqslant \left(\frac{\varepsilon d}{2n}\right)^{-2\left|E_1(H_{(1)})\cap E_1(H_{(2)})\right|-\left|E_1(H_{(1)})\cap E_{\geqslant 2}(H_{(2)})\right|-\left|E_{\geqslant 2}(H_{(1)})\cap E_1(H_{(2)})\right|} \cdot \left(\frac{d}{n}\right)^{|E_1(H_{(1)})\cap E_1(H_{(2)})|-|E_{\geqslant 2}(H_{(1)})\cap E_{\geqslant 2}(H_{(2)})|} \\
&= \left(\frac{\varepsilon^2 d}{4n}\right)^{-\left|E_1(H_{(1)})\cap E_1(H_{(2)})\right|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{-\left|E_1(H_{(1)})\cap E_{\geqslant 2}(H_{(2)})\right|-\left|E_{\geqslant 2}(H_{(1)})\cap E_1(H_{(2)})\right|} \cdot \left(\frac{d}{n}\right)^{-|E_{\geqslant 2}(H_{(1)})\cap E_{\geqslant 2}(H_{(2)})|} \\
&\overset{(*)}{\leqslant} \left(\frac{\varepsilon^2 d}{4n}\right)^{-\left|E_1(H_{(1)})\cap E_1(H_{(2)})\right|-\left|E_1(H_{(1)})\cap E_{\geqslant 2}(H_{(2)})\right|-\left|E_{\geqslant 2}(H_{(1)})\cap E_1(H_{(2)})\right|-|E_{\geqslant 2}(H_{(1)})\cap E_{\geqslant 2}(H_{(2)})|} \\
&= \left(\frac{\varepsilon^2 d}{4n}\right)^{-\left|E(H_{(1)})\cap E(H_{(2)})\right|}.
\end{aligned}$$

where the last inequality follows from the fact that $\varepsilon \leqslant 2$. By combining this with Eq. (B.1.31), we get

$$\frac{\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]}{\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right] \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right]} \leqslant \left(1 + \frac{4}{n^{\frac{1}{16}}}\right) \cdot \frac{P_s^H}{P_s^{H_{(1)}} P_s^{H_{(2)}}} \cdot \left(\frac{4n}{\varepsilon^2 d}\right)^{\left|E(H_{(1)})\cap E(H_{(2)})\right|}. \tag{B.1.34}$$

Now from Lemma B.30, we have

$$\begin{aligned}
P_s^H &= \prod_{v \in V(H)} P_s\left(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\right) \\
&= \left(\prod_{v \in V(H_{(1)}) \setminus V(H_{(2)})} P_s\left(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\right)\right) \cdot \left(\prod_{v \in V(H_{(2)}) \setminus V(H_{(1)})} P_s\left(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\right)\right) \\
&\qquad \times \left(\prod_{v \in V(H_{(1)}) \cap V(H_{(2)})} P_s\left(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\right)\right).
\end{aligned}$$

Now for every $v \in V(H)$, we have

$$d_1^H(v) + d_{\geqslant 2}^H(v) = d_{G(H)}(v) \geqslant d_{G(H_{(1)})}(v) = d_1^{H_{(1)}}(v) + d_{\geqslant 2}^{H_{(1)}}(v).$$

Similarly, we gave

$$d_1^H(v) + d_{\geqslant 2}^H(v) \geqslant d_1^{H_{(2)}}(v) + d_{\geqslant 2}^{H_{(2)}}(v).$$

From Lemma B.30, we know that the function P_s is non-decreasing. Therefore,

$$\begin{aligned}
P_s^H &\leqslant \left(\prod_{v\in V(H_{(1)})\setminus V(H_{(2)})} P_s\left(\Delta - d_1^{H_{(1)}}(v) - d_{\geqslant 2}^{H_{(1)}}\right)\right)\cdot\left(\prod_{v\in V(H_{(2)})\setminus V(H_{(1)})} P_s\left(\Delta - d_1^{H_{(2)}}(v) - d_{\geqslant 2}^{H_{(2)}}(v)\right)\right) \\
&\qquad\times\left(\prod_{v\in V(H_{(1)})\cap V(H_{(2)})} \min\left\{P_s\left(\Delta - d_1^{H_{(1)}}(v) - d_{\geqslant 2}^{H_{(1)}}\right), P_s\left(\Delta - d_1^{H_{(2)}}(v) - d_{\geqslant 2}^{H_{(2)}}(v)\right)\right\}\right) \\
&= \frac{\left(\prod_{v\in V(H_{(1)})} P_s\left(\Delta - d_1^{H_{(1)}}(v) - d_{\geqslant 2}^{H_{(1)}}\right)\right)\cdot\left(\prod_{v\in V(H_{(2)})} P_s\left(\Delta - d_1^{H_{(2)}}(v) - d_{\geqslant 2}^{H_{(2)}}(v)\right)\right)}{\prod_{v\in V(H_{(1)})\cap V(H_{(2)})} \max\left\{P_s\left(\Delta - d_1^{H_{(1)}}(v) - d_{\geqslant 2}^{H_{(1)}}\right), P_s\left(\Delta - d_1^{H_{(2)}}(v) - d_{\geqslant 2}^{H_{(2)}}(v)\right)\right\}} \\
&= \frac{P_s^{H_{(1)}} P_s^{H_{(2)}}}{\prod_{v\in V(H_{(1)})\cap V(H_{(2)})} \max\left\{P_s\left(\Delta - d_1^{H_{(1)}}(v) - d_{\geqslant 2}^{H_{(1)}}\right), P_s\left(\Delta - d_1^{H_{(2)}}(v) - d_{\geqslant 2}^{H_{(2)}}(v)\right)\right\}},
\end{aligned}$$

where the last equality follows from applying Lemma B.30 to $H_{(1)}$ and $H_{(2)}$. Therefore,

$$\frac{P_s^H}{P_s^{H_{(1)}} P_s^{H_{(2)}}} \leqslant \prod_{v\in V(H_{(1)})\cap V(H_{(2)})} F_v, \tag{B.1.35}$$

where

$$F_v = \min\left\{\frac{1}{P_s\left(\Delta - d_1^{H_{(1)}}(v) - d_{\geqslant 2}^{H_{(1)}}\right)}, \frac{1}{P_s\left(\Delta - d_1^{H_{(2)}}(v) - d_{\geqslant 2}^{H_{(2)}}(v)\right)}\right\}.$$

Now for every vertex $v \in V(H_{(1)}) \cap V(H_{(2)})$, we have:

- If $d_1^{H_{(1)}}(v) + d_{\geqslant 2}^{H_{(1)}}) \leqslant \frac{3\Delta}{4}$ or $d_1^{H_{(2)}}(v) + d_{\geqslant 2}^{H_{(2)}}) \leqslant \frac{3\Delta}{4}$, then Lemma B.30 implies that

$$F_v \leqslant \frac{1}{1-\frac{\eta}{2}}. \tag{B.1.36}$$

- If $d_1^{H_{(1)}}(v) + d_{\geqslant 2}^{H_{(1)}}) > \frac{3\Delta}{4}$ and $d_1^{H_{(2)}}(v) + d_{\geqslant 2}^{H_{(2)}}) > \frac{3\Delta}{4}$, then Lemma B.30 implies that

$$F_v \leqslant e^{4d}. \tag{B.1.37}$$

Now for every set E of edges, let $d_{\mathsf{E}}(v)$ be the number of edges in E which are incident to v. We have:

$$d_{E(H_{(1)})\cup E(H_{(2)})}(v) = d_{G(H)}(v) = d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta,$$

$$d_{E(H_{(1)})}(v) = d_{G(H_{(1)})}(v) = d_1^{H_{(1)}}(v) + d_{\geqslant 2}^{H_{(1)}}(v) > \frac{3\Delta}{4},$$

and

$$d_{E(H_{(2)})}(v) = d_{G(H_{(2)})}(v) = d_1^{H_{(2)}}(v) + d_{\geqslant 2}^{H_{(2)}}(v) > \frac{3\Delta}{4}.$$

Therefore, if $d_1^{H_{(1)}}(v) + d_{\geqslant 2}^{H_{(1)}}) > \frac{3\Delta}{4}$ and $d_1^{H_{(2)}}(v) + d_{\geqslant 2}^{H_{(2)}}) > \frac{3\Delta}{4}$, we have

$$d_{E(H_{(1)})\cap E(H_{(2)})}(v) = d_{E(H_{(1)})}(v) + d_{E(H_{(2)})}(v) - d_{E(H_{(1)})\cup E(H_{(2)})}(v) \geqslant \frac{\Delta}{2}. \tag{B.1.38}$$

Now define the sets

$$V_{H_{(1)},H_{(2)}}^{d=0} = \left\{v \in V(H_{(1)}) \cap V(H_{(2)}) : \ d_{E(H_{(1)})\cap E(H_{(2)})}(v) = 0\right\}$$

and

$$V_{H_{(1)},H_{(2)}}^{d>0} = \left\{v \in V(H_{(1)}) \cap V(H_{(2)}) : \ d_{E(H_{(1)})\cap E(H_{(2)})}(v) > 0\right\}.$$

Clearly, $\left\{V_{H_{(1)},H_{(2)}}^{d=0}, V_{H_{(1)},H_{(2)}}^{d>0}\right\}$ is a partition of $V(H_{(1)})\cap V(H_{(2)})$. Therefore, from Eq. (B.1.35) and Eq. (B.1.36), we get

$$\begin{aligned}\frac{P_s^H}{P_s^{H_{(1)}}P_s^{H_{(2)}}} &\leqslant \left(\prod_{v\in V_{H_{(1)},H_{(2)}}^{d=0}} F_v\right)\cdot\left(\prod_{v\in V_{H_{(1)},H_{(2)}}^{d>0}} F_v\right) \leqslant \frac{1}{\left(1-\frac{\eta}{2}\right)^{\left|V_{H_{(1)},H_{(2)}}^{d=0}\right|}}\cdot\prod_{v\in V_{H_{(1)},H_{(2)}}^{d>0}} F_v \\ &= \frac{1}{\left(1-\frac{\eta}{2}\right)^{\left|V_{H_{(1)},H_{(2)}}^{d=0}\right|}}\cdot\prod_{v\in V_{H_{(1)},H_{(2)}}^{d>0}} \tilde{F}^{d_{E(H_{(1)})\cap E(H_{(2)})}(v)},\end{aligned} \tag{B.1.39}$$

where

$$\tilde{F} = \max_{v\in V_{H_{(1)},H_{(2)}}^{d>0}} F_v^{1/d_{E(H_{(1)})\cap E(H_{(2)})}(v)}.$$

Now from Eq. (B.1.36), Eq. (B.1.37) and Eq. (B.1.38), we get

$$\tilde{F} \leqslant \max\left\{\frac{1}{1-\frac{\eta}{2}}, e^{\frac{4d}{\Delta/2}}\right\} \leqslant \frac{e^{8d/\Delta}}{1-\frac{\eta}{2}}.$$

On the other hand, since

$$V_{H_{(1)},H_{(2)}}^{d>0} = \bigcup_{uv\in E(H_{(1)})\cap E(H_{(2)})} \{u,v\},$$

it is easy to see that

$$\prod_{v\in V^{d>0}_{H_{(1)},H_{(2)}}} \tilde{F}^{d_{E(H_{(1)})\cap E(H_{(2)})}(v)} = \tilde{F}^{2|E(H_{(1)})\cap E(H_{(2)})|} \leqslant \left(\frac{e^{8d/\Delta}}{1-\frac{\eta}{2}}\right)^{2|E(H_{(1)})\cap E(H_{(2)})|} \leqslant \left(\frac{e^{16d/\Delta}}{1-\eta}\right)^{|E(H_{(1)})\cap E(H_{(2)})|}.$$

By combining this with Eq. (B.1.34) and Eq. (B.1.39), we get

$$\frac{\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]}{\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right]\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right]} \leqslant \frac{\left(1+\frac{4}{n^{\frac{1}{16}}}\right)}{\left(1-\frac{\eta}{2}\right)^{\left|V^{d=0}_{H_{(1)},H_{(2)}}\right|}}\cdot\left(\frac{4n\cdot e^{16d/\Delta}}{\varepsilon^2 d(1-\eta)}\right)^{|E(H_{(1)})\cap E(H_{(2)})|}. \tag{B.1.40}$$

Now since $H_{(1)}, H_{(2)} \in \mathrm{NBSAW}^*_{s,t}$ and $H = H_{(1)} \oplus H_{(2)}$ is (s,t)-pleasant, we can have at most one vertex in $V^{d=0}_{H_{(1)},H_{(2)}}$. Therefore,

$$\frac{\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]}{\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right]\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right]} \leqslant \frac{\left(1+\frac{4}{n^{\frac{1}{16}}}\right)}{1-\frac{\eta}{2}}\cdot\left(\frac{4n\cdot e^{16d/\Delta}}{\varepsilon^2 d(1-\eta)}\right)^{|E(H_{(1)})\cap E(H_{(2)})|} \leqslant \left(1+\frac{\eta}{2}\right)\cdot\left(\frac{4n\cdot e^{16d/\Delta}}{\varepsilon^2 d(1-\eta)}\right)^{|E(H_{(1)})\cap E(H_{(2)})|},$$

where the last inequality is true for n large enough. □

Lemma B.37. *Let $H = H' \oplus H''$, where $H' = H'_{(1)} \oplus H'_{(2)}$ and $H'' = H''_{(1)} \oplus H''_{(2)}$ for some $H'_{(1)}, H'_{(2)}, H'_{(1)}, H'_{(2)} \in \mathrm{NBSAW}^*_{s,t}$, where $t = K\log n$. Assume that*

- *$V(H') \cap V(H'') = \varnothing$, $V(H'_{(1)}) \cap V(H'_{(2)}) \neq \varnothing$ and $V(H''_{(1)}) \cap V(H''_{(2)}) \neq \varnothing$.*
- *H' and H'' are $(s,2t)$-pleasant multigraphs.*

If $A > \max\left\{1, 400K, \frac{100}{K}\right\}$, then

- *If $|E'''_{\geqslant 2}(H)| \leqslant \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ and $d^H_1(v) + d^H_{\geqslant 2}(v) \leqslant \Delta$ for every $v \in V(H)$, then for n large enough, we have*
$$\mathbb{E}\left[\overline{\mathbf{Y}}_H\right] \leqslant \left(1+\frac{4}{n^{\frac{1}{16}}}\right)\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(1)}}\right]\cdot\mathbb{E}\left[\overline{\mathbf{Y}}_{H_{(2)}}\right].$$
- *If $|E'''_{\geqslant 2}(H)| > \frac{\log n}{12\log\left(\frac{2}{\varepsilon}\right)}$ or there exists $v \in V(H)$ such that $d^H_1(v) + d^H_{\geqslant 2}(v) > \Delta$, then for n large enough, we have*
$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\right]\right| \leqslant \frac{4}{n^{\frac{1}{12}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

Proof. The lemma follows immediately from Lemma B.32 and Lemma B.34. □

B.2 Bounds for the centered matrix

In this section, we will study $\left|\mathbb{E}\left[\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t\right)\right]\right|$.

Definition B.38. For every $H \in \mathrm{BSAW}_{s,t}$, we denote the s self-avoiding-walks that form H as $W_1(H), \ldots, W_t(H)$, and we denote the set $\{W_1(H), \ldots, W_t(H)\}$ as $\mathcal{W}(H)$.

Definition B.39. For every self-avoiding-walk W, define

$$\mathbf{x}_W = \prod_{uv \in W} \mathbf{x}_u \mathbf{x}_v.$$

It is easy to see that $\mathbf{x}_W = \mathbf{x}_i \mathbf{x}_j$, where i and j are the end-vertices of W. In other words, for every $i, j \in [n]$ and every $W \in \mathrm{SAW}^s_{ij}$, we have $\mathbf{x}_W = \mathbf{x}_i \mathbf{x}_j$.

Let us now analyze $\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t\right)$:

$$\begin{aligned}
\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t\right) &= \sum_{\substack{i_1,\ldots,i_{t+1} \in [n]: \\ i_{t+1} = i_1}} \prod_{l \in [t]} \left[\left(\frac{1}{|\mathrm{SAW}^s_{i_l i_{l+1}}|} \left(\frac{2n}{\varepsilon \cdot d}\right)^s \sum_{W \in \mathrm{SAW}^s_{i_l i_l + 1}} \overline{\mathbf{Y}}_W \right) - \mathbf{x}_{i_l i_{l+1}} \right] \\
&= \sum_{\substack{i_1,\ldots,i_{t+1} \in [n]: \\ i_{t+1} = i_1}} \prod_{l \in [t]} \left[\frac{1}{|\mathrm{SAW}^s_{i_l i_{l+1}}|} \left(\frac{2n}{\varepsilon \cdot d}\right)^s \cdot \sum_{W \in \mathrm{SAW}^s_{i_l i_l + 1}} \left(\overline{\mathbf{Y}}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{i_l i_{l+1}} \right) \right] \\
&= \left[\frac{(n-s+1)!}{n!} \left(\frac{2n}{\varepsilon \cdot d}\right)^s \right]^t \cdot \sum_{\substack{i_1,\ldots,i_{t+1} \in [n]: \\ i_{t+1} = i_1}} \prod_{l \in [t]} \left[\sum_{W \in \mathrm{SAW}^s_{i_l i_l + 1}} \left(\overline{\mathbf{Y}}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W \right) \right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t\right) &= (1 \pm o(1)) \cdot n^t \left(\frac{2}{\varepsilon \cdot d}\right)^{st} \cdot \sum_{\substack{i_1,\ldots,i_{t+1} \in [n]: \\ i_{t+1} = i_1}} \sum_{\substack{W_1 \in \mathrm{SAW}^s_{i_1 i_2}, \\ W_2 \in \mathrm{SAW}^s_{i_2 i_3}, \\ \ldots, \\ W_t \in \mathrm{SAW}^s_{i_t i_{t+1}}}} \prod_{l \in [t]} \left[\overline{\mathbf{Y}}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W \right] \\
&= (1 \pm o(1)) \cdot n^t \left(\frac{2}{\varepsilon \cdot d}\right)^{st} \cdot \sum_{H \in \mathrm{BSAW}_{s,t}} \prod_{l \in [t]} \left[\overline{\mathbf{Y}}_{W_l(H)} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W_l(H)} \right] \\
&= (1 \pm o(1)) \cdot n^t \left(\frac{2}{\varepsilon \cdot d}\right)^{st} \cdot \sum_{H \in \mathrm{BSAW}_{s,t}} \hat{\mathbf{Y}}_H,
\end{aligned} \tag{B.2.1}$$

where

$$\hat{\mathbf{Y}}_H := \prod_{W \in \mathcal{W}(H)} \left[\overline{\mathbf{Y}}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W \right]. \tag{B.2.2}$$

If we compare Eq. (B.2.1) with the non-centered case, we can see that $\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t\right)$ has the same expression as $\mathrm{Tr}\left(Q^{(s)}(\overline{\mathbf{Y}})^t\right)$, except that $\overline{\mathbf{Y}}_H$ is replaced with $\hat{\mathbf{Y}}_H$. The next section is dedicated for the proof of an upper bound on $\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$ for every $H \in \mathrm{BSAW}_{s,t}$. This will allow us to prove an upper bound on $\left|\mathbb{E}\left[\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t\right)\right]\right|$.

B.2.1 An upper bound for every block self-avoiding-walk

Before proving the upper bound on $\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$, we will first informally analyze the same quantity but for the non-truncated case. The main reason for doing so is that the non-truncated case is much simpler, and the analysis of the non-truncated case contains many of the elements of the proof of the truncated case. Therefore, understanding the non-truncated case will be helpful later in understanding the much more complicated truncated case.

B.2.1.1 Warm-up with the non-truncated case

We would like to upper bound the following quantity:

$$\left|\mathbb{E}\left[\prod_{W \in \mathcal{W}(H)} \left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \middle| \mathbf{x}\right]\right|.$$

The following lemma shows that the expectation of every multiplicand in the above expression is exactly zero.

Lemma B.40. *For every $i, j \in [n]$ and every $W \in SAW^s_{ij}$, we have*

$$\mathbb{E}[\mathbf{Y}_W|\mathbf{x}] = \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W = \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_i \mathbf{x}_j.$$

Proof. We have:

$$\begin{aligned}\mathbb{E}[\mathbf{Y}_W|\mathbf{x}] &= \mathbb{E}\left[\prod_{uv \in W} \mathbf{Y}_{uv} \middle| \mathbf{x}\right] = \prod_{uv \in W} \mathbb{E}[\mathbf{Y}_{uv}|\mathbf{x}] = \prod_{uv \in W} \left(\frac{\varepsilon d \mathbf{x}_u \mathbf{x}_v}{2n}\right) \\ &= \left(\frac{\varepsilon d}{2n}\right)^s \prod_{uv \in W} \mathbf{x}_u \mathbf{x}_v = \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W = \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_i \mathbf{x}_j.\end{aligned}$$

□

Lemma B.40 implies that for every $W \in \mathcal{W}(H)$, we have:

$$\mathbb{E}\left[\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W \middle| \mathbf{x}\right] = 0.$$

Therefore, if there is one walk $W' \in \mathcal{W}(H)$ such that all the edges in W' have multiplicity 1 in H, then $\mathbf{Y}_{W'} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W'}$ is conditionally independent from $\left(\mathbf{Y}_{W''} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W''}\right)_{W'' \in \mathcal{W}(H)\setminus\{W'\}}$, hence

$$\begin{aligned}
&\mathbb{E}\left[\prod_{W\in\mathcal{W}(H)}\left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right)\middle|\mathbf{x}\right] \\
&\qquad = \mathbb{E}\left[\mathbf{Y}_{W'} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W'}\middle|\mathbf{x}\right] \cdot \mathbb{E}\left[\prod_{W''\in\mathcal{W}(H)\setminus\{W'\}}\left(\mathbf{Y}_{W''} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W''}\right)\middle|\mathbf{x}\right] = 0.
\end{aligned} \tag{B.2.3}$$

On the other hand, if every walk $W \in \mathcal{W}(H)$ has at least one edge of multiplicity $\geqslant 2$, we can do the following:

$$\begin{aligned}
&\mathbb{E}\left[\prod_{W\in\mathcal{W}(H)}\left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right)\middle|\mathbf{x}\right] \\
&\qquad = \sum_{\mathcal{W}\subseteq\mathcal{W}(H)} (-1)^{|\mathcal{W}(H)|-|\mathcal{W}|} \cdot \mathbb{E}\left[\left(\prod_{W\in\mathcal{W}}\mathbf{Y}_W\right)\cdot\left(\prod_{W\in\mathcal{W}(H)\setminus\mathcal{W}}\left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right)\middle|\mathbf{x}\right] \\
&\qquad = \sum_{\mathcal{W}\subseteq\mathcal{W}(H)} (-1)^{t-|\mathcal{W}|} \cdot \left(\prod_{W\in\mathcal{W}(H)\setminus\mathcal{W}}\left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right)\cdot \mathbb{E}\left[\prod_{W\in\mathcal{W}}\mathbf{Y}_W\middle|\mathbf{x}\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left|\mathbb{E}\left[\prod_{W\in\mathcal{W}(H)}\left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right)\middle|\mathbf{x}\right]\right| &\leqslant \sum_{\mathcal{W}\subseteq\mathcal{W}(H)}\left(\frac{\varepsilon d}{2n}\right)^{s(|\mathcal{W}(H)|-|\mathcal{W}|)} \cdot \left|\mathbb{E}\left[\prod_{W\in\mathcal{W}}\mathbf{Y}_W\middle|\mathbf{x}\right]\right| \\
&= \sum_{\mathcal{W}\subseteq\mathcal{W}(H)}\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}}}\middle|\mathbf{x}\right]\right|,
\end{aligned} \tag{B.2.4}$$

where

$$H_{\mathcal{W}} = \bigoplus_{W\in\mathcal{W}} W.$$

Recall that in the non-truncated case, for every edge $uv \in E(H)$, we have

$$\mathbb{E}\big[\mathbf{Y}_{uv}^{m_H(uv)}\big|\mathbf{x}\big] = \begin{cases} \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v d}{2n} & \text{if } m_H(uv) = 1, \\ \big(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\big)\frac{d}{n} + O\Big(\frac{1}{n^2}\Big) & \text{if } m_H(uv) \geqslant 2. \end{cases}$$

Let $\mathcal{W} \in \mathcal{W}(H)$ and let $W \in \mathcal{W}$. Define $\mathcal{W}' = \mathcal{W} \setminus \{W\}$. We have the following possibilities:

(a) $W \cap E(H_{\mathcal{W}'}) = \varnothing$, in which case we have

$$\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}}}\big|\mathbf{x}\big] = \mathbb{E}\big[\mathbf{Y}_W\big|\mathbf{x}\big] \cdot \mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}'}}\big|\mathbf{x}\big] = \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W \cdot \mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}'}}\big|\mathbf{x}\big],$$

which implies that

$$\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \big|\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}}}\big|\mathbf{x}\big]\big| = \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}'|)} \cdot \big|\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}'}}\big|\mathbf{x}\big]\big|. \tag{B.2.5}$$

(b) $W \cap E(H_{\mathcal{W}'}) \neq \varnothing$. In this case, let $\mathsf{E} = W \cap E(H_{\mathcal{W}'})$ and partition E into

$$\mathsf{E}_1 = W \cap E_1(H_{\mathcal{W}'}) = \big\{e \in W \cap E(H_{\mathcal{W}'}) : \ e \text{ has multiplicity 1 in } H_{\mathcal{W}'}\big\},$$

and

$$\mathsf{E}_{\geqslant 2} = W \cap E_{\geqslant 2}(H_{\mathcal{W}'}) = \big\{e \in W \cap E(H_{\mathcal{W}'}) : \ e \text{ has multiplicity at least 2 in } H_{\mathcal{W}'}\big\}.$$

It is easy to see that we have:

$$\begin{aligned} \big|\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}}}\big|\mathbf{x}\big]\big| &= (1 \pm o(1)) \cdot \left(\frac{\varepsilon d}{2n}\right)^{|W|-|\mathsf{E}|} \cdot \left[\prod_{uv \in \mathsf{E}_1} \left[\Big(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\Big)\frac{d}{n} + O\Big(\frac{1}{n^2}\Big)\right]\right] \cdot \frac{\big|\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}'}}\big|\mathbf{x}\big]\big|}{\big(\frac{\varepsilon d}{2n}\big)^{|\mathsf{E}_1|}} \\ &= (1 \pm o(1))\left(\frac{\varepsilon d}{2n}\right)^{s-|\mathsf{E}_1|-|\mathsf{E}|} \cdot \left[\prod_{uv \in \mathsf{E}_1} \Big(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\Big)\frac{d}{n}\right] \cdot \big|\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}'}}\big|\mathbf{x}\big]\big|. \end{aligned}$$

Therefore,

$$\begin{aligned} &\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \big|\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}}}\big|\mathbf{x}\big]\big| \\ &\qquad = (1 \pm o(1))\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|+1)-|\mathsf{E}_1|-|\mathsf{E}|} \cdot \left[\prod_{uv \in \mathsf{E}_1} \Big(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\Big)\frac{d}{n}\right] \cdot \big|\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}'}}\big|\mathbf{x}\big]\big| \\ &\qquad = (1 \pm o(1))\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}'|)-|\mathsf{E}_1|-|\mathsf{E}|} \cdot \left[\prod_{uv \in \mathsf{E}_1} \Big(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\Big)\frac{d}{n}\right] \cdot \big|\mathbb{E}\big[\mathbf{Y}_{H_{\mathcal{W}'}}\big|\mathbf{x}\big]\big| \end{aligned}$$

$$\geqslant (1 \pm o(1))\left(\frac{\varepsilon d}{2n}\right)^{-|\mathsf{E}_1|-|\mathsf{E}|} \cdot \left(\left(1-\frac{\varepsilon}{2}\right)\frac{d}{n}\right)^{|\mathsf{E}_1|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}'|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}'}}|\mathbf{x}\right]\right|,$$

which implies that

$$\begin{aligned}\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}'|)} &\cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}'}}|\mathbf{x}\right]\right| \\ &\leqslant (1 \pm o(1))\left(\frac{\frac{\varepsilon}{2}}{1-\frac{\varepsilon}{2}}\right)^{|\mathsf{E}_1|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|\mathsf{E}|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}}}|\mathbf{x}\right]\right| \qquad \text{(B.2.6)} \\ &= O\left(\frac{1}{n^{|E|}}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}}}|\mathbf{x}\right]\right|.\end{aligned}$$

From Eq. (B.2.5) and Eq. (B.2.6), we conclude that if n is large enough, then we always have

$$\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}'|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}'}}|\mathbf{x}\right]\right| \leqslant \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}}}|\mathbf{x}\right]\right|. \qquad \text{(B.2.7)}$$

Remark B.41. Note that in Eq. (B.2.6) and in the above equation, we implicitly used $\frac{\varepsilon}{2-\varepsilon} = o(n)$. This means that if ε is close to 2, then n should be large in order for the above equation to be true. We used $\frac{\varepsilon}{2-\varepsilon} = o(n)$ here because this is an informal discussion, and using $\frac{\varepsilon}{2-\varepsilon} = o(n)$ will help us in illustrating the proof strategy in a simple way. However, when we formally compute the upper bound for the truncated case, we will not use $\frac{\varepsilon}{2-\varepsilon} = o(n)$. The main reason why we avoided using $\frac{\varepsilon}{2-\varepsilon} = o(n)$ in the formal proof is because we do not want to require n to be larger than necessary in order for our results to hold. More precisely, if $d \gg 1$ and ε is close to 2, the weak recovery problem should be easy, and we should not require n to be too large.

If $W \cap E(H_{\mathcal{W}'}) \neq \emptyset$, then Eq. (B.2.6) implies that

$$\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}'|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}'}}|\mathbf{x}\right]\right| \leqslant O\left(\frac{1}{n}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}}}|\mathbf{x}\right]\right|.$$

Now if every walk in $\mathcal{W}(H)$ contains at least one edge of multiplicity $\geqslant 2$ in H, then for every $W \in \mathcal{W}(H)$, we have

$$\begin{aligned}\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}(H)\setminus\{W\}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}(H)\setminus\{W\}}}|\mathbf{x}\right]\right| &\leqslant O\left(\frac{1}{n}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}(H)|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}(H)}}|\mathbf{x}\right]\right| \\ &= O\left(\frac{1}{n}\right) \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H}|\mathbf{x}\right]\right|.\end{aligned}$$

On the other hand, for every $\mathcal{W} \subsetneq \mathcal{W}(H)$, there exists $W \in \mathcal{W}(H)$ such that $\mathcal{W} \subseteq \mathcal{W}(H) \setminus \{W\}$. From Eq. (B.2.7) we can deduce that

$$\left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}}}|\mathbf{x}\right]\right| \leqslant \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}(H)\setminus\{W\}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}(H)\setminus\{W\}}}|\mathbf{x}\right]\right|$$

$$\leqslant O\left(\frac{1}{n}\right) \cdot \left|\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}\right]\right|.$$

If we put this in Eq. (B.2.4), we get

$$\begin{aligned}
\left|\mathbb{E}\left[\prod_{W \in \mathcal{W}(H)} \left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \middle| \mathbf{x}\right]\right| &\leqslant \left|\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}\right]\right| + \sum_{\mathcal{W} \subsetneq \mathcal{W}(H)} \left(\frac{\varepsilon d}{2n}\right)^{s(t-|\mathcal{W}|)} \cdot \left|\mathbb{E}\left[\mathbf{Y}_{H_{\mathcal{W}}} \middle| \mathbf{x}\right]\right| \\
&\leqslant \left|\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}\right]\right| + \sum_{\mathcal{W} \subsetneq \mathcal{W}(H)} O\left(\frac{1}{n}\right) \cdot \left|\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}\right]\right| \\
&\leqslant \left(1 + O\left(\frac{2^{|\mathcal{W}(H)|} - 1}{n}\right)\right) \cdot \left|\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}\right]\right| \\
&= \left(1 + O\left(\frac{2^t - 1}{n}\right)\right) \cdot \left|\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}\right]\right| \\
&= (1 + o(1)) \cdot \left|\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}\right]\right|,
\end{aligned} \tag{B.2.8}$$

where the last equality follows from the fact that $2^t \leqslant 2^{K \log n} = n^K = o(n)$, assuming that $K < 1$.

In summary, if there is at least one walk in $\mathcal{W}(H)$ such that all its edges have multiplicity 1 in H, then

$$\left|\mathbb{E}\left[\prod_{W \in \mathcal{W}(H)} \left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \middle| \mathbf{x}\right]\right| = 0.$$

On the other hand, if every walk in $\mathcal{W}(H)$ contains at least one edge of multiplicity $\geqslant 2$ in H, then

$$\left|\mathbb{E}\left[\prod_{W \in \mathcal{W}(H)} \left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \middle| \mathbf{x}\right]\right| \leqslant (1 + o(1)) \cdot \left|\mathbb{E}\left[\mathbf{Y}_H \middle| \mathbf{x}\right]\right|.$$

An approximate version of this phenomenon occurs in the truncated case. We will show that if there is at least one walk in $\mathcal{W}(H)$ such that all its edges have multiplicity 1 in H and all its vertices have low degrees in H, then $\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$ will be very small. In all the other cases, we will show that $\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$ is not too large compared to $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$.

As we will see, most nice block self-avoiding-walks have many walks whose edges are all of multiplicity 1. This means that the multigraphs that contributed significantly in the case of the non-centered matrix $\mathbb{E}\left[\mathrm{Tr}\left(Q^s \left(\overline{\mathbf{Y}}\right)^t\right)\right]$, will contribute very little in the case of the centered matrix $\mathbb{E}\left[\mathrm{Tr}\left(\left(Q^{(s)}\left(\overline{\mathbf{Y}}\right) - \mathbf{x}\mathbf{x}^\mathsf{T}\right)^t\right)\right]$.

On the other hand, multigraphs that contributed negligibly in the case of the non-centered matrix $\mathbb{E}\left[\mathrm{Tr}\left(Q^s \left(\overline{\mathbf{Y}}\right)^t\right)\right]$, will contribute a comparable amount in the case of the cen-

tered matrix $\mathbb{E}\left[\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\intercal\right)^t\right)\right]$. This essentially means that $\mathbb{E}\left[\mathrm{Tr}\left(\left(Q^{(s)}(\overline{\mathbf{Y}}) - \mathbf{x}\mathbf{x}^\intercal\right)^t\right)\right]$ is negligible with respect to $\mathbb{E}\left[\mathrm{Tr}\left(Q^s(\overline{\mathbf{Y}})^t\right)\right]$.

B.2.1.2 Analyzing walks of multiplicity 1

The discussion in the previous section motivates the following definition:

Definition B.42. Let $H \in \mathrm{BSAW}_{s,t}$. For every $W \in \mathcal{W}(H)$, we say that W is of multiplicity 1 in H if *every* edge in W is of multiplicity 1 in H. Define

$$\mathcal{W}_1(H) = \left\{W \in \mathcal{W}(H) : \ W \text{ is of multiplicity 1 in } H\right\},$$

and

$$\mathcal{W}_{\geqslant 2}(H) = \mathcal{W}(H) \setminus \mathcal{W}_1(H).$$

In the previous section, we used the fact that if H contains a walk of multiplicity 1, then

$$\left|\mathbb{E}\left[\prod_{W \in \mathcal{W}(H)} \left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \middle| \mathbf{x}\right]\right| = 0.$$

We would like to show something similar for the truncated case. Recall that if an edge has both its end-vertices in $\mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$,[23] then there is a significant probability that it will be a safe edge[24], in which case it will behave similarly to the non-truncated case. This motivates the following definition:

Definition B.43. A walk $W \in \mathcal{W}_1(H)$ is said to be *reassuring* if $V(W) \subseteq \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$. We denote the set of reassuring walks in $\mathcal{W}_1(H)$ as $\mathcal{W}_{1r}(H)$.

A walk $W \in \mathcal{W}_1(H)$ that is not reassuring is said to be *disturbing*. We denote the set of disturbing walks in $\mathcal{W}_1(H)$ as $\mathcal{W}_{1d}(H)$.

Clearly, $\left\{\mathcal{W}_{1r}(H), \mathcal{W}_{1d}(H)\right\}$ is a partition of $\mathcal{W}_1(H)$.

Roughly speaking, if H contains a reassuring walk, then with significant probability this walk will behave similarly to the truncated case. This will cause $\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right|$ to be small.

Definition B.44. Recall Definition B.2 and let $H \in \mathrm{BSAW}_{s,t}$. For every walk $W \in \mathcal{W}(H)$, if $V(W)$ is completely $(\mathbf{G}, H)$-*safe*, we say that W is $(\mathbf{G}, H)$-*walk-safe*. We say that W is $(\mathbf{G}, H)$-*walk-unsafe* if it is not $(\mathbf{G}, H)$-*walk-safe*.

We say that a subset $\mathcal{W}$ of $\mathcal{W}(H)$ is *completely* $(\mathbf{G}, H)$-*walk-safe* if all the walks in it are $(\mathbf{G}, H)$-*walk-safe*. Similarly, we say that $\mathcal{W}$ is *completely* $(\mathbf{G}, H)$-*walk-unsafe* if all the walks in it are $(\mathbf{G}, H)$-*walk-unsafe*.

If $\mathbf{G}$ and H are clear from the context, we drop $(\mathbf{G}, H)$ and simply write walk-safe, completely walk-safe, walk-unsafe, and completely walk-unsafe.

[23] Recall Definition B.4 and Definition B.5

[24] Recall Definition B.2

We emphasize that in order for a set of walks to be completely walk-unsafe, it is not necessary that the set of vertices forming the walks in it is completely unsafe: It is sufficient that every walk contains at least one unsafe vertex.

Definition B.45. Let $H \in \mathrm{BSAW}_{s,t}$. For every $\mathcal{W} \subset \mathcal{W}(H)$, we define

$$H_{\mathcal{W}} = \bigoplus_{W \in \mathcal{W}} W,$$

and so

$$\overline{\mathbf{Y}}_{H_{\mathcal{W}}} = \overline{\mathbf{Y}}_{\bigoplus_{W \in \mathcal{W}} W} = \prod_{W \in \mathcal{W}} \overline{\mathbf{Y}}_W.$$

Lemma B.46. *Let $H \in BSAW_{s,t}$ be such that $t = K \log n$. We have*

$$\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right| \leqslant \frac{n^{\frac{2K}{A}}}{2^{3As \cdot |\mathcal{W}_{1r}(H)|}} \cdot \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|} \cdot 2^{|\mathcal{W}_1(H)|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}_1(H)|} \cdot \sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}),$$

where

$$F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) = \left(\frac{3d}{n}\right)^{s(|\mathcal{W}_{\geqslant 2}(H)| - |\mathcal{W}_{\geqslant 2}|)} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}| \middle| \mathbf{x}\right],$$

and $\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}$ is as in Definition B.14

Proof. We only provide a high level description of the proof here. The detailed proof can be found in Appendix B.3.2.1.

We start by upper bounding the probability that no reassuring walk is walk-safe. Then, we show that in the event that there is at least one reassuring walk that is walk-safe, the conditional expectation will be zero.

Now for the case that no reassuring walk is walk-safe, we write an equation that is similar to Eq. (B.2.4), and upper bound each summand using the same techniques that allowed us to prove Lemma B.15.

See Appendix B.3.2.1 for the details. □

B.2.1.3 Analyzing walks of multiplicity at least 2

Lemma B.47. *Let $H \in BSAW_{s,t}$ be such that $t = K \cdot \log n$ and let $F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})$ be as in Lemma B.46. If $K \leqslant \frac{1}{100}$ and n is large enough, we have*

$$\sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) \leqslant \frac{2 \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}(H)})|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}^b(H)|}}{\prod_{v \in \mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d_{\geqslant 2}^H(v) - \Delta\right)}} \cdot \prod_{uv \in E_{\geqslant 2}^a(H)} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right], \tag{B.2.9}$$

where $F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})$ is as in Lemma B.46.

Proof. We use Lemma B.16 and perform calculations that are similar to those in Eq. (B.2.8). The detailed proof can be found in Appendix B.3.2.2. □

B.2.1.4 Proof of the upper bound for every block self-avoiding walk

Definition B.48. For every $H \in \mathrm{BSAW}_{s,t}$, define the following quantity:

$$\hat{U}_H(\mathbf{x}) = 2n^{\frac{K}{A}} \cdot \frac{2^{|E_1^a(H)|}}{2^{As\cdot|\mathcal{W}_{1r}(H)|}} \cdot \overline{U}_H(\mathbf{x}),$$

where $\overline{U}_H(\mathbf{x})$ is as in Definition B.7.

Lemma B.49. *If $A > \max\{100K, 1\}$, $K \leqslant \frac{1}{100}$ and n is large enough, then for every $H \in BSAW_{s,t}$, we have*

$$\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right| \leqslant \hat{U}_H(\mathbf{x}),$$

where $\hat{U}_H(\mathbf{x})$ is as in Definition B.48.

Proof. From Lemma B.46 and Lemma B.47, we have

$$\begin{aligned}
&\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right| \\
&\quad\leqslant \frac{2^{|\mathcal{W}_1(H)|+1}}{2^{3As\cdot|\mathcal{W}_{1r}(H)|}} \cdot \frac{n^{\frac{2K}{A}} \cdot \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}_1(H)|+|E_1(H_{\mathcal{W}_{\geqslant 2}(H)})|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}^b(H)|}}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d_{\geqslant 2}^H(v)-\Delta\right)}} \\
&\qquad\qquad \times \prod_{uv\in E_{\geqslant 2}^a(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right] \\
&\quad= \frac{2^{|\mathcal{W}_1(H)|+1}}{2^{3As\cdot|\mathcal{W}_{1r}(H)|}} \cdot \frac{n^{\frac{2K}{A}} \cdot \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}^b(H)|}}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d_{\geqslant 2}^H(v)-\Delta\right)}} \cdot \prod_{uv\in E_{\geqslant 2}^a(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right] \\
&\quad= \frac{2^{|\mathcal{W}_1(H)|+1}}{2^{3As\cdot|\mathcal{W}_{1r}(H)|}} \cdot \overline{U}_H(\mathbf{x}).
\end{aligned}$$

Now since $A > \max\{100K, 1\}$ and $s \geqslant 1$, we have

$$3As \cdot |\mathcal{W}_{1r}(H)| \geqslant As \cdot |\mathcal{W}_{1r}(H)| + |\mathcal{W}_{1r}(H)|,$$

hence

$$\begin{aligned}
\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right| &\leqslant \frac{2^{|\mathcal{W}_{1r}(H)|+|\mathcal{W}_{1d}(H)|+1}}{2^{As\cdot|\mathcal{W}_{1r}(H)|+|\mathcal{W}_{1r}(H)|}} \cdot \overline{U}_H(\mathbf{x}) \\
&= 2 \cdot \frac{2^{|\mathcal{W}_{1d}(H)|}}{2^{As\cdot|\mathcal{W}_{1r}(H)|}} \cdot \overline{U}_H(\mathbf{x}).
\end{aligned}$$

Observe that for every $W \in \mathcal{W}_{1d}(H)$, there exists at least one edge $e \in W$ such that $e \in E_1^a(H)$ or $e \in E_1^d(H)$, where $E_1^d(H)$ is as in Eq. (B.1.8). Therefore,

$$|\mathcal{W}_{1d}(H)| \leqslant |E_1^a(H)| + |E_1^d(H)|,$$

which implies that

$$2^{|\mathcal{W}_{1d}(H)|} \leqslant 2^{|E_1^a(H)|} \cdot 2^{|E_1^d(H)|} \leqslant 2^{|E_1^a(H)|} \cdot \left(\frac{6}{\varepsilon}\right)^{|E_1^d(H)|} \overset{(*)}{\leqslant} 2^{|E_1^a(H)|} \cdot n^{\frac{K}{A}},$$

where $(*)$ follows from Lemma B.51. We conclude that

$$\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right| \leqslant 2n^{\frac{K}{A}} \cdot \frac{2^{|E_1^a(H)|}}{2^{As\cdot|\mathcal{W}_{1r}(H)|}} \cdot \overline{U}_H(\mathbf{x}) = \hat{U}_H(\mathbf{x}).$$

□

B.3 Proofs of technical lemmas for the trace bounds

B.3.1 Proofs of technical lemmas for the non-centered matrix

B.3.1.1 Upper bound on the probability of having unsafe vertices

Proof of Lemma B.9. By the union bound, we have

$$\begin{aligned}
&\mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{4}\middle|\mathbf{x}\right] \\
&\leqslant \sum_{\substack{S\subseteq[n]\setminus V(H):\\ |S|=\lceil\Delta/4\rceil}} \mathbb{P}\left[\{\forall u\in S, uv\in\mathbf{G}\}\middle|\mathbf{x}\right] = \sum_{\substack{S\subseteq[n]\setminus V(H):\\ |S|=\lceil\Delta/4\rceil}} \prod_{u\in S}\left[\left(1+\frac{\varepsilon}{2}\mathbf{x}_{uv}\right)\frac{d}{n}\right] \\
&\leqslant \sum_{\substack{S\subseteq[n]\setminus V(H):\\ |S|=\lceil\Delta/4\rceil}} \left(\frac{2d}{n}\right)^{|S|} = \binom{n-|V(H)|}{\lceil\Delta/4\rceil}\left(\frac{2d}{n}\right)^{\lceil\Delta/4\rceil} \leqslant \binom{n}{\lceil\Delta/4\rceil}\left(\frac{2d}{n}\right)^{\lceil\Delta/4\rceil} \\
&\leqslant \frac{n^{\lceil\Delta/4\rceil}(2d)^{\lceil\Delta/4\rceil}}{\lceil\Delta/4\rceil!n^{\lceil\Delta/4\rceil}} \leqslant \frac{(2d)^{\lceil\Delta/4\rceil}}{(\lceil\Delta/4\rceil/2)^{\lceil\Delta/4\rceil/2}} \leqslant \frac{(2d)^{2\lceil\Delta/4\rceil}}{(\lceil\Delta/4\rceil/2)^{\lceil\Delta/4\rceil/2}}.
\end{aligned}$$

Now from Eq. (B.1.1), we have $\Delta \geqslant 128e^4d^4$ and so $\lceil\Delta/4\rceil/2 \geqslant 16e^4d^4$. Therefore,

$$\mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\middle|\mathbf{x}\right] \leqslant \frac{(2d)^{2\lceil\Delta/4\rceil}}{(16e^4d^4)^{\lceil\Delta/4\rceil/2}} = \frac{(2d)^{2\lceil\Delta/4\rceil}}{(2ed)^{2\lceil\Delta/2\rceil}} = e^{-2\lceil\Delta/4\rceil} \leqslant e^{-\Delta/2}.$$

From Eq. (B.1.1), we also have

$$\frac{\Delta}{2} \geqslant \log(2As) + 6A\tau s^2\cdot\log 2 + 4A^2\tau^2s^2\left(\log\frac{6}{\varepsilon}\right)^2$$

$$\geqslant \log(2As) + 6A\tau s^2 \cdot \log 2 + 4\tau^2 s^2 \left(\log \frac{6}{\varepsilon}\right)$$
$$= \log\left[2As \cdot 2^{6A\tau s^2}\left(\frac{6}{\varepsilon}\right)^{4\tau^2 s^2}\right].$$

Therefore,

$$\mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\middle|\mathbf{x}\right] \leqslant \frac{1}{2As \cdot 2^{6A\tau s^2}}\left(\frac{\varepsilon}{6}\right)^{4\tau^2 s^2}.$$

□

Proof of Lemma B.11. If S is completely crossing, then for every vertex in S, we have at least one H-cross-edge that is present in $\mathbf{G}$, and which is incident to it. Therefore, we have at least $\lceil |S|/2 \rceil$ H-cross-edges that are present in $\mathbf{G}$. Since we have at most $|V(H)|^2 \leqslant s^2t^2$ H-cross-edges, the number of collections of H-cross-edges of size $\lceil |S|/2 \rceil$ is at most

$$\binom{s^2t^2}{\lceil |S|/2 \rceil} \leqslant (s^2t^2)^{\lceil |S|/2 \rceil}.$$

Given $\mathbf{x}$, the conditional probability that any particular edge is present in $\mathbf{G}$ is at most

$$\left(1 + \frac{\varepsilon}{2}\right)\frac{d}{n} \leqslant \frac{2d}{n}.$$

Therefore, given $\mathbf{x}$, the conditional probability that S is completely crossing can be upper bounded by:

$$\mathbb{P}\left[\left\{S \text{ is completely } H\text{-crossing in } \mathbf{G}\right\}\middle|\mathbf{x}\right] \leqslant (s^2t^2)^{\lceil |S|/2 \rceil}\left(\frac{2d}{n}\right)^{\lceil |S|/2 \rceil}$$
$$= \left(\frac{2ds^2t^2}{n}\right)^{\lceil |S|/2 \rceil} \leqslant \left(\frac{2ds^2t^2}{n}\right)^{|S|/2},$$

where the last inequality is true for n large enough. □

Proof of Lemma B.12. First notice that if $v \in \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$, then

$$d^H_1(v) \leqslant \tau \leqslant \frac{1}{4} \cdot \left[\log(2As) + 12A\tau s^2 \cdot \log 2 + 8A^2\tau^2 s^2 \left(\log \frac{6}{\varepsilon}\right)^2\right] \leqslant \frac{\Delta}{4},$$

and

$$d^H_{\geqslant 2}(v) \leqslant \tau \leqslant \frac{\Delta}{4}.$$

Therefore,

$$d_H(v) = d^H_1(v) + d^H_{\geqslant 2}(v) \leqslant \frac{\Delta}{2},$$

which implies that

$$\begin{aligned}\mathbb{P}\big[\{V \text{ is completely unsafe}\}\big|\mathbf{x}\big] &= \mathbb{P}\big[\{\forall v \in V, d_{\mathbf{G}-G(H)}(v) > \Delta - d_{G(H)}(v)\}\big|\mathbf{x}\big] \\ &\leqslant \mathbb{P}\left[\left\{\forall v \in V, d_{\mathbf{G}-G(H)}(v) > \Delta - \frac{\Delta}{2}\right\}\middle|\mathbf{x}\right] \\ &= \mathbb{P}\left[\left\{\forall v \in V, d_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\right\}\middle|\mathbf{x}\right].\end{aligned}$$

Now for every $S \subseteq V$, define the events

$$\mathcal{E}_{S,H,c\text{-}cf} = \Big\{S \text{ is completely } H\text{-cross-free in } \mathbf{G}\Big\},$$

and

$$\mathcal{E}_{S,H,c\text{-}c} = \Big\{S \text{ is completely } H\text{-crossing in } \mathbf{G}\Big\}.$$

We have:

$$\begin{aligned}&\mathbb{P}\big[\{V \text{ is completely unsafe}\}\big|\mathbf{x}\big] \\ &\quad\leqslant \mathbb{P}\left[\left\{\forall v \in V, d_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\right\}\middle|\mathbf{x}\right] \\ &\quad= \sum_{S\subseteq V} \mathbb{P}\left[\left\{\forall v \in V, d_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\right\}\middle|\mathbf{x}, \mathcal{E}_{S,H,c\text{-}cf} \cap \mathcal{E}_{V\setminus S,H,c\text{-}c}\right]\mathbb{P}\big[\mathcal{E}_{S,H,c\text{-}cf} \cap \mathcal{E}_{V\setminus S,H,c\text{-}c}\big|\mathbf{x}\big] \\ &\quad\leqslant \sum_{S\subseteq V} \mathbb{P}\left[\left\{\forall v \in S, d_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\right\}\middle|\mathbf{x}, \mathcal{E}_{S,H,c\text{-}cf} \cap \mathcal{E}_{V\setminus S,H,c\text{-}c}\right]\mathbb{P}\big[\mathcal{E}_{V\setminus S,H,c\text{-}c}\big|\mathbf{x}\big] \\ &\quad\leqslant \sum_{S\subseteq V} \mathbb{P}\left[\left\{\forall v \in S, d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\right\}\middle|\mathbf{x}, \mathcal{E}_{S,H,c\text{-}cf} \cap \mathcal{E}_{V\setminus S,H,c\text{-}c}\right]\left(\frac{2ds^2t^2}{n}\right)^{(|V|-|S|)/2},\end{aligned}$$

where the last inequality follows from Lemma B.11 and the fact that if v is cross-free, then $d^i_{\mathbf{G}-G(H)}(v) = 0$ and so $d_{\mathbf{G}-G(H)}(v) = d^o_{\mathbf{G}-G(H)}(v)$.

Now since $(d^o_{\mathbf{G}-G(H)}(v))_{v\in S}$ are conditionally mutually independent given $\mathbf{x}$, and since they are conditionally independent from $\mathcal{E}_{S,H,c\text{-}cf} \cap \mathcal{E}_{V\setminus S,H,c\text{-}c}$ given $\mathbf{x}$, we have

$$\mathbb{P}\big[\{V \text{ is completely unsafe}\}\big|\mathbf{x}\big] \leqslant \sum_{S\subseteq V}\left(\prod_{v\in S}\mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\middle|\mathbf{x}\right]\right)\left(\frac{2ds^2t^2}{n}\right)^{(|V|-|S|)/2}.$$

Now from Lemma B.9 we get

$$\mathbb{P}\big[\{V \text{ is completely unsafe}\}\big|\mathbf{x}\big] \leqslant \sum_{S\subseteq V}\left(\frac{\eta}{2}\right)^{|S|}\left(st\sqrt{\frac{2d}{n}}\right)^{|V|-|S|} = \left(\frac{\eta}{2} + st\sqrt{\frac{2d}{n}}\right)^{|V|} \leqslant \eta^{|V|},$$

where the last inequality is true for n large enough. □

B.3.1.2 Upper bound on the contribution of edges of multiplicity 1

In order to prove Lemma B.15, we need a few lemmas.

The following lemma proves a result that is similar to Lemma B.15, but instead of using the best upper bound for all the edges in $E_1^b(H)$, we use loose upper bounds for the edges that are incident to $\big(\mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)\big) \setminus U$, where U is some subset of $\mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$. This lemma will be useful later in situations where we cannot guarantee that these edges are likely to be safe.

Lemma B.50. *Let H be a multigraph such that $|V(H)| \leqslant st = sK\log n$. Recall Definition B.4, Definition B.5 and Definition B.6, and let $E_1^b(H)$ and $E_1^d(H)$ be as in Eq. (B.1.7) and Eq. (B.1.8), respectively.*

Let $\mathcal{S}(H) = \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$ and let $U \subseteq \mathcal{S}(H)$. Let $\mathcal{E}$ be an event such that:

(a) The event $\mathcal{E}$ depends only on $\mathbf{x}$ and $\mathbf{G} - E_1(H)$, i.e., the event $\mathcal{E}$ is $\sigma(\mathbf{x}, \mathbf{G} - E_1(H))$-measurable. In other words, if we condition on $\mathbf{x}$, then $\mathcal{E}$ depends only $(\mathbb{1}_{uv\in\mathbf{G}})_{u,v\in[n]:\, uv\notin E_1(H)}$. This implies that given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent from $(\mathbb{1}_{uv\in\mathbf{G}})_{uv\in E_1(H)}$.

(b) For every $V \subseteq U$, we have

$$\mathbb{P}\big[\{V \text{ is completely unsafe}\}\big|\mathbf{x}, \mathcal{E}\big] \leqslant \eta^{|V|}.$$

(c) If we are given $\mathbf{x}$ and $\mathcal{E}$, then for every $V \subseteq U$, the event

$$\{V \text{ is completely safe}\} \cap \{U \setminus V \text{ is completely unsafe}\},$$

is conditionally independent from $(\mathbb{1}_{\{uv\in\mathbf{G}\}})_{uv\in E(H)}$.

Then,

$$\big|\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x}, \mathcal{E}\big]\big| \leqslant n^{\frac{K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+|E_1^d(H)|+\tau(|\mathcal{S}(H)|-|U|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\big[|\tilde{\mathbf{Y}}_{\geqslant 2}^H|\big|\mathbf{x}, \mathcal{E}\big],$$

where $\tilde{\mathbf{Y}}_{\geqslant 2}^H$ is as in Definition B.14.

Proof. It is easy to see that $\{E_1^a(H), E_1^b(H), E_1^d(H)\}$ is a partition of $E_1(H)$.

For every $V \subseteq U$, define the events

$$\mathcal{E}_{V,H,c\text{-}s} = \Big\{V \text{ is completely } H\text{-safe in } \mathbf{G}\Big\},$$

and

$$\mathcal{E}_{V,H,c\text{-}us} = \Big\{V \text{ is completely } H\text{-unsafe in } \mathbf{G}\Big\}.$$

We have:

$$\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x}, \mathcal{E}\big] = \sum_{V\subseteq U} \mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}\big] \cdot \mathbb{P}\big[\mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}\big|\mathbf{x}, \mathcal{E}\big].$$

Now let $V \subseteq U$ and suppose that $\mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}$ occurs, i.e., V is completely safe and $U \setminus V$ is completely unsafe. Since $\mathcal{S}(H) \setminus V \subseteq \mathcal{S}(H) \subseteq \mathcal{S}_1(H)$, there are at most $\tau \cdot |\mathcal{S}(H) \setminus V|$ edges of multiplicity 1 in H that are incident to vertices in $\mathcal{S}(H) \setminus V$. In particular, there are at most $\tau \cdot |\mathcal{S}(H) \setminus V|$ edges in $E_1^b(H)$ which are incident to vertices in $\mathcal{S}(H) \setminus V$.

Since every edge in $E_1^b(H)$ has both its ends in $\mathcal{S}(H)$, there are at least $|E_1^b(H)| - \tau \cdot |\mathcal{S}(H) \setminus V|$ edges in $E_1^b(H)$ which have both their end-vertices in V. This means that there are at least $|E_1^b(H)| - \tau \cdot |\mathcal{S}(H) \setminus V|$ safe edges in $E_1^b(H)$. For every $V \subseteq U$, let S_V be an arbitrary subset of $E_1^b(H)$ containing exactly $\max\{0, |E_1^b(H)| - \tau \cdot |\mathcal{S}(H) \setminus V|\}$ safe edges. Note that the choice of S_V depends only on the structure of H, and does not depend on the random sample $(\mathbf{G}, \mathbf{x})$.

Property (a) implies that we can apply Lemma B.13 to the event $\mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}$ and the edges in S_V. We get:

$$\begin{aligned}
&\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}\right] \\
&\qquad = \left(\prod_{uv \in S_V} \mathbb{E}[\mathbf{Y}_{uv} | \mathbf{x}]\right) \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H-S_V} \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}\right] \\
&\qquad = \left(\prod_{uv \in S_V} \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v d}{2n}\right) \cdot \mathbb{E}\left[\overline{\mathbf{Y}}_{H-S_V} \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right]\right| &\leqslant \sum_{V \subseteq U} \left(\frac{\varepsilon d}{2n}\right)^{|S_V|} \left|\mathbb{E}\left[\overline{\mathbf{Y}}_{H-S_V} \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}\right]\right| \\
&\qquad \times \mathbb{P}\left[\mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us} \middle| \mathbf{x}, \mathcal{E}\right] \\
&\leqslant \sum_{V \subseteq U} \left(\frac{\varepsilon d}{2n}\right)^{|S_V|} \mathbb{E}\left[|\overline{\mathbf{Y}}_{H-S_V}| \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}\right] \\
&\qquad \times \mathbb{P}\left[\mathcal{E}_{U\setminus V,H,c\text{-}us} \middle| \mathbf{x}, \mathcal{E}\right] \qquad \text{(B.3.1)} \\
&\leqslant \sum_{V \subseteq U} \left(\frac{\varepsilon d}{2n}\right)^{|S_V|} \mathbb{E}\left[|\overline{\mathbf{Y}}_{H-S_V}| \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U\setminus V,H,c\text{-}us}\right] \cdot \eta^{|U|-|V|},
\end{aligned}$$

where the last inequality follows from Property (b).

Now we upper bound $|\overline{\mathbf{Y}}_{H-S_V}|$ as follows:

$$\begin{aligned}
|\overline{\mathbf{Y}}_{H-S_V}| &= |\mathbf{Y}_{H-S_V}| \cdot \mathbb{1}_{\mathcal{E}^c_{V(H-S_V)}} \\
&= |\mathbf{Y}_{H-S_V}| \cdot \prod_{uv \in E(H)\setminus S_V} \left(\mathbb{1}_{\{d_{\mathbf{G}}(u) \leqslant \Delta\}} \cdot \mathbb{1}_{\{d_{\mathbf{G}}(v) \leqslant \Delta\}}\right)
\end{aligned}$$

$$
\begin{aligned}
&\leqslant |\mathbf{Y}_{H-S_V}| \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left(\mathbb{1}_{\{d_{\mathbf{G}}(u) \leqslant \Delta\}} \cdot \mathbb{1}_{\{d_{\mathbf{G}}(v) \leqslant \Delta\}}\right) \\
&\leqslant |\mathbf{Y}_{H-S_V}| \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left(\mathbb{1}_{\{d_{\mathbf{G} \cap E_{\geqslant 2}(H)}(u) \leqslant \Delta\}} \cdot \mathbb{1}_{\{d_{\mathbf{G} \cap E_{\geqslant 2}(H)}(v) \leqslant \Delta\}}\right) \\
&= |\mathbf{Y}_{(H-S_V) \cap E_1(H)}| \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left(\mathbf{Y}_{uv}^{m_H(uv)} \mathbb{1}_{\{d_{\mathbf{G} \cap E_{\geqslant 2}(H)}(u) \leqslant \Delta\}} \cdot \mathbb{1}_{\{d_{\mathbf{G} \cap E_{\geqslant 2}(H)}(v) \leqslant \Delta\}}\right) \\
&= |\mathbf{Y}_{E_1(H) \setminus S_V}| \cdot |\tilde{\mathbf{Y}}^H_{\geqslant 2}|.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[|\overline{\mathbf{Y}}_{H-S_V}| \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U \setminus V,H,c\text{-}us}\right] \\
&\qquad \leqslant \mathbb{E}\left[|\mathbf{Y}_{E_1(H) \setminus S_V}| \cdot |\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U \setminus V,H,c\text{-}us}\right] \\
&\qquad \overset{(*)}{=} \mathbb{E}\left[|\mathbf{Y}_{E_1(H) \setminus S_V}| \cdot |\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right] \\
&\qquad \overset{(\dagger)}{=} \mathbb{E}\left[|\mathbf{Y}_{E_1(H) \setminus S_V}| \middle| \mathbf{x}, \mathcal{E}\right] \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right] \\
&\qquad \overset{(\ddagger)}{=} \mathbb{E}\left[|\mathbf{Y}_{E_1(H) \setminus S_V}| \middle| \mathbf{x}\right] \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right],
\end{aligned}
$$

where $(*)$ follows from Property (c), $(\dagger)$ and $(\ddagger)$ follow from the fact that $\mathcal{E}$ is $\sigma(\mathbf{x}, \mathbf{G} - E_1(H))$-measurable. Hence,

$$
\begin{aligned}
\mathbb{E}\left[|\overline{\mathbf{Y}}_{H-S_V}| \middle| \mathbf{x}, \mathcal{E} \cap \mathcal{E}_{V,H,c\text{-}s} \cap \mathcal{E}_{U \setminus V,H,c\text{-}us}\right] &\leqslant \left(\prod_{uv \in E_1(H) \setminus S_V} \mathbb{E}\left[|\mathbf{Y}_{uv}| \middle| \mathbf{x}\right]\right) \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right] \\
&\leqslant \left[\prod_{uv \in E_1(H) \setminus S_V} \left(2 + \frac{\varepsilon}{2}\right) \frac{d}{n}\right] \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right] \qquad \text{(B.3.2)} \\
&\leqslant \left(\frac{3d}{n}\right)^{|E_1(H) \setminus S_V|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right],
\end{aligned}
$$

Combining Eq. (B.3.1) and Eq. (B.3.2), we get:

$$
\begin{aligned}
\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right]\right| &\leqslant \sum_{V \subseteq U} \left(\frac{\varepsilon d}{2n}\right)^{|S_V|} \cdot \eta^{|U|-|V|} \left(\frac{3d}{n}\right)^{|E_1(H) \setminus S_V|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right] \\
&\overset{(\wr)}{=} \left(\frac{3d}{n}\right)^{|E_1^a(H)| + |E_1^d(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right] \cdot \sum_{V \subseteq U} \left(\frac{\varepsilon d}{2n}\right)^{|S_V|} \cdot \eta^{|U|-|V|} \left(\frac{3d}{n}\right)^{|E_1^b(H) \setminus S_V|} \\
&= \left(\frac{3d}{n}\right)^{|E_1^a(H)| + |E_1^d(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right] \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1^b(H)|} \cdot \sum_{V \subseteq U} \eta^{|U|-|V|} \left(\frac{6}{\varepsilon}\right)^{|E_1^b(H) \setminus S_V|},
\end{aligned}
$$

where $(\wr)$ follows from the fact that $\left\{E_1^a(H), E_1^b(H) \setminus S_V, E_1^d(H)\right\}$ is a partition of $E_1(H) \setminus S_V$.

Now since $|S_V| \geqslant |E_1^b(H)| - \tau \cdot |\mathcal{S}(H) \setminus V|$, we have

$$|E_1^b(H) \setminus S_V| \leqslant \tau \cdot |\mathcal{S}(H) \setminus V|.$$

Therefore,

$$\begin{aligned}
&\sum_{V \subseteq U} \eta^{|U|-|V|} \left(\frac{6}{\varepsilon}\right)^{|E_1^b(H) \setminus S_V|} \\
&\qquad \leqslant \sum_{V \subseteq U} \eta^{|U|-|V|} \left(\frac{6}{\varepsilon}\right)^{\tau \cdot (|\mathcal{S}(H)|-|V|)} = \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)} \sum_{V \subseteq U} \left[\eta \left(\frac{6}{\varepsilon}\right)^{\tau}\right]^{|U|-|V|} \\
&\qquad = \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)} \left[1 + \eta \left(\frac{6}{\varepsilon}\right)^{\tau}\right]^{|U|} \leqslant \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)} \left[1 + \frac{1}{As}\left(\frac{\varepsilon}{6}\right)^{\tau} \left(\frac{6}{\varepsilon}\right)^{\tau}\right]^{st} \\
&\qquad = \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)} \left(1 + \frac{1}{As}\right)^{st} \leqslant e^{\frac{1}{As} st} \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)} \\
&\qquad = e^{\frac{1}{A} \cdot K \cdot \log n} \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)} = n^{\frac{K}{A}} \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)}.
\end{aligned}$$

We conclude that

$$\begin{aligned}
&\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right]\right| \\
&\qquad \leqslant n^{\frac{K}{A}} \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)} \cdot \left(\frac{3d}{n}\right)^{|E_1^a(H)|+|E_1^d(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H| \middle| \mathbf{x}, \mathcal{E}\right] \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1^b(H)|} \\
&\qquad = n^{\frac{K}{A}} \left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1^a(H)|+|E_1^d(H)|+|E_1^b(H)|} \cdot \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+|E_1^d(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H| \middle| \mathbf{x}, \mathcal{E}\right] \\
&\qquad = n^{\frac{K}{A}} \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+|E_1^d(H)|+\tau(|\mathcal{S}(H)|-|U|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H| \middle| \mathbf{x}, \mathcal{E}\right].
\end{aligned}$$

□

Lemma B.51. *Let H be a multigraph with at most $st = sK \log n$ vertices and at most st multi-edges, we have*

$$\left(\frac{6}{\varepsilon}\right)^{|E_1^d(H)|} \leqslant n^{\frac{K}{A}}.$$

Proof. Recall that

$$E_1^d(H) = \left\{uv \in E_1(H) \setminus E_1^a(H) : \ u \notin \mathcal{S}_{\geqslant 2}(H) \text{ or } v \notin \mathcal{S}_{\geqslant 2}(H)\right\}.$$

Since every edge $uv \in E_1^d(H)$ satisfies $uv \in E_1(H) \setminus E_1^a(H)$, we must have $u \in \mathcal{S}_1(H)$ and $v \in \mathcal{S}_1(H)$. On the other hand, every edge in $E_1^d(H)$ is incident to at least one vertex in

$V(H) \setminus \mathcal{S}_{\geqslant 2}(H) = \mathcal{I}_{\geqslant 2}(H) \cup \mathcal{L}_{\geqslant 2}(H)$. Therefore, every edge in $E_1^d(H)$ is a multiplicity-1 edge that is incident to at least one vertex in $\mathcal{S}_1(H) \cap (\mathcal{I}_{\geqslant 2}(H) \cup \mathcal{L}_{\geqslant 2}(H))$.

Since we have at most st edges in $G(H)$, we have

$$\frac{\Delta}{4} \cdot |\mathcal{I}_{\geqslant 2}(H) \cup \mathcal{L}_{\geqslant 2}(H)| \leqslant \sum_{v \in \mathcal{I}_{\geqslant 2}(H) \cup \mathcal{L}_{\geqslant 2}(H)} d_{G(H)}(v) \leqslant \sum_{v \in V(H)} d_{G(H)}(v) \leqslant 2st. \tag{B.3.3}$$

Therefore,

$$\begin{aligned}\left|\mathcal{S}_1(H) \cap \left(\mathcal{I}_{\geqslant 2}(H) \cup \mathcal{L}_{\geqslant 2}(H)\right)\right| &\leqslant |\mathcal{I}_{\geqslant 2}(H) \cup \mathcal{L}_{\geqslant 2}(H)| \\ &\leqslant \frac{8st}{\Delta} \overset{(*)}{\leqslant} \frac{8st}{8\tau^2 A^2 s^2 \left(\log \frac{6}{\varepsilon}\right)^2} \leqslant \frac{t}{A\tau\left(\log \frac{6}{\varepsilon}\right)},\end{aligned}$$

where $(*)$ follows from Eq. (B.1.1). Now since every edge in $E_1^d(H)$ is is incident to at least one vertex in $\mathcal{S}_1(H) \cap (\mathcal{I}_{\geqslant 2}(H) \cup \mathcal{L}_{\geqslant 2}(H))$, and since every vertex in $\mathcal{S}_1(H)$ is incident to at most τ multiplicity-1 edges, we conclude that

$$|E_1^d(H)| \leqslant \frac{t}{A\tau\left(\log \frac{6}{\varepsilon}\right)} \cdot \tau \leqslant \frac{t}{A \log \frac{6}{\varepsilon}},$$

and

$$\left(\frac{6}{\varepsilon}\right)^{|E_1^d(H)|} \leqslant e^{\frac{t}{A \log \frac{6}{\varepsilon}} \log \frac{6}{\varepsilon}} = e^{\frac{1}{A} \cdot K \log n} = n^{\frac{K}{A}}.$$

□

Lemma B.52. *Let H be a multigraph with at most $st = sK \log n$ vertices and at most st multi-edges, and let $\mathcal{S}(H) = \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$. Assume that $U \subseteq \mathcal{S}(H)$ and $\mathcal{E}$ satisfy the conditions of Lemma B.50. Furthermore, assume that $\mathcal{E}$ satisfies the following additional condition:*

(d) Given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent from $(\mathbb{1}_{\{uv \in \mathbf{G}\}})_{uv \in E_{\geqslant 2}(H)}$.

Then,

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right]\right| \leqslant n^{\frac{2K}{A}} \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)| + \tau(|\mathcal{S}(H)| - |U|)} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H| \middle| \mathbf{x}\right].$$

Proof. Let $E_1^b(H)$ and $E_1^d(H)$ be as in Eq. (B.1.7) and Eq. (B.1.8), respectively. From Lemma B.50, we know that

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right]\right| \leqslant n^{\frac{K}{A}} \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)| + |E_1^d(H)| + \tau(|\mathcal{S}(H)| - |U|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H| \middle| \mathbf{x}, \mathcal{E}\right].$$

On the other hand, Property (d) implies that $\mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H| \middle| \mathbf{x}, \mathcal{E}\right] = \mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H| \middle| \mathbf{x}\right]$. Therefore,

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right]\right| \leqslant n^{\frac{K}{A}} \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)| + |E_1^d(H)| + \tau(|\mathcal{S}(H)| - |U|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H| \middle| \mathbf{x}\right]. \tag{B.3.4}$$

Combining this with Lemma B.51, we get

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}\right]\right| \leqslant n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+\tau(|\mathcal{S}(H)|-|U|)}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left|\mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}|\middle|\mathbf{x}\right]\right|.$$

□

Now we are ready to prove Lemma B.15.

Proof of Lemma B.15. Let $\mathcal{S}(H) = \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$. By using Lemma B.12, it can be easily seen that if we take $U = \mathcal{S}(H)$ and $\mathcal{E}$ to be an "almost sure event", i.e., $\mathbb{P}[\mathcal{E}] = 1$, then the conditions of Lemma B.50 and Lemma B.52 are satisfied. Therefore,

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right| &= \left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}\right]\right| \\ &\leqslant n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left|\mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}|\middle|\mathbf{x}\right]\right|.\end{aligned}$$

□

B.3.1.3 Upper bound on the contribution of edges of multiplicity at least 2

Proof of Lemma B.16. Recall the definitions of $E^a_{\geqslant 2}(H)$ and $E^b_{\geqslant 2}(H)$ from Definition B.6.

We have

$$\begin{aligned}|\tilde{\mathbf{Y}}^H_{\geqslant 2}| &= \prod_{uv\in E_{\geqslant 2}(H)}\left|\tilde{\mathbf{Y}}^{m_H(uv)}_{uv,E_{\geqslant 2}(H)}\right| \\ &= \prod_{uv\in E_{\geqslant 2}(H)}\left|\mathbf{Y}^{m_H(uv)}_{uv}\right|\cdot\mathbb{1}_{\{d_{\mathbf{G}\cap E_{\geqslant 2}(H)}(u)\leqslant\Delta\}}\cdot\mathbb{1}_{\{d_{\mathbf{G}\cap E_{\geqslant 2}(H)}(v)\leqslant\Delta\}} \\ &\leqslant \left(\prod_{uv\in E^a_{\geqslant 2}(H)}\left|\mathbf{Y}^{m_H(uv)}_{uv}\right|\right)\cdot\left(\prod_{uv\in E^b_{\geqslant 2}(H)}\left|\mathbf{Y}^{m_H(uv)}_{uv}\right|\right)\cdot\left(\prod_{v\in\mathcal{L}_{\geqslant 2}(H)}\mathbb{1}_{\{d_{\mathbf{G}\cap E^b_{\geqslant 2}(H)}(v)\leqslant\Delta\}}\right) \\ &= \left|\mathbf{Y}_{H^a_{\geqslant 2}}\right|\cdot\left|\mathbf{Y}_{H^b_{\geqslant 2}}\right|\cdot\mathbb{1}_{\mathcal{E}^{b,H}_{\geqslant 2}},\end{aligned}$$

where

$$\begin{aligned}\mathbf{Y}_{H^a_{\geqslant 2}} &= \prod_{uv\in E^a_{\geqslant 2}(H)}\mathbf{Y}^{m_H(uv)}_{uv}, \\ \mathbf{Y}_{H^b_{\geqslant 2}} &= \prod_{uv\in E^b_{\geqslant 2}(H)}\mathbf{Y}^{m_H(uv)}_{uv},\end{aligned}$$

and

$$\mathcal{E}^{b,H}_{\geqslant 2} = \bigcap_{v\in\mathcal{L}_{\geqslant 2}(H)}\left\{d_{\mathbf{G}\cap E^b_{\geqslant 2}(H)}(v)\leqslant\Delta\right\}.$$

Therefore,

$$\begin{aligned}\mathbb{E}\big[|\tilde{\mathbf{Y}}^H_{\geqslant 2}|\big|\mathbf{x}\big] &= \mathbb{E}\big[|\mathbf{Y}_{H^a_{\geqslant 2}}|\cdot|\mathbf{Y}_{H^b_{\geqslant 2}}|\cdot\mathbb{1}_{\mathcal{E}^{b,H}_{\geqslant 2}}\big|\mathbf{x}\big]\\ &\overset{(*)}{=} \mathbb{E}\big[|\mathbf{Y}_{H^a_{\geqslant 2}}|\big|\mathbf{x}\big]\cdot\mathbb{E}\big[|\mathbf{Y}_{H^b_{\geqslant 2}}|\cdot\mathbb{1}_{\mathcal{E}^{b,H}_{\geqslant 2}}\big|\mathbf{x}\big],\end{aligned} \tag{B.3.5}$$

where $(*)$ follows from the fact that given $\mathbf{x}$, the random variable $\mathbf{Y}_{H^a_{\geqslant 2}}$ is conditionally independent from $(\mathbf{Y}_{H^b_{\geqslant 2}}, \mathbb{1}_{\mathcal{E}^{b,H}_{\geqslant 2}})$. This is because the presence or absence of edges in $E^a_{\geqslant 2}(H)$ do not affect the degree $d_{\mathbf{G}\cap E^b_{\geqslant 2}(H)}(v)$ of any vertex $v \in \mathcal{L}_{\geqslant 2}(H)$.

Now let us evaluate $\mathbb{E}\big[|\mathbf{Y}_{H^a_{\geqslant 2}}|\big|\mathbf{x}\big]$ and $\mathbb{E}\big[|\mathbf{Y}_{H^b_{\geqslant 2}}|\cdot\mathbb{1}_{\mathcal{E}^{b,H}_{\geqslant 2}}\big|\mathbf{x}\big]$. We have

$$\begin{aligned}\mathbb{E}\big[|\mathbf{Y}_{H^a_{\geqslant 2}}|\big|\mathbf{x}\big] &= \prod_{uv\in E^a_{\geqslant 2}(H)}\mathbb{E}\big[|\mathbf{Y}_{uv}|^{m_H(uv)}\big|\mathbf{x}\big]\\ &= \prod_{uv\in E^a_{\geqslant 2}(H)}\left[\Big(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\Big)\frac{d}{n}\Big(1-\frac{d}{n}\Big)^{m_H(uv)} + \Big(1-\Big(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\Big)\frac{d}{n}\Big)\Big(\frac{d}{n}\Big)^{m_H(uv)}\right]\\ &\leqslant \prod_{uv\in E^a_{\geqslant 2}(H)}\left[\Big(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\Big)\frac{d}{n}+\frac{d^2}{n^2}\right].\end{aligned} \tag{B.3.6}$$

On the other hand,

$$\begin{aligned}&\mathbb{E}\big[|\mathbf{Y}_{H^b_{\geqslant 2}}|\cdot\mathbb{1}_{\mathcal{E}^{b,H}_{\geqslant 2}}\big|\mathbf{x}\big]\\ &= \sum_{\substack{S\subseteq E^b_{\geqslant 2}(H):\\ \forall v\in\mathcal{L}_{\geqslant 2}(H), d_S(v)\leqslant\Delta}}\left[\prod_{uv\in S}\Big(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\Big)\frac{d}{n}\Big(1-\frac{d}{n}\Big)^{m_H(uv)}\right]\\ &\qquad\qquad\times\left[\prod_{uv\in E^b_{\geqslant 2}(H)\setminus S}\Big(1-\Big(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\Big)\frac{d}{n}\Big)\Big(\frac{d}{n}\Big)^{m_H(uv)}\right]\\ &\leqslant \sum_{\substack{S\subseteq E^b_{\geqslant 2}(H):\\ \forall v\in\mathcal{L}_{\geqslant 2}(H),\, d_S(v)\leqslant\Delta}}\left[\prod_{uv\in S}\frac{2d}{n}\right]\cdot\left[\prod_{uv\in E^b_{\geqslant 2}(H)\setminus S}\Big(\frac{2d}{n}\Big)^2\right] = \sum_{\substack{S\subseteq E^b_{\geqslant 2}(H):\\ \forall v\in\mathcal{L}_{\geqslant 2}(H),\, d_S(v)\leqslant\Delta}}\Big(\frac{2d}{n}\Big)^{|S|+2|E^b_{\geqslant 2}(H)|-2|S|}\\ &= \Big(\frac{2d}{n}\Big)^{|E^b_{\geqslant 2}(H)|}\sum_{\substack{S\subseteq E^b_{\geqslant 2}(H):\\ \forall v\in\mathcal{L}_{\geqslant 2}(H),\, d_S(v)\leqslant\Delta}}\Big(\frac{2d}{n}\Big)^{|E^b_{\geqslant 2}(H)|-|S|}.\end{aligned} \tag{B.3.7}$$

Now define

$$E^{b,i}_{\geqslant 2}(H) = \big\{uv\in E^b_{\geqslant 2}(H):\ u\in\mathcal{L}_{\geqslant 2}(H) \text{ and } v\in\mathcal{L}_{\geqslant 2}(H)\big\}.$$

The superscript i indicates that the edges in $E^{b,i}_{\geqslant 2}(H)$ are internal to $\mathcal{L}_{\geqslant 2}(H)$, i.e., both end-vertices are in $\mathcal{L}_{\geqslant 2}(H)$. Furthermore, for every $v \in \mathcal{L}_{\geqslant 2}(H)$, define:

$$E^{b}_{\geqslant 2}(v,H) = \left\{uv : \ uv \in E^{b}_{\geqslant 2}(H)\right\},$$

$$E^{b,i}_{\geqslant 2}(v,H) = \left\{uv : \ uv \in E^{b}_{\geqslant 2}(H) \text{ and } u \in \mathcal{L}_{\geqslant 2}(H)\right\},$$

and

$$E^{b,o}_{\geqslant 2}(v,H) = \left\{uv : \ uv \in E^{b}_{\geqslant 2}(H) \text{ and } u \notin \mathcal{L}_{\geqslant 2}(H)\right\}.$$

The superscript o in $E^{b,o}_{\geqslant 2}(v,H)$ indicates that the edges in $E^{b,o}_{\geqslant 2}(v,H)$ go from v to $V(H) \setminus \mathcal{L}_{\geqslant 2}(H)$, i.e., they go to the "outside" of $\mathcal{L}_{\geqslant 2}(H)$. We have the following:

- For every $v \in \mathcal{L}_{\geqslant 2}(H)$, $\left\{E^{b,i}_{\geqslant 2}(v,H), E^{b,o}_{\geqslant 2}(v,H)\right\}$ is a partition of $E^{b}_{\geqslant 2}(v,H)$.
- For every $v \in \mathcal{L}_{\geqslant 2}(H)$, we have $|E^{b}_{\geqslant 2}(v,H)| = d^{H}_{\geqslant 2}(v) > \Delta$.
- $\left\{E^{b,i}_{\geqslant 2}(H)\right\} \cup \left\{E^{b,o}_{\geqslant 2}(v,H) : v \in \mathcal{L}_{\geqslant 2}(H)\right\}$ is a partition of $E^{b}_{\geqslant 2}(H)$.

Now for every $S \subseteq E^{b}_{\geqslant 2}(H)$, define

$$S^i = S \cap E^{b,i}_{\geqslant 2}(H),$$

and for every $v \in \mathcal{L}_{\geqslant 2}(H)$, define

$$S_v = S \cap E^{b}_{\geqslant 2}(v,H), \quad S^i_v = S \cap E^{b,i}_{\geqslant 2}(v,H) \quad \text{and} \quad S^o_v = S \cap E^{b,o}_{\geqslant 2}(v,H).$$

It is easy to see that $S^i_v = S_v \cap E^{b,i}_{\geqslant 2}(v,H) = S^i \cap E^{b,i}_{\geqslant 2}(v,H)$ and $S^o_v = S_v \cap E^{b,o}_{\geqslant 2}(v,H)$.

Now for two sequences of sets $(A_i)_{i\in I}$ and $(B_i)_{i\in I}$ that are indexed by the same index set I, we write $(A_i)_{i\in I} \subseteq (B_i)_{i\in I}$ to indicate that $A_i \subseteq B_i$ for all $i \in I$. We have

$$\begin{aligned}
&\sum_{\substack{S\subseteq E^{b}_{\geqslant 2}(H):\\ \forall v\in\mathcal{L}_{\geqslant 2}(H),\, d_S(v)\leqslant\Delta}} \left(\frac{2d}{n}\right)^{|E^{b}_{\geqslant 2}(H)|-|S|} \\
&\qquad = \sum_{\substack{S^i\subseteq E^{b,i}_{\geqslant 2}(H),\\ (S^o_v)_{v\in\mathcal{L}_2(H)}\subseteq(E^{b,o}_{\geqslant 2}(v,H))_{v\in\mathcal{L}_2(H)}:\\ \forall v\in\mathcal{L}_{\geqslant 2}(H),\, |S^i_v|+|S^o_v|\leqslant\Delta}} \left(\frac{2d}{n}\right)^{|E^{b}_{\geqslant 2}(H)|-|S^i|-\sum_{v\in\mathcal{L}_{\geqslant 2}(H)}|S^o_v|} \\
&\qquad = \sum_{\substack{S^i\subseteq E^{b,i}_{\geqslant 2}(H),\\ (S^o_v)_{v\in\mathcal{L}_2(H)}\subseteq(E^{b,o}_{\geqslant 2}(v,H))_{v\in\mathcal{L}_2(H)}:\\ \forall v\in\mathcal{L}_{\geqslant 2}(H),\, |S^i_v|+|S^o_v|\leqslant\Delta}} \left(\frac{2d}{n}\right)^{|E^{b,i}_{\geqslant 2}(H)|-|S^i|} \prod_{v\in\mathcal{L}_{\geqslant 2}(H)} \left(\frac{2d}{n}\right)^{|E^{b,o}_{\geqslant 2}(v,H)|-|S^o_v|},
\end{aligned}$$

hence,

$$\sum_{\substack{S\subseteq E^b_{\geqslant 2}(H):\\ \forall v\in\mathcal{L}_{\geqslant 2}(H),\, d_S(v)\leqslant\Delta}} \left(\frac{2d}{n}\right)^{|E^b_{\geqslant 2}(H)|-|S|}$$

$$= \sum_{S^i\subseteq E^{b,i}_{\geqslant 2}(H)} \left(\frac{2d}{n}\right)^{|E^{b,i}_{\geqslant 2}(H)|-|S^i|} \prod_{v\in\mathcal{L}_{\geqslant 2}(H)} \left[\sum_{\substack{S^o_v\subseteq E^{b,o}_{\geqslant 2}(v,H):\\ |S^i_v|+|S^o_v|\leqslant\Delta}} \left(\frac{2d}{n}\right)^{|E^{b,o}_{\geqslant 2}(v,H)|-|S^o_v|}\right]$$

$$\leqslant \sum_{S^i\subseteq E^{b,i}_{\geqslant 2}(H)} \left(\sqrt{\frac{2d}{n}}\right)^{2|E^{b,i}_{\geqslant 2}(H)|-2|S^i|} \prod_{v\in\mathcal{L}_{\geqslant 2}(H)} \left[\sum_{\substack{S^o_v\subseteq E^{b,o}_{\geqslant 2}(v,H):\\ |S^i_v|+|S^o_v|\leqslant\Delta}} \left(\sqrt{\frac{2d}{n}}\right)^{|E^{b,o}_{\geqslant 2}(v,H)|-|S^o_v|}\right]$$

$$= \sum_{S^i\subseteq E^{b,i}_{\geqslant 2}(H)} \left(\sqrt{\frac{2d}{n}}\right)^{\sum_{v\in\mathcal{L}_{\geqslant 2}(H)} 2(|E^{b,i}_{\geqslant 2}(v,H)|-|S^i_v|)} \prod_{v\in\mathcal{L}_{\geqslant 2}(H)} \left[\sum_{\substack{S^o_v\subseteq E^{b,o}_{\geqslant 2}(v,H):\\ |S^i_v|+|S^o_v|\leqslant\Delta}} \left(\sqrt{\frac{2d}{n}}\right)^{|E^{b,o}_{\geqslant 2}(v,H)|-|S^o_v|}\right].$$

Therefore,

$$\sum_{\substack{S\subseteq E^b_{\geqslant 2}(H):\\ \forall v\in\mathcal{L}_{\geqslant 2}(H),\, d_S(v)\leqslant\Delta}} \left(\frac{2d}{n}\right)^{|E^b_{\geqslant 2}(H)|-|S|}$$

$$\leqslant \sum_{S^i\subseteq E^{b,i}_{\geqslant 2}(H)} \prod_{v\in\mathcal{L}_{\geqslant 2}(H)} \left[\sum_{\substack{S^o_v\subseteq E^{b,o}_{\geqslant 2}(v,H):\\ |S^i_v|+|S^o_v|\leqslant\Delta}} \left(\sqrt{\frac{2d}{n}}\right)^{|E^{b,o}_{\geqslant 2}(v,H)|-|S^o_v|+|E^{b,i}_{\geqslant 2}(v,H)|-|S^i_v|}\right]$$

$$\leqslant \prod_{v\in\mathcal{L}_{\geqslant 2}(H)} \left[\sum_{S^i_v\subseteq E^{b,i}_{\geqslant 2}(v,H)} \sum_{\substack{S^o_v\subseteq E^{b,o}_{\geqslant 2}(v,H):\\ |S^i_v|+|S^o_v|\leqslant\Delta}} \left(\sqrt{\frac{2d}{n}}\right)^{|E^{b,o}_{\geqslant 2}(v,H)|-|S^o_v|+|E^{b,i}_{\geqslant 2}(v,H)|-|S^i_v|}\right]$$

$$= \prod_{v\in\mathcal{L}_{\geqslant 2}(H)} \left[\sum_{\substack{S_v\subseteq E^b_{\geqslant 2}(v,H):\\ |S_v|\leqslant\Delta}} \left(\sqrt{\frac{2d}{n}}\right)^{|E^b_{\geqslant 2}(v,H)|-|S_v|}\right]. \tag{B.3.8}$$

Now for every $v \in \mathcal{L}_{\geqslant 2}(H)$, we have

$$\sum_{\substack{S_v \subseteq E^b_{\geqslant 2}(v,H): \\ |S_v| \leqslant \Delta}} \left(\sqrt{\frac{2d}{n}}\right)^{|E^b_{\geqslant 2}(v,H)|-|S_v|} \leqslant (\Delta+1)\binom{|E^b_{\geqslant 2}(v,H)|}{\Delta}\left(\sqrt{\frac{2d}{n}}\right)^{|E^b_{\geqslant 2}(v,H)|-\Delta}$$
$$\leqslant (\Delta+1)\binom{st}{\Delta}\left(\sqrt{\frac{2d}{n}}\right)^{|E^b_{\geqslant 2}(v,H)|-\Delta} \leqslant (\Delta+1)\frac{(st)^\Delta}{\Delta!}\left(\sqrt{\frac{2d}{n}}\right)^{d^H_{\geqslant 2}(v)-\Delta},$$

By combining this with Eq. (B.3.7) and Eq. (B.3.8), we get

$$\begin{aligned}
\mathbb{E}\left[|\mathbf{Y}_{H^b_{\geqslant 2}}| \cdot \mathbb{1}_{\mathcal{E}^{b,H}_{\geqslant 2}} \middle| \mathbf{x}\right] &\leqslant \left(\frac{2d}{n}\right)^{|E^b_{\geqslant 2}(H)|} \prod_{v \in \mathcal{L}_{\geqslant 2}(H)} \left[(\Delta+1)\frac{(st)^\Delta}{\Delta!}\left(\sqrt{\frac{2d}{n}}\right)^{d^H_{\geqslant 2}(v)-\Delta}\right] \\
&\overset{(\dagger)}{\leqslant} \left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H)|} \prod_{v \in \mathcal{L}_{\geqslant 2}(H)} \left[2^{d^H_{\geqslant 2}(v)} \cdot (\Delta+1)\frac{(st)^\Delta}{\Delta!}\left(\sqrt{\frac{2d}{n}}\right)^{d^H_{\geqslant 2}(v)-\Delta}\right] \\
&= \left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H)|} \prod_{v \in \mathcal{L}_{\geqslant 2}(H)} \left[2^\Delta \cdot (\Delta+1)\frac{(st)^\Delta}{\Delta!}\left(2\sqrt{\frac{2d}{n}}\right)^{d^H_{\geqslant 2}(v)-\Delta}\right] \\
&\overset{(\ddagger)}{\leqslant} \left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H)|} \frac{1}{\prod_{v \in \mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d^H_{\geqslant 2}(v)-\Delta\right)}},
\end{aligned}$$

where (†) follows from the fact that $|E^b_{\geqslant 2}(H)| \leqslant \sum_{v \in \mathcal{L}_{\geqslant 2}(H)} d^H_{\geqslant 2}(v)$, and (‡) is true for n large enough[25]. By combining this with Eq. (B.3.5) and Eq. (B.3.6), we get

$$\begin{aligned}
\mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}\right] &\leqslant \frac{1}{\prod_{v \in \mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d^H_{\geqslant 2}(v)-\Delta\right)}} \left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H)|} \prod_{uv \in E^a_{\geqslant 2}(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right] \\
&= \frac{1}{\prod_{v \in \mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d^H_{\geqslant 2}(v)-\Delta\right)}} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \prod_{uv \in E^a_{\geqslant 2}(H)} \left[1+\frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2} + \frac{d}{n}\right].
\end{aligned}$$

□

[25] Recall that $d^H_{\geqslant 2}(v) > \Delta$ for all $v \in \mathcal{L}_{\geqslant 2}(H)$

B.3.1.4 Upper bounds for the contribution of the not-well-behaved event

We will prove Lemma B.26 in three steps. We start by proving an upper bound on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b}^c\right] \cdot \mathbb{P}[\mathcal{E}_{H,b}^c]\right|$, where $\mathcal{E}_{H,b}$ is as in Eq. (B.1.13).

Lemma B.53. *Let H be a multigraph with at most $st = sK\log n$ vertices and let $\mathcal{E}_{H,b}$ be the event that $\mathbf{x}$ is approximately balanced on $[n] \setminus V(H)$. If n is large enough, then*

$$\mathbb{P}[\mathcal{E}_{H,b}^c] \leqslant 2e^{-\frac{9}{8}\sqrt{n}}.$$

Proof. Define

$$S_{[n]\setminus V(H)} = \sum_{v\in[n]\setminus V(H)} \mathbf{x}_v,$$

where the sum is performed according to the arithmetic of integers in $\mathbb{Z}$. It is easy to see that if $\mathcal{E}_{H,b}^c$ occurs, then $|S_{[n]\setminus V(H)}| > n - |V(H)| - 2\left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$. Therefore,

$$\begin{aligned}\mathbb{P}[\mathcal{E}_{H,b}^c] &\leqslant \mathbb{P}\left[|S_{[n]\setminus V(H)}| \geqslant n - |V(H)| - 2\left\lceil \tfrac{n}{2} - n^{\frac{3}{4}} \right\rceil\right] \leqslant \mathbb{P}\left[|S_{[n]\setminus V(H)}| \geqslant n - st - 2\left(\tfrac{n}{2} - n^{\frac{3}{4}} + 1\right)\right] \\ &= \mathbb{P}\left[|S_{[n]\setminus V(H)}| \geqslant 2\left(n^{\frac{3}{4}} - \frac{st}{2} - 1\right)\right] \leqslant \mathbb{P}\left[|S_{[n]\setminus V(H)}| \geqslant \frac{3}{2}\cdot n^{\frac{3}{4}}\right],\end{aligned}$$

where the last inequality is true for n large enough. By applying Hoeffding's inequality, we get:

$$\mathbb{P}[\mathcal{E}_{H,b}^c] \leqslant 2\cdot e^{-2\frac{\left(\frac{3}{2}\cdot n^{\frac{3}{4}}\right)^2}{(n-|V(H)|)\cdot(1-(-1))^2}} = 2e^{-\frac{18n^{\frac{3}{2}}}{16(n-|V(H)|)}} \leqslant 2e^{-\frac{9n^{\frac{3}{2}}}{8n}} = 2e^{-\frac{9}{8}\sqrt{n}}.$$

□

Lemma B.54. *Let H be a multigraph with at most $st = sK\log n$ vertices and at most st multi-edges, and assume that $E_1^a(H) = \varnothing$. If n is large enough, then*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b}^c\right] \cdot \mathbb{P}[\mathcal{E}_{H,b}^c]\right| \leqslant e^{-\sqrt{n}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

Proof. Since $E_1^a(H) = \varnothing$, it follows from Definition B.7 and Lemma B.17 that for n large enough, we have

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right| &\leqslant \overline{U}_H(\mathbf{x}) \\ &= n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|} \frac{1}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H)} n^{\frac{1}{4}\left(d_{\geqslant 2}^H(v)-\Delta\right)}} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|\mathbb{E}_{\geqslant 2}(H)|} \prod_{uv\in E_{\geqslant 2}^a(H)} \left[1 + \frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2} + \frac{3d}{\sqrt{n}}\right].\end{aligned}$$

$$\overset{(*)}{\leqslant} n^{\frac{2K}{A}} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} 3^{|E^a_{\geqslant 2}(H)|} \leqslant n^{\frac{2K}{A}} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot 3^{st},$$

where $(*)$ is true for n large enough. Since $t = K \cdot \log n$, we have $3^{st} = e^{(\log 3) sK \log n} = n^{sK \log 3}$. Therefore,

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}\right]\right| \leqslant n^{\frac{2K}{A} + sK \log 3} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}.$$

Now since the event $\mathcal{E}^c_{H,b}$ depends only on $\mathbf{x}$, i.e., it is $\sigma(\mathbf{x})$-measurable, we deduce that for n large enough, we have

$$\begin{aligned}
\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{H,b}\right] \cdot \mathbb{P}[\mathcal{E}^c_{H,b}]\right| &\leqslant n^{\frac{2K}{A} + sK \log 3} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot \mathbb{P}[\mathcal{E}^c_{H,b}] \\
&\overset{(\dagger)}{\leqslant} n^{\frac{2K}{A} + sK \log 3} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot 2e^{-\frac{9}{8}\sqrt{n}} \\
&\overset{(\ddagger)}{\leqslant} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot e^{-\sqrt{n}},
\end{aligned}$$

where $(\dagger)$ follows from Lemma B.53 and $(\ddagger)$ is true for n large enough. □

Next, we prove an upper bound on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\overline{g}}]\right|$, where $\mathcal{E}_{H,b\overline{g}}$ is as in Eq. (B.1.17). We will need the following definition:

Definition B.55. Assume that $\mathbf{x}$ is approximately balanced on $[n] \setminus V(H)$. We say that a vertex $v \in V(H)$ is $(\mathbf{G}, \mathbf{x}, H)$-*lucky* if it satisfies the following two conditions:

(1) There is no H-cross-edge that is present in $\mathbf{G}$, and which is incident to it. In other words[26], $d^i_{\mathbf{G}-G(H)}(v) = 0$.

(2) In the sampled graph $\mathbf{G}$, the vertex v is not adjacent to any vertex in $\left([n] \setminus V(H)\right) \setminus V_b(H, \mathbf{x})$.

We say that $v \in V(H)$ is $(\mathbf{G}, \mathbf{x}, H)$-*unlucky* if it is not $(\mathbf{G}, \mathbf{x}, H)$-*lucky*.

A subset of $V(H)$ is said to be *completely* $(\mathbf{G}, \mathbf{x}, H)$-*lucky* if all the vertices in it are $(\mathbf{G}, \mathbf{x}, H)$-*lucky*. We say that it is *completely* $(\mathbf{G}, \mathbf{x}, H)$-*unlucky* if all the vertices in it are $(\mathbf{G}, \mathbf{x}, H)$-*unlucky*.

If $\mathbf{G}, \mathbf{x}$ and H are clear from the context, we drop $(\mathbf{G}, \mathbf{x}, H)$ and simply write lucky, unlucky, completely lucky, and completely unlucky.

[26] Recall Definition B.8

Lemma B.56. *Let H be a multigraph with at most $st = sK\log n$ vertices. If n is large enough, then for every $S \subseteq V(H)$, the conditional probability that S is completely $(\mathbf{G}, \mathbf{x}, H)$-unlucky given $\mathbf{x}$ that is approximately balanced on $[n] \setminus V(H)$ can be upper bounded by:*

$$\mathbb{P}\left[\left\{S \text{ is completely } (\mathbf{G}, \mathbf{x}, H)\text{-unlucky}\right\} \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] \leqslant \frac{1}{n^{|S|/5}}.$$

Proof. Given $\mathbf{x}$, the conditional probability that any particular edge is present in $\mathbf{G}$ is at most

$$\left(1 + \frac{\varepsilon}{2}\right)\frac{d}{n} \leqslant \frac{2d}{n}.$$

A vertex can be unlucky either because it is incident to an H-cross-edge, or because it is adjacent to a vertex in $\big([n] \setminus V(H)\big) \setminus V_b(H, \mathbf{x})$.

Since $\left|\big([n] \setminus V(H)\big) \setminus V_b(H, \mathbf{x})\right| \leqslant 2\left\lceil n^{\frac{3}{4}}\right\rceil$, it follows from the union bound that, given $\mathbf{x}$ that is approximately balanced on $[n] \setminus V(H)$, the conditional probability that any particular vertex $v \in S$ is adjacent to some vertex in $\big([n] \setminus V(H)\big) \setminus V_b(H, \mathbf{x})$ is at most

$$2\left\lceil n^{\frac{3}{4}}\right\rceil\frac{2d}{n} \leqslant \frac{8n^{\frac{3}{4}}d}{n} = \frac{8d}{n^{\frac{1}{4}}}.$$

Let $S' \subseteq S$ be the set of vertices in S that are not adjacent to any vertex in $\big([n] \setminus V(H)\big) \setminus V_b(H, \mathbf{x})$. If S is completely $(\mathbf{G}, \mathbf{x}, H)$-unlucky, then for every $v \in S'$, we have at least one H-cross-edge that is present in $\mathbf{G}$, and which is incident to it. Therefore, we have at least $\left\lceil \frac{|S'|}{2} \right\rceil$ H-cross-edges that are present in $\mathbf{G}$. Since we have at most $|V(H)|^2 \leqslant s^2t^2$ H-cross-edges, the number of collections of H-cross-edges of size $\left\lceil \frac{|S'|}{2} \right\rceil$ is at most

$$\binom{s^2t^2}{\left\lceil \frac{|S'|}{2} \right\rceil} \leqslant (s^2t^2)^{\left\lceil \frac{|S'|}{2} \right\rceil}.$$

Therefore, for n large enough, given $\mathbf{x}$ that is approximately balanced on $[n] \setminus V(H)$, the conditional probability that S is completely unlucky can be upper bounded by:

$$\begin{aligned}
&\mathbb{P}\left[\left\{S \text{ is completely } (\mathbf{G}, \mathbf{x}, H)\text{-unlucky}\right\} \middle| \mathcal{E}_{H,b}\right] \\
&\quad\leqslant \sum_{S' \subseteq S}\left(\frac{8d}{n^{\frac{1}{4}}}\right)^{|S|-|S'|} \cdot (s^2t^2)^{\left\lceil \frac{|S'|}{2} \right\rceil}\left(\frac{2d}{n}\right)^{\left\lceil \frac{|S'|}{2} \right\rceil} \overset{(*)}{\leqslant} \sum_{S' \subseteq S}\left(\frac{8d}{n^{\frac{1}{4}}}\right)^{|S|-|S'|} \cdot \left(\frac{2ds^2t^2}{n}\right)^{\frac{|S'|}{2}} \\
&\quad= \sum_{S' \subseteq S}\left(\frac{8d}{n^{\frac{1}{4}}}\right)^{|S|-|S'|} \cdot \left(\frac{\sqrt{2d} \cdot st}{n^{\frac{1}{2}}}\right)^{|S'|} \leqslant \sum_{S' \subseteq S}\left(\frac{8dst}{n^{\frac{1}{4}}}\right)^{|S|-|S'|} \cdot \left(\frac{8dst}{n^{\frac{1}{4}}}\right)^{|S'|} = \sum_{S' \subseteq S}\left(\frac{8dst}{n^{\frac{1}{4}}}\right)^{|S|} \\
&\quad= 2^{|S|}\left(\frac{8dst}{n^{\frac{1}{4}}}\right)^{|S|} = \left(\frac{16dst}{n^{\frac{1}{4}}}\right)^{|S|} \overset{(\dagger)}{\leqslant} \frac{1}{n^{|S|/5}},
\end{aligned}$$

where $(*)$ and $(\dagger)$ are true for n large enough. □

Lemma B.57. *Let H be a multigraph with at most $st = sK\log n$ vertices and at most st multi-edges. Assume that $\mathcal{L}_1(H) = \varnothing$. Let $V \subseteq \mathcal{S}_{\geqslant 2}(H)$, and for every $v \in V$, define*

$$d^o_{\mathbf{G}-G(H),b}(v) = \left|\left\{u \in V_b(H,\mathbf{x}) :\ uv \in \mathbf{G}\right\}\right|,$$

and

$$d^o_{\mathbf{G}-G(H),\bar{b}}(v) = \left|\left\{u \in \left([n] \setminus V(H)\right) \setminus V_b(H,\mathbf{x}) :\ uv \in \mathbf{G}\right\}\right|.$$

Clearly[27]*, $d^o_{\mathbf{G}-G(H)}(v) = d^o_{\mathbf{G}-G(H),b}(v) + d^o_{\mathbf{G}-G(H),\bar{b}}(v)$. Let $\mathcal{E}$ be an event such that:*

- *$\mathcal{E}$ implies that $\mathbf{x}$ is approximately balanced on $[n] \setminus V(H)$, i.e., $\mathcal{E}_{H,b} \subseteq \mathcal{E}$.*
- *$\mathcal{E}$ implies that V is completely $(\mathbf{G}, H, \mathbf{x})$-lucky, i.e., $d^i_{\mathbf{G}-G(H)}(v) = d^o_{\mathbf{G}-G(H),\bar{b}}(v) = 0$ for every $v \in V$.*
- *Given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent from $\left(d^o_{\mathbf{G}-G(H),b}(v)\right)_{v\in V}$.*

Then

$$\mathbb{P}\left[\left\{V \text{ is completely } H\text{-unsafe in } \mathbf{G}\right\}\middle|\mathbf{x},\mathcal{E}\right] \leqslant \left(\frac{\eta}{2}\right)^{|V|},$$

where η is as in Lemma B.9.

Proof. Since $d^i_{\mathbf{G}-G(H)}(v) = d^o_{\mathbf{G}-G(H),\bar{b}}(v) = 0$ for every $v \in V$, then v is unsafe if and only if $d^o_{\mathbf{G}-G(H),b}(v) > \Delta - d_{G(H)}(v)$. Therefore, for every $V \subseteq V(H)$, we have

$$\begin{aligned}
\mathbb{P}\left[\left\{V \text{ is completely unsafe}\right\}\middle|\mathbf{x},\mathcal{E}\right] &= \mathbb{P}\left[\left\{d^o_{\mathbf{G}-G(H),b}(v) > \Delta - d_{G(H)}(v),\ \forall v \in V\right\}\middle|\mathbf{x},\mathcal{E}\right] \\
&\overset{(*)}{=} \mathbb{P}\left[\left\{d^o_{\mathbf{G}-G(H),b}(v) > \Delta - d_{G(H)}(v),\ \forall v \in V\right\}\middle|\mathbf{x}\right] \\
&= \prod_{v\in V} \mathbb{P}\left[d^o_{\mathbf{G}-G(H),b}(v) > \Delta - d_{G(H)}(v)\middle|\mathbf{x}\right] \\
&\leqslant \prod_{v\in V} \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \Delta - d_{G(H)}(v)\middle|\mathbf{x}\right],
\end{aligned}$$

where $(*)$ follows from the fact that given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent from $\left(d^o_{\mathbf{G}-G(H),b}(v)\right)_{v\in V(H)}$.

Now since $V \subset \mathcal{S}_{\geqslant 2}(H)$ and since $\mathcal{L}_1(H) = \varnothing$, we have $d_{G(H)}(v) = d^H_1(v) + d^H_{\geqslant 2}(v) \leqslant \frac{\Delta}{4} + \tau \leqslant \frac{\Delta}{2}$. Therefore,

$$\begin{aligned}
\mathbb{P}\left[\left\{V \text{ is completely unsafe}\right\}\middle|\mathbf{x},\mathcal{E}\right] &\leqslant \prod_{v\in V} \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \Delta - d_{G(H)}(v)\middle|\mathbf{x}\right] \\
&\leqslant \prod_{v\in V} \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2}\middle|\mathbf{x}\right] \overset{(\dagger)}{\leqslant} \left(\frac{\eta}{2}\right)^{|V|},
\end{aligned} \tag{B.3.9}$$

where $(\dagger)$ follows from Lemma B.9. ∎

[27] Recall Definition B.8

Lemma B.58. *Let H be a multigraph with at most $st = sK\log n$ vertices and at most st multi-edges. Assume that $\mathcal{L}_1(H) = \mathcal{L}_{\geqslant 2}(H) = \varnothing$, and that $E_{\geqslant 2}(H)$ forms a forest. If $A > \max\{100K, 1\}$ and n is large enough, then we have*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\overline{g}}]\right| \leqslant \frac{1}{n^{\frac{1}{6}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},$$

where $\mathcal{E}_{H,b\overline{g}}$ is an in Eq. (B.1.17).

Proof. Recall Definition B.4, Definition B.5 and Definition B.6. Since $\mathcal{L}_1(H) = \mathcal{L}_{\geqslant 2}(H) = \varnothing$, we have $E_1^a(H) = \varnothing$, $E_{\geqslant 2}^b(H) = \varnothing$, and $E_{\geqslant 2}(H) = E_{\geqslant 2}^a(H)$. Furthermore, $\mathcal{S}_1(H) = V(H)$ and so $\mathcal{S}(H) := \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H) = \mathcal{S}_{\geqslant 2}(H)$.

For every $V \subseteq V(H)$, define the events

$$\mathcal{E}_{V,H,c\text{-}l} = \Big\{V \text{ is completely } (\mathbf{G}, \mathbf{x}, H)\text{-lucky}\Big\},$$

$$\mathcal{E}_{V,H,c\text{-}ul} = \Big\{V \text{ is completely } (\mathbf{G}, \mathbf{x}, H)\text{-unlucky}\Big\}.$$

If $\mathcal{E}_{H,g}^c$ occurs then there exists at least one vertex $v \in V(H)$ that is unlucky. Hence, $\{\mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul} : U \subsetneq V(H)\}$ forms a partition of the event $\mathcal{E}_{H,g}^c$. Therefore,

$$\begin{aligned}
&\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\overline{g}}|\mathbf{x}]\right| \\
&= \left|\sum_{U\subsetneq V(H)} \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul}\right] \cdot \mathbb{P}\left[\mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul} \middle| \mathbf{x}\right]\right| \\
&\leqslant \sum_{U\subsetneq V(H)} \left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul}\right]\right| \cdot \mathbb{P}\left[\mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul} \middle| \mathbf{x}\right] \\
&\leqslant \sum_{U\subsetneq V(H)} \left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul}\right]\right| \cdot \mathbb{P}\left[\mathcal{E}_{V(H)\setminus U,H,c\text{-}ul} \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] \\
&\leqslant \sum_{U\subsetneq V(H)} \left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul}\right]\right| \cdot \frac{1}{n^{\frac{1}{5}|V(H)\setminus U|}},
\end{aligned} \tag{B.3.10}$$

where the last inequality follows from Lemma B.56.

Now fix $U \subsetneq V(H)$. Lemma B.57 implies that for every $V \subseteq U \cap \mathcal{S}(H)$, we have

$$\mathbb{P}\left[\{V \text{ is completely } H\text{-unsafe in } \mathbf{G}\} \middle| \mathbf{x}, \mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul}\right] \leqslant \left(\frac{\eta}{2}\right)^{|V|} \leqslant \eta^{|V|}.$$

It is now easy to see that the conditions of Lemma B.50 are satisfied for $U \cap \mathcal{S}(H) \subset \mathcal{S}(H)$

and $\mathcal{E} = \mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul}$. Therefore,

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b} \cap \mathcal{E}_{U,H,c\text{-}l} \cap \mathcal{E}_{V(H)\setminus U,H,c\text{-}ul}\right]\right| &= \left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right]\right| \\ &\leqslant n^{\frac{K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+|E_1^d(H)|+\tau(|\mathcal{S}(H)|-|U\cap\mathcal{S}(H)|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}\right] \qquad \text{(B.3.11)} \\ &\leqslant n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{\tau(|\mathcal{S}(H)|-|U\cap\mathcal{S}(H)|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}\right],\end{aligned}$$

where the last equality follows from Lemma B.51, the fact that $E_1^a(H) = \varnothing$, and the fact that given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent from $\tilde{\mathbf{Y}}^H_{\geqslant 2}$.

Notice that

$$|\mathcal{S}(H)| - |U \cap \mathcal{S}(H)| = |\mathcal{S}(H) \setminus U| \leqslant |V(H) \setminus U| = |V(H)| - |U|. \qquad \text{(B.3.12)}$$

Now by combining Eq. (B.3.10) and Eq. (B.3.11) and Eq. (B.3.12), we get

$$\begin{aligned}&\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\overline{g}}|\mathbf{x}]\right| \\ &\qquad \leqslant \sum_{U\subsetneq V(H)} n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{\tau(|V(H)|-|U|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}\right] \cdot \frac{1}{n^{\frac{1}{5}(|V(H)|-|U|)}} \\ &\qquad = n^{\frac{2K}{A}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}\right] \sum_{U\subsetneq V(H)} \left(\frac{6^\tau}{\varepsilon^\tau \cdot n^{\frac{1}{5}}}\right)^{|V(H)|-|U|} \\ &\qquad = n^{\frac{2K}{A}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}\right] \cdot \left[\left(1 + \frac{6^\tau}{\varepsilon^\tau \cdot n^{\frac{1}{5}}}\right)^{|V(H)|} - 1\right].\end{aligned}$$

Now since $|V(H)| \leqslant st = sK\log n$, we have

$$\left(1 + \frac{6^\tau}{\varepsilon^\tau \cdot n^{\frac{1}{5}}}\right)^{|V(H)|} \leqslant \left(1 + \frac{6^\tau}{\varepsilon^\tau \cdot n^{\frac{1}{5}}}\right)^{st} \leqslant e^{\frac{6^\tau}{\varepsilon^\tau \cdot n^{\frac{1}{5}}} st} = 1 + O\left(\frac{6^\tau \cdot st}{\varepsilon^\tau \cdot n^{\frac{1}{5}}}\right).$$

Thus,

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,b\overline{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\overline{g}}|\mathbf{x}]\right| \leqslant O\left(\frac{6^\tau \cdot st \cdot n^{\frac{2K}{A}}}{\varepsilon^\tau \cdot n^{\frac{1}{5}}}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}\right].$$

By using Lemma B.16 and the fact that $E^b_{\geqslant 2}(H) = \varnothing$, $E^a_{\geqslant 2}(H) = E_{\geqslant 2}(H)$, and $\mathcal{L}_{\geqslant 2}(H) = \varnothing$, we get

$$\mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}\right] \leqslant \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \prod_{uv\in E_{\geqslant 2}(H)} \left[1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2} + \frac{d}{n}\right].$$

Therefore,

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{H,b\overline{g}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,b\overline{g}}|\mathbf{x}]\right| \leqslant O\left(\frac{6^\tau\cdot st\cdot n^{\frac{2K}{A}}}{\varepsilon^\tau\cdot n^{\frac{1}{5}}}\right)\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\prod_{uv\in E_{\geqslant 2}(H)}\left[1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}+\frac{d}{n}\right].$$

If $A>\max\{100K,1\}$, and n is large enough, we get

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{H,b\overline{g}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,b\overline{g}}|\mathbf{x}]\right| \leqslant \frac{1}{2n^{\frac{1}{6}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\prod_{uv\in E_{\geqslant 2}(H)}\left[1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}+\frac{d}{n}\right].$$

We conclude that

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathcal{E}_{H,b\overline{g}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,b\overline{g}}]\right| &\leqslant \frac{1}{2n^{\frac{1}{6}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\cdot\mathbb{E}\left[\prod_{uv\in E_{\geqslant 2}(H)}\left[1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}+\frac{d}{n}\right]\right]\\ &\overset{(\ddagger)}{=}\frac{1}{2n^{\frac{1}{6}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\cdot\left[\prod_{uv\in E_{\geqslant 2}(H)}\left(1+\frac{d}{n}\right)\right]\\ &=\frac{1}{2n^{\frac{1}{6}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\cdot\left(1+O\left(\frac{dst}{n}\right)\right)\\ &\overset{(\wr)}{\leqslant}\frac{1}{n^{\frac{1}{6}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},\end{aligned}$$

where (‡) follows from Lemma B.22 and the fact that $E_{\geqslant 2}(H)$ forms a forest, and ($\wr$) is true for n large enough. □

Now we will prove an upper bound on $\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathcal{E}_{H,bg\overline{d}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,bg\overline{d}}]\right|$, where $\mathcal{E}_{H,bg\overline{d}}$ is as in Eq. (B.1.18).

Lemma B.59. *Let H be a multigraph with at most $st=sK\log n$ vertices and at most st multi-edges. Assume that $\mathcal{L}_1(H)=\mathcal{L}_{\geqslant 2}(H)=\varnothing$, and that $E_{\geqslant 2}(H)$ forms a forest. If $A>\max\{100K,1\}$ and n is large enough, then we have*

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathcal{E}_{H,bg\overline{d}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,bg\overline{d}}]\right| \leqslant \frac{1}{n^{\frac{1}{2}}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},$$

where $\mathcal{E}_{H,bg\overline{d}}$ is an in Eq. (B.1.18).

Proof. Recall Definition B.4, Definition B.5 and Definition B.6. It is easy to see that if we take $U=\mathcal{S}(H):=\mathcal{S}_1(H)\cap\mathcal{S}_{\geqslant 2}(H)=V(H)$ and $\mathcal{E}=\mathcal{E}_{H,bg\overline{d}}$, then the conditions of Lemma B.50 are satisfied. Therefore,

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{H,bg\overline{d}}\right]\right|$$

$$\leqslant n^{\frac{K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+|E_1^d(H)|+\tau(|\mathcal{S}(H)|-|U|)}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H|\middle|\mathbf{x},\mathcal{E}_{H,bg\bar{d}}\right]$$
$$\leqslant n^{\frac{2K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H|\middle|\mathbf{x},\mathcal{E}_{H,bg\bar{d}}\right],$$

where the last inequality follows from Lemma B.51 and the fact that $E_1^a(H)=\varnothing$ and $U=\mathcal{S}(H)$.

Therefore,

$$\begin{aligned}&\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{H,bg\bar{d}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,bg\bar{d}}|\mathbf{x}]\right|\\&\qquad\leqslant n^{\frac{2K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\mathbb{E}\left[|\tilde{\mathbf{Y}}_{\geqslant 2}^H|\middle|\mathbf{x},\mathcal{E}_{H,bg\bar{d}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,bg\bar{d}}|\mathbf{x}]\\&\qquad= n^{\frac{2K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\cdot\mathbb{E}\left[|\mathbf{Y}_{\geqslant 2}^H|\middle|\mathbf{x},\mathcal{E}_{H,bg\bar{d}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,bg\bar{d}}|\mathbf{x}],\end{aligned}\tag{B.3.13}$$

where the last equality follows[28] from the fact that $\mathcal{L}_{\geqslant 2}(H)=\varnothing$, which means that $\tilde{\mathbf{Y}}_{\geqslant 2}^H=\mathbf{Y}_{\geqslant 2}^H$, where

$$\mathbf{Y}_{\geqslant 2}^H=\prod_{uv\in E_{\geqslant 2}(H)}\mathbf{Y}_{uv}^{m_H(uv)}.$$

Now for every $E\subseteq E_{\geqslant 2}(H)$, define the events:

$$\mathcal{E}_{E,c\text{-}d}=\left\{e\in\mathbf{G},\ \forall e\in E\right\}\quad\text{and}\quad\mathcal{E}_{E,c\text{-}\bar{d}}=\left\{e\notin\mathbf{G},\ \forall e\in E\right\}.$$

Let $\mathcal{E}_{H,bg}$ be as in Eq. (B.1.16). Since $\left\{\mathcal{E}_{E,c\text{-}d}\cap\mathcal{E}_{E_{\geqslant 2}(H)\setminus E,c\text{-}\bar{d}}:\ E\subsetneq E_{\geqslant 2}(H)\right\}$ is a partition of $\mathcal{E}_{H,d}^c$ and since $\mathcal{E}_{H,bg\bar{d}}=\mathcal{E}_{H,bg}\cap\mathcal{E}_{H,d}^c$, we can write:

$$\begin{aligned}&\mathbb{E}\left[|\mathbf{Y}_{\geqslant 2}^H|\middle|\mathbf{x},\mathcal{E}_{H,bg\bar{d}}\right]\cdot\mathbb{P}[\mathcal{E}_{H,bg\bar{d}}|\mathbf{x}]\\&=\mathbb{E}\left[|\mathbf{Y}_{\geqslant 2}^H|\middle|\mathbf{x},\mathcal{E}_{H,bg}\cap\mathcal{E}_{H,d}^c\right]\cdot\mathbb{P}[\mathcal{E}_{H,bg}\cap\mathcal{E}_{H,d}^c|\mathbf{x}]\\&=\sum_{E\subsetneq E_{\geqslant 2}(H)}\mathbb{E}\left[|\mathbf{Y}_{\geqslant 2}^H|\middle|\mathbf{x},\mathcal{E}_{H,bg}\cap\mathcal{E}_{E,c\text{-}d}\cap\mathcal{E}_{E_{\geqslant 2}(H)\setminus E,c\text{-}\bar{d}}\right]\cdot\mathbb{P}\left[\mathcal{E}_{H,bg}\cap\mathcal{E}_{E,c\text{-}d}\cap\mathcal{E}_{E_{\geqslant 2}(H)\setminus E,c\text{-}\bar{d}}\middle|\mathbf{x}\right]\\&\leqslant\sum_{E\subsetneq E_{\geqslant 2}(H)}\mathbb{E}\left[|\mathbf{Y}_{\geqslant 2}^H|\middle|\mathbf{x},\mathcal{E}_{H,bg}\cap\mathcal{E}_{E,c\text{-}d}\cap\mathcal{E}_{E_{\geqslant 2}(H)\setminus E,c\text{-}\bar{d}}\right]\cdot\mathbb{P}[\mathcal{E}_{E,c\text{-}d}\cap\mathcal{E}_{E_{\geqslant 2}(H)\setminus E,c\text{-}\bar{d}}|\mathbf{x}]\\&=\sum_{E\subsetneq E_{\geqslant 2}(H)}\prod_{uv\in E}\left[\left(1-\frac{d}{n}\right)^{m_H(uv)}\cdot\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\cdot\prod_{uv\in E_{\geqslant 2}(H)\setminus E}\left[\left(1-\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right)\cdot\left(\frac{d}{n}\right)^{m_H(uv)}\right]\\&\leqslant\sum_{E\subsetneq E_{\geqslant 2}(H)}\prod_{uv\in E}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\cdot\prod_{uv\in E_{\geqslant 2}(H)\setminus E}\left(\frac{d}{n}\right)^2\end{aligned}$$

[28] We could have used $|\tilde{\mathbf{Y}}_{\geqslant 2}^H|\leqslant|\mathbf{Y}_{\geqslant 2}^H|$, which is true in general.

$$= \prod_{uv\in E_{\geqslant 2}(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d}{n^2}\right] - \prod_{uv\in E_{\geqslant 2}(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right].$$

By combining this with Eq. (B.3.13), we get

$$\begin{aligned}
&\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{H,bg\bar{d}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,bg\bar{d}}|\mathbf{x}]\right| \\
&\qquad \leqslant n^{\frac{2K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left[\prod_{uv\in E_{\geqslant 2}(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d}{n^2}\right] - \prod_{uv\in E_{\geqslant 2}(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,bg\bar{d}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,bg\bar{d}}]\right| \\
&\qquad \leqslant n^{\frac{2K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[\prod_{uv\in E_{\geqslant 2}(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d}{n^2}\right] - \prod_{uv\in E_{\geqslant 2}(H)} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right] \\
&\qquad \overset{(*)}{=} n^{\frac{2K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left[\prod_{uv\in E_{\geqslant 2}(H)} \left(\frac{d}{n}+\frac{d}{n^2}\right) - \prod_{uv\in E_{\geqslant 2}(H)} \left(\frac{d}{n}\right)\right] \\
&\qquad \leqslant n^{\frac{2K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left[\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\left(1+O\left(\frac{dst}{n}\right)\right) - \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\right] \\
&\qquad = O\left(\frac{dst\cdot n^{\frac{2K}{A}}}{n}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \overset{(\dagger)}{\leqslant} \frac{1}{\sqrt{n}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},
\end{aligned}$$

where $(*)$ follows from Lemma B.22 and the fact thatr $E_{\geqslant 2}(H)$ is a forest, and $(\dagger)$ is true when $A > \max\{100K, 1\}$ and n is large enough. □

Now we are ready to prove Lemma B.26

Proof of Lemma B.26. Since $\left\{\mathcal{E}^c_{H,b}, \mathcal{E}_{H,b\bar{g}}, \mathcal{E}_{H,bg\bar{d}}\right\}$ is a partition of $\mathcal{E}^c_{wb,H}$, we get:

$$\begin{aligned}
\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{wb,H}\right] \cdot \mathbb{P}[\mathcal{E}^c_{wb,H}] = \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{H,b}\right] \cdot \mathbb{P}[\mathcal{E}^c_{H,b}] + \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b\bar{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\bar{g}}] \\
+ \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,bg\bar{d}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,bg\bar{d}}].
\end{aligned}$$

It follows from Lemma B.54, Lemma B.58 and Lemma B.59 that:

$$\begin{aligned}
\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{wb,H}\right] \cdot \mathbb{P}[\mathcal{E}^c_{wb,H}]\right| &\leqslant \left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}^c_{H,b}\right] \cdot \mathbb{P}[\mathcal{E}^c_{H,b}]\right| + \left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,b\bar{g}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,b\bar{g}}]\right| \\
&\quad + \left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathcal{E}_{H,bg\bar{d}}\right] \cdot \mathbb{P}[\mathcal{E}_{H,bg\bar{d}}]\right| \\
&\leqslant \left(e^{-\sqrt{n}} + \frac{1}{n^{\frac{1}{6}}} + \frac{1}{\sqrt{n}}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}
\end{aligned}$$

$$\leqslant \frac{2}{n^{\frac{1}{6}}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},$$

where the last inequality is true for n large enough. □

B.3.1.5 Upper bound on the negligible part of the contribution of the well-behaved event

In order to prove Lemma B.27, we need the following lemma:

Lemma B.60. *If $P^{wb}_{2,H}(\mathbf{x})$ is as in Eq. (B.1.24), then for n large enough, we have*

$$|P^{wb}_{2,H}(\mathbf{x})| \leqslant \frac{1}{n^{\frac{1}{5}}} \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right].$$

Proof. If $\mathbf{x}$ is not approximately balanced on $[n] \setminus V(H)$, then

$$|P^{wb}_{2,H}(\mathbf{x})| = 0 \leqslant \frac{1}{n^{\frac{1}{5}}} \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right].$$

Now assume that $\mathbf{x}$ is approximately balanced on $[n] \setminus V(H)$, and let $E_{\overline{bc}}(H, \mathbf{x})$ be as in Eq. (B.1.22). We have:

$$\begin{aligned}
|P^{wb}_{2,H}(\mathbf{x})| &= \left| \sum_{\substack{S \subseteq E_{\overline{bc}}(H,\mathbf{x}):\\ S \neq \varnothing}} \prod_{uv \in S} \left[-\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right] \right| \cdot \left[\prod_{uv \in E_{\geqslant 2}(H)} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right] \\
&\leqslant \sum_{\substack{S \subseteq E_{\overline{bc}}(H,\mathbf{x}):\\ S \neq \varnothing}} \left(\frac{2d}{n}\right)^{|S|} \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right].
\end{aligned}$$

On the other hand, for n large enough, we have:

$$|E_{\overline{bc}}(H, \mathbf{x})| = |E_c(H)| + |V(H)| \cdot \left|\left([n] \setminus V(H)\right) \setminus V_b(H, \mathbf{x})\right| \leqslant s^2t^2 + 2st \cdot n^{3/4} \leqslant 3s^2t^2 \cdot n^{3/4}.$$

Hence,

$$\sum_{\substack{S \subseteq E_{\overline{bc}}(H,\mathbf{x}):\\ S \neq \varnothing}} \left(\frac{2d}{n}\right)^{|S|} = \left(1 + \frac{2d}{n}\right)^{|E_{\overline{bc}}(H,\mathbf{x})|} - 1 = O\left(\frac{2d \cdot |E_{\overline{bc}}(H, \mathbf{x})|}{n}\right) \leqslant O\left(\frac{6ds^2t^2 \cdot n^{\frac{3}{4}}}{n}\right) \leqslant \frac{1}{n^{\frac{1}{5}}},$$

where the last inequality is true for n large enough. Therefore, we have

$$|P^{wb}_{2,H}(\mathbf{x})| \leqslant \frac{1}{n^{\frac{1}{5}}} \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right].$$

□

Now we are ready to prove Lemma B.27.

Proof of Lemma B.27. Lemma B.57 implies that for every $V \subseteq \mathcal{S}(H)$, we have

$$\mathbb{P}\left[\left\{V \text{ is completely } H\text{-unsafe in } \mathbf{G}\right\} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \leqslant \left(\frac{\eta}{2}\right)^{|V|} \leqslant \eta^{|V|}.$$

It is now easy to see that if we take $\mathcal{E} = \mathcal{E}_{wb,H}$ and $U = \mathcal{S}(H)$, then the conditions of Lemma B.50 are satisfied. Therefore,

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right]\right| &\leqslant n^{\frac{K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+|E_1^d(H)|+\tau(|\mathcal{S}(H)|-|U|)} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \\ &\overset{(*)}{\leqslant} n^{\frac{2K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \overset{(\dagger)}{\leqslant} n^{\frac{K}{A}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|},\end{aligned}$$

where $(*)$ follows from Lemma B.51 and the fact that $E_1^a(H) = \varnothing$ and $U = \mathcal{S}(H)$, and $(\dagger)$ follows from the fact that $|\tilde{\mathbf{Y}}^H_{\geqslant 2}| \leqslant 1$. Combining this with Lemma B.60, we get:

$$\begin{aligned}\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \cdot P^{wb}_{2,H}(\mathbf{x})\right| &\leqslant \frac{n^{\frac{2K}{A}}}{n^{\frac{1}{5}}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \prod_{uv \in E_{\geqslant 2}(H)}\left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right] \\ &\leqslant \frac{1}{n^{\frac{1}{6}}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \prod_{uv \in E_{\geqslant 2}(H)}\left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right],\end{aligned}$$

where the last inequality is true if $A > \max\{100K, 1\}$ and n is large enough. Therefore,

$$\begin{aligned}\mathbb{E}\left[\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \cdot P^{wb}_{2,H}(\mathbf{x})\right|\right] &\leqslant \frac{1}{n^{\frac{1}{6}}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \cdot \mathbb{E}\left[\prod_{uv \in E_{\geqslant 2}(H)}\left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right]\right] \\ &\overset{(\ddagger)}{=} \frac{1}{n^{\frac{1}{6}}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \prod_{uv \in E_{\geqslant 2}(H)}\left(\frac{d}{n}\right) = \frac{1}{n^{\frac{1}{6}}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},\end{aligned}$$

where $(\ddagger)$ follows from Lemma B.22 and the fact that $E_{\geqslant 2}(H)$ is a forest. □

B.3.1.6 Tight bounds on the significant part of the contribution of the well-behaved event in the case of pleasant multigraphs

In order to prove Lemma B.29, we need a few definitions and lemmas.

Definition B.61. For every $v \in V(H)$, let $\mathbf{D}^H_v$ be the number of edges in $\mathbf{G}$ between v and $V_b(H, \mathbf{x})$. We denote $(\mathbf{D}^H_v)_{v \in V(H)}$ as $\mathbf{D}^H$.

Note that if $\mathcal{E}_{wb,H}$ occurs, then[29]

$$d_{\mathbf{G}-G(H)}(v) = d^o_{\mathbf{G}-G(H)}(v) = \mathbf{D}^H_v.$$

Furthermore, $|V^+_b(H,\mathbf{x})| = \left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$ and $|V^-_b(H,\mathbf{x})| = \left\lceil \frac{n}{2} - n^{\frac{3}{4}} \right\rceil$, where

$$V^+_b(H,\mathbf{x}) = \left\{u \in V_b(H,\mathbf{x}) : \ \mathbf{x}_u = +1\right\},$$

and

$$V^-_b(H,\mathbf{x}) = \left\{u \in V_b(H,\mathbf{x}) : \ \mathbf{x}_u = -1\right\}.$$

It is easy to see that given $\mathbf{x}$ that is approximately balanced on $[n] \setminus V(H)$, the random variables $\{\mathbf{D}^H_v\}_{v\in V(H)}$ are conditionally mutually independent, and they are conditionally independent of $(\mathcal{E}_{H,g}, \mathcal{E}_{H,d})$.

Lemma B.62. *Let H be a multigraph with at most $st = sK\log n$ vertices. For every $v \in V(H)$, and every $\mathsf{d} = (\mathsf{d}_v)_{v\in V(H)} \in \mathbb{N}^{V(H)}$, we have:*

$$\mathbb{P}\left[\mathbf{D}^H = \mathsf{d}\middle|\mathbf{x}, \mathcal{E}_{wb,H}\right] = \mathbb{P}\left[\mathbf{D}^H = \mathsf{d}\middle|\mathcal{E}_{wb,H}\right] = \mathbb{P}\left[\mathbf{D}^H = \mathsf{d}\middle|\mathbf{x}, \mathcal{E}_{H,b}\right] = \mathbb{P}\left[\mathbf{D}^H = \mathsf{d}\middle|\mathcal{E}_{H,b}\right] = P_{wb}(\mathsf{d}),$$

where

$$P_{wb}(\mathsf{d}) = \prod_{v\in V(H)} P_{wb}(\mathsf{d}_v)$$

and

$$P_{wb}(\mathsf{d}_v) = \sum_{a=0}^{\mathsf{d}_v}\left[\binom{\left\lceil \frac{n}{2} - n^{\frac{3}{4}}\right\rceil}{a} \cdot \left[\left(1+\frac{\varepsilon}{2}\right)\frac{d}{n}\right]^a \cdot \left[1 - \left(1+\frac{\varepsilon}{2}\right)\frac{d}{n}\right]^{\left\lceil \frac{n}{2} - n^{\frac{3}{4}}\right\rceil - a}\right]$$
$$\times \left[\binom{\left\lceil \frac{n}{2} - n^{\frac{3}{4}}\right\rceil}{\mathsf{d}_v - a} \cdot \left[\left(1-\frac{\varepsilon}{2}\right)\frac{d}{n}\right]^{\mathsf{d}_v - a} \cdot \left[1 - \left(1-\frac{\varepsilon}{2}\right)\frac{d}{n}\right]^{\left\lceil \frac{n}{2} - n^{\frac{3}{4}}\right\rceil - \mathsf{d}_v + a}\right].$$

Furthermore, for every $v \in V(H)$, we have

$$\mathbb{P}\left[\mathbf{D}^H_v = \mathsf{d}_v\middle|\mathbf{x}, \mathcal{E}_{H,b}\right] = \mathbb{P}\left[\mathbf{D}^H_v = \mathsf{d}_v\middle|\mathcal{E}_{H,b}\right] = P_{wb}(\mathsf{d}_v).$$

In other words, $\mathbb{P}\left[\mathbf{D}^H = \mathsf{d}\middle|\mathbf{x}, \mathcal{E}_{wb,H}\right]$ depends only on $\mathsf{d} = (\mathsf{d}_v)_{v\in V(H)}$, i.e., it does not depend on $\mathbf{x}$, and it depends on $\mathbf{G}$ only through $\mathbf{D}^H = (\mathbf{D}^H_v)_{v\in V(H)}$. Furthermore, given $\mathcal{E}_{wb,H}$, the random variables $(\mathbf{D}^H_v)_{v\in V(H)}$ are conditionally mutually independent. The same is true if we condition on $\mathcal{E}_{H,b}$ instead of $\mathcal{E}_{wb,H}$.

Note that if $\mathsf{d}_v > 2\left\lceil \frac{n}{2} - n^{\frac{3}{4}}\right\rceil$, then $P_{wb}(\mathsf{d}_v) = 0$.

[29] Recall Definition B.8

Proof. We have

$$\begin{aligned}\mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] &= \mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{H,b} \cap \mathcal{E}_{H,g} \cap \mathcal{E}_{H,d}\right] \\ &\overset{(*)}{=} \mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] = \prod_{v \in V(H)} \mathbb{P}\left[\mathbf{D}_v^H = \mathsf{d}_v \middle| \mathbf{x}, \mathcal{E}_{H,b}\right],\end{aligned}$$

where $(*)$ follows from the fact that given $\mathbf{x}$ that is approximately balanced on $[n] \setminus V(H)$, the random variable $\mathbf{D}^H$ is conditionally independent from $(\mathcal{E}_{H,g}, \mathcal{E}_{H,d})$.

Now for every $v \in V(H)$, we have

$$\begin{aligned}&\mathbb{P}\left[\mathbf{D}_v^H = \mathsf{d}_v \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] \\ &= \sum_{\substack{U \subseteq V_b(H,\mathbf{x}): \\ |U| = \mathsf{d}_v}} \left[\prod_{u \in U} \mathbb{P}[uv \in \mathbf{G} | \mathbf{x}, \mathcal{E}_{H,b}]\right] \cdot \left[\prod_{u \in V_b(H,\mathbf{x}) \setminus U} \mathbb{P}[uv \notin \mathbf{G} | \mathbf{x}, \mathcal{E}_{H,b}]\right] \\ &\quad = \sum_{a=0}^{\mathsf{d}_v} \left(\sum_{\substack{U^+ \subseteq V_b^+(H,\mathbf{x}): \\ |U^+| = a}} \left[\prod_{u \in U^+} \mathbb{P}[uv \in \mathbf{G} | \mathbf{x}, \mathcal{E}_{H,b}]\right] \cdot \left[\prod_{u \in V_b^+(H,\mathbf{x}) \setminus U^+} \mathbb{P}[uv \notin \mathbf{G} | \mathbf{x}, \mathcal{E}_{H,b}]\right] \right) \\ &\qquad \times \left(\sum_{\substack{U^- \subseteq V_b^-(H,\mathbf{x}): \\ |U^-| = \mathsf{d}_v - a}} \left[\prod_{u \in U^-} \mathbb{P}[uv \in \mathbf{G} | \mathbf{x}, \mathcal{E}_{H,b}]\right] \cdot \left[\prod_{u \in V_b^-(H,\mathbf{x}) \setminus U^-} \mathbb{P}[uv \notin \mathbf{G} | \mathbf{x}, \mathcal{E}_{H,b}]\right] \right),\end{aligned}$$

hence,

$$\begin{aligned}\mathbb{P}\left[\mathbf{D}_v^H = \mathsf{d}_v \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] &= \sum_{a=0}^{\mathsf{d}_v} \left[\binom{\lceil \frac{n}{2} - n^{\frac{3}{4}} \rceil}{a} \left[\left(1 + \frac{\varepsilon \mathbf{x}_v}{2}\right)\frac{d}{n}\right]^a \cdot \left[1 - \left(1 + \frac{\varepsilon \mathbf{x}_v}{2}\right)\frac{d}{n}\right]^{\lceil \frac{n}{2} - n^{\frac{3}{4}} \rceil - a} \right] \\ &\qquad \times \left[\binom{\lceil \frac{n}{2} - n^{\frac{3}{4}} \rceil}{\mathsf{d}_v - a} \left[\left(1 - \frac{\varepsilon \mathbf{x}_v}{2}\right)\frac{d}{n}\right]^{\mathsf{d}_v - a} \cdot \left[1 - \left(1 - \frac{\varepsilon \mathbf{x}_v}{2}\right)\frac{d}{n}\right]^{\lceil \frac{n}{2} - n^{\frac{3}{4}} \rceil - \mathsf{d}_v + a} \right] \\ &= \sum_{a=0}^{\mathsf{d}_v} \left[\binom{\lceil \frac{n}{2} - n^{\frac{3}{4}} \rceil}{a} \left[\left(1 + \frac{\varepsilon}{2}\right)\frac{d}{n}\right]^a \cdot \left[1 - \left(1 + \frac{\varepsilon}{2}\right)\frac{d}{n}\right]^{\lceil \frac{n}{2} - n^{\frac{3}{4}} \rceil - a} \right] \\ &\qquad \times \left[\binom{\lceil \frac{n}{2} - n^{\frac{3}{4}} \rceil}{\mathsf{d}_v - a} \left[\left(1 - \frac{\varepsilon}{2}\right)\frac{d}{n}\right]^{\mathsf{d}_v - a} \cdot \left[1 - \left(1 - \frac{\varepsilon}{2}\right)\frac{d}{n}\right]^{\lceil \frac{n}{2} - n^{\frac{3}{4}} \rceil - \mathsf{d}_v + a} \right] \\ &= P_{wb}(\mathsf{d}_v) = \mathbb{P}\left[\mathbf{D}_v^H = \mathsf{d}_v \middle| \mathcal{E}_{H,b}\right],\end{aligned}$$

where the last equality follows from the fact that $\mathbb{P}\left[\mathbf{D}_v^H = \mathsf{d}_v \middle| \mathbf{x}, \mathcal{E}_{H,b}\right]$ does not depend on $\mathbf{x}$. We conclude that

$$\mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] = \prod_{v \in V(H)} P_{wb}(\mathsf{d}_v) = P_{wb}(\mathsf{d}).$$

Furthermore, since $\mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] = \mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{H,b}\right]$ does not depend on $\mathbf{x}$, we have

$$\mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] = \mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathcal{E}_{wb,H}\right] = \mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] = \mathbb{P}\left[\mathbf{D}^H = \mathsf{d} \middle| \mathcal{E}_{H,b}\right].$$

□

Lemma B.63. *Let H be a multigraph with at most $st = sK \log n$ vertices. We have*

$$\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] = \sum_{\mathsf{d} \in \mathbb{N}^{V(H)}} \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\right] \cdot P_{wb}(\mathsf{d}),$$

where $P_{wb}(\mathsf{d})$ is as in Lemma B.62.

Proof. We have:

$$\begin{aligned}\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] &= \sum_{\mathsf{d} \in \mathbb{N}^{V(H)}} \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\right] \cdot \mathbb{P}[\mathbf{D}^H = \mathsf{d} | \mathbf{x}, \mathcal{E}_{wb,H}] \\ &= \sum_{\mathsf{d} \in \mathbb{N}^{V(H)}} \mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\right] \cdot P_{wb}(\mathsf{d}).\end{aligned}$$

□

In the following, we will fix $\mathsf{d} \in \mathbb{N}^{V(H)}$ and focus on studying $\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\right]$.

Definition B.64. Let H be an arbitrary multigraph and let $\mathsf{d} \in \mathbb{N}^{V(H)}$. For every $v \in V(E_1(H))$, we define the d-*criticality* of v in H as $c^H_{\mathsf{d}}(v) = \Delta - d^H_{\geqslant 2}(v) - \mathsf{d}_v$. If $d^H_1(v) \leqslant c^H_{\mathsf{d}}(v)$, we say that v is d-*safe in* H. If $d^H_1(v) > c^H_{\mathsf{d}}(v)$, we say that v is d-*unsafe in* H.

An edge $uv \in E_1(H)$ is said to be d-*safe* if both u and v are d-safe. We denote the set of d-safe edges in $E_1(H)$ as S^H_{d}.

Remark B.65. If $\mathbf{D} = \mathsf{d}$ and $\mathcal{E}_{wb,H}$ occurs, then since all the edges of multiplicity at least 2 are present in $\mathbf{G}$, we can see that a vertex $v \in V(E_1(H))$ causes truncation if and only if

$$\left|\left\{e \in E_1(H) \cap \mathbf{G} : \ e \text{ is incident to } v\right\}\right| > c^H_{\mathsf{d}}(v).$$

As can be easily seen, the criticality of a vertex v is the maximum number of edges from $\left\{e \in E_1(H) : \ e \text{ is incident to } v\right\}$ which can be present in $\mathbf{G}$ without causing truncation.

If v is d-safe, then we are sure that v does not cause truncation. If uv is a d-safe vertex, then we are sure that its presence in $\mathbf{G}$ will not cause truncation.

Definition B.66. Let H be an arbitrary multigraph and let $\mathsf{d} \in \mathbb{N}^{V(H)}$. For every $\mathsf{E} \subseteq E_1(H)$ and every $\mathsf{d} \in \mathbb{N}^{V(H)}$, we say that E is d-*structurally safe* if for every $v \in V(E_1(H))$ we have $d_{\mathsf{E}}(v) \leqslant c^H_{\mathsf{d}}(v)$. For every $\mathsf{E} \subseteq E_1(H)$, define:

$$\mathcal{S}^H_{\mathsf{d}}(\mathsf{E}) = \big\{\mathsf{S} \subseteq \mathsf{E} : \ \mathsf{S} \text{ is } \mathsf{d}\text{-structurally safe}\big\}.$$

Remark B.67. Let H be an (s,t)-pleasant multigraph and let $H^{(1)}, \dots, H^{(r_H)}$ be the agreeable components of H. Since $H^{(1)}, \dots, H^{(r_H)}$ are vertex-disjoint. We can deduce from this that

$$\mathcal{S}^H_{\mathsf{d}}(E_1(H)) = \big\{\mathsf{S}_1 \cup \ldots \cup \mathsf{S}_{r_H} : \ \forall i \in [r_H],\ \mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}\big(E_1\big(H^{(i)}\big)\big)\big\},$$

and

$$\big\{\mathsf{E}_{\geqslant 2} : \mathsf{E}_{\geqslant 2} \subseteq E_{\geqslant 2}\big(H^{(*)}\big)\big\} = \Big\{\mathsf{E}^{(1)}_{\geqslant 2} \cup \ldots \cup \mathsf{E}^{(r_H)}_{\geqslant 2} : \ \forall i \in [r_H],\ \mathsf{E}^{(i)}_{\geqslant 2} \subseteq E_{\geqslant 2}\big(H^{(i)}\big)\Big\},$$

where $H^{(*)} = \bigcup_{i \in [r_H]} H^{(i)}$.

Lemma B.68. *Let H be an (s,t)-pleasant multigraph and let $H^{(1)}, \dots, H^{(r_H)}$ be its agreeable components*[30]*. For every $\mathsf{d} \in \mathbb{N}^{V(H)}$, we have*

$$\begin{aligned}\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \cdot P^{wb}_{1,H}(\mathbf{x})\Big] \\ = \left(\frac{\tilde{d}}{n}\right)^{|E_1(H)|} \left(1 - \frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot \mathbb{P}\big[\mathcal{E}_{H,b}\big] \cdot \prod_{i \in [r_H]} J_{H,\mathsf{d},i},\end{aligned}$$

where

$$\begin{aligned}J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} \sum_{\mathsf{S}'_i \subseteq \mathsf{S}_i} \sum_{\mathsf{S}''_i \subseteq E_1(H^{(i)}) \setminus \mathsf{S}_i} \sum_{\mathsf{E}^{(i)}_{\geqslant 2} \subseteq E_{\geqslant 2}(H^{(i)})} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}'_i| + |\mathsf{E}^{(i)}_{\geqslant 2}|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|\mathsf{S}''_i|} \\ \times (-1)^{|E_1(H^{(i)})| - |\mathsf{S}_i| - |\mathsf{S}''_i|} \cdot \mathbb{E}\left[\prod_{uv \in \mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v\right],\end{aligned} \tag{B.3.14}$$

$$\tilde{d} = d\left(1 - \frac{d}{n}\right), \qquad \text{and} \qquad \tilde{\varepsilon} = \frac{\varepsilon d^2}{2\tilde{d}}.$$

Proof. Let $\mathbf{x}$ be such that $\mathcal{E}_{H,b}$ occurs[31]. From Remark B.65 we can see that if $\mathcal{E}_{wb,H}$ occurs and $\mathbf{D}^H = \mathsf{d}$, then H is not truncated if and only if $E(\mathbf{G}) \cap E_1(H) \in \mathcal{S}^H_{\mathsf{d}}(E_1(H))$. Now using the fact that given $\mathbf{x}$, the random variables $(\mathbb{1}_{uv \in \mathbf{G}})_{uv \in E_1(H)}$ are conditionally independent

[30] See Definition B.21

[31] Note that $\mathcal{E}_{H,b}$ depends only on $(\mathbf{x}_u)_{u \in [n] \setminus V(H)}$, i.e., it is $\sigma\big(\{\mathbf{x}_u : \ u \in [n] \setminus V(H)\}\big)$-measurable.

from the events $\mathcal{E}_{wb,H}$ and $\{\mathbf{D}^H = \mathsf{d}\}$, and the fact that given $\mathcal{E}_{wb,H}$, we have $\mathbf{Y}_{uv} = 1 - \frac{d}{n}$ for every $uv \in E_{\geqslant 2}(H)$, we deduce that:

$$\begin{aligned}
&\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \\
&= \sum_{\mathsf{S} \in \mathcal{S}^H_{\mathsf{d}}(E_1(H))} \prod_{uv \in \mathsf{S}} \left[\left(1 - \frac{d}{n}\right) \cdot \mathbb{P}[uv \in \mathbf{G} | \mathbf{x}]\right] \cdot \prod_{uv \in E_1(H) \setminus \mathsf{S}} \left[\left(-\frac{d}{n}\right) \cdot \mathbb{P}[uv \notin \mathbf{G} | \mathbf{x}]\right] \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left(1 - \frac{d}{n}\right)^{m_H(uv)}.
\end{aligned} \tag{B.3.15}$$

Now for every $uv \in E_1(H)$, we have

$$\left(1 - \frac{d}{n}\right) \cdot \mathbb{P}[uv \in \mathbf{G} | \mathbf{x}] = \left(1 - \frac{d}{n}\right) \cdot \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \frac{d}{n} = \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \frac{\tilde{d}}{n}. \tag{B.3.16}$$

On the other hand,

$$\begin{aligned}
\left(-\frac{d}{n}\right) \cdot \mathbb{P}[e \notin \mathbf{G} | \mathbf{x}] &= \left(-\frac{d}{n}\right) \cdot \left[1 - \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \frac{d}{n}\right] = -\frac{d}{n}\left(1 - \frac{d}{n} - \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v d}{2n}\right) \\
&= -\frac{\tilde{d}}{n} + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v d^2}{2n^2} = -\frac{\tilde{d}}{n}\left(1 - \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v d^2}{2\tilde{d}n}\right) = -\frac{\tilde{d}}{n}\left(1 - \frac{\tilde{\varepsilon} \mathbf{x}_u \mathbf{x}_v}{n}\right).
\end{aligned} \tag{B.3.17}$$

By combining Eq. (B.1.23), Eq. (B.3.15), Eq. (B.3.16) and Eq. (B.3.16), we get

$$\begin{aligned}
&\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \cdot P^{wb}_{1,H}(\mathbf{x}) \\
&= \sum_{\mathsf{S} \in \mathcal{S}^H_{\mathsf{d}}(E_1(H))} \prod_{uv \in \mathsf{S}} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \frac{\tilde{d}}{n}\right] \cdot \prod_{uv \in E_1(H) \setminus \mathsf{S}} \left[-\frac{\tilde{d}}{n}\left(1 - \frac{\tilde{\varepsilon} \mathbf{x}_u \mathbf{x}_v}{n}\right)\right] \\
&\qquad \times \prod_{uv \in E_{\geqslant 2}(H)} \left[\left(1 - \frac{d}{n}\right)^{m_H(uv)} \cdot \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \frac{d}{n}\right] \cdot \mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x}).
\end{aligned}$$

Now for every $\mathsf{E} \subseteq E(H)$, define $m_H(\mathsf{E}) = \sum_{uv \in \mathsf{E}} m_H(uv)$. We have:

$$\begin{aligned}
&\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \cdot P^{wb}_{1,H}(\mathbf{x}) \\
&= \left(\frac{\tilde{d}}{n}\right)^{|E_1(H)|} \left(1 - \frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \\
&\quad \times \sum_{\mathsf{S} \in \mathcal{S}^H_{\mathsf{d}}(E_1(H))} \prod_{uv \in \mathsf{S}} \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \cdot \prod_{uv \in E_1(H) \setminus \mathsf{S}} \left(-1 + \frac{\tilde{\varepsilon} \mathbf{x}_u \mathbf{x}_v}{n}\right) \cdot \prod_{uv \in E_{\geqslant 2}(H)} \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \cdot \mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x}) \\
&= \left(\frac{\tilde{d}}{n}\right)^{|E_1(H)|} \left(1 - \frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \\
&\qquad \times \sum_{\mathsf{S} \in \mathcal{S}^H_{\mathsf{d}}(E_1(H))} \left[\sum_{\mathsf{S}' \subseteq \mathsf{S}} \prod_{uv \in \mathsf{S}'} \left(\frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\right] \cdot \left[\sum_{\mathsf{S}'' \subseteq E_1(H) \setminus \mathsf{S}} (-1)^{|E_1(H)| - |\mathsf{S}| - |\mathsf{S}''|} \prod_{uv \in \mathsf{S}''} \left(\frac{\tilde{\varepsilon} \mathbf{x}_u \mathbf{x}_v}{n}\right)\right]
\end{aligned}$$

$$\times \left[\sum_{\mathsf{E}_{\geqslant 2} \subseteq E_{\geqslant 2}(H)} \prod_{uv \in \mathsf{E}_{\geqslant 2}} \left(\frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2} \right) \right] \cdot \mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x})$$

$$= \left(\frac{\tilde{d}}{n}\right)^{|E_1(H)|} \left(1 - \frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}$$

$$\times \sum_{\mathsf{S} \in \mathcal{S}_{\mathsf{d}}^H(E_1(H))} \sum_{\mathsf{S}' \subseteq \mathsf{S}} \sum_{\mathsf{S}'' \subseteq E_1(H) \setminus \mathsf{S}} \sum_{\mathsf{E}_{\geqslant 2} \subseteq E_{\geqslant 2}(H)} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}'| + |\mathsf{E}_{\geqslant 2}|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|\mathsf{S}''|} \cdot (-1)^{|E_1(H)| - |\mathsf{S}| - |\mathsf{S}''|}$$

$$\times \mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x}) \cdot \prod_{uv \in \mathsf{S}' \cup \mathsf{S}'' \cup \mathsf{E}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v.$$

Now since $\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x})$ depends only on $(\mathbf{x}_u)_{u \in [n] \setminus V(H)}$, it is independent from $(\mathbf{x}_v)_{v \in V(H)}$. Therefore, for every $\mathsf{S} \in \mathcal{S}_{\mathsf{d}}^H(E_1(H))$, $\mathsf{S}' \subseteq \mathsf{S}$, $\mathsf{S}'' \subseteq E_1(H) \setminus \mathsf{S}$ and $\mathsf{E}_{\geqslant 2} \subseteq E_{\geqslant 2}(H)$, we have:

$$\mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x}) \cdot \prod_{uv \in \mathsf{S}' \cup \mathsf{S}'' \cup \mathsf{E}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v \right] = \mathbb{E}\left[\mathbb{1}_{\mathcal{E}_{H,b}}(\mathbf{x}) \right] \cdot \mathbb{E}\left[\prod_{uv \in \mathsf{S}' \cup \mathsf{S}'' \cup \mathsf{E}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v \right]$$

$$= \mathbb{P}\left[\mathcal{E}_{H,b} \right] \cdot \mathbb{E}\left[\prod_{uv \in \mathsf{S}' \cup \mathsf{S}'' \cup \mathsf{E}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v \right].$$

Notice that if $\mathsf{E}_{\geqslant 2}$ contains an edge outside $H^{(*)}$, then $\mathsf{S}' \cup \mathsf{S}'' \cup \mathsf{E}_{\geqslant 2}$ cannot be a union of edge-disjoint cycles[32] (see Definition B.21). Lemma B.22 now implies that if $\mathsf{E}_{\geqslant 2} \not\subseteq E_{\geqslant 2}(H^{(*)})$, then

$$\mathbb{E}\left[\prod_{uv \in \mathsf{S}' \cup \mathsf{S}'' \cup \mathsf{E}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v \right] = 0.$$

Therefore,

$$\mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d} \right] \cdot P_{1,H}^{wb}(\mathbf{x}) \right]$$

$$= \left(\frac{\tilde{d}}{n}\right)^{|E_1(H)|} \left(1 - \frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot \mathbb{P}\left[\mathcal{E}_{H,b} \right]$$

$$\times \sum_{\mathsf{S} \in \mathcal{S}_{\mathsf{d}}^H(E_1(H))} \sum_{\mathsf{S}' \subseteq \mathsf{S}} \sum_{\mathsf{S}'' \subseteq E_1(H) \setminus \mathsf{S}} \sum_{\mathsf{E}_{\geqslant 2} \subseteq E_{\geqslant 2}(H^{(*)})} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}'| + |\mathsf{E}_{\geqslant 2}|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|\mathsf{S}''|} \cdot (-1)^{|E_1(H)| - |\mathsf{S}| - |\mathsf{S}''|}$$

$$\times \mathbb{E}\left[\prod_{uv \in \mathsf{S}' \cup \mathsf{S}'' \cup \mathsf{E}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v \right]. \tag{B.3.18}$$

[32] Recall that $H^{(*)} = \bigcup_{i \in [r_H]} H^{(i)}$.

Now since H is (s,t)-pleasant, the agreeable components $H^{(1)},\dots,H^{(r_H)}$ are vertex disjoint. Combining this with Remark B.67, it is easy to see that we can rewrite Eq. (B.3.18) as follows:

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big]\\
&=\left(\frac{\tilde{d}}{n}\right)^{|E_1(H)|}\left(1-\frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\cdot\mathbb{P}\big[\mathcal{E}_{H,b}\big]\\
&\qquad\times\prod_{i\in[r_H]}\Bigg(\sum_{\mathsf{S}_i\in\mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))}\sum_{\mathsf{S}'_i\subseteq\mathsf{S}_i}\sum_{\mathsf{S}''_i\subseteq E_1(H^{(i)})\setminus\mathsf{S}_i}\sum_{\mathsf{E}^{(i)}_{\geqslant 2}\subseteq E_{\geqslant 2}(H^{(i)})}\left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}'_i|+|\mathsf{E}^{(i)}_{\geqslant 2}|}\cdot\left(\frac{\tilde{\varepsilon}}{n}\right)^{|\mathsf{S}''_i|}\\
&\qquad\qquad\times(-1)^{|E_1(H^{(i)})|-|\mathsf{S}_i|-|\mathsf{S}''_i|}\cdot\mathbb{E}\left[\prod_{uv\in\mathsf{S}'_i\cup\mathsf{S}''_i\cup\mathsf{E}^{(i)}_{\geqslant 2}}\mathbf{x}_u\mathbf{x}_v\right]\Bigg).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big]\\
&\qquad=\left(\frac{\tilde{d}}{n}\right)^{|E_1(H)|}\left(1-\frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\cdot\mathbb{P}\big[\mathcal{E}_{H,b}\big]\cdot\prod_{i\in[r_H]}J_{H,\mathsf{d},i}.
\end{aligned}$$

□

We now turn to study the value of $J_{H,\mathsf{d},i}$. The following lemma will be useful.

Lemma B.69. *Let $\mathsf{E}\subseteq E_1(H)$ and let $\{\mathsf{E}',\mathsf{E}''\}$ be a partition of E. Let*

$$f:\{\mathsf{S}'':\mathsf{S}''\subseteq\mathsf{E}''\}\to\mathbb{R}$$

be an arbitrary function on the subsets of E''. If E' contains a d-safe edge uv then

$$\sum_{\mathsf{S}\in\mathcal{S}^H_{\mathsf{d}}(\mathsf{E})}(-1)^{|\mathsf{E}'|-|\mathsf{S}\cap\mathsf{E}'|}\cdot f(\mathsf{S}\cap\mathsf{E}'')=0.$$

Proof. Since $uv\in\mathsf{E}$ is d-safe, we can write

$$\mathcal{S}^H_{\mathsf{d}}(\mathsf{E})=\mathcal{S}^H_{\mathsf{d}}(\mathsf{E}\setminus\{uv\})\cup\big\{\mathsf{S}\cup\{uv\}:\ \mathsf{S}\in\mathcal{S}^H_{\mathsf{d}}(\mathsf{E}\setminus\{uv\})\big\}.$$

Therefore,

$$\sum_{\mathsf{S}\in\mathcal{S}^H_{\mathsf{d}}(\mathsf{E})}(-1)^{|\mathsf{E}'|-|\mathsf{S}\cap\mathsf{E}'|}\cdot f(\mathsf{S}\cap\mathsf{E}'')$$

$$
\begin{aligned}
&= \sum_{\mathsf{S}\in\mathcal{S}_{\mathsf{d}}^{H}(\mathsf{E}\setminus\{uv\})} \left[(-1)^{|\mathsf{E}'|-|\mathsf{S}\cap\mathsf{E}'|}\cdot f(\mathsf{S}\cap\mathsf{E}'') + (-1)^{|\mathsf{E}'|-|(\mathsf{S}\cup\{uv\})\cap\mathsf{E}'|}\cdot f\big((\mathsf{S}\cup\{uv\})\cap\mathsf{E}''\big)\right] \\
&= \sum_{\mathsf{S}\in\mathcal{S}_{\mathsf{d}}^{H}(\mathsf{E}\setminus\{uv\})} \left[(-1)^{|\mathsf{E}'|-|\mathsf{S}\cap\mathsf{E}'|}\cdot f(\mathsf{S}\cap\mathsf{E}'') + (-1)^{|\mathsf{E}'|-|(\mathsf{S}\cap\mathsf{E}')\cup\{uv\}|}\cdot f\big(\mathsf{S}\cap\mathsf{E}''\big)\right] \\
&= \sum_{\mathsf{S}\in\mathcal{S}_{\mathsf{d}}^{H}(\mathsf{E}\setminus\{uv\})} (-1)^{|\mathsf{E}'|-|\mathsf{S}\cap\mathsf{E}'|}[1+(-1)]\cdot f\big(\mathsf{S}\cap\mathsf{E}''\big) = 0.
\end{aligned}
$$

□

Lemma B.70. *Let H be an (s,t)-pleasant multigraph. Let $H^{(i)}$ be a type-1 agreeable component of H and let $J_{H,\mathsf{d},i}$ be as in Lemma B.68. We have:*

- *If $E_1(H^{(i)})$ contains a d-safe edge, then*[33]

$$
J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i\in\mathcal{S}_{\mathsf{d}}^{H}(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|+|E'''_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|}.
$$

- *If $E_1(H^{(i)})$ does not contain a d-safe edge, then*

$$
|J_{H,\mathsf{d},i}| \leqslant 4\cdot 2^{|E_1(H^{(i)})|}.
$$

Proof. Since $H^{(i)}$ is a cycle, Lemma B.22 implies that

$$
\mathbb{E}\left[\prod_{uv\in\mathsf{S}'_i\cup\mathsf{S}''_i\cup\mathsf{E}^{(i)}_{\geqslant 2}} \mathbf{x}_u\mathbf{x}_v\right] \neq 0
$$

if and only if $\mathsf{S}'_i\cup\mathsf{S}''_i\cup\mathsf{E}^{(i)}_{\geqslant 2}\in\big\{\varnothing, E\big(H^{(i)}\big)\big\}$, in which case we have

$$
\mathbb{E}\left[\prod_{uv\in\mathsf{S}'_i\cup\mathsf{S}''_i\cup\mathsf{E}^{(i)}_{\geqslant 2}} \mathbf{x}_u\mathbf{x}_v\right] = 1.
$$

Notice the following:

- If $\mathsf{S}'_i\cup\mathsf{S}''_i\cup\mathsf{E}^{(i)}_{\geqslant 2}=\varnothing$, then $\mathsf{S}'_i=\mathsf{S}''_i=\mathsf{E}^{(i)}_{\geqslant 2}=\varnothing$, in which case we have

$$
\big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i| - |\mathsf{S}''_i| = \big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i|,
$$

and

$$
|\mathsf{S}'_i| + \big|\mathsf{E}^{(i)}_{\geqslant 2}\big| = |\mathsf{S}''_i| = 0.
$$

[33] Recall that for a type-1 agreeable component $H^{(i)}$, we have $E'''_{\geqslant 2}(H^{(i)}) = E_{\geqslant 2}(H^{(i)})$. See Definition B.28.

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = E(H^{(i)})$, then $\mathsf{S}'_i = \mathsf{S}_i$, $\mathsf{S}''_i = E_1\big(H^{(i)}\big) \setminus \mathsf{S}_i$, and $\mathsf{E}^{(i)}_{\geqslant 2} = E_{\geqslant 2}(H^{(i)})$, in which case we have

$$\big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i| - |\mathsf{S}''_i| = 0,$$

$$|\mathsf{S}'_i| + \big|\mathsf{E}^{(i)}_{\geqslant 2}\big| = |\mathsf{S}_i| + \big|E_{\geqslant 2}\big(H^{(i)}\big)\big|,$$

and

$$|\mathsf{S}''_i| = \big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i|.$$

Applying this to Eq. (B.3.14), we get

$$J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} \left[(-1)^{|E_1(H^{(i)})|-|\mathsf{S}_i|} + \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|+|E_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|}\right].$$

It follows from Lemma B.69 that if there exists a d-safe edge in $E_1\big(H^{(i)}\big)$, then

$$J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|+|E_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|}.$$

On the other hand, if $E_1\big(H^{(i)}\big)$ does not contain any d-safe edge, then

$$|J_{H,\mathsf{d},i}| \leqslant \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} (1+1) = 2 \cdot |\mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))| \leqslant 2 \cdot 2^{|E_1(H^{(i)})|} \leqslant 4 \cdot 2^{|E_1(H^{(i)})|}.$$

□

Lemma B.71. *Let H be an (s,t)-pleasant multigraph. Let $H^{(i)}$ be a type-2 agreeable component of H and let $J_{H,\mathsf{d},i}$ be as in Lemma B.68. We have:*

- *If each cycle*[34] *of $E(H^{(i)})$ contains a d-safe edge of multiplicity 1, then*[35]

$$J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|+|E'''_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|}.$$

- *If there exists at least one cycle of $E(H^{(i)})$ that does not contain any d-safe edge of multiplicity 1, then*

$$|J_{H,\mathsf{d},i}| \leqslant 4 \cdot 2^{|E_1(H^{(i)})|}.$$

[34] There are two cycles in $E(H^{(i)})$.

[35] Recall that for a type-2 agreeable component $H^{(i)}$, we have $E'''_{\geqslant 2}(H^{(i)}) = E_{\geqslant 2}(H^{(i)})$. See Definition B.28.

Proof. Let $E'\big(H^{(i)}\big)$ and $E''\big(H^{(i)}\big)$ be the two cycles of $H^{(i)}$ as in definition Definition B.20. Define

$$\begin{aligned} E'_1\big(H^{(i)}\big) &= E_1\big(H^{(i)}\big) \cap E'\big(H^{(i)}\big), \\ E''_1\big(H^{(i)}\big) &= E_1\big(H^{(i)}\big) \cap E''\big(H^{(i)}\big), \\ E'_{\geqslant 2}\big(H^{(i)}\big) &= E_{\geqslant 2}\big(H^{(i)}\big) \cap E'\big(H^{(i)}\big), \end{aligned}$$

and

$$E''_{\geqslant 2}\big(H^{(i)}\big) = E_{\geqslant 2}\big(H^{(i)}\big) \cap E''\big(H^{(i)}\big).$$

Since $E'(H^{(i)})$ and $E''(H^{(i)})$ are two edge-disjoint cycles that intersect only in one vertex, it follows from Lemma B.22 that

$$\mathbb{E}\left[\prod_{uv \in \mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v\right] \neq 0$$

if and only if $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} \in \big\{\varnothing, E'(H^{(i)}), E''(H^{(i)}), E(H^{(i)})\big\}$, in which case we have

$$\mathbb{E}\left[\prod_{uv \in \mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v\right] = 1.$$

Notice the following:

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = \varnothing$, then $\mathsf{S}'_i = \mathsf{S}''_i = \mathsf{E}^{(i)}_{\geqslant 2} = \varnothing$, in which case we have

  $$\begin{aligned} \big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i| - |\mathsf{S}''_i| &= \big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i| \\ &= \big|E'_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E'_1\big(H^{(i)}\big)\big| + \big|E''_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E''_1\big(H^{(i)}\big)\big|, \end{aligned}$$

  and

  $$|\mathsf{S}'_i| + \big|\mathsf{E}^{(i)}_{\geqslant 2}\big| = |\mathsf{S}''_i| = 0.$$

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = E'\big(H^{(i)}\big)$, then $\mathsf{S}'_i = \mathsf{S}_i \cap E'_1\big(H^{(i)}\big)$, $\mathsf{S}''_i = E'_1\big(H^{(i)}\big) \setminus \mathsf{S}_i$ and $\mathsf{E}^{(i)}_{\geqslant 2} = E'_{\geqslant 2}\big(H^{(i)}\big)$, in which case we have

  $$\begin{aligned} &\big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i| - |\mathsf{S}''_i| \\ &\qquad = \big|E'_1\big(H^{(i)}\big)\big| + \big|E''_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E'_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E''_1\big(H^{(i)}\big)\big| - \big|E'_1\big(H^{(i)}\big) \setminus \mathsf{S}_i\big| \\ &\qquad = \big|E''_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E''_1\big(H^{(i)}\big)\big|, \end{aligned}$$

  $$|\mathsf{S}'_i| + \big|\mathsf{E}^{(i)}_{\geqslant 2}\big| = \big|\mathsf{S}_i \cap E'_1\big(H^{(i)}\big)\big| + \big|E'_{\geqslant 2}\big(H^{(i)}\big)\big|,$$

  and

  $$|\mathsf{S}''_i| = \big|E'_1\big(H^{(i)}\big) \setminus \mathsf{S}_i\big| = \big|E'_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E'_1\big(H^{(i)}\big)\big|.$$

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = E''\big(H^{(i)}\big)$, then $\mathsf{S}'_i = \mathsf{S}_i \cap E''_1\big(H^{(i)}\big)$, $\mathsf{S}''_i = E''_1\big(H^{(i)}\big) \setminus \mathsf{S}_i$ and $\mathsf{E}^{(i)}_{\geqslant 2} = E''_{\geqslant 2}\big(H^{(i)}\big)$, in which case we have

$$\begin{aligned}
&\big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i| - |\mathsf{S}''_i| \\
&\qquad = \big|E'_1\big(H^{(i)}\big)\big| + \big|E''_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E'_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E''_1\big(H^{(i)}\big)\big| - \big|E''_1\big(H^{(i)}\big) \setminus \mathsf{S}_i\big| \\
&\qquad = \big|E'_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E'_1\big(H^{(i)}\big)\big|,
\end{aligned}$$

$$|\mathsf{S}'_i| + \big|\mathsf{E}^{(i)}_{\geqslant 2}\big| = \big|\mathsf{S}_i \cap E''_1\big(H^{(i)}\big)\big| + \big|E''_{\geqslant 2}\big(H^{(i)}\big)\big|,$$

and

$$|\mathsf{S}''_i| = \big|E''_1\big(H^{(i)}\big) \setminus \mathsf{S}_i\big| = \big|E''_1\big(H^{(i)}\big)\big| - \big|\mathsf{S}_i \cap E''_1\big(H^{(i)}\big)\big|.$$

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = E\big(H^{(i)}\big)$, then $\mathsf{S}'_i = \mathsf{S}_i$, $\mathsf{S}''_i = E_1\big(H^{(i)}\big) \setminus \mathsf{S}_i$ and $\mathsf{E}^{(i)}_{\geqslant 2} = E_{\geqslant 2}(H^{(i)})$, in which case we have

$$\big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i| - |\mathsf{S}''_i| = 0,$$

$$|\mathsf{S}'_i| + \big|\mathsf{E}^{(i)}_{\geqslant 2}\big| = |\mathsf{S}_i| + \big|E_{\geqslant 2}\big(H^{(i)}\big)\big|,$$

and

$$|\mathsf{S}''_i| = \big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i|.$$

Applying this to Eq. (B.3.14), we get

$$\begin{aligned}
J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} \Bigg[& (-1)^{|E'_1(H^{(i)})| - |\mathsf{S}_i \cap E'_1(H^{(i)})|} \cdot (-1)^{|E''_1(H^{(i)})| - |\mathsf{S}_i \cap E''_1(H^{(i)})|} \\
& + (-1)^{|E''_1(H^{(i)})| - |\mathsf{S}_i \cap E''_1(H^{(i)})|} \cdot \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i \cap E'_1(H^{(i)})| + |E'_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E'_1(H^{(i)})| - |\mathsf{S}_i \cap E'_1(H^{(i)})|} \\
& + (-1)^{|E'_1(H^{(i)})| - |\mathsf{S}_i \cap E'_1(H^{(i)})|} \cdot \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i \cap E''_1(H^{(i)})| + |E''_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E''_1(H^{(i)})| - |\mathsf{S}_i \cap E''_1(H^{(i)})|} \\
& \qquad + \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i| + |E_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})| - |\mathsf{S}_i|} \Bigg].
\end{aligned}$$

It follows from Lemma B.69 that if there exists a d-safe edge in $E'_1\big(H^{(i)}\big)$ and a d-safe edge in $E''_1\big(H^{(i)}\big)$, then

$$J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i| + |E_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})| - |\mathsf{S}_i|}.$$

On the other hand, if $E_1'(H^{(i)})$ does not contain any d-safe edge or $E_1''(H^{(i)})$ does not contain any d-safe edge, then

$$|J_{H,\mathsf{d},i}| \leqslant \sum_{\mathsf{S}_i \in \mathcal{S}_{\mathsf{d}}^H(E_1(H^{(i)}))} (1+1+1+1) = 4 \cdot |\mathcal{S}_{\mathsf{d}}^H(E_1(H^{(i)}))| \leqslant 4 \cdot 2^{|E_1(H^{(i)})|}.$$

□

Lemma B.72. *Let H be an (s,t)-pleasant multigraph. Let $H^{(i)}$ be a type-3 agreeable component of H and let $J_{H,\mathsf{d},i}$ be as in Lemma B.68. Let $E'(H^{(i)})$ and $E''(H^{(i)})$ be two cycles of $H^{(i)}$ as in definition Definition B.20, i.e., $E'(H^{(i)}) \cap E''(H^{(i)})$ is a simple path of edges of multiplicity at least 2. Define*

$$E'''(H^{(i)}) = \left(E'(H^{(i)}) \cup E''(H^{(i)})\right) \setminus \left(E'(H^{(i)}) \cap E''(H^{(i)})\right),$$

$$E'''_{\geqslant 2}(H^{(i)}) = E_{\geqslant 2}(H^{(i)}) \cap E'''(H^{(i)}) = E_{\geqslant 2}(H^{(i)}) \setminus \left(E'(H^{(i)}) \cap E''(H^{(i)})\right),$$

$$E_1'(H^{(i)}) = E_1(H^{(i)}) \cap E'(H^{(i)}),$$

and

$$E_1''(H^{(i)}) = E_1(H^{(i)}) \cap E''(H^{(i)}).$$

We have:

- *If there exists a d-safe edge in $E_1'(H^{(i)})$ and a d-safe edge in $E_1''(H^{(i)})$, then*

$$J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}_{\mathsf{d}}^H(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|+|E'''_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|}.$$

- *If $E_1'(H^{(i)})$ does not contain any d-safe edge or $E_1''(H^{(i)})$ does not contain any d-safe edge, then*

$$|J_{H,\mathsf{d},i}| \leqslant 4 \cdot 2^{|E_1(H^{(i)})|}.$$

Proof. Define

$$E'_{\geqslant 2}(H^{(i)}) = E_{\geqslant 2}(H^{(i)}) \cap E'(H^{(i)}),$$

and

$$E''_{\geqslant 2}(H^{(i)}) = E_{\geqslant 2}(H^{(i)}) \cap E''(H^{(i)}).$$

The only subsets of $E(H^{(i)})$ consisting of edge-disjoint unions of cycles are $\varnothing$, $E'(H^{(i)})$, $E''(H^{(i)})$ and $E'''(H^{(i)})$. It follows from Lemma B.22 that

$$\mathbb{E}\left[\prod_{uv \in \mathsf{S}_i' \cup \mathsf{S}_i'' \cup \mathsf{E}_{\geqslant 2}^{(i)}} \mathbf{x}_u \mathbf{x}_v\right] \neq 0$$

if and only if $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} \in \left\{\varnothing, E'(H^{(i)}), E''(H^{(i)}), E'''(H^{(i)})\right\}$, in which case we have

$$\mathbb{E}\left[\prod_{uv \in \mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2}} \mathbf{x}_u \mathbf{x}_v\right] = 1.$$

Notice the following:

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = \varnothing$, then $\mathsf{S}'_i = \mathsf{S}''_i = \mathsf{E}^{(i)}_{\geqslant 2} = \varnothing$, in which case we have

$$\begin{aligned}\left|E_1\left(H^{(i)}\right)\right| - |\mathsf{S}_i| - |\mathsf{S}''_i| &= \left|E_1\left(H^{(i)}\right)\right| - |\mathsf{S}_i| \\ &= \left|E'_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E'_1\left(H^{(i)}\right)\right| + \left|E''_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E''_1\left(H^{(i)}\right)\right|,\end{aligned}$$

  and

$$|\mathsf{S}'_i| + \left|\mathsf{E}^{(i)}_{\geqslant 2}\right| = |\mathsf{S}''_i| = 0.$$

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = E'\left(H^{(i)}\right)$, then $\mathsf{S}'_i = \mathsf{S}_i \cap E'_1\left(H^{(i)}\right)$, $\mathsf{S}''_i = E'_1\left(H^{(i)}\right) \setminus \mathsf{S}_i$ and $\mathsf{E}^{(i)}_{\geqslant 2} = E'_{\geqslant 2}\left(H^{(i)}\right)$, in which case we have

$$\begin{aligned}&\left|E_1\left(H^{(i)}\right)\right| - |\mathsf{S}_i| - |\mathsf{S}''_i| \\ &\quad= \left|E'_1\left(H^{(i)}\right)\right| + \left|E''_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E'_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E''_1\left(H^{(i)}\right)\right| - \left|E'_1\left(H^{(i)}\right) \setminus \mathsf{S}_i\right| \\ &\quad= \left|E''_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E''_1\left(H^{(i)}\right)\right|,\end{aligned}$$

$$|\mathsf{S}'_i| + \left|\mathsf{E}^{(i)}_{\geqslant 2}\right| = \left|\mathsf{S}_i \cap E'_1\left(H^{(i)}\right)\right| + \left|E'_{\geqslant 2}\left(H^{(i)}\right)\right|,$$

  and

$$|\mathsf{S}''_i| = \left|E'_1\left(H^{(i)}\right) \setminus \mathsf{S}_i\right| = \left|E'_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E'_1\left(H^{(i)}\right)\right|.$$

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = E''\left(H^{(i)}\right)$, then $\mathsf{S}'_i = \mathsf{S}_i \cap E''_1\left(H^{(i)}\right)$, $\mathsf{S}''_i = E''_1\left(H^{(i)}\right) \setminus \mathsf{S}_i$ and $\mathsf{E}^{(i)}_{\geqslant 2} = E''_{\geqslant 2}\left(H^{(i)}\right)$, in which case we have

$$\begin{aligned}&\left|E_1\left(H^{(i)}\right)\right| - |\mathsf{S}_i| - |\mathsf{S}''_i| \\ &\quad= \left|E'_1\left(H^{(i)}\right)\right| + \left|E''_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E'_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E''_1\left(H^{(i)}\right)\right| - \left|E''_1\left(H^{(i)}\right) \setminus \mathsf{S}_i\right| \\ &\quad= \left|E'_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E'_1\left(H^{(i)}\right)\right|,\end{aligned}$$

$$|\mathsf{S}'_i| + \left|\mathsf{E}^{(i)}_{\geqslant 2}\right| = \left|\mathsf{S}_i \cap E''_1\left(H^{(i)}\right)\right| + \left|E''_{\geqslant 2}\left(H^{(i)}\right)\right|,$$

  and

$$|\mathsf{S}''_i| = \left|E''_1\left(H^{(i)}\right) \setminus \mathsf{S}_i\right| = \left|E''_1\left(H^{(i)}\right)\right| - \left|\mathsf{S}_i \cap E''_1\left(H^{(i)}\right)\right|.$$

- If $\mathsf{S}'_i \cup \mathsf{S}''_i \cup \mathsf{E}^{(i)}_{\geqslant 2} = E'''\big(H^{(i)}\big)$, then $\mathsf{S}'_i = \mathsf{S}_i$, $\mathsf{S}''_i = E_1\big(H^{(i)}\big) \setminus \mathsf{S}_i$ and $\mathsf{E}^{(i)}_{\geqslant 2} = E'''_{\geqslant 2}(H^{(i)})$, in which case we have

$$\big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i| - |\mathsf{S}''_i| = 0,$$

$$|\mathsf{S}'_i| + \big|\mathsf{E}^{(i)}_{\geqslant 2}\big| = |\mathsf{S}_i| + \big|E'''_{\geqslant 2}\big(H^{(i)}\big)\big|,$$

and

$$|\mathsf{S}''_i| = \big|E_1\big(H^{(i)}\big)\big| - |\mathsf{S}_i|.$$

Applying this to Eq. (B.3.14), we get

$$\begin{aligned} J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} \Bigg[& (-1)^{|E'_1(H^{(i)})| - |\mathsf{S}_i \cap E'_1(H^{(i)})|} \cdot (-1)^{|E''_1(H^{(i)})| - |\mathsf{S}_i \cap E''_1(H^{(i)})|} \\ & + (-1)^{|E''_1(H^{(i)})| - |\mathsf{S}_i \cap E''_1(H^{(i)})|} \cdot \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i \cap E'_1(H^{(i)})| + |E'_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\widetilde{\varepsilon}}{n}\right)^{|E'_1(H^{(i)})| - |\mathsf{S}_i \cap E'_1(H^{(i)})|} \\ & + (-1)^{|E'_1(H^{(i)})| - |\mathsf{S}_i \cap E'_1(H^{(i)})|} \cdot \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i \cap E''_1(H^{(i)})| + |E''_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\widetilde{\varepsilon}}{n}\right)^{|E''_1(H^{(i)})| - |\mathsf{S}_i \cap E''_1(H^{(i)})|} \\ & + \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i| + |E'''_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\widetilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})| - |\mathsf{S}_i|} \Bigg]. \end{aligned}$$

It follows from Lemma B.69 that if there exists a d-safe edge in $E'_1\big(H^{(i)}\big)$ and a d-safe edge in $E''_1\big(H^{(i)}\big)$, then

$$J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i| + |E'''_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\widetilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})| - |\mathsf{S}_i|}.$$

On the other hand, if $E'_1\big(H^{(i)}\big)$ does not contain any d-safe edge or $E''_1\big(H^{(i)}\big)$ does not contain any d-safe edge, then

$$|J_{H,\mathsf{d},i}| \leqslant \sum_{\mathsf{S}_i \in \mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))} (1+1+1+1) = 4 \cdot |\mathcal{S}^H_{\mathsf{d}}(E_1(H^{(i)}))| \leqslant 4 \cdot 2^{|E_1(H^{(i)})|}.$$

□

Lemma B.73. *Let H be an (s,t)-pleasant multigraph. Let $H^{(i)}$ be an agreeable component of H and let $J_{H,\mathsf{d},i}$ be as in Lemma B.68. We have:*

- *If every cycle of $E(H^{(i)})$ contains a* d*-safe edge of multiplicity 1, then*

$$J_{H,\mathsf{d},i} = \sum_{\mathsf{S}_i \in \mathcal{S}_{\mathsf{d}}^H(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|+|E'''_{\geqslant 2}(H^{(i)})|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|}.$$

- *If there exists at least one cycle of $E(H^{(i)})$ that does not contain any* d*-safe edge of multiplicity 1, then*

$$|J_{H,\mathsf{d},i}| \leqslant 4 \cdot 2^{|E_1(H^{(i)})|}.$$

Proof. This is a direct corollary of Lemma B.70, Lemma B.71 and Lemma B.72. □

Lemma B.74. *Let H be an (s,t)-pleasant multigraph and let $H^{(i)}$ be an agreeable component of H. We have:*

- *If all the edges of $E_1\big(H^{(i)}\big)$ are* d*-safe, then*

$$\sum_{\mathsf{S}_i \in \mathcal{S}_{\mathsf{d}}^H(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|} = \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|}.$$

- *If there exists an edge in $E_1\big(H^{(i)}\big)$ that is not* d*-safe, then*

$$\sum_{\mathsf{S}_i \in \mathcal{S}_{\mathsf{d}}^H(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|} \leqslant \frac{1}{\sqrt{n}} \cdot \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|}.$$

Proof. Notice that

$$\frac{\varepsilon}{2} + \frac{\tilde{\varepsilon}}{n} = \frac{\varepsilon}{2} + \frac{\varepsilon d^2}{2\tilde{d}n} = \frac{\varepsilon}{2}\left(1 + \frac{d^2}{d\big(1-\frac{d}{n}\big)n}\right) = \frac{\varepsilon}{2}\left(1 + \frac{d}{n-d}\right) = \frac{\varepsilon}{2} \cdot \frac{n}{n-d} = \frac{\varepsilon d}{2\tilde{d}}. \tag{B.3.19}$$

- If all the edges of $E_1\big(H^{(i)}\big)$ are d-safe, then all subsets of $E_1\big(H^{(i)}\big)$ are d-structurally safe, hence

$$\begin{aligned}\sum_{\mathsf{S}_i \in \mathcal{S}_{\mathsf{d}}^H(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|} &= \sum_{\mathsf{S}_i \subseteq E_1(H^{(i)})} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|} \\ &= \left(\frac{\varepsilon}{2} + \frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|} = \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|}.\end{aligned}$$

- If there exists an edge in $E_1\big(H^{(i)}\big)$ that is not d-safe, then $E_1\big(H^{(i)}\big)$ is not d-structurally safe, hence

$$\begin{aligned}
&\sum_{\mathsf{S}_i \in \mathcal{S}_{\mathsf{d}}^H(E_1(H^{(i)}))} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|} \\
&\qquad \leqslant \sum_{\mathsf{S}_i \subsetneq E_1(H^{(i)})} \left(\frac{\varepsilon}{2}\right)^{|\mathsf{S}_i|} \cdot \left(\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|-|\mathsf{S}_i|} = \left(\frac{\varepsilon}{2}+\frac{\tilde{\varepsilon}}{n}\right)^{|E_1(H^{(i)})|} - \left(\frac{\varepsilon}{2}\right)^{|E_1(H^{(i)})|} \\
&\qquad \overset{(*)}{=} \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|} - \left(\frac{\varepsilon}{2}\right)^{|E_1(H^{(i)})|} = \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|} \left(1 - \left(\frac{\tilde{d}}{d}\right)^{|E_1(H^{(i)})|}\right) \\
&\qquad = \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|} \left(1 - \left(1-\frac{d}{n}\right)^{|E_1(H^{(i)})|}\right) \leqslant \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|} \left(1 - \left(1-\frac{d}{n}\right)^{st}\right) \\
&\qquad = \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|} \left(1 - \left(1 - O\left(\frac{std}{n}\right)\right)\right) \overset{(\dagger)}{\leqslant} \frac{1}{\sqrt{n}} \cdot \left(\frac{\varepsilon d}{2\tilde{d}}\right)^{|E_1(H^{(i)})|},
\end{aligned}$$

  where $(*)$ follows from Eq. (B.3.19) and $(\dagger)$ is true for n large enough.

□

Lemma B.75. *Let H be an (s,t)-pleasant multigraph, where $t = K\log n$. Let $E'''_{\geqslant 2}(H)$ be as in Definition B.28. We have the following:*

- *If $S_{\mathsf{d}}^H = E_1(H)$, i.e., if all the edges of $E_1(H)$ are d-safe, then*

$$\begin{aligned}
&\left(1-\frac{1}{\sqrt{n}}\right) \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \\
&\quad \leqslant \mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \cdot P_{1,H}^{wb}(\mathbf{x})\Big] \leqslant \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}.
\end{aligned}$$

- *If $S_{\mathsf{d}}^H \neq E_1(H)$ and every cycle in $E(H)$ contains a d-edge of multiplicity 1, then for n large enough, we have*

$$0 \leqslant \mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \cdot P_{1,H}^{wb}(\mathbf{x})\Big] \leqslant \frac{1}{\sqrt{n}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}.$$

- *If there is a cycle in $E(H)$ that does not contain any d-safe edge of multiplicity 1, then for n large enough, we have*

$$\left|\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H \big| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \cdot P_{1,H}^{wb}(\mathbf{x})\Big]\right| \leqslant \left(\frac{16}{\varepsilon}\right)^{st} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}.$$

Proof. From Lemma B.68 we have

$$\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big] = \left(\frac{\tilde{d}}{n}\right)^{|E_1(H)|}\left(1-\frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|}\cdot\mathbb{P}\big[\mathcal{E}_{H,b}\big]\cdot\prod_{i\in[r_H]}J_{H,\mathsf{d},i},$$

We will provide tight bounds on $\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big]$. In order to do this, we will distinguish between three cases:

- If $S^H_{\mathsf{d}}=E_1(H)$, then for every $i\in[r_H]$, all the edges in $E_1\big(H^{(i)}\big)$ are d-safe. Now from Lemma B.68, Lemma B.73 and Lemma B.74, and from the fact that $|E_1(H)|=\sum_{i\in[r_H]}\big|E_1\big(H^{(i)}\big)\big|$ and $|E'''_{\geqslant 2}(H)|=\sum_{i\in[r_H]}\big|E'''_{\geqslant 2}\big(H^{(i)}\big)\big|$, we can deduce that

$$\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big] = \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(1-\frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}\cdot\mathbb{P}\big[\mathcal{E}_{H,b}\big].$$

  Now from Lemma B.53 we have

$$\left(1-\frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))}\cdot\mathbb{P}\big[\mathcal{E}_{H,b}\big]\geqslant\left(1-\frac{d}{n}\right)^{st}\cdot\left(1-2e^{\frac{9}{8}\sqrt{n}}\right) \geqslant\left(1-O\left(\frac{dst}{n}\right)\right)\cdot\left(1-2e^{\frac{9}{8}\sqrt{n}}\right)\geqslant 1-\frac{1}{\sqrt{n}},$$

  where the last inequality is true for n large enough. Therefore, if $S^H_{\mathsf{d}}=E_1(H)$, we have

$$\left(1-\frac{1}{\sqrt{n}}\right)\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}$$
$$\leqslant\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big]\leqslant\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}.$$

- If $S^H_{\mathsf{d}}\neq E_1(H)$ and every cycle in $E(H)$ has at least one d-safe edge of multiplicity 1, define

$$\mathcal{I}^{H,\mathsf{d}}_{E\not\subseteq S}=\big\{i\in[r_H]:\ E_1\big(H^{(i)}\big)\not\subseteq S^H_{\mathsf{d}}\big\}.$$

  Since $S^H_{\mathsf{d}}\neq E_1(H)$, then we must have $\mathcal{I}^{H,\mathsf{d}}_{E\not\subseteq S}\neq\varnothing$. Now from Lemma B.68, Lemma B.73, Lemma B.74, and from the fact that $|E_1(H)|=\sum_{i\in[r_H]}\big|E_1\big(H^{(i)}\big)\big|$ and $|E'''_{\geqslant 2}(H)|=\sum_{i\in[r_H]}\big|E'''_{\geqslant 2}\big(H^{(i)}\big)\big|$, we can deduce that

$$
\begin{aligned}
0 &\leqslant \mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \cdot P^{wb}_{1,H}(\mathbf{x})\Big] \\
&\leqslant \left(\frac{1}{\sqrt{n}}\right)^{|\mathcal{I}^{H,\mathsf{d}}_{E\not\subseteq S}|} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(1-\frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \cdot \mathbb{P}\big[\mathcal{E}_{H,b}\big] \\
&\leqslant \frac{1}{\sqrt{n}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}.
\end{aligned}
$$

- If there is a cycle in $E(H)$ that does not contain any d-safe edge of multiplicity 1, define

$$
\mathcal{I}^H_{\mathsf{d}} = \Big\{ i \in [r_H] : \ H^{(i)} \text{ contains a cycle that does not contain any } \mathsf{d}\text{-safe edge of multiplicity } 1 \Big\}.
$$

From Lemma B.68, Lemma B.73, Lemma B.74, from the fact that $|E_1(H)| = \sum_{i\in[r_H]} \big|E_1\big(H^{(i)}\big)\big|$, and from the definition of $\mathcal{I}^H_{\mathsf{d}}$, we can deduce that

$$
\begin{aligned}
&\Big|\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\big] \cdot P^{wb}_{1,H}(\mathbf{x})\Big]\Big| \\
&\leqslant \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(1-\frac{d}{n}\right)^{m_H(E_{\geqslant 2}(H))} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot \mathbb{P}\big[\mathcal{E}_{H,b}\big] \cdot \prod_{i\in\mathcal{I}^H_{\mathsf{d}}} \left[4 \cdot \left(\frac{4}{\varepsilon}\right)^{|E_1(H^{(i)})|}\right] \\
&\leqslant \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot \prod_{i\in\mathcal{I}^H_{\mathsf{d}}} \left[\left(\frac{16}{\varepsilon}\right)^{|E_1(H^{(i)})|}\right] \leqslant \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|} \cdot \left(\frac{16}{\varepsilon}\right)^{|E_1(H)|} \\
&\leqslant \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \cdot \left(\frac{16}{\varepsilon}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \\
&\leqslant \left(\frac{16}{\varepsilon}\right)^{st} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}.
\end{aligned}
$$

□

Lemma B.76. *Let H be an arbitrary multigraph with at most $st = sK\log n$ vertices. If $A \geqslant 1$, then for every $v \in \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$, we have*

$$
\mathbb{P}\big[\mathbf{D}^H_v + d^H_1(v) + d^H_{\geqslant 2}(v) > \Delta\big|\mathbf{x}, \mathcal{E}_{wb,H}\big] \leqslant \frac{1}{2} \cdot \left(\frac{2\varepsilon}{33}\right)^{4A^2 s}.
$$

Proof. We have

$$
\mathbb{P}\big[\mathbf{D}^H_v + d^H_1(v) + d^H_{\geqslant 2}(v) > \Delta\big|\mathbf{x}, \mathcal{E}_{wb,H}\big]
$$

$$
\begin{aligned}
&\leqslant \mathbb{P}\left[\mathbf{D}_v^H + \tau + \frac{\Delta}{4} > \Delta \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \leqslant \mathbb{P}\left[\mathbf{D}_v^H > \frac{\Delta}{2} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \\
&= \mathbb{P}\left[\mathbf{D}_v^H > \frac{\Delta}{2} \middle| \mathbf{x}, \mathcal{E}_{H,b} \cap \mathcal{E}_{H,g} \cap \mathcal{E}_{H,d}\right] \overset{(*)}{=} \mathbb{P}\left[\mathbf{D}_v^H > \frac{\Delta}{2} \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] \\
&\overset{(\dagger)}{\leqslant} \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) > \frac{\Delta}{2} \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] \overset{(\ddagger)}{\leqslant} \frac{1}{2As \cdot 2^{6A\tau s^2}} \left(\frac{\varepsilon}{6}\right)^{4s^2\tau^2} \overset{(\wr)}{\leqslant} \frac{1}{2 \cdot 2^{6A^2 s}} \left(\frac{\varepsilon}{6}\right)^{4A^2 s} \\
&= \frac{1}{2} \cdot \left(\frac{\varepsilon}{6 \cdot 2^{3/2}}\right)^{4A^2 s} \leqslant \frac{1}{2} \cdot \left(\frac{2\varepsilon}{33}\right)^{4A^2 s},
\end{aligned}
$$

where $(*)$ follows from the fact that given $\mathbf{x}$ and $\mathcal{E}_{H,b}$, the random variable $\mathbf{D}_v^H$ is conditionally independent from $(\mathcal{E}_{H,g}, \mathcal{E}_{H,d})$, $(\dagger)$ follows from the fact that $\mathbf{D}_v^H \leqslant d^o_{\mathbf{G}-G(H)}(v)$, and $(\ddagger)$ follows from Lemma B.9 and the fact that $\mathcal{E}_{H,b}$ is $\sigma(\mathbf{x})$-measurable. $(\wr)$ follows from the fact that $\tau = As \log \frac{6}{\varepsilon} \geqslant A \geqslant 1$ and the fact that $s \geqslant 1$. □

Lemma B.77. *Let H be an (s,t)-pleasant multigraph where $t = K \log n$ vertices. Define the set*

$$
\mathsf{D}_{cycles}^{H,safe} = \left\{\mathsf{d} \in \mathbb{N}^{V(H)} : \textit{ Every cycle of } H \textit{ has at least one } \mathsf{d}\textit{-safe edge of multiplicity } 1\right\}.
$$

We have:

$$
\mathbb{P}\left[\mathbf{D}^H \notin \mathsf{D}_{cycles}^{H,safe} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \leqslant 3st \cdot \left(\frac{2\varepsilon}{33}\right)^{Ast}.
$$

Proof. Let C be an arbitrary cycle of H. Since C is a cycle, it is easy to see that there is a subset $C' \subset E_1(C)$ of size at least

$$
|C'| \geqslant \left\lfloor \frac{|E_1(C)|}{2} \right\rfloor \geqslant \frac{|E_1(C)|}{2} - 1,
$$

in such a way that no two edges in C' are incident to each other. Let

$$
\begin{aligned}
C'' &= \left\{uv \in C' : \ d^H_{\geqslant 2}(u) \leqslant \frac{\Delta}{4} \text{ and } d^H_{\geqslant 2}(v) \leqslant \frac{\Delta}{4}\right\} \\
&= \left\{e \in C' : \ e \text{ is not incident to any vertex in } \mathcal{I}_{\geqslant 2}(H)\right\}.
\end{aligned}
$$

Since there are no edges in C' that are incident to each other, it is easy to see that every vertex in $\mathcal{I}_{\geqslant 2}(H)$ is incident to at most one vertex in $E_1(C)$. Therefore, $|C''| \geqslant |C'| - |\mathcal{I}_{\geqslant 2}(H)|$. Now from Eq. (B.3.3) and Eq. (B.1.1) we have

$$
|\mathcal{I}_{\geqslant 2}(H)| \leqslant \frac{8st}{\Delta} \leqslant \frac{8st}{40Asd} = \frac{t}{5Ad} \leqslant \frac{t}{5A}.
$$

Hence,

$$
|C''| \geqslant |C'| - |\mathcal{I}_{\geqslant 2}(H)| \geqslant \frac{|E_1(C)|}{2} - 1 - \frac{t}{5A}.
$$

Recall that every cycle of an (s,t) pleasant multigraph contains at least $\frac{t}{A}$ edges of multiplicity 1, hence $|E_1(C)| \geqslant \frac{t}{A}$. Therefore,

$$|C''| \geqslant \frac{t}{2A} - 1 - \frac{t}{5A} \geqslant \frac{t}{4A}. \tag{B.3.20}$$

If C does not contain any $\mathbf{D}^H$-safe of multiplicity 1, then for every $uv \in C'' \subset E_1(C)$, either u is $\mathbf{D}^H$-unsafe or v is unsafe, i.e., we must have $\mathbf{D}^H_u + d^H_{\geqslant 2}(u) > \Delta - 2$ or $\mathbf{D}^H_v + d^H_{\geqslant 2}(v) > \Delta - 2$. Therefore,

$$\begin{aligned}
&\mathbb{P}\Big[C \text{ does not contain any } \mathbf{D}^H\text{-safe of multiplicity } 1 \Big| \mathbf{x}, \mathcal{E}_{wb,H}\Big] \\
&\leqslant \mathbb{P}\Big[\big\{\forall uv \in C'', \mathbf{D}^H_u + d^H_{\geqslant 2}(u) > \Delta - 2 \text{ or } \mathbf{D}^H_v + d^H_{\geqslant 2}(v) > \Delta - 2\big\}\Big| \mathbf{x}, \mathcal{E}_{wb,H}\Big] \\
&\overset{(*)}{=} \prod_{uv \in C''} \mathbb{P}\Big[\big\{\mathbf{D}^H_u + d^H_{\geqslant 2}(u) > \Delta - 2 \text{ or } \mathbf{D}^H_v + d^H_{\geqslant 2}(v) > \Delta - 2\big\}\Big| \mathbf{x}, \mathcal{E}_{wb,H}\Big] \\
&\leqslant \prod_{uv \in C''} \Big(\mathbb{P}\Big[\mathbf{D}^H_u + d^H_1(u) + d^H_{\geqslant 2}(u) > \Delta \Big| \mathbf{x}, \mathcal{E}_{wb,H}\Big] + \mathbb{P}\Big[\mathbf{D}^H_v + d^H_1(v) + d^H_{\geqslant 2}(v) > \Delta \Big| \mathbf{x}, \mathcal{E}_{wb,H}\Big]\Big) \\
&\overset{(\dagger)}{=} \prod_{uv \in C''} \left(\frac{1}{2} \cdot \left(\frac{2\varepsilon}{33}\right)^{4A^2 s} + \frac{1}{2} \cdot \left(\frac{2\varepsilon}{33}\right)^{4A^2 s}\right) = \left(\frac{2\varepsilon}{33}\right)^{4A^2 s \cdot |C''|} \overset{(\star)}{\leqslant} \left(\frac{2\varepsilon}{33}\right)^{4A^2 s \cdot \frac{t}{4A}} = \left(\frac{2\varepsilon}{33}\right)^{Ast},
\end{aligned}$$

where $(*)$ follows from the fact that given $\mathbf{x}$ and $\mathcal{E}_{wb,H}$, the random variables $\left(\mathbf{D}^H_v\right)_{v \in V(H)}$ are conditionally mutually independent and from the fact that for every two different edges $u_1v_1, u_2v_2 \in C''$, we have[36] $\{u_1, v_1\} \cap \{u_2, v_2\} = \varnothing$. $(\dagger)$ follows from Lemma B.76. $(\star)$ follows from Eq. (B.3.20).

Now since H is (s,t)-pleasant, there are r_H agreeable components of H. Furthermore, each agreeable component contains 1, 2 or 3 cycles depending on whether the agreeable component is of type 1, 2 or 3, respectively. Since the cycles of H are exactly those of its agreeable components, we conclude that there are at most $3r_H \leqslant 3st$ cycles in H. We conclude that

$$\mathbb{P}\left[\mathbf{D}^H \notin \mathsf{D}^{H,\text{safe}}_{\text{cycles}} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right] \leqslant 3st \cdot \left(\frac{2\varepsilon}{33}\right)^{Ast}.$$

□

Lemma B.78. *Let H be an (s,t)-pleasant multigraph where $t = K \log n$. If $A > \max\{1, \frac{100}{K}\}$ and n is large enough, then*

$$\sum_{\substack{\mathsf{d} \in \mathbb{N}^{V(H)}: \\ \mathsf{d} \notin \mathsf{D}^{H,safe}_{cycles}}} \left|\mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}_{wb,H}, \mathbf{D}^H = \mathsf{d}\right] \cdot P^{wb}_{1,H}(\mathbf{x})\right]\right| \cdot P_{wb}(\mathsf{d}) \leqslant \frac{1}{\sqrt{n}} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|} \left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|},$$

where $P_{wb}(\mathsf{d}) = \mathbb{P}\left[D^H = \mathsf{d} \middle| \mathbf{x}, \mathcal{E}_{wb,H}\right]$ is as in Lemma B.62.

[36] This follows from the fact that there are no edges in C'' that are incident to each other.

Proof. From Lemma B.75, we have

$$\begin{aligned}
&\sum_{\substack{\mathsf{d}\in\mathbb{N}^{V(H)}:\\ \mathsf{d}\notin \mathsf{D}^{H,\mathrm{safe}}_{\mathrm{cycles}}}} \left|\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big]\right|\cdot P_{wb}(\mathsf{d})\\
&\qquad\leqslant \sum_{\substack{\mathsf{d}\in\mathbb{N}^{V(H)}:\\ \mathsf{d}\notin \mathsf{D}^{H,\mathrm{safe}}_{\mathrm{cycles}}}} \left(\frac{16}{\varepsilon}\right)^{st}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}\cdot P_{wb}(\mathsf{d})\\
&\qquad= \sum_{\substack{\mathsf{d}\in\mathbb{N}^{V(H)}:\\ \mathsf{d}\notin \mathsf{D}^{H,\mathrm{safe}}_{\mathrm{cycles}}}} \left(\frac{16}{\varepsilon}\right)^{st}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}\mathbb{P}\big[D^H=\mathsf{d}\big|\mathbf{x},\mathcal{E}_{wb,H}\big]\\
&\qquad= \left(\frac{16}{\varepsilon}\right)^{st}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}\cdot\mathbb{P}\Big[\mathbf{D}^H\notin \mathsf{D}^{H,\mathrm{safe}}_{\mathrm{cycles}}\Big|\mathbf{x},\mathcal{E}_{wb,H}\Big]\\
&\qquad\leqslant \left(\frac{16}{\varepsilon}\right)^{st}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}\cdot 3st\cdot\left(\frac{2\varepsilon}{33}\right)^{Ast},
\end{aligned}\tag{B.3.21}$$

where the last inequality follows from Lemma B.77. Now notice that

$$\begin{aligned}
\left(\frac{16}{\varepsilon}\right)^{st}\cdot 3st\cdot\left(\frac{2\varepsilon}{33}\right)^{Ast}&\overset{(*)}{\leqslant}\left(\frac{16}{\varepsilon}\right)^{Ast}\cdot 3st\cdot\left(\frac{2\varepsilon}{33}\right)^{Ast}=3st\cdot\left(\frac{32}{33}\right)^{Ast}\leqslant 3st\cdot\left(\frac{32}{33}\right)^{At}\\
&=3st\cdot\left(\frac{32}{33}\right)^{AK\log n}=\frac{3st}{n^{AK\log\frac{33}{32}}}\leqslant\frac{3st}{n^{AK\cdot\frac{1}{100}}}\overset{(\dagger)}{\leqslant}\frac{1}{\sqrt{n}},
\end{aligned}$$

where $(*)$ and $(\dagger)$ are true if $A>\max\{1,\frac{100}{K}\}$ and n is large enough. By combining this with Eq. (B.3.21), we get the lemma. □

Now we are ready to prove Lemma B.29.

Proof of Lemma B.29. From Lemma B.63, we have

$$\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big]=\sum_{\substack{\mathsf{d}\in\mathbb{N}^{V(H)}:\\ L^H(\mathsf{d})=1}}\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P^{wb}_{1,H}(\mathbf{x})\Big]\cdot P_{wb}(\mathsf{d}).$$

Now for every $\mathsf{d}\in\mathbb{N}^{V(H)}$, define

$$L^H_s(\mathsf{d})=\mathbb{1}_{\{\forall v\in V(H),\ \mathsf{d}_v+d^H_1(v)+d^H_{\geqslant 2}(v)\leqslant\Delta\}}.$$

Note that $L^H_s(\mathsf{d})=1$ if and only if $S^H_{\mathsf{d}}=E_1(H)$. Note also that if $L^H_s(\mathsf{d})=1$, then we must have $\mathsf{d}\in\mathsf{D}^{H,\mathrm{safe}}_{\mathrm{cycles}}$. Therefore,

$$\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H}\big]\cdot P_{1,H}^{wb}(\mathbf{x})\Big]=\sum_{\substack{\mathsf{d}\in\mathbb{N}^{V(H)}:\\ L_s^H(\mathsf{d})=1}}\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P_{1,H}^{wb}(\mathbf{x})\Big]\cdot P_{wb}(\mathsf{d})$$
$$+\sum_{\substack{\mathsf{d}\in\mathsf{D}_{\text{cycles}}^{H,\text{safe}}:\\ L_s^H(\mathsf{d})=0}}\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P_{1,H}^{wb}(\mathbf{x})\Big]\cdot P_{wb}(\mathsf{d})$$
$$+\sum_{\substack{\mathsf{d}\in\mathbb{N}^{V(H)}:\\ \mathsf{d}\notin\mathsf{D}_{\text{cycles}}^{H,\text{safe}}}}\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P_{1,H}^{wb}(\mathbf{x})\Big]\cdot P_{wb}(\mathsf{d}).\tag{B.3.22}$$

If $L_s^H(\mathsf{d})=1$, then $S_{\mathsf{d}}=E_1(H)$ and Lemma B.75 implies that

$$\left(1-\frac{1}{\sqrt{n}}\right)\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E''_{\geqslant 2}(H)|}$$
$$\leqslant\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P_{1,H}^{wb}(\mathbf{x})\Big]\leqslant\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E''_{\geqslant 2}(H)|}.\tag{B.3.23}$$

On the other hand, if $\mathsf{d}\in\mathsf{D}_{\text{cycles}}^{H,\text{safe}}$ and $L_s^H(\mathsf{d})=0$, then Lemma B.75 implies that for n large enough, we have

$$0\leqslant\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H},\mathbf{D}^H=\mathsf{d}\big]\cdot P_{1,H}^{wb}(\mathbf{x})\Big]\leqslant\frac{1}{\sqrt{n}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E''_{\geqslant 2}(H)|}.\tag{B.3.24}$$

By combining Eq. (B.3.22), Eq. (B.3.23),Eq. (B.3.24) and Lemma B.78, we get

$$\begin{aligned}&\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{wb,H}\big]\cdot P_{1,H}^{wb}(\mathbf{x})\Big]\\&\leqslant\sum_{\substack{\mathsf{d}\in\mathbb{N}^{V(H)}:\\ L_s^H(\mathsf{d})=1}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E''_{\geqslant 2}(H)|}\cdot P_{wb}(\mathsf{d})\\&\quad+\sum_{\substack{\mathsf{d}\in\mathsf{D}_{\text{cycles}}^{H,\text{safe}}:\\ L_s^H(\mathsf{d})=0}}\frac{1}{\sqrt{n}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E''_{\geqslant 2}(H)|}\cdot P_{wb}(\mathsf{d})\\&\quad+\frac{1}{\sqrt{n}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E''_{\geqslant 2}(H)|}.\end{aligned}$$

Now from Lemma B.62 we have $P_{wb}(\mathsf{d})=\mathbb{P}\big[\mathbf{D}^H=\mathsf{d}\big|\mathcal{E}_{H,b}\big]$. Therefore,

$$\mathbb{E}\Big[\mathbb{E}\big[\overline{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{H,b}\big]\cdot P_{1,H}^{wb}(\mathbf{x})\Big]$$

$$
\begin{aligned}
&\leqslant \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \cdot \mathbb{P}\left[L_s^H(\mathbf{D}^H)=1\middle|\mathcal{E}_{H,b}\right] \\
&\quad + \frac{1}{\sqrt{n}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \cdot \mathbb{P}\left[\mathbf{D}^H \in \mathsf{D}^{H,\text{safe}}_{\text{cycles}} \text{ and } L_s^H(\mathbf{D}^H)=0\middle|\mathcal{E}_{H,b}\right] \\
&\quad + \frac{1}{\sqrt{n}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \\
&\leqslant \left(\mathbb{P}\left[L_s^H(\mathbf{D}^H)=1\middle|\mathcal{E}_{H,b}\right]+\frac{2}{\sqrt{n}}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}.
\end{aligned}
$$

Now recall from Lemma B.62 that $\mathbb{P}\left[\mathbf{D}^H=\mathsf{d}\middle|\mathcal{E}_{H,b}\right]=\mathbb{P}\left[\mathbf{D}^H=\mathsf{d}\middle|\mathcal{E}_{H,b}\right]$. Therefore,

$$
\begin{aligned}
\mathbb{P}\left[L_s^H(\mathbf{D}^H)=1\middle|\mathcal{E}_{H,b}\right] &= \mathbb{P}\left[\left\{\forall v\in V(H),\ \mathbf{D}_v^H+d_1^H(v)+d_{\geqslant 2}^H(v)\leqslant\Delta\right\}\middle|\mathcal{E}_{H,b}\right] \\
&= P_s^H.
\end{aligned}
\tag{B.3.25}
$$

Similarly, from Eq. (B.3.22), Eq. (B.3.23),Eq. (B.3.24), Eq. (B.3.25) and Lemma B.78, we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{wb,H}\right]\cdot P_{1,H}^{wb}(\mathbf{x})\right] \\
&\geqslant \left(\sum_{\substack{\mathsf{d}\in\mathbb{N}^{V(H)}:\\ L_s^H(\mathsf{d})=1}}\left(1-\frac{1}{\sqrt{n}}\right)\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}\cdot P_{wb}(\mathsf{d})\right) \\
&\qquad\qquad - \frac{1}{\sqrt{n}}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \\
&= \left[\left(1-\frac{1}{\sqrt{n}}\right)\cdot\mathbb{P}\left[L_s^H(\mathbf{D}^H)=1\middle|\mathcal{E}_{H,b}\right]-\frac{1}{\sqrt{n}}\right]\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|} \\
&\geqslant \left(P_s^H-\frac{2}{\sqrt{n}}\right)\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H)|+|E'''_{\geqslant 2}(H)|}\left(\frac{d}{n}\right)^{|E_{\geqslant 2}(H)|-|E'''_{\geqslant 2}(H)|}.
\end{aligned}
$$

□

B.3.1.7 Probability of safety in pleasant multigraphs

Proof of Lemma B.30. Let H be an (s,t)-pleasant multigraph with $t = K\cdot\log n$. From Lemma B.62, we know that given $\mathcal{E}_{H,b}$, the random variables $(\mathbf{D}_v^H)_{v\in V(H)}$ are conditionally mutually independent, hence we can rewrite P_s^H as follows:

$$
P_s^H = \mathbb{P}\left[\left\{\forall v\in V(H),\ \mathbf{D}_v^H+d_1^H(v)+d_{\geqslant 2}^H(v)\leqslant\Delta\right\}\middle|\mathcal{E}_{H,b}\right]
$$

$$= \prod_{v \in V(H)} \mathbb{P}\left[\mathbf{D}_v^H + d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta \middle| \mathcal{E}_{H,b}\right].$$

Now for every $v \in V(H)$, we have:

$$\begin{aligned}\mathbb{P}\left[\mathbf{D}_v^H + d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta \middle| \mathcal{E}_{H,b}\right] &= \mathbb{P}\left[\mathbf{D}_v^H \leqslant \Delta - d_1^H(v) - d_{\geqslant 2}^H(v) \middle| \mathcal{E}_{H,b}\right] \\ &= \sum_{\substack{\mathsf{d}_v \in \mathbb{N}: \\ \mathsf{d}_v \leqslant \Delta - d_1^H(v) - d_{\geqslant 2}^H(v)}} \mathbb{P}\left[\mathbf{D}_v^H = \mathsf{d}_v \middle| \mathcal{E}_{H,b}\right] \\ &= P_s\left(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\right),\end{aligned}$$

where $P_s(\ell)$ is defined for every $\ell \in \mathbb{Z}$ as

$$P_s(\ell) = \sum_{\substack{\mathsf{d}_v \in \mathbb{N}: \\ \mathsf{d}_v \leqslant \mathsf{d}}} P_{wb}(\mathsf{d}_v),$$

and $P_{wb}(\mathsf{d}_v)$ is as in Lemma B.62. Clearly, P_s is a non-decreasing function.

If $\ell < 0$, then by picking an arbitrary $v \in V(H)$, we get

$$P_s(\ell) = \mathbb{P}\left[\mathbf{D}_v^H \leqslant \ell \middle| \mathcal{E}_{H,b}\right] = 0.$$

On the other hand, if $\ell \geqslant 0$, we have

$$\begin{aligned}P_s(\ell) = \mathbb{P}\left[\mathbf{D}_v^H \leqslant \ell \middle| \mathcal{E}_{H,b}\right] &\overset{(*)}{=} \mathbb{P}\left[\mathbf{D}_v^H \leqslant \ell \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] \\ &\overset{(\dagger)}{\geqslant} \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) \leqslant \ell \middle| \mathbf{x}, \mathcal{E}_{H,b}\right] \overset{(\ddagger)}{=} \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) \leqslant \ell \middle| \mathbf{x}\right],\end{aligned} \tag{B.3.26}$$

where $(*)$ follows from the fact that given $\mathcal{E}_{wb,H}$, the random variable $\mathbf{D}^H$ is conditionally independent from $\mathbf{x}$ (see Lemma B.62). $(\dagger)$ follows from the fact that $\mathbf{D}_v^H \leqslant d^o_{\mathbf{G}-G(H)}(v)$ for every $v \in V(H)$. $(\ddagger)$ follows from the fact that the event $\mathcal{E}_{H,b}$ is $\sigma(\mathbf{x})$-measurable. For every $\ell \geqslant 0$, we can further lower bound $P_s(\ell)$ as follows:

$$\begin{aligned}P_s(\ell) &\geqslant \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) \leqslant \ell \middle| \mathbf{x}\right] \geqslant \mathbb{P}\left[d^o_{\mathbf{G}-G(H)}(v) = 0 \middle| \mathbf{x}\right] = \prod_{u \in [n] \setminus V(H)} \mathbb{P}\left[uv \notin \mathbf{G} \middle| \mathbf{x}\right] \\ &= \prod_{u \in [n] \setminus V(H)} \left[1 - \left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n}\right] \geqslant \left(1 - \frac{2d}{n}\right)^{n - |V(H)|} \geqslant \left(1 - \frac{2d}{n}\right)^n \\ &= \left(\frac{1}{1 + \frac{2d}{n-2d}}\right)^n \overset{(\star)}{\geqslant} \left(\frac{1}{1 + \frac{4d}{n}}\right)^n \geqslant e^{-4d},\end{aligned}$$

where $(\star)$ is true for n large enough.

Now if $\ell \geqslant \frac{\Delta}{4}$, we get from Eq. (B.3.26) and Lemma B.9 that

$$P_s(\ell) \geqslant \mathbb{P}\left[d^o_{\mathbf{G}-G(H)} \leqslant \ell \middle| \mathbf{x}\right] \geqslant \mathbb{P}\left[d^o_{\mathbf{G}-G(H)} \leqslant \frac{\Delta}{4} \middle| \mathbf{x}\right] \geqslant 1 - \frac{\eta}{2}.$$

□

Proof of Lemma B.31. From Lemma B.30, we have

$$P_s^H = \prod_{v \in V(H)} P_s\big(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\big).$$

If there exists $v \in V(H)$ such that $d_1^H(v) + d_{\geqslant 2}^H(v) > \Delta$, then $P_s\big(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\big) = 0$ and $P_s^H = 0$.

Now assume that $d_1^H(v) + d_{\geqslant 2}^H(v) \leqslant \Delta$ for every $v \in V(H)$. Since H is (s,t)-pleasant, we have $\mathcal{L}_{\geqslant 2}(H) = \varnothing$, so $V(H) = \mathcal{S}_{\geqslant 2}(H) \cup \mathcal{I}_{\geqslant 2}(H)$. Notice the following:

- If $v \in \mathcal{S}_{\geqslant 2}(H)$, then $d_{\geqslant 2}^H(v) \leqslant \frac{\Delta}{4}$. Now since H is (s,t)-pleasant, we have $d_1^H(v) \leqslant 4$, and so
$$\Delta - d_1^H(v) - d_{\geqslant 2}^H(v) \geqslant \Delta - 4 - \frac{\Delta}{4} \geqslant \frac{\Delta}{2}.$$
Lemma B.30 and Lemma B.9 now imply that
$$\begin{aligned} P_s\big(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\big) &\geqslant 1 - \frac{\eta}{2} \geqslant 1 - \frac{1}{2As}\left(\frac{\varepsilon}{6}\right)^{\tau} \\ &\geqslant 1 - \frac{1}{2As} = \frac{1}{1 + \frac{1}{2As-1}} \overset{(*)}{\geqslant} \frac{1}{1 + \frac{1}{As}}, \end{aligned} \tag{B.3.27}$$
where $(*)$ is true if $A > 1$.
- If $v \in \mathcal{I}_{\geqslant 2}(H)$, we will use the assumption $\Delta - d_1^H(v) - d_{\geqslant 2}^H(v) \geqslant 0$. Lemma B.30 now implies that
$$P_s\big(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\big) \geqslant e^{-4d}. \tag{B.3.28}$$

Now From Lemma B.30, we have

$$\begin{aligned} P_s^H &= \prod_{v \in V(H)} P_s\big(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\big) \\ &= \left(\prod_{v \in \mathcal{S}_{\geqslant 2}(H)} P_s\big(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\big)\right) \cdot \left(\prod_{v \in \mathcal{I}_{\geqslant 2}(H)} P_s\big(\Delta - d_1^H(v) - d_{\geqslant 2}^H(v)\big)\right) \\ &\overset{(\dagger)}{\geqslant} \frac{1}{\left(1 + \frac{1}{As}\right)^{|\mathcal{S}_{\geqslant 2}(H)|}} \cdot e^{-4d \cdot |\mathcal{I}_{\geqslant 2}(H)|} \overset{(\ddagger)}{\geqslant} \frac{1}{\left(1 + \frac{1}{As}\right)^{st}} \cdot e^{-4d \cdot \frac{t}{4Ad}} \geqslant \frac{1}{e^{\frac{t}{A}}} \cdot e^{-\frac{t}{A}} \\ &= \frac{1}{e^{\frac{2t}{A}}} = \frac{1}{e^{\frac{2K \log n}{A}}} = \frac{1}{n^{\frac{2K}{A}}}, \end{aligned}$$

where $(\dagger)$ follows from Eq. (B.3.27) and Eq. (B.3.28). $(\ddagger)$ follows from Eq. (B.3.3) and Eq. (B.1.1), which imply that $|\mathcal{I}_{\geqslant 2}(H)| \leqslant \frac{8st}{\Delta} \leqslant \frac{8st}{40Asd} \leqslant \frac{t}{4Ad}$. □

B.3.2 Proofs of technical lemmas for the centered matrix

B.3.2.1 Analyzing walks of multiplicity 1

In order to prove Lemma B.46, we need two lemmas. The following lemma shows that it is unlikely that the set of reassuring walks is completely walk-unsafe.

Lemma B.79. *If $A > \max\{100K, 1\}$ and n is large enough, then given $\mathbf{x}$, the conditional probability that $\mathcal{W}_{1r}(H)$ is completely walk-unsafe can be upper bounded by:*

$$\mathbb{P}\Big[\big\{\mathcal{W}_{1r}(H) \text{ is completely walk-unsafe}\big\}\Big|\mathbf{x}\Big] \leqslant (s\eta)^{\frac{1}{2s\tau}|\mathcal{W}_{1r}(H)|},$$

where η is as in Eq. (B.1.3).

Proof. Since every $W \in \mathcal{W}_{1r}(H)$ contains $s+1$ vertices, and since $d_1^H(v) \leqslant \tau$ for every vertex $v \in V(W)$, we can see that $W \in \mathcal{W}_{1r}(H)$ intersects at most $(s+1)(\tau-1)$ others walks in $\mathcal{W}_{1r}(H)$. Therefore, we can find a subset $\mathcal{W} \subseteq \mathcal{W}_{1r}(H)$ such that:

- For every $W_1, W_2 \in \mathcal{W}$, we have $V(W_1) \cap V(W_2) = \varnothing$.
- $|\mathcal{W}| \geqslant \frac{1}{(s+1)(\tau-1)+1}|\mathcal{W}_{1r}(H)| \geqslant \frac{1}{2s\tau}|\mathcal{W}_{1r}(H)|$.

Now if $\mathcal{W}_{1r}(H)$ is completely walk-unsafe, then $\mathcal{W}$ is completely walk-unsafe, and so every walk $W \in \mathcal{W}$ is walk-unsafe. Therefore, for every $W \in \mathcal{W}$, the set $V(W)$ is not completely safe, which means that there exists at least one vertex $v_W \in V(W)$ which is unsafe. Now since $V(W_1) \cap V(W_2) = \varnothing$ for every $W_1, W_2 \in \mathcal{W}$, we have $v_{W_1} \neq v_{W_2}$ for every $W_1, W_2 \in \mathcal{W}$. Therefore, the set $V_{\mathcal{W}} = \big\{v_W : W \in \mathcal{W}\big\} \subseteq \bigcup_{W \in \mathcal{W}_{1r}} V(W)$ satisfies the following:

- $|V_{\mathcal{W}}| = |\mathcal{W}| \geqslant \frac{1}{2s\tau}|\mathcal{W}_{1r}(H)|$.
- $|V_{\mathcal{W}} \cap V(W)| = 1$ for every $W \in \mathcal{W}$.
- $V_{\mathcal{W}}$ is completely unsafe.

Now since $V(W) \subseteq \mathcal{S}_1(H) \cap \mathcal{S}_{\geqslant 2}(H)$ for every $W \in \mathcal{W}_{1r}(H)$, it follows from Lemma B.12 that for every $V \subseteq \bigcup_{W \in \mathcal{W}_{1r}} V(W)$ satisfying $|V| = |\mathcal{W}|$, we have

$$\mathbb{P}\Big[\big\{V \text{ is completely unsafe}\big\}\Big|\mathbf{x}\Big] \leqslant \eta^{|V|} = \eta^{|\mathcal{W}|}.$$

On the other hand, since there are $s^{|\mathcal{W}|}$ subsets $V \subseteq \bigcup_{W \in \mathcal{W}_{1r}} V(W)$ which satisfy $|V| = |\mathcal{W}|$ and $|V_{\mathcal{W}} \cap V(W)| = 1$ for every $W \in \mathcal{W}$, we conclude that

$$\mathbb{P}\Big[\big\{\mathcal{W}_{1r}(H) \text{ is completely walk-unsafe}\big\}\Big|\mathbf{x}\Big] \leqslant s^{|\mathcal{W}|} \cdot \eta^{|\mathcal{W}|} = (s\eta)^{|\mathcal{W}|} \leqslant (s\eta)^{\frac{1}{2s\tau}|\mathcal{W}_{1r}(H)|},$$

where the last inequality follows from the fact that if $A > \max\{100K, 1\}$ then $s\eta < 1$ (see the definition of η in Eq. (B.1.3)). □

The following lemma shows that in the event that there is one walk of multiplicity 1 that is walk-safe, the conditional expectation of $\hat{\mathbf{Y}}_H$ given this event and given $\mathbf{x}$ will be zero. This can be seen as the truncated version of Eq. (B.2.3).

Lemma B.80. *Let $W \in \mathcal{W}_1(H)$, and let $\mathcal{E}$ be an event satisfying:*

- *$\mathcal{E}$ implies that W is $(\mathbf{G}, H)$-walk-safe, i.e., $\forall G \in \mathcal{E}$, the walk W is (G, H)-walk-safe.*
- *Given $\mathbf{x}$ and $\mathcal{E}$, the random variable $(\mathbb{1}_{\{uv\in\mathbf{G}\}})_{uv\in E(W)}$ is conditionally independent from $(\overline{\mathbf{Y}}_{u'v'})_{u'v'\in E(H)\setminus E(W)}$.*
- *Given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent of $(\mathbb{1}_{\{uv\in\mathbf{G}\}})_{uv\in E(W)}$.*

We have

$$\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right] = 0.$$

Proof. We have

$$\begin{aligned}
\mathbb{E}\left[\hat{\mathbf{Y}}_H \middle| \mathbf{x}, \mathcal{E}\right] &= \mathbb{E}\left[\left(\overline{\mathbf{Y}}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \cdot \prod_{W'\in\mathcal{W}(H)\setminus\{W\}} \left(\overline{\mathbf{Y}}_{W'} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W'}\right) \middle| \mathbf{x}, \mathcal{E}\right] \\
&\overset{(*)}{=} \mathbb{E}\left[\left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \cdot \prod_{W'\in\mathcal{W}(H)\setminus\{W\}} \left(\overline{\mathbf{Y}}_{W'} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W'}\right) \middle| \mathbf{x}, \mathcal{E}\right] \\
&\overset{(\dagger)}{=} \mathbb{E}\left[\left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \middle| \mathbf{x}, \mathcal{E}\right] \cdot \mathbb{E}\left[\prod_{W'\in\mathcal{W}(H)\setminus\{W\}} \left(\overline{\mathbf{Y}}_{W'} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W'}\right) \middle| \mathbf{x}, \mathcal{E}\right] \\
&\overset{(\ddagger)}{=} \mathbb{E}\left[\left(\mathbf{Y}_W - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_W\right) \middle| \mathbf{x}\right] \cdot \mathbb{E}\left[\prod_{W'\in\mathcal{W}(H)\setminus\{W\}} \left(\overline{\mathbf{Y}}_{W'} - \left(\frac{\varepsilon d}{2n}\right)^s \mathbf{x}_{W'}\right) \middle| \mathbf{x}, \mathcal{E}\right] \\
&\overset{(\wr)}{=} 0,
\end{aligned}$$

where $(*)$ follows from the fact that if W is walk-safe, then $\overline{\mathbf{Y}}_W = \mathbf{Y}_W$, $(\dagger)$ follows from the fact that given $\mathbf{x}$ and $\mathcal{E}$, the random variable $(\mathbb{1}_{\{uv\in\mathbf{G}\}})_{uv\in E(W)}$ is conditionally independent from $(\overline{\mathbf{Y}}_{u'v'})_{u'v'\in E(H)\setminus E(W)}$, $(\ddagger)$ follows from the fact that given $\mathbf{x}$, the event $\mathcal{E}$ is conditionally independent of $(\mathbb{1}_{\{uv\in\mathbf{G}\}})_{uv\in E(W)}$, and $(\wr)$ follows from Lemma B.40. □

Now we are ready to prove Lemma B.46.

Proof of Lemma B.46. For every $\mathcal{W} \subseteq \mathcal{W}_{1r}(H)$, define the following events:

$$\mathcal{E}_{\mathcal{W},H,c\text{-}ws} = \left\{\mathcal{W} \text{ is completely } (\mathbf{G}, H)\text{-walk-safe}\right\},$$

and

$$\mathcal{E}_{\mathcal{W},H,c\text{-}wus} = \big\{\mathcal{W} \text{ is completely } (\mathbf{G},H)\text{-walk-unsafe}\big\}.$$

We have:

$$\begin{aligned}&\mathbb{E}\big[\hat{\mathbf{Y}}_H\big|\mathbf{x}\big]\\ &= \sum_{\mathcal{W}\subseteq\mathcal{W}_{1r}(H)} \mathbb{E}\big[\hat{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{\mathcal{W},H,c\text{-}ws}\cap\mathcal{E}_{\mathcal{W}_{1r}(H)\setminus\mathcal{W},H,c\text{-}wus}\big]\cdot\mathbb{P}\big[\mathcal{E}_{\mathcal{W},H,c\text{-}ws}\cap\mathcal{E}_{\mathcal{W}_{1r}(H)\setminus\mathcal{W},H,c\text{-}wus}\big].\end{aligned}$$

Now from Lemma B.80, we know that for every $\mathcal{W}\subseteq\mathcal{W}_{1r}(H)$ satisfying $\mathcal{W}\neq\emptyset$, we have

$$\mathbb{E}\big[\hat{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{\mathcal{W},H,c\text{-}ws}\cap\mathcal{E}_{\mathcal{W}_{1r}(H)\setminus\mathcal{W},H,c\text{-}wus}\big]=0.$$

Therefore,

$$\mathbb{E}\big[\hat{\mathbf{Y}}_H\big|\mathbf{x}\big]=\mathbb{E}\big[\hat{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\big]\cdot\mathbb{P}\big[\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\big]. \tag{B.3.29}$$

Now from Lemma B.79, we have

$$\mathbb{P}\big[\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\big]\leqslant(s\eta)^{\frac{1}{2s\tau}|\mathcal{W}_{1r}(H)|}=\Big[(s\eta)^{\frac{1}{2s\tau}}\Big]^{|\mathcal{W}_{1r}(H)|}. \tag{B.3.30}$$

On the other hand, we have

$$\begin{aligned}&\mathbb{E}\big[\hat{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\big]\\ &\quad=\mathbb{E}\left[\prod_{W\in\mathcal{W}(H)}\left(\overline{\mathbf{Y}}_W-\left(\frac{\varepsilon d}{2n}\right)^s\mathbf{x}_W\right)\middle|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\right]\\ &\quad=\sum_{\mathcal{W}\subseteq\mathcal{W}(H)}\left[\prod_{W\in\mathcal{W}(H)\setminus\mathcal{W}}\left(-\left(\frac{\varepsilon d}{2n}\right)^s\mathbf{x}_W\right)\right]\cdot\mathbb{E}\left[\prod_{W\in\mathcal{W}}\overline{\mathbf{Y}}_W\middle|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\right]\\ &\quad=\sum_{\mathcal{W}\subseteq\mathcal{W}(H)}\left[\prod_{W\in\mathcal{W}(H)\setminus\mathcal{W}}\left(-\left(\frac{\varepsilon d}{2n}\right)^s\mathbf{x}_W\right)\right]\cdot\mathbb{E}\Big[\overline{\mathbf{Y}}_{H_{\mathcal{W}}}\Big|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\Big].\end{aligned}$$

Therefore,

$$\Big|\mathbb{E}\big[\hat{\mathbf{Y}}_H\big|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\big]\Big|\leqslant\sum_{\mathcal{W}\subseteq\mathcal{W}(H)}\left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}(H)\setminus\mathcal{W}|}\cdot\Big|\mathbb{E}\Big[\overline{\mathbf{Y}}_{H_{\mathcal{W}}}\Big|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\Big]\Big|. \tag{B.3.31}$$

Now fix $\mathcal{W}\subseteq\mathcal{W}(H)$ and let $\mathcal{S}(H_{\mathcal{W}})=\mathcal{S}_1(H_{\mathcal{W}})\cap\mathcal{S}_{\geqslant 2}(H_{\mathcal{W}})$. If we take $\mathcal{E}=\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}$ and $U=\mathcal{S}(H_{\mathcal{W}})\setminus\left(\bigcup_{W\in\mathcal{W}_{1r}(H)}V(W)\right)$, it is easy to see that the conditions of Lemma B.50 and Lemma B.52 are satisfied. Therefore,

$$\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\right]\right| \leqslant n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_{1a}(H_{\mathcal{W}})|+\tau(|\mathcal{S}(H_{\mathcal{W}})|-|U|)}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}})|}\cdot\mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}})}|\middle|\mathbf{x}\right].$$

Now observe that $E_{1a}(H_{\mathcal{W}}) \subseteq E_{1a}(H) \cup \big(E_1(H_{\mathcal{W}}) \cap E_{\geqslant 2}(H)\big)$ and

$$E_1(H_{\mathcal{W}}) \cap E_{\geqslant 2}(H) \subseteq \bigcup_{W\in\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}} W,$$

which implies that

$$|E_{1a}(H_{\mathcal{W}})| \leqslant |E_{1a}(H)| + s|\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}|.$$

On the other hand, we have

$$|S(H_{\mathcal{W}})| - |U| \leqslant \left|\bigcup_{W\in\mathcal{W}_{1r}(H)} V(W)\right| \leqslant \sum_{W\in\mathcal{W}_{1r}(H)}|V(W)| = |\mathcal{W}_{1r}(H)|\cdot(s+1) \leqslant 2s\cdot|\mathcal{W}_{1r}(H)|.$$

Therefore,

$$\begin{aligned}
&\left|\mathbb{E}\left[\overline{\mathbf{Y}}_H\middle|\mathbf{x},\mathcal{E}_{\mathcal{W}_{1r}(H),H,c\text{-}wus}\right]\right| \\
&\qquad\leqslant n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+s|\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}|+2s\tau\cdot|\mathcal{W}_{1r}(H)|}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}})|}\cdot\mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}})}|\middle|\mathbf{x}\right] \\
&\qquad\leqslant n^{\frac{2K}{A}}\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+s|\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}|}\cdot\left[\left(\frac{6}{\varepsilon}\right)^{2s\tau}\right]^{|\mathcal{W}_{1r}(H)|}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}})|}\cdot\mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}})}|\middle|\mathbf{x}\right]
\end{aligned}$$

By combining this with Eq. (B.3.29) and Eq. (B.3.30) and Eq. (B.3.31), we get

$$\begin{aligned}
\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H\middle|\mathbf{x}\right]\right| \leqslant \sum_{\mathcal{W}\subseteq\mathcal{W}(H)}&\left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}(H)\setminus\mathcal{W}|}\cdot n^{\frac{2K}{A}}\cdot\left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|+s|\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}|} \\
&\times\left[(s\eta)^{\frac{1}{2s\tau}}\cdot\left(\frac{6}{\varepsilon}\right)^{2s\tau}\right]^{|\mathcal{W}_{1r}(H)|}\cdot\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}})|}\cdot\mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}})}|\middle|\mathbf{x}\right].
\end{aligned}$$

Now if $A > \max\{100K, 1\}$, we get from Eq. (B.1.3) that

$$\eta = \frac{1}{As\cdot 2^{6A\tau s^2}}\left(\frac{\varepsilon}{6}\right)^{4s^2\tau^2} \leqslant \frac{1}{s\cdot 2^{6A\tau s^2}}\left(\frac{\varepsilon}{6}\right)^{4s^2\tau^2},$$

hence

$$(s\eta)^{\frac{1}{2s\tau}}\cdot\left(\frac{6}{\varepsilon}\right)^{2s\tau} \leqslant \left(\frac{1}{2^{6A\tau s^2}}\left(\frac{\varepsilon}{6}\right)^{4s^2\tau^2}\right)^{\frac{1}{2s\tau}}\cdot\left(\frac{6}{\varepsilon}\right)^{2s\tau} = \frac{1}{2^{3As}}\cdot\left(\frac{\varepsilon}{6}\right)^{2s\tau}\cdot\left(\frac{6}{\varepsilon}\right)^{2s\tau} = \frac{1}{2^{3As}}.$$

Therefore,

$$\left|\mathbb{E}\left[\hat{\mathbf{Y}}_H|\mathbf{x}\right]\right| \leqslant \frac{n^{\frac{2K}{A}}}{2^{3As\cdot|\mathcal{W}_{1r}(H)|}} \cdot \left(\frac{6}{\varepsilon}\right)^{|E_1^a(H)|} \cdot \sum_{\mathcal{W}\subseteq\mathcal{W}(H)} \left(\frac{6}{\varepsilon}\right)^{s|\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}|} \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}(H)\setminus\mathcal{W}|+|E_1(H_{\mathcal{W}})|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}})}||\mathbf{x}\right]. \tag{B.3.32}$$

Now we have

$$\begin{aligned}
&\sum_{\mathcal{W}\subseteq\mathcal{W}(H)} \left(\frac{6}{\varepsilon}\right)^{s|\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}|} \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}(H)\setminus\mathcal{W}|+|E_1(H_{\mathcal{W}})|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}})}||\mathbf{x}\right] \\
&= \sum_{\mathcal{W}_1\subseteq\mathcal{W}_1(H)} \sum_{\mathcal{W}_{\geqslant 2}\subseteq\mathcal{W}_{\geqslant 2}(H)} \left(\frac{6}{\varepsilon}\right)^{s|\mathcal{W}_{\geqslant 2}(H)\setminus(\mathcal{W}_1\cup\mathcal{W}_{\geqslant 2})|} \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}(H)\setminus(\mathcal{W}_1\cup\mathcal{W}_{\geqslant 2})|+|E_1(H_{\mathcal{W}_1\cup\mathcal{W}_{\geqslant 2}})|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}_1\cup\mathcal{W}_{\geqslant 2}})}||\mathbf{x}\right],
\end{aligned} \tag{B.3.33}$$

Now for every $\mathcal{W}_1 \subseteq \mathcal{W}_1(H)$ and every $\mathcal{W}_{\geqslant 2} \subset \mathcal{W}_{\geqslant 2}(H)$, we have

$$\left|\mathcal{W}_{\geqslant 2}(H) \setminus \left(\mathcal{W}_1 \cup \mathcal{W}_{\geqslant 2}\right)\right| = |\mathcal{W}_{\geqslant 2}(H) \setminus \mathcal{W}_{\geqslant 2}| = |\mathcal{W}_{\geqslant 2}(H)| - |\mathcal{W}_{\geqslant 2}|, \tag{B.3.34}$$

$$E_1(H_{\mathcal{W}_1\cup\mathcal{W}_{\geqslant 2}}) = E_1(H_{\mathcal{W}_1}) \cup E_1(H_{\mathcal{W}_{\geqslant 2}}),$$
$$E_1(H_{\mathcal{W}_1}) \cap E_1(H_{\mathcal{W}_{\geqslant 2}}) = \varnothing,$$

and

$$|E_1(H_{\mathcal{W}_1})| = s|\mathcal{W}_1|,$$

hence

$$|E_1(H_{\mathcal{W}_1\cup\mathcal{W}_{\geqslant 2}})| = s|\mathcal{W}_1| + |E_1(H_{\mathcal{W}_{\geqslant 2}})|.$$

On the other hand,

$$\begin{aligned}
|\mathcal{W}(H) \setminus (\mathcal{W}_1 \cup \mathcal{W}_{\geqslant 2})| &= |\mathcal{W}(H)| - |\mathcal{W}_1| - |\mathcal{W}_{\geqslant 2}| \\
&= |\mathcal{W}_1(H)| + |\mathcal{W}_{\geqslant 2}(H)| - |\mathcal{W}_1| - |\mathcal{W}_{\geqslant 2}|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&s|\mathcal{W}(H)\setminus(\mathcal{W}_1 \cup \mathcal{W}_{\geqslant 2})| + |E_1(H_{\mathcal{W}_1\cup\mathcal{W}_{\geqslant 2}})| \\
&\quad = s|\mathcal{W}_1(H)| + s|\mathcal{W}_{\geqslant 2}(H)| - s|\mathcal{W}_1| - s|\mathcal{W}_{\geqslant 2}| + s|\mathcal{W}_1| + |E_1(H_{\mathcal{W}_{\geqslant 2}})| \\
&\quad = s|\mathcal{W}_1(H)| + s(|\mathcal{W}_{\geqslant 2}(H)| - |\mathcal{W}_{\geqslant 2}|) + |E_1(H_{\mathcal{W}_{\geqslant 2}})|.
\end{aligned} \tag{B.3.35}$$

Furthermore, we have

$$E_{\geqslant 2}(H_{\mathcal{W}_1\cup\mathcal{W}_{\geqslant 2}}) = E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}). \tag{B.3.36}$$

By combining Eq. (B.3.34), Eq. (B.3.33), Eq. (B.3.35), and Eq. (B.3.36), we get

$$\sum_{\mathcal{W}\subseteq\mathcal{W}(H)} \left(\frac{6}{\varepsilon}\right)^{s|\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}|} \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}(H)\setminus\mathcal{W}|+|E_1(H_{\mathcal{W}})|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}})}||\mathbf{x}\right]$$

$$= \sum_{\mathcal{W}_1 \subseteq \mathcal{W}_1(H)} \sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} \left(\frac{6}{\varepsilon}\right)^{s(|\mathcal{W}_{\geqslant 2}(H)|-|\mathcal{W}_{\geqslant 2}|)} \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}_1(H)|+s(|\mathcal{W}_{\geqslant 2}(H)|-|\mathcal{W}_{\geqslant 2}|)+|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}| \middle| \mathbf{x}\right]$$

$$= \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}_1(H)|} \cdot \left[\sum_{\mathcal{W}_1 \subseteq \mathcal{W}_1(H)} 1\right] \cdot \sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} \left(\frac{3d}{n}\right)^{s(|\mathcal{W}_{\geqslant 2}(H)|-|\mathcal{W}_{\geqslant 2}|)} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}| \middle| \mathbf{x}\right]$$

$$= 2^{|\mathcal{W}_1(H)|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}_1(H)|} \cdot \sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} \left(\frac{3d}{n}\right)^{s(|\mathcal{W}_{\geqslant 2}(H)|-|\mathcal{W}_{\geqslant 2}|)} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \mathbb{E}\left[|\tilde{\mathbf{Y}}_{E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}| \middle| \mathbf{x}\right]$$

$$= 2^{|\mathcal{W}_1(H)|} \cdot \left(\frac{\varepsilon d}{2n}\right)^{s|\mathcal{W}_1(H)|} \cdot \sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}).$$

By combining this with Eq. (B.3.32) , we get the lemma. □

B.3.2.2 Analyzing walks of multiplicity at least 2

In order to prove Lemma B.47, we need a few definitions and lemmas.

Definition B.81. Let $H \in \mathrm{BSAW}_{s,t}$. For every $\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)$, define

$$K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) = \frac{\left(\frac{3d}{n}\right)^{s(|\mathcal{W}_{\geqslant 2}(H)|-|\mathcal{W}_{\geqslant 2}|)} \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}})|}}{\prod_{v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} n^{\frac{1}{4}\left(d^{H_{\mathcal{W}_{\geqslant 2}}}_{\geqslant 2}(v)-\Delta\right)}} \cdot \left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right].$$

Lemma B.82. *Let $H \in BSAW_{s,t}$. For every $\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)$, we have*

$$F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) \leqslant K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}).$$

Proof. This is a direct corollary from Lemma B.16 and the fact that $E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) = E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \cup E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$. □

Lemma B.82 implies that $\sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) \leqslant \sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})$. In the following few lemmas, we will prove an upper bound on $\sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})$. This will yield an upper bound on $\sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})$.

The following lemma compares $K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})$ and $K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})$ in the case where $\mathcal{W}_{\geqslant 2}$ and $\mathcal{W}'_{\geqslant 2}$ differ by exactly one walk.

Lemma B.83. *If $H \in BSAW_{s,t}$ and n is large enough, then for every $\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)$ and every $W \in \mathcal{W}_{\geqslant 2}$, if we define $\mathcal{W}'_{\geqslant 2} = \mathcal{W}_{\geqslant 2} \setminus \{W\}$, then*

$$\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})}$$

$$\leqslant \frac{\left(1+\frac{1}{\sqrt{n}}\right)\cdot\left(\frac{3d}{n}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap W|}}{\left(\frac{2\sqrt{n}}{9d}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}\cdot\left(\frac{n\sqrt{n}}{2d}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}\cdot\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}.$$

Proof. We have

$$\prod_{v\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} n^{\frac{1}{4}\left(d^{H_{\mathcal{W}_{\geqslant 2}}}_{\geqslant 2}(v)-\Delta\right)} = n^{-\frac{\Delta}{4}|\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}\cdot n^{\sum_{v\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\frac{1}{4}d^{H_{\mathcal{W}_{\geqslant 2}}}_{\geqslant 2}(v)}}.$$

Furthermore,

$$\sum_{v\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} d_{H_{\mathcal{W}_{\geqslant 2}}}(v) = 2\cdot|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})| + |E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|,$$

where

$$E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) = \left\{uv\in E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}):\ u\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \text{ and } v\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right\},$$

and

$$E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) = \left\{uv\in E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}):\ u\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \text{ and } v\notin\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right\}.$$

Therefore,

$$\prod_{v\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} n^{\frac{1}{4}\left(d^{H_{\mathcal{W}_{\geqslant 2}}}_{\geqslant 2}(v)-\Delta\right)} = n^{-\frac{\Delta}{4}|\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|+\frac{1}{2}|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|+\frac{1}{4}|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|},$$

hence

$$\begin{aligned}
&K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})\\
&= \frac{\left(\frac{3d}{n}\right)^{s(|\mathcal{W}_{\geqslant 2}(H)|-|\mathcal{W}_{\geqslant 2}|)}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}})|}}{n^{-\frac{\Delta}{4}|\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|+\frac{1}{2}|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|+\frac{1}{4}|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}}\cdot\left(\frac{d}{n}\right)^{|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}\cdot\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]\\
&= n^{\frac{\Delta}{4}|\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}\cdot\left(\frac{3d}{n}\right)^{s(|\mathcal{W}_{\geqslant 2}(H)|-|\mathcal{W}_{\geqslant 2}|)}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}})|}\cdot\left(\frac{d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}\cdot\left(\frac{d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}\\
&\qquad\times\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right].
\end{aligned}$$

If we apply the above equation to $\mathcal{W}'_2$ and use the fact that $|\mathcal{W}'_{\geqslant 2}| = |\mathcal{W}_{\geqslant 2}| - 1$, we get

$$\begin{aligned}
&K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})\\
&= n^{\frac{\Delta}{4}|\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}\cdot\left(\frac{3d}{n}\right)^{s+s(|\mathcal{W}_{\geqslant 2}(H)|-|\mathcal{W}_{\geqslant 2}|)}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})|}\cdot\left(\frac{d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}\cdot\left(\frac{d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}
\end{aligned}$$

$$\times \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \frac{d}{n} + \frac{d^2}{n^2} \right].$$

Therefore,

$$\begin{aligned}
&\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} \\
&= n^{\frac{\Delta}{4}\left(|\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| - |\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|\right)} \cdot \left(\frac{3d}{n}\right)^s \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})| - |E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| - |E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \\
&\qquad \times \left(\frac{d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| - |E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \frac{\prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \frac{d}{n} + \frac{d^2}{n^2} \right]}{\prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right) \frac{d}{n} + \frac{d^2}{n^2} \right]}.
\end{aligned} \tag{B.3.37}$$

Now since $\mathcal{W}'_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}$, the multigraph $H_{\mathcal{W}'_{\geqslant 2}}$ is a submultigraph of $H_{\mathcal{W}_{\geqslant 2}}$, hence $d^{H_{\mathcal{W}'_{\geqslant 2}}}_{\geqslant 2}(v) \leqslant d^{H_{\mathcal{W}'_{\geqslant 2}}}_{\geqslant 2}(v)$ for every $v \in V(H_{\mathcal{W}'_{\geqslant 2}})$. This means that $\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$. Therefore,

$$\begin{aligned}
&\Delta \cdot \left(|\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})| - |\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|\right) \\
&= \sum_{v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \Delta \geqslant \sum_{v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} d^{H_{\mathcal{W}'_{\geqslant 2}}}_{\geqslant 2}(v) \\
&= 2 \cdot \left|\left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) : \ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ and } v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\}\right| \\
&\quad + \left|\left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) : \ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ and } v \notin \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\}\right| \\
&= 2 \cdot \left|\left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) : \ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ and } v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\}\right| \\
&\quad + \left|\left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) : \ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ and } v \notin \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \text{ and } v \notin \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\}\right| \\
&\quad + \left|\left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) : \ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ and } v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\}\right| \\
&\overset{(*)}{=} 2 \cdot \left|\left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) : \ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ and } v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\}\right| \\
&\quad + \left|\left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) : \ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ and } v \notin \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \text{ and } v \notin \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\}\right| \\
&\quad + \left|\left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) : \ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ and } v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \text{ and } v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\}\right| \\
&= 2 \cdot \left|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right| + \left|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right| + \left|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right|,
\end{aligned}$$

where $(*)$ follows from the fact that $\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$. By combining this with Eq. (B.3.37), we get

$$\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} \leqslant n^{-\frac{1}{2}|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|-\frac{1}{4}|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|-\frac{1}{4}|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}$$
$$\times \left(\frac{3d}{n}\right)^s \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})|-|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}$$
$$\times \left(\frac{2d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \frac{\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}{\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}. \tag{B.3.38}$$

Now we have:

$$\begin{aligned}
&\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right] \\
&= \prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right] \cdot \prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\setminus E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right] \\
&\leqslant \prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\frac{2d}{n}+\frac{d^2}{n^2}\right] \cdot \prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\setminus E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right] \\
&= \prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\frac{2d}{n}\left(1+\frac{d}{2n}\right)\right] \cdot \prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\setminus E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right] \\
&\leqslant \left(1+O\left(\frac{dst}{2n}\right)\right)\cdot\left(\frac{2d}{n}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\setminus E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right] \\
&\leqslant \left(1+\frac{1}{\sqrt{n}}\right)\cdot\left(\frac{2d}{n}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{2d}{n}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \\
&\qquad\qquad \times \prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\setminus E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right],
\end{aligned}$$

where the last inequality is true for n large enough. By combining this with Eq. (B.3.38), we get

$$\begin{aligned}
&\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} \\
&\leqslant \left(1+\frac{1}{\sqrt{n}}\right)\cdot n^{-\frac{1}{4}|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}\cdot\left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}\cdot\left(\frac{2d}{n^{\frac{5}{4}}}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}
\end{aligned}$$

$$\times \left(\frac{3d}{n}\right)^{s} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})|-|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{2d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}$$
$$\times \frac{\prod_{uv\in E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\setminus E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}{\prod_{uv\in E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}.$$

Hence,

$$\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} \leqslant \left(1+\frac{1}{\sqrt{n}}\right)\cdot n^{-\frac{1}{4}|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{3d}{n}\right)^{s} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})|-|E_1(H_{\mathcal{W}_{\geqslant 2}})|}$$
$$\times \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \cdot \left(\frac{2d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}$$
$$\times \frac{\prod_{uv\in E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\setminus E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}{\prod_{uv\in E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}.$$

Now since $H_{\mathcal{W}'_{\geqslant 2}}$ is a submultigraph of $H_{\mathcal{W}_{\geqslant 2}}$, we have $E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) = E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \cup E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$, which implies that

$$E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \cup E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}).$$

This means that

$$E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \setminus E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \subseteq E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}).$$

On the other hand, since $E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \cap E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) = \varnothing$, we have

$$E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \left(E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \setminus E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right) = E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}).$$

Therefore,

$$\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} \leqslant \left(1+\frac{1}{\sqrt{n}}\right)\cdot n^{-\frac{1}{4}|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{3d}{n}\right)^{s} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})|-|E_1(H_{\mathcal{W}_{\geqslant 2}})|}$$
$$\times \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \cdot \frac{\left(\frac{2d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}}{\prod_{uv\in E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}.$$

Now since $\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$ and $E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$, we have

$$\begin{aligned} E^{b}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) &= \left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) :\ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \text{ or } v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right\} \\ &\subseteq \left\{uv \in E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) :\ u \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \text{ or } v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right\} = E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}). \end{aligned} \tag{B.3.39}$$

On the other hand, since $E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq E^{b}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$ and since $\left\{E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}), E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right\}$ is a partition of $E^{b}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$, we have

$$\begin{aligned} E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) &= E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \\ &= \left(E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right) \cup \left(E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} &\leqslant \left(1+\frac{1}{\sqrt{n}}\right) \cdot n^{-\frac{1}{4}|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{3d}{n}\right)^{s} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})|-|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \\ &\quad \times \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \cdot \left(\frac{2d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \\ &\quad \times \frac{\left(\frac{2d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|-|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}}{\prod\limits_{uv \in E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1+\frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right]} \\ &= \left(1+\frac{1}{\sqrt{n}}\right) \cdot \left(\frac{3d}{n}\right)^{s} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})|-|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \\ &\quad \times \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \frac{\left(\frac{2d}{n^{\frac{5}{4}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|-|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}}{\prod\limits_{uv \in E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1+\frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right]}. \end{aligned}$$

Now since $E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \subseteq E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$ and $E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \cap E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) = \emptyset$, we have

$$E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \subseteq E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}),$$

and so

$$|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})| - |E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| \leqslant 0.$$

Hence, for n large enough, we have

$$\begin{aligned} \frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} &\leqslant \left(1+\frac{1}{\sqrt{n}}\right) \cdot \left(\frac{3d}{n}\right)^{s} \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})|-|E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|-|E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \\ &\quad \times \left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \frac{\left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|-|E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}}{\prod\limits_{uv \in E^{a}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^{a}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1+\frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right]}. \end{aligned}$$

Now observe that

$$
\begin{aligned}
&|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| \\
&= \left|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \left(E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cup E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right)\right| \\
&= \left|\left(E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right) \setminus E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right| \\
&\overset{(\dagger)}{=} |E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| - |E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| \\
&= |E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| + |E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| - |E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| - |E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| \\
&= |E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| + |E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| \\
&\qquad\qquad - |E^{b,i}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| - |E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,i}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})| - |E^{b,o}_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^{b,o}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|,
\end{aligned}
$$

where (†) follows from the fact that $E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$ and $E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) = \varnothing$, which imply that $E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})$. Therefore,

$$
\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} \leqslant \left(1 + \frac{1}{\sqrt{n}}\right) \cdot \left(\frac{3d}{n}\right)^s \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})| - |E_1(H_{\mathcal{W}_{\geqslant 2}})|} \cdot \frac{\left(\frac{2d}{n^{\frac{3}{2}}}\right)^{-|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}}{\prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right]}.
$$

Now since $E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \subseteq E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})$, we have

$$
\begin{aligned}
\left(E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right) \cap E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) &\subseteq \left(E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right) \cap E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \\
&\subseteq E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \cap E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) = \varnothing,
\end{aligned}
$$

hence,

$$
\begin{aligned}
E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) &= \left(E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right) \setminus E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \\
&= E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus \left(E^a_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cup E^b_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\right) \\
&= E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}).
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} &\leqslant \frac{\left(1 + \frac{1}{\sqrt{n}}\right) \cdot \left(\frac{3d}{n}\right)^s \cdot \left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}'_{\geqslant 2}})| - |E_1(H_{\mathcal{W}_{\geqslant 2}})|}}{\left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \cdot \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right]} \\
&\leqslant \frac{\left(1 + \frac{1}{\sqrt{n}}\right) \cdot \left(\frac{3d}{n}\right)^{s + |E_1(H_{\mathcal{W}'_{\geqslant 2}})| - |E_1(H_{\mathcal{W}_{\geqslant 2}})|}}{\left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \cdot \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right]}.
\end{aligned}
\tag{B.3.40}
$$

Now notice that $H_{\mathcal{W}_{\geqslant 2}}$ is obtained from $H_{\mathcal{W}'_{\geqslant 2}}$ by adding the self-avoiding-walk W of length s. This means that

$$E_1(H_{\mathcal{W}_{\geqslant 2}}) = \big(E_1(H_{\mathcal{W}'_{\geqslant 2}}) \setminus W\big) \cup \big(W \setminus E(H_{\mathcal{W}'_{\geqslant 2}})\big).$$

Therefore,

$$\begin{aligned} s + |E_1(H_{\mathcal{W}'_{\geqslant 2}})| - |E_1(H_{\mathcal{W}_{\geqslant 2}})| &= |W| + |E_1(H_{\mathcal{W}'_{\geqslant 2}})| - |E_1(H_{\mathcal{W}'_{\geqslant 2}}) \setminus W| - |W \setminus E(H_{\mathcal{W}'_{\geqslant 2}})| \\ &= \big|E_1(H_{\mathcal{W}'_{\geqslant 2}}) \cap W\big| + \big|E(H_{\mathcal{W}'_{\geqslant 2}}) \cap W\big| \\ &= \big|E_1(H_{\mathcal{W}'_{\geqslant 2}}) \cap W\big| + \big|E_1(H_{\mathcal{W}'_{\geqslant 2}}) \cap W\big| + \big|E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap W\big| \\ &= 2 \cdot \big|E_1(H_{\mathcal{W}'_{\geqslant 2}}) \cap W\big| + \big|E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap W\big|. \end{aligned}$$

Now it is easy to see that

$$\begin{aligned} E_1(H_{\mathcal{W}'_{\geqslant 2}}) \cap W &= E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \\ &= \big(E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\big) \cup \big(E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\big). \end{aligned}$$

Thus,

$$\begin{aligned} s + |E_1(H_{\mathcal{W}'_{\geqslant 2}})| - |E_1(H_{\mathcal{W}_{\geqslant 2}})| = 2 \cdot \big|E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\big| + 2 \cdot \big|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\big| \\ + \big|E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap W\big|. \end{aligned}$$

By combining this with Eq. (B.3.40), it follows that for n large enough, we have

$$\begin{aligned} &\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})} \\ &\leqslant \frac{\left(1 + \frac{1}{\sqrt{n}}\right) \cdot \left(\frac{3d}{n}\right)^{2|E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| + 2|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| + |E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap W|}}{\left(\frac{2d}{n^{\frac{3}{2}}}\right)^{|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \cdot \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right]} \\ &= \frac{\left(1 + \frac{1}{\sqrt{n}}\right) \cdot \left(\frac{3d}{n}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}}) \cap W|}}{\left(\frac{2\sqrt{n}}{9d}\right)^{|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \cdot \left(\frac{n^2}{9d^2}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|} \cdot \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right]}. \end{aligned}$$

Now since

$$|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| = |E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})| + |E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|,$$

we get

$$\frac{K_{\mathcal{W}'_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})}$$

$$\leqslant \frac{\left(1+\frac{1}{\sqrt{n}}\right)\cdot\left(\frac{3d}{n}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})\cap W|}}{\left(\frac{2\sqrt{n}}{9d}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}\cdot\left(\frac{n\sqrt{n}}{2d}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})|}\cdot\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}'_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}.$$

□

The following lemma proves an upper bound on $K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})$ for every $\mathcal{W}_{\geqslant 2}\subsetneq\mathcal{W}_{\geqslant 2}(H)$.

Lemma B.84. *If $H\in BSAW_{s,t}$ and n is large enough, then for every $\mathcal{W}_{\geqslant 2}\subsetneq\mathcal{W}_{\geqslant 2}(H)$, we have*

$$K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})\leqslant\frac{9d}{\sqrt{n}}\cdot\hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}),$$

where

$$\hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x})=\frac{\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}(H)})|}}{\prod_{v\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})}n^{\frac{1}{4}\left(d^{H_{\mathcal{W}_{\geqslant 2}(H)}}_{\geqslant 2}(v)-\Delta\right)}}\cdot\left(\frac{2d}{n}\right)^{|E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})|}\cdot\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{3d^2}{n\sqrt{n}}\right].\tag{B.3.41}$$

Proof. Let $k=|\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}_2|$, and let $W_1,\dots,W_k\in\mathcal{W}_{\geqslant 2}(H)$ be such that

$$\mathcal{W}_{\geqslant 2}(H)\setminus\mathcal{W}_2=\{W_1,\dots,W_k\}.$$

Now define $\mathcal{W}^{(0)}_{\geqslant 2}=\mathcal{W}_{\geqslant 2}$ and for every $1\leqslant m\leqslant k$, define

$$\mathcal{W}^{(m)}_{\geqslant 2}=\mathcal{W}_{\geqslant 2}\cup\{W_1,\dots,W_m\}.$$

Clearly, $\mathcal{W}^{(k)}_{\geqslant 2}=\mathcal{W}_{\geqslant 2}(H)$, and for every $1\leqslant m\leqslant k$, we have $\mathcal{W}^{(m)}_{\geqslant 2}=\mathcal{W}^{(m-1)}_{\geqslant 2}\cup\{W_m\}$.

From Lemma B.83, we know that for every $1\leqslant m\leqslant k$, we have

$$\frac{K_{\mathcal{W}^{(m-1)}_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}^{(m)}_{\geqslant 2}}(\mathbf{x})}\leqslant\frac{\left(1+\frac{1}{\sqrt{n}}\right)\cdot\left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})|}\cdot\left(\frac{2d}{n\sqrt{n}}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})|}}{\left(\frac{n}{3d}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})\cap W_m|}\cdot\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}.$$

Therefore,

$$\begin{aligned}\frac{K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x})}&=\frac{K_{\mathcal{W}^{(0)}_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}^{(k)}_{\geqslant 2}}(\mathbf{x})}=\prod_{m=1}^{k}\frac{K_{\mathcal{W}^{(m-1)}_{\geqslant 2}}(\mathbf{x})}{K_{\mathcal{W}^{(m)}_{\geqslant 2}}(\mathbf{x})}\\&\leqslant\frac{\left(1+\frac{1}{\sqrt{n}}\right)^k\cdot\prod_{m=1}^{k}\left[\left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})|}\cdot\left(\frac{2d}{n\sqrt{n}}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})|}\right]}{\prod_{m=1}^{k}\left[\left(\frac{n}{3d}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})\cap W_m|}\cdot\prod_{uv\in E^a_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}})\setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})}\left[\left(1+\frac{\varepsilon\mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]\right]}.\end{aligned}\tag{B.3.42}$$

Now since

$$\mathcal{W}_{\geqslant 2}^{(0)} \subseteq \mathcal{W}_{\geqslant 2}^{(1)} \subseteq \ldots \subseteq \mathcal{W}_{\geqslant 2}^{(m)},$$

we have

$$E_{\geqslant 2}(\mathcal{W}_{\geqslant 2}^{(0)}) \subseteq E_{\geqslant 2}(\mathcal{W}_{\geqslant 2}^{(1)}) \subseteq \ldots \subseteq E_{\geqslant 2}(\mathcal{W}_{\geqslant 2}^{(m)}),$$

and

$$\begin{aligned}\prod_{m=1}^{k}\left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}^{(m)}})\setminus E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}^{(m-1)}})|} &= \prod_{m=1}^{k}\left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}^{(m)}})|-|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}^{(m-1)}})|} \\ &= \left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}^{(k)}})|-|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}^{(0)}})|} \\ &= \left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})|-|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|}.\end{aligned} \tag{B.3.43}$$

On the other hand, we have

$$\left(1+\frac{1}{\sqrt{n}}\right)^{k} = 1 + O\left(\frac{k}{\sqrt{n}}\right) \leqslant 1 + O\left(\frac{|\mathcal{W}_{\geqslant 2}(H)|}{\sqrt{n}}\right) \leqslant 1 + O\left(\frac{|\mathcal{W}(H)|}{\sqrt{n}}\right) = 1 + O\left(\frac{t}{\sqrt{n}}\right) \leqslant 2, \tag{B.3.44}$$

where the last inequality is true for n large enough.

Furthermore, from Definition B.81, we have

$$\begin{aligned}K_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) &= \frac{\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}(H)})|}}{\prod_{v\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})} n^{\frac{1}{4}\left(d_{\geqslant 2}^{H_{\mathcal{W}_{\geqslant 2}(H)}}(v)-\Delta\right)}} \cdot \left(\frac{2d}{n}\right)^{|E_{\geqslant 2}^{b}(H_{\mathcal{W}_{\geqslant 2}(H)})|} \cdot \prod_{uv\in E_{\geqslant 2}^{a}(H_{\mathcal{W}_{\geqslant 2}(H)})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right] \\ &= R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \prod_{uv\in E_{\geqslant 2}^{a}(H_{\mathcal{W}_{\geqslant 2}(H)})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right],\end{aligned} \tag{B.3.45}$$

where

$$R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) = \frac{\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H_{\mathcal{W}_{\geqslant 2}(H)})|}}{\prod_{v\in\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})} n^{\frac{1}{4}\left(d_{\geqslant 2}^{H_{\mathcal{W}_{\geqslant 2}(H)}}(v)-\Delta\right)}} \cdot \left(\frac{2d}{n}\right)^{|E_{\geqslant 2}^{b}(H_{\mathcal{W}_{\geqslant 2}(H)})|}.$$

Therefore,

$$\begin{aligned}&\frac{K_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x})}{\prod_{m=1}^{k}\prod_{uv\in E_{\geqslant 2}^{a}(H_{\mathcal{W}_{\geqslant 2}^{(m)}})\setminus E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}^{(m-1)}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]} \\ &= R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \frac{\prod_{uv\in E_{\geqslant 2}^{a}(H_{\mathcal{W}_{\geqslant 2}(H)})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}{\prod_{m=1}^{k}\prod_{uv\in E_{\geqslant 2}^{a}(H_{\mathcal{W}_{\geqslant 2}^{(m)}})\setminus E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}^{(m-1)}})}\left[\left(1+\frac{\varepsilon \mathbf{x}_u\mathbf{x}_v}{2}\right)\frac{d}{n}+\frac{d^2}{n^2}\right]}\end{aligned}$$

$$= R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) \setminus \left[\bigcup_{m=1}^{k} \left(E^a_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})\right)\right]} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right].$$

By combining this with Eq. (B.3.42), Eq. (B.3.43), Eq. (B.3.44), we get

$$\begin{aligned}
K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) &\leqslant 2 \cdot \left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) \setminus E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \\
&\quad \times \prod_{m=1}^{k} \left[\left(\frac{3d}{n}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}}) \cap W_m|} \cdot \left(\frac{2d}{n\sqrt{n}}\right)^{|E^a_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})|}\right] \\
&\quad \times \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) \setminus \left[\bigcup_{m=1}^{k} \left(E^a_{\geqslant 2}(H_{\mathcal{W}^{(m)}_{\geqslant 2}}) \setminus E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}})\right)\right]} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2}\right] \\
&\leqslant 2 \cdot \left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) \setminus E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \prod_{m=1}^{k} \left(\frac{3d}{n}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}}) \cap W_m|} \\
&\quad \times \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{d^2}{n^2} + \frac{2d}{n\sqrt{n}}\right] \\
&\leqslant 2 \cdot \left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) \setminus E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})|} \cdot R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \prod_{m=1}^{k} \left(\frac{3d}{n}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}}) \cap W_m|} \\
&\quad \times \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right] \\
&\leqslant 2 \cdot \left(\frac{9d}{2\sqrt{n}}\right)^{|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) \setminus E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})| + \sum_{m=1}^{k} |E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}}) \cap W_m|} \cdot R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \\
&\quad \times \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right].
\end{aligned}$$

Now observe that

$$\begin{aligned}
&|E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) \setminus E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}})| + \sum_{m=1}^{k} |E_{\geqslant 2}(H_{\mathcal{W}^{(m-1)}_{\geqslant 2}}) \cap W_m| \\
&\qquad \geqslant |E_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) \setminus E_{\geqslant 2}(H_{\mathcal{W}^{(k-1)}_{\geqslant 2}})| + |E_{\geqslant 2}(H_{\mathcal{W}^{(k-1)}_{\geqslant 2}}) \cap W_k| \\
&\qquad = |E_1(H_{\mathcal{W}^{(k-1)}_{\geqslant 2}}) \cap W_k| + |E_{\geqslant 2}(H_{\mathcal{W}^{(k-1)}_{\geqslant 2}}) \cap W_k| = |E(H_{\mathcal{W}^{(k-1)}_{\geqslant 2}}) \cap W_k| \geqslant 1,
\end{aligned}$$

where the last inequality follows from the fact that $W_k \in \mathcal{W}_{\geqslant 2}(H)$, which means that there is at least one edge in W_k that already appears in $E(H_{\mathcal{W}^{(k-1)}_{\geqslant 2}})$.

Therefore, for n large enough, we have

$$K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) \leqslant 2 \cdot \left(\frac{9d}{2\sqrt{n}}\right) \cdot R_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \prod_{uv \in E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)})} \left[\left(1 + \frac{\varepsilon \mathbf{x}_u \mathbf{x}_v}{2}\right)\frac{d}{n} + \frac{3d^2}{n\sqrt{n}}\right] = \frac{9d}{\sqrt{n}} \cdot \hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}).$$

□

Now we are ready to prove Lemma B.47.

Proof of Lemma B.47. We know from Lemma B.84 that for every $\mathcal{W}_{\geqslant 2} \subsetneq \mathcal{W}_{\geqslant 2}(H)$, we have

$$K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) \leqslant \frac{9d}{\sqrt{n}} \cdot \hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}).$$

On the other hand, it is easy to see that

$$K_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \leqslant \hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}).$$

Now from Lemma B.82 we get

$$\begin{aligned}
&\sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} F_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) \\
&\qquad \leqslant \sum_{\mathcal{W}_{\geqslant 2} \subseteq \mathcal{W}_{\geqslant 2}(H)} K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) = K_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) + \sum_{\mathcal{W}_{\geqslant 2} \subsetneq \mathcal{W}_{\geqslant 2}(H)} K_{\mathcal{W}_{\geqslant 2}}(\mathbf{x}) \\
&\qquad \leqslant \hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \left(1 + \sum_{\mathcal{W}_{\geqslant 2} \subsetneq \mathcal{W}_{\geqslant 2}(H)} \frac{9d}{\sqrt{n}}\right) = \hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \left[1 + \frac{9d}{\sqrt{n}}\left(2^{|\mathcal{W}_{\geqslant 2}(H)|} - 1\right)\right] \\
&\qquad \leqslant \hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \left(1 + \frac{9d}{\sqrt{n}} \cdot e^t\right) = \hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}) \cdot \left(1 + \frac{9d}{\sqrt{n}} \cdot n^K\right) \leqslant 2\hat{K}_{\mathcal{W}_{\geqslant 2}(H)}(\mathbf{x}),
\end{aligned} \tag{B.3.46}$$

where the last inequality is true if $K \leqslant \frac{1}{100}$ and n is large enough.

By noticing that $E^a_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) = E^a_{\geqslant 2}(H)$, $E^b_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) = E^b_{\geqslant 2}(H)$, $\mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) = \mathcal{L}_{\geqslant 2}(H)$ and that $d^{H_{\mathcal{W}_{\geqslant 2}(H)}}_{\geqslant 2}(v) = d^H_{\geqslant 2}(v)$ for every $v \in \mathcal{L}_{\geqslant 2}(H_{\mathcal{W}_{\geqslant 2}(H)}) = \mathcal{L}_{\geqslant 2}(H)$, if we combine Eq. (B.3.46) with Eq. (B.3.41), we get Eq. (B.2.9). □

B.4 Tools for block self-avoiding walks

B.4.1 Splitting the expectation of block self-avoiding walks

In this section we prove Fact 4.65, Fact 4.94 and Fact 4.58.

Fact B.85 (Restatement of Fact 4.65)**.** *Let $\mathcal{B} = \{B_1, \dots, B_z\}$ be a collection of disjoint connected graphs on at least 2 vertices. Then for any x, $H \in \mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ and $j \in [z]$*

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{4}n^{-1/25A}\left(\frac{6}{\varepsilon}\right)^{2\ell_j+2q_j} \mathbb{E}\,\overline{U}_{H(V,V\setminus B_j)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus B_j)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(B_j)}(\mathbf{x})\,,$$
$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \geqslant \frac{1}{4}n^{-1/25A}\,\mathbb{E}\,\overline{U}_{H(V,V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(B_i)}(\mathbf{x})\,.$$

Proof. The first inequality follows observing that there are at most $\ell_i + q_i$ edges in $E_1^a(H)$ incident to $H(B_i)$ and observing that the edges in the multiway cut separating $H(B_1), \dots, H(B_z)$ have multiplicity 1. The second inequality is a consequence of Definition 4.55 and the fact that for any $B_i \in \mathcal{B}$, the edges in the cut $H(V \setminus B_i, B_i)$ have multiplicity 1. □

Fact B.86 (Restatement of Fact 4.94)**.** *Consider the settings of Theorem 4.44. Let $u, v \in [n]$, let $\mathcal{B} = \{B_1, \dots, B_z\}$ be a collection of disjoint connected graphs on at least 2 vertices and let B^{uv} be a (possibly empty) graph disjoint from any graph in $\mathcal{B}$. Then for any x, $H \in \mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uv})$ and $i \in [z+1]$*

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{4}n^{-1/25A}\left(\frac{6}{\varepsilon}\right)^{2\ell_i+2q_i} \mathbb{E}\,\overline{U}_{H(V,V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(B_i)}(\mathbf{x})\,,$$
$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \geqslant \frac{1}{4}n^{-1/25A}\,\mathbb{E}\,\overline{U}_{H(V,V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus B_i)}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(B_i)}(\mathbf{x})\,.$$

Proof. The proof is similar to the one of Fact 4.65. There are at most $\ell_i + q_i$ edges in $E_1^a(H)$ incident to $H(B_i)$ and observing that the edges in the multiway cut separating $H(B_1), \dots, H(B_z)$ have multiplicity 1. The second inequality follows by Definition 4.55 and definition of $\mathcal{M}_{\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+1}}(\mathcal{B}, B^{uv})$. □

Fact B.87 (Restatement of Fact 4.58)**.** *Consider the settings of Theorem 4.43 and let $\Psi \geqslant 0$. Let $H \in BSAW_{s,t}$ be a multigraph on at most $O(t)$ vertices and let H^* be an induced sub-multigraph of H satisfying:*

1. *the maximum ($\geqslant 2$)-degree in H^* is $\Psi \geqslant 0$,*
2. *all the edges in the cut $H(V(H), V(H) \setminus V(H^*))$ have multiplicity one in H.*

We denote $\ell, q \geqslant 0$ as the number of multiplicity-1 edges in H^* and $H(V(H), V(H) \setminus V(H^*))$ respectively. Let Z be a set of vertices in $V(H^*)$ such that $H(V(H^*) \setminus Z)$ has no multiplicity-2 cycles. Then

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{4}n^{-1/25A}\left(\frac{6}{\varepsilon}\right)^{2\ell+2q} \mathbb{E}\,\overline{U}_{H(V,V\setminus V(H^*))}(\mathbf{x}) \cdot \mathbb{E}\,\overline{U}_{H(V\setminus V(H^*))}(\mathbf{x})$$
$$\cdot \left(1+\frac{\varepsilon}{2}\right)^{|Z|\cdot\Psi} \cdot \left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|} \cdot \prod_{v\in V(H^*)} \left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v)-\tau,0\right\}}$$

$$\cdot \frac{1}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|}\prod_{e\in E^a_{\geqslant 2}(H^*)}\left[\frac{d}{n}+\frac{3d^2}{n\sqrt{n}}\right],$$

and

$$\begin{aligned}\mathbb{E}\,\overline{U}_H(\mathbf{x}) \geqslant & \frac{1}{4}n^{-1/25A}\,\mathbb{E}\,\overline{U}_{H(V,V\setminus V(H^*))}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(V\setminus V(H^*))}(\mathbf{x})\\ &\cdot\left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|}\cdot\prod_{v\in V(H^*)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d^H_1(v)-\tau,0\right\}}\\ &\cdot\frac{1}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|}\prod_{e\in E^a_{\geqslant 2}(H^*)}\left[\frac{d}{n}+\frac{3d^2}{n\sqrt{n}}\right],\end{aligned}$$

Proof. Consider the first inequality. We denote multigraph H' as the multigraph obtained by removing all edges incident to vertice in Z. Then there is no multiplicity-2 cycles in H. By the Fact B.86(or Fact B.85), we have

$$\mathbb{E}\,\overline{U}_H(\mathbf{x}) \leqslant \frac{1}{4}n^{-1/25A}\left(\frac{6}{\varepsilon}\right)^{2\ell+2q}\mathbb{E}\,\overline{U}_{H(V,V\setminus V(H^*))}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(V\setminus V(H^*))}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H*}(\mathbf{x})$$

It remains to bound $\mathbb{E}\,\overline{U}_{H*}(\mathbf{x})$. We note that

$$\begin{aligned}\mathbb{E}\,\overline{U}_{H*}(\mathbf{x}) =& \left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|}\cdot\prod_{v\in V(H^*)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d^H_1(v)-\tau,0\right\}}\\ &\cdot\frac{1}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|}\\ &\mathbb{E}\prod_{e\in E^a_{\geqslant 2}(H^*)}\left[\frac{d}{n}(1+\frac{\varepsilon}{2}x_ix_j)+\frac{3d^2}{n\sqrt{n}}\right]\\ \leqslant&\left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|}\cdot\prod_{v\in V(H^*)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d^H_1(v)-\tau,0\right\}}\\ &\cdot\frac{1}{\prod\limits_{v\in\mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|}\\ &(1+\frac{\varepsilon}{2})^{|Z|\cdot\Psi}\,\mathbb{E}\prod_{e\in E^a_{\geqslant 2}(H^*)\cap E(H')}\left[\frac{d}{n}(1+\frac{\varepsilon}{2}x_ix_j)+\frac{3d^2}{n\sqrt{n}}\right]\\ =&\left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|}\cdot\prod_{v\in V(H^*)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d^H_1(v)-\tau,0\right\}}\end{aligned}$$

$$
\begin{aligned}
&\cdot \frac{1}{\prod_{v\in\mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|}\\
&\cdot (1+\frac{\varepsilon}{2})^{|Z|\cdot\Psi}\prod_{e\in E^a_{\geqslant 2}(H^*)\cap E(H')}\left(\frac{d}{n}+\frac{3d^2}{n\sqrt{n}}\right)\\
\leqslant&\left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|}\cdot\prod_{v\in V(H^*)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v)-\tau,0\right\}}\\
&\cdot \frac{1}{\prod_{v\in\mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|}\\
&\cdot (1+\frac{\varepsilon}{2})^{|Z|\cdot\Psi}\prod_{e\in E^a_{\geqslant 2}(H^*)}\left(\frac{d}{n}+\frac{3d^2}{n\sqrt{n}}\right)
\end{aligned}
$$

The first inequality follows from by observing that observing that $|E(H^*)\setminus E(H')|\leqslant |Z|\cdot\Psi$. The second inequality follows since H' does not contain any multiplicity-2 cycle. The first bound thus follows.

For the second inequality, we still use Fact B.86.

$$
\mathbb{E}\,\overline{U}_H(\mathbf{x})\leqslant\frac{1}{4}n^{-1/25A}\,\mathbb{E}\,\overline{U}_{H(V,V\setminus V(H^*))}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H(V\setminus V(H^*))}(\mathbf{x})\cdot\mathbb{E}\,\overline{U}_{H^*}(\mathbf{x})
$$

We note that for any H^*

$$
\mathbb{E}\prod_{e\in E^a_{\geqslant 2}(H^*)}\left(\frac{d}{n}(1+\frac{\varepsilon}{2}x_ix_j)+\frac{3d^2}{n\sqrt{n}}\right)\geqslant\prod_{e\in E^a_{\geqslant 2}(H^*)}\left(\frac{d}{n}+\frac{3d^2}{n\sqrt{n}}\right)
$$

Therefore we have

$$
\begin{aligned}
\mathbb{E}\,\overline{U}_{H^*}(\mathbf{x})=&\left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|}\cdot\prod_{v\in V(H^*)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v)-\tau,0\right\}}\\
&\cdot\frac{1}{\prod_{v\in\mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|}\\
&\mathbb{E}\prod_{e\in E^a_{\geqslant 2}(H^*)}\left[\frac{d}{n}(1+\frac{\varepsilon}{2}x_ix_j)+\frac{3d^2}{n\sqrt{n}}\right]\\
\geqslant&\left(\frac{2d}{n}\right)^{\left|E^b_{\geqslant 2}(H^*)\right|}\cdot\prod_{v\in V(H^*)}\left(\frac{6}{\varepsilon}\right)^{\max\left\{2d_1^H(v)-\tau,0\right\}}\\
&\cdot\frac{1}{\prod_{v\in\mathcal{L}_{\geqslant 2}(H^*)} n^{\frac{1}{4}\left(d^{H^*}_{\geqslant 2}(v)-\Delta\right)}}\left(\frac{\varepsilon d}{2n}\right)^{|E_1(H^*)|}
\end{aligned}
$$

$$\prod_{e \in E^a_{\geqslant 2}(H^*)} \left[\frac{d}{n} + \frac{3d^2}{n\sqrt{n}} \right]$$

The claim thus follows. □

B.4.2 Counting block self-avoiding walks

B.4.2.1 Bounding number of BSAW given a subgraph

In this section we provide the counting arguments needed in Section 4.5. We reuse the notation of such section and we assume the premises of Theorem 4.43 and Theorem 4.44 to hold.

Lemma B.88. *Let B_i be a connected subgraph of the underlying graph of a block self-avoiding walk $H \in M_{s,t}$ with m_i vertices. Let $H(B_i)$ be the multigraph on $V(B_i)$ that is induced by $H(B_i)$. Suppose that:*

- *The cut $H(V(B_i), V(H) \setminus V(B_i))$ consists of ℓ_i edges of multiplicity 1.*
- *All edges in $H(B_i)$ of multiplicity $\geqslant 2$ are in B_i.*
- *The number of edges of multiplicity 1 in $H(B_i)$ is q_i.*
- *The number of edges of multiplicity 2 in $H(B_i)$ is h_i.*
- *The edges of multiplicity larger than 2 satisfy*

$$\sum_{\substack{e \in H(B_i) \\ m_H(e) \geqslant 3}} m_H(e) = p_i.$$

Then we have

$$\sum_{\substack{v \in V(B_i) \\ d^H(v) \geqslant 3}} \left(\frac{m_H(v)}{2} - 1 \right) \leqslant (8h_i + 4p_i)/s + 11q_i + 8\ell_i + 4p_i + 6(h_i - m_i + 1)\,,$$

where $d^H(v)$ is the degree of v in the underlying graph $G(H)$, i.e., without counting multiplicities, and

$$m_H(v) = \sum_{\substack{e \in E(H), \\ e \text{ is incident to } v}} m_H(e).$$

Proof. Let B^* be a spanning tree of B_i. We start by deriving an upper bound on the number of leaves of B^*. Notice that if a vertex $v \in V(B^*)$ is a leaf of B^*, then it must satisfy at least one of the following three conditions:

(a) v is incident to an edge in the cut $H(V(B_i), V(H) \setminus V(B_i))$.

(b) v is incident to an edge in $G(H(B_i)) \setminus B^*$.

(c) v is a pivot of H, i.e., v is an end-vertex of one of the s self-avoiding walks forming H.

We will now upper-bound the number of leaves of each kind:

- Since each edge in $H(V(B_i), V(H) \setminus V(B_i))$ is incident to exactly one vertex in $V(B_i)$, there are at most ℓ_i leaves of B^* satisfying Condition (a).
- Since there are at most $\left(q_i + h_i + \frac{p_i}{3}\right) - (m_i - 1)$ edges in $G(H(B_i)) \setminus B^*$, and since each edge is incident to exactly two vertices, there are at most $2q_i + 2h_i + \frac{2}{3}p_i - 2(m_i - 1)$ leaves of B^* satisfying Condition (b).
- Each pivot vertex must be an end-vertex of a self-avoiding walk of H. There are at most $(q_i + 2h_i + p_i)/s$ self-avoiding walks of H that lie entirely in $H(B_i)$, and there are at most ℓ_i self-avoiding walks of H that intersect $H(B_i)$ without being entirely in $H(B_i)$. Hence, $H(B_i)$ intersects at most $(q_i + 2h_i + p_i)/s + \ell_i$ self-avoiding walks of H. Now since each self-avoiding walk contains exactly 2 end-vertices, we deduce that the number of pivots of H in $V(B_i)$ is at most

$$2(q_i + 2h_i + p_i)/s + 2\ell_i. \tag{B.4.1}$$

We now conclude that there are at most $2(q_i + 2h_i + p_i)/s + 2\ell_i$ leaves of B^* satisfying Condition (c).

Therefore, the number of leaves of B^* is at most

$$\begin{aligned}\ell_i + 2q_i + 2h_i + \frac{2}{3}p_i - 2(m_i - 1) + 2(q_i + 2h_i + p_i)/s + 2\ell_i \\ \leqslant (4h_i + 2p_i)/s + 3\ell_i + 4q_i + p_i + 2(h_i - m_i + 1)\end{aligned}$$

Let $d^{B^*}(v)$ be the degree of a vertex $v \in B^*$. It is easy to see that the number of leaves in B^* is equal to

$$2 + \sum_{\substack{v \in V(B_i) \\ d^{B^*}(v) \geqslant 3}} (d^{B^*}(v) - 2).$$

Therefore,

$$\sum_{\substack{v \in V(B_i) \\ d^{B^*}(v) \geqslant 3}} (d^{B^*}(v) - 2) \leqslant (4h_i + 2p_i)/s + 3\ell_i + 4q_i + p_i + 2(h_i - m_i + 1). \tag{B.4.2}$$

Now define

$$m_{H,2}(v) = \sum_{\substack{e \in E(H), \\ e \text{ is incident to } v, \\ m_H(e) = 2}} m_H(e)\,,$$

and

$$m_{H,\neq 2}(v) = \sum_{\substack{e\in E(H),\\ e \text{ is incident to } v,\\ m_H(e)\neq 2}} m_H(e)\,.$$

Clearly, $m_H(v) = m_{H,2}(v) + m_{H,\neq 2}(v)$. Therefore,

$$\begin{aligned}
\sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3}} \left(\frac{m_H(v)}{2} - 1\right) &= \sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3}} \left(\frac{m_{H,2}(v)}{2} - 1\right) + \sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3}} \frac{m_{H,\neq 2}(v)}{2} \\
&\overset{(a)}{\leqslant} \sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3}} \left(d^{B_i}(v) - 1\right) + \sum_{v\in V(B_i)} \frac{m_{H,\neq 2}(v)}{2} \\
&= \sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3}} \left(d^{B^*}(v) - 1\right) + \sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3}} \left(d^{B_i}(v) - d^{B^*}(v)\right) + \frac{\ell_i}{2} + \sum_{\substack{e\in H(B_i)\\ m_H(e)\neq 2}} m_H(e) \\
&= \sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3}} \left(d^{B^*}(v) - 1\right) + 2\big(|E(B_i)| - |E(B^*)|\big) + \frac{\ell_i}{2} + q_i + p_i \\
&\leqslant \sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3}} (d^{B^*}(v) - 1) + 2\left(h_i + \frac{p_i}{3} - (m_i - 1)\right) + \frac{\ell_i}{2} + q_i + p_i \\
&\leqslant \sum_{\substack{v\in V(B_i)\\ d^{B^*}(v)\geqslant 3}} (d^{B^*}(v) - 1) + \left(\sum_{\substack{v\in V(B_i)\\ d^H(v)\geqslant 3\\ d^{B^*}(v)=2}} 1\right) + 2(h_i - m_i + 1) + \ell_i + q_i + 2p_i \\
&\overset{(b)}{\leqslant} 2\cdot \sum_{\substack{v\in V(B_i)\\ d^{B^*}(v)\geqslant 3}} (d^{B^*}(v) - 2) + (2q_i + \ell_i) + 2(h_i - m_i + 1) + \ell_i + q_i + 2p_i \\
&\overset{(c)}{\leqslant} (8h_i + 4p_i)/s + 6\ell_i + 8q_i + 2p_i + 4(h_i - m_i + 1) \\
&\qquad\qquad + 2(h_i - m_i + 1) + 2\ell_i + 3q_i + 2p_i \\
&= (8h_i + 4p_i)/s + 11q_i + 8\ell_i + 4p_i + 6(h_i - m_i + 1)\,,
\end{aligned}$$

where (a) follows from the fact that every edge of multiplicity 2 and incident to $v \in V(B_i)$ must be in B_i, hence $m_{H,2}(v) \leqslant 2d^{B_i}(v)$. (b) follows from the fact that if $v \in V(B_i)$ satisfies $d^H(v) \geqslant 3$ and $d^{B_i}(v) = 2$, then v must be incident to an edge of multiplicity 1, i.e., v must be incident to an edge in $E(H(B_i)) \setminus E(B_i)$, or an edge in the cut $H(V(B_i), V(H) \setminus V(B_i))$. There are q_i edges in $E(H(B_i)) \setminus E(B_i)$, each of which is incident to two vertices in $V(B_i)$, and there are ℓ_i edges in the cut $H(V(B_i), V(H) \setminus V(B_i))$, each of which is incident to one vertex in $V(B_i)$. (c) follows from (B.4.2). □

Lemma B.89. *Let $\mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ be as defined in Definition 4.62. We denote $\ell = \sum_{i=1}^z \ell_i$, and define q, p, h in the same way. Then, the number of block self-avoiding walks in the set $\mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ is at most*

$$2n^{st-2h-p-q-\frac{\ell}{2}+m} \cdot m^{2p+2q+\ell} \cdot (st)^{3\frac{\ell}{2}+q+2} \cdot \prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)}.$$

Proof. We would like to derive an upper bound on the number of block self-avoiding walks in $\mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$. We will do this by analyzing how we can construct a block self-avoiding walk $H \in \mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$. In order to construct H, we will need to make some choices, and by counting the number of possibilities of each choice, we can derive an upper-bound on the size of $\mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$.

We first choose the vertices of the respective copies $B'_1, B'_2, \ldots, B'_z$ of $B_1, B_2, \ldots, B_z$ inside H. There are at most n^m possibilities for choosing these vertices.

We denote the union of the z disjoint graphs $B'_1, B'_2, \ldots, B'_z$ as

$$B' = \bigcup_{i=1}^{z} B'_i.$$

The size of multiway cut $H\left(V(B'_1), \ldots, V(B'_z), V(H) \setminus \left(\bigcup_{j\in[z]} V(B'_z)\right)\right)$ is bounded by ℓ, and all the edges in the cuts have multiplicity 1. Let $H(B')$ be the multi-graph that is induced by H on $V(B')$. There are q multiplicity-1 edges and h multiplicity-2 edges in $H(B')$. The edges of multiplicity at least 3 in $H(B')$ satisfy the following equation:

$$\sum_{\substack{e\in H(B')\\ m_H(e)\geqslant 3}} m_H(e) = p.$$

Next, we choose the multi-graph $H(B')$ that is induced by H on $V(B')$, i.e., we have to choose multiplicities for the edges of B' (which must be at least 2), and we have to choose q edges of multiplicity 1 whose both end-vertices must lie in $V(B')$. Note that there are p multi-edges in $H(B')$ that correspond to edges of multiplicity at least 3, so specifying these multi-edges is equivalent to specifying the multiplicities of the edges of B', because once we specify the edges of B' of multiplicity at least 3, any remaining edge in B' must have multiplicity-2. Since both end-vertices of these multi-edges lie in $V(B')$ and since $|V(B')| = m$, there are at most m^{2p} ways to specify the p multi-edges. For choosing the q multiplicity-1 edges, there are at most m^{2q} ways of doing so.

The remaining of the multi-graph H consists of edges of multiplicity-1. More precisely, since H is a block self-avoiding walk, the remaining edges of H can be partitioned into a number of disjoint walks of multiplicity 1: Each time we exit $V(B')$, we exit through one of the cut edges, we go through a walk of multiplicity 1, and then re-enter $V(B')$ through

another cut edge. We call these walks as *outside walks*. We call the first edge of each outside walk as an *outgoing cut-edge*, and we call the last edge of each outside walk as a *returning cut-edge*. The remaining edges of the outside walks are called *outside edges*.

Note that there might be some cut-edges that are outgoing and returning at the same time. This can happen if a cut-edge is incident to two disjoint connected components B_i and B_j with $i \neq j$. We call such cut-edges as *bridge cut-edges*. Let b be the number of bridge cut-edges. It is easy to see the following:

- The total number of cut-edges is equal to $\ell - b$.
- The total number of non-bridge cut-edges is $\ell - 2b$.
- The total number of non-bridge outgoing cut-edges is $\frac{\ell-2b}{2} = \frac{\ell}{2} - b$. Similarly, the total number of non-bridge returning cut-edges is $\frac{\ell}{2} - b$.
- There are b outside walks of length 1. These correspond to bridge cut-edges.
- There are $\frac{\ell-2b}{2} = \frac{\ell}{2} - b$ outside walks of length at least 2. We call these outside walks as *non-bridge outside walks*.

As we can see, the total number of outside walks is $\frac{\ell}{2}$.

Since both end-vertices of each bridge cut-edge must lie inside $V(B')$, there are at most m^{2b} ways of choosing them.

Now since one end-vertex of each non-bridge outgoing cut-edge must lie in $V(B')$ while the other must like outside $V(B')$, there are at most nm ways to choose each non-bridge outgoing cut-edge. Since there are $\frac{\ell}{2} - b$ non-bridge outgoing cut-edges, we have at most $(nm)^{\frac{\ell}{2}-b}$ ways to choose them. Now for each non-bridge outgoing cut-edge, we specify the length of the corresponding non-bridge outside walk. Since the length of each outside walk is at most st, there are at most $(st)^{\frac{\ell}{2}-b}$ ways for choosing the lengths of all non-bridge outside walks.

Next, we choose the outside edges. Since we have already specified the non-bridge outgoing cut-edges, this already specifies one end-vertex of the first outside edge in each of the $\frac{\ell}{2} - b$ non-bridge outside walks, so we can specify the first outside edge by only specifying the second end-vertex. Similarly, we can iteratively specify all the outside edges by successively specifying the second end-vertex for each outside edge. Since there are $st - 2h - p - q - \ell + b$ outside edges[37], there are at most $n^{st-2h-p-q-\ell+b}$ ways for specifying them.

Now we turn to specifying the non-bridge returning cut-edges. Since we have already specified the lengths of the $\frac{\ell}{2} - b$ non-bridge outside walks, we know when we have reached the last outside edge, and so the next edge in the walk must be a returning non-bridge cut-edge. Since the last outside edge of each non-bridge outside walk is already specified, this determines one end-vertex of the returning cut-edge and we only need to specify the

[37]Recall that the total number of cut edges is $\ell - b$.

other end-vertex that lies inside $V(B')$. Now since there are $\frac{\ell}{2} - b$ non-bridge returning cut-edges, there are at most $m^{\frac{\ell}{2}-b}$ ways to specify them.

So far, we have completely specified the multi-graph structure of H. However, the block self-avoiding walk structure is more than that. We have to specify an Eulerian walk of H which can be divided into s subwalks that are self-avoiding.

Now before starting to specify the vertices $v_1, \ldots, v_{st}$ in the sequence of the block self-avoiding walk H, we will further specify where exactly will the q multiplicity-1 edges of $H(B')$ and the $\ell - b$ cut-edges appear in the sequence $(v_1v_2),\ (v_2, v_3), \ldots, (v_{st-1}, v_{st}),\ (v_{st}, v_1)$, i.e., for each edge e among these edges, we will specify an index $k_e \in \{1, \ldots, st\}$ such that $e = (v_{k_e}, v_{k_e+1})$ if $k_e < st$, and $e = (v_{st}, v_1)$ if $k_e = st$. Since there are st choices for each index, there are at most $(st)^{q+\ell-b}$ possibilities to choose all these indices. Let $\mathcal{K}$ be the set of indices that we obtain.

We start by specifying the first and second vertices[38], which we denote as v_1 and v_2, respectively, and from there we iteratively specify each next vertex in the walk. We have at most $(st)^2$ choices for choosing the pair (v_1, v_2).

Assume that we have already specified the first k vertices $v_1, \ldots, v_k$ of the walk, where $k \geqslant 2$. The next vertex v_{k+1} must be adjacent to v_k in H. Notice the following:

- If $v_k \notin V(B')$, then since we know both v_{k-1} and v_k, we can deduce which non-bridge outside walk contains the edge (v_{k-1}, v_k). Now given that all the non-bridge outside walks have been specified, there is a unique choice for the vertex v_{k+1}: It is the next vertex in the outside walk.[39]

- If $k \in \mathcal{K}$, then the edge (v_k, v_{k+1}) has already been fixed to be one specific edge: It is either a cut-edge, or an edge of multiplicity 1 in $H(B')$. Therefore, there is a unique choice for v_{k+1}.

- If $v_k \in V(B_i)$ for some $i \in [z]$ and $k \notin \mathcal{K}$, the next edge must be an edge of multiplicity at least 2. Therefore, there are at most $d^H_{\geqslant 2}(v_k) \leqslant \Psi_i$ choices for v_{k+1}. We note that each vertex v is visited in H for $\frac{m_H(v)}{2}$ times, where

$$m_H(v) = \sum_{\substack{e \in E(H), \\ e \text{ is incident to } v}} m_H(e)\,.$$

  Furthermore, when a vertex $v \in V(B')$ is visited for the last time, there is only one choice for the next vertex.

[38] We specified the first two vertices instead of only specifying the first vertex in order to know where to go in case $v_1 \notin V(B')$. This will be made clear when we discuss how we choose the next vertex v_{k+1} assuming that we already specified $v_1 \ldots, v_k$.

[39] It is important to realize here that when we specified the non-bridge outside walks, we did not only specify the structure of the graph that is formed by the cut-edges and the outside edges: We also specified how the block self-avoiding walk H will pass through these edges and in what order.

Therefore, once we have fixed the multigraph, the number b of bridge cut-edges, the $\frac{\ell}{2} - b$ non-bridge outside walks, the indices of the cut-edges and the multiplicity-1 edges of $H(B')$ in the walk, and the first two vertices, the number of remaining choices to completely specify the block self-avoiding walk H is at most:

$$\prod_{i=1}^{z} \prod_{v \in V(B_i')} \Psi_i^{\frac{m_H(v)}{2}-1} = \prod_{i=1}^{z} \Psi_i^{\sum_{v \in V(B_i')} \frac{m_H(v)}{2}-1}$$
$$\leqslant \prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)},$$

where the last inequality follows from Lemma B.88.

Now since $0 \leqslant b \leqslant \frac{\ell}{2}$, we conclude that the size of $\mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})$ can be upper-bounded as follows:

$$\begin{aligned}
&\left|\mathcal{M}_{s,t,\{\mathcal{F}_i,\Psi_i\}_{i=1}^z}(\mathcal{B})\right| \\
&\leqslant \sum_{b=0}^{\lfloor \frac{\ell}{2} \rfloor} n^m \cdot m^{2p} \cdot m^{2q} \cdot m^{2b} \cdot (nm)^{\frac{\ell}{2}-b} \cdot (st)^{\frac{\ell}{2}-b} \cdot n^{st-2h-p-q-\ell+b} \cdot m^{\frac{\ell}{2}-b} \cdot (st)^{q+\ell-b} \cdot (st)^2 \\
&\qquad \times \prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)} \\
&= \left(\sum_{b=0}^{\lfloor \frac{\ell}{2} \rfloor} (st)^{-2b}\right) n^{st-2h-p-q-\frac{\ell}{2}+m} \cdot m^{2p+2q+\ell} \cdot (st)^{3\frac{\ell}{2}+q+2} \cdot \prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)} \\
&\leqslant \frac{1}{1-(st)^{-2}} \cdot n^{st-2h-p-q-\frac{\ell}{2}+m} \cdot m^{2p+2q+\ell} \cdot (st)^{3\frac{\ell}{2}+q+2} \cdot \prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)} \\
&\leqslant 2n^{st-2h-p-q-\frac{\ell}{2}+m} \cdot m^{2p+2q+\ell} \cdot (st)^{3\frac{\ell}{2}+q+2} \cdot \prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)},
\end{aligned}$$

where the last inequality is true when $st \geqslant 2$.

□

B.4.2.2 Counting nice block self-avoiding walks

We split the proof of Lemma 4.82 into two steps.

Lemma B.90. *We denote $NBSAW_{s,t,m,z,t_F}$ as the subset of $NBSAW_{s,t,m,z}$, in which all nice block self-avoiding walks contain t_F pivoting vertices incident to any multiplicity-2 edge. Then there are at most $n^{\ell+m-z}(20s)^{2t_F}\binom{\ell}{z}\binom{m+z-1}{z}\binom{t+z-1}{z}$ different multi-graphs associated with the block self-avoiding walks in $NBSAW_{s,t,m,z,t_F}$*

Proof. For any $H \in \mathrm{NBSAW}_{s,t,m,z}$, we denote the forest formed by all the multiplicity-2 edges in H as F_H. Then we note the leaves in F_H must be pivots in H, and there are t pivots in H. Thus we have $z \leqslant t$. We denote the length of the cycle as ℓ. Since there are $m - z$ multiplicity-2 edges, we have $\ell = st - 2(m - z)$. We pick the vertices in the cycle and choose the roots of trees in F_H. There are at most $n^\ell \binom{\ell}{z}$ such choices.

Next we generate each of the z trees. We first fix the number of vertices and leaves in each tree. Since there are at most t leaves and m vertices in F_H, we have at most $\binom{m+z-1}{z}\binom{t+z-1}{z}$ choices, which is also bounded by $(2es)^t$.

Then we bound the number of choices for each tree B_i in the F_H given the number of vertices m_i and the number of leaves p_i. Using lemma Fact B.105, there are at most $\binom{2m_i}{2t_i}2^{2t_i}$ shapes for a tree with m_i vertices and p_i leaves. Multiplying together we have the number of choices bounded by

$$\prod_{i=1}^{z}\binom{2m_i}{2p_i}2^{2p_i} \leqslant 2^{2p}\prod_{i=1}^{z}\binom{2m_i}{2p_i}$$

where $p = \sum_{i=1}^{z} p_i$ is the total number of leaves in the forest. The inequality follows from the fact that each leaf must be a pivot. Further we observe that $m_i \leqslant s(t_i + 2)$, where $t_i \geqslant z_i$ is the number of pivots contained in the tree B_i. This follows since the tree is at most incident to $t_i + 2$ blocks of self-avoiding walks in H. Thus we have

$$\prod_{i=1}^{z}\binom{2m_i}{2p_i} \leqslant \prod_{i=1}^{z}\binom{2m_i}{2t_i} \leqslant \prod_{i=1}^{z}\left(\frac{em_i}{t_i}\right)^{2t_i} \leqslant \prod_{i=1}^{z}(10s)^{2t_i}$$

We denote $t_F = \sum_{i=1}^{z} t_i$ as the number of pivots contained in the forest. Then we have

$$\prod_{i=1}^{z}\binom{2m_i}{2p_i} \leqslant (10s)^{2t_F}$$

Finally there are n^{m-z} ways of choosing vertices in the trees which are not contained by the multiplicity-1 cycle. Therefore in all, we have at most $n^{\ell+m-z}(20s)^{2t_F}\binom{\ell}{z}\binom{m+z-1}{z}\binom{t+z-1}{z}$ choices for the multigraph associated with such block self-avoiding walks. This leads to the claim. □

The following lemma is then a simple corollary:

Lemma B.91. *There are at most $n^{st-m+z}(20s)^{4t}$ different multi-graphs associated with the block self-avoiding walks in $\mathrm{NBSAW}_{s,t,m,z}$*

Proof. From the above lemma Lemma B.90, we already have bound $n^{\ell+m-z}(20s)^{2t_F}\binom{\ell}{z}\binom{m+z-1}{z}\binom{t+z-1}{z}$. Next we note that $z \leqslant t$, thus we have $\binom{\ell}{z} \leqslant (es)^t$ and $\binom{t+z-1}{z} \leqslant 2^{2t}$. Further we have $m \leqslant st$, thus $\binom{m+z-1}{z} \leqslant (2es)^t$. Therefore for $z \leqslant t_F \leqslant t$, we have

$$n^{\ell+m-z}(20s)^{2t_F}\binom{\ell}{z}\binom{m+z-1}{z}\binom{t+z-1}{z} \leqslant n^{\ell+m-z}(20s)^{2t_F}(2es)^{2t}$$

Finally by summing t_F from z to t, we have the claim. □

Finally we bound the number of nice block self-avoiding walks. For proving this, we observe a simple fact which will be useful for counting nice block self-avoiding walks here and future subsections. We say that the sequence of H enters a tree T at step k if the k-th vertex is not in T and the $k+1$-th vertex is the root of T. Similarly we say that the sequence of H leaves a tree T at step k if the k-th vertex is the root of T and the $k+1$-th vertex is not in T.

Fact B.92. *For each tree formed by a set of multiplicity-2 edges in the nice block self-avoiding walk H, once the sequence of H enters the tree through a cut, then all of the edges in the tree must be visited twice before the sequence of H leaves the tree through any cut edge.*

Proof. We note that the cutting edges of any multiplicity-2 tree are either one multiplicity-2 edge or two multiplicity-1 edges. Thus if the sequence of H enters the tree through a cut and then leaves through a cut, there is no cutting edge to take such that the sequence can enter the tree again. Therefore all of the edges in this tree must be visited between the entering cut edge and leaving cut edge. □

Now we conclude the bound on the number of nice block self-avoiding walks.

Lemma B.93. *The size of $NBSAW_{s,t,m,z}$ is bounded by*

$$\Delta^{3t}(40s)^{4t}n^{st-m+z}$$

, where $\Delta \leqslant st$ is the upper bound on the maximum degree $\geqslant 2$ in the nice block self-avoiding walks.

Proof. By lemma Lemma B.91, there are at most $n^{st-m+z}(20s)^{4t}$ multigraphs associated with block self-avoiding walks in the set $\mathrm{NBSAW}_{s,t,\ell,\Delta}$. Then fixing the associated multi-graph, we choose the block self-avoiding walk.

By truncation, the maximum degree-$\geqslant 2$ of vertices is bounded by Δ. There are at most t leaves in the forest of multiplicity-2 edges, and other vertices have degree at least 2. Therefore as in the proof of Lemma B.88, for one of the z trees B, we have

$$\sum_{\substack{v\in V(B)\\ \deg_B(v)\geqslant 3}} (\deg_B(v)-1) \leqslant 2t+z \leqslant 3t$$

By the above fact, for each root of the forest, all of the edges in associated tree must be visited twice in the sequence of H before the next adjacent vertex in the cycle is taken unless it is root of the tree containing the first vertex in the block self-avoiding walk. Further we denote r as the root of B, then $\deg_B(r)$ is smaller than the number of leaves in B.

For choosing the block self-avoiding walk based on the multigraph, we first decide the first edge in the walk, which takes at most Δst choices. Next if the first edge is contained in a multiplicity-2 tree B, then we choose the the cut edge leaving B and its index in the sequence of H, which takes $2st$ choices. After these two steps, the multiplicities of edges in

each index of the sequence of H is fixed. Finally we generate the sequence of H respecting these previous two steps. Then the number of choices for the block self-avoiding walk given the associated multigraph is bounded by

$$\Delta st \cdot 2st \prod_B \left[\prod_{\substack{v \in V(B) \\ \mathrm{ded}_B(v) \geqslant 3}} (\deg_B(v) - 1) \right] \leqslant 2^t \Delta 3t$$

In all, we have the total number of such block self-avoiding walks bounded by $\Delta^{3t}(40s)^{4t} n^{\ell - z + m}$ □

The lemma Lemma B.94 is a simple corollary

Lemma B.94 (Restatement of Lemma 4.82)**.** *Consider the settings of Theorem 4.43. Let $r \geqslant 0$ be an integer. For n large enough*

$$\sum_{\substack{m,z \geqslant 0 \\ m-z=r}} \left| NBSAW^*_{s,t,m,z} \right| \leqslant \Delta^{5t+\Delta} n^{st-r} .$$

Proof. Since $\Delta \geqslant s^2$, we have $\Delta^{3t}(40s)^{4t} n^{\ell - z + m} \leqslant (40)^{4t} \Delta^{5t} n^{\ell - r}$ No we sum m, z from 0 to t,

$$\sum_{\substack{m,z \geqslant 0 \\ m-z=r}} \left| \mathrm{NBSAW}^*_{s,t,m,z} \right| \leqslant t^2 (40)^{4t} \Delta^{5t} n^{st-r} \leqslant \Delta^{6t} .$$

□

B.4.2.3 Counting refined set of nice block self-avoiding walk

Here we prove Lemma 4.83

Lemma B.95 (Restatement of Lemma 4.83)**.** *Consider the settings of Theorem 4.43. Let $m, z \geqslant 0$ be integers such that $0 \leqslant m - z < \frac{t\sqrt{s}}{2}$. Then for n large enough*

$$\left| N_{w,q,m,z} \right| \leqslant 2^{ws} \cdot (1+\delta)^{\frac{t}{10}} \cdot n^{-q} \cdot \left| NBSAW^*_{s,t,d^H \leqslant \Delta, m', z'} \right|$$

for some m', z' such that $m' - z' = m - z - q$. Moreover, it holds that $q > t - 2t/\sqrt{s}$.

Proof. Let $H \in \mathrm{N}_{w,q,m,z}$. Denote by q^* the number of pivots in H that are incident to an edge of multiplicity 2. We have

$$|E_{\geqslant 2}(H)| = m - z \geqslant q + (q^* - q) \cdot s/2$$

since if a pivot has edges of multiplicity 2 and it is not a leaf, then one of the adjacent walks must have all its edges of multiplicity 2. By definition of $\mathrm{N}_{w,q,m,z}$, we must have

$q^* \geqslant t - 2t/s$ since otherwise $w \geqslant t/s$ and the set is empty. Suppose now that $q \leqslant q^* - t/\sqrt{s}$. By the above reasoning we have $m - z \geqslant \frac{t\sqrt{s}}{2}$ contradicting our assumption. Thus it must be that $q > q^* - t/\sqrt{s} \geqslant t - 2t/\sqrt{s}$.

Now we need to bound the number of different ways in which we can choose w walks to be in $W_1(H)$, q^* pivots among t and q leaves among q^* pivots. This can be bounded by

$$\binom{q^*}{t-q^*-w}\binom{w}{t-q^*-w}\binom{q^*}{q} \leqslant 2^{t/\sqrt{s}} \cdot \binom{q^*}{t-q^*-w}^2 \leqslant 2^{t/\sqrt{s}} \cdot \binom{q^*}{w}^2 \leqslant s^{10t/s} \leqslant (1+\delta)^{t/10}. \tag{B.4.3}$$

We are almost ready to carry out the counting argument. [T: To update from here] First we introduce additional notation. For a nice block self-avoiding walk H_1 with edges u_1u_2, u_2u_3, u_3, u_4 of multiplicity 1. We say that H_2 is obtained from H_1 folding u_2 if H_2 obtained from H replacing the edges u_2u_3, u_3u_4 with the edges u_2u_1, u_2u_4. Notice that if $d^{H_2}(u) = d^{H_1}(u) + 1$, thus if $d^{H_1}(u) < \Delta$ then H_2 is a nice block self-avoiding walk. Moreover if $|E_{\geqslant 2}(H_1)| = m_1$ then $|E_{\geqslant 2}(H_2)| = m_1 + 1$.

Now notice that each $H \in \mathrm{N}_{w,q,m}$ can be obtained from some H^* in $\mathrm{NBSAW}^*_{s,t} \cap \mathcal{M}(\emptyset)$ through a sequence of m foldings. Moreover, we may assume that the first q folding are those fixing the leaves of H. Let H^{**} be the nice block self-avoiding walk obtained from H^* by folding the q leaves of H and let $S^*\ast$ be the sequence of folding that results in H from H^{**}.

Applying $S^*\ast$ to H^* we obtain a nice block self-avoiding walk H' in $\mathrm{NNBSAW}^c_{s,t}$ with $|E_{\geqslant 2}(H')| = |E_{\geqslant 2}(H)| - q$. Moreover Eq. (B.4.3) bounds the number of ways that a walk H' may arise with this process from non-isomorphic $H \in \mathrm{N}_{w,q,m}$. The result follows.

□

B.4.2.4 Counting nice walks with large degree

Here we prove Lemma 4.70.

Lemma B.96. *Consider the settings of Theorem 4.43. Define the set*

$$NBSAW^*_{s,t,m,z,\ell_1,\ell_2} := \Big\{H \in NBSAW^*_{s,t,m,z} \;\Big|\; \big|\{v \in v(H) \;|\; d^H(v) = \Delta+1\}\big| = \ell_1\,, \\ \big|\{v \in v(H) \;|\; d^H(v) = \Delta+2\}\big| = \ell_2\Big\}$$

Then for n large enough, there exists $t \geqslant \ell \geqslant \ell_1 + 2\ell_2$ such that

$$\left|NBSAW^*_{s,t,m,z,\ell_1,\ell_2}\right| \leqslant (1+o(1)) \cdot 2^t \cdot n^{-\ell} \left|NBSAW^*_{s,t,d^H\leqslant\Delta,m,z+\ell}\right|.$$

Proof. Our plan is to show that for each $H \in \mathrm{NBSAW}^*_{s,t,m,z,\ell_1,\ell_2}$, we can transform it into at least n^ℓ different $H' \in \mathrm{NBSAW}^*_{s,t,d^H\leqslant\Delta,m,z+\ell}$. Further we show that for any $H' \in \mathrm{NBSAW}^*_{s,t,d^H\leqslant\Delta,m,z+\ell}$, it can be generated by at most 2^t different $H \in \mathrm{NBSAW}^*_{s,t,m,z,\ell_1,\ell_2}$.

For each vertex v in H with degree larger than Δ, it must be contained in the multiplicity-1 cycle of H. For each such vertex u, we denote the last multiplicity-2 edge incident to it in the sequence as (u, v). Further we denote the second multiplicity-1 edge incident to it in the sequence as (u, u'). Then our transformation strategy is to replace the last (u, v) edge in H with (v, w_u) and (u, u') edge with (w_u, u) where $w_u \in [n] \setminus V(H)$. We further require that for different u, w_u are different.

We note that the nice block self-avoiding walk after transformation is also a nice block self-avoiding walk. Further for each vertex with degree larger than Δ, its degree is reduced by 1(but we need to be cautious that some vertices of degree Δ can have degree $\Delta + 1$ after the transformation). Finally after performing transformation for one vertex v, the number of connected components of multiplicity-2 edges(i.e the value of z) is increased by 1 and the number of different vertices in the nice block self-avoiding walk is also increased by 1.

We keep iterating on such transformation until every vertex has degree $\leqslant \Delta$. Such process terminates after at most t iterations because for nice block self-avoiding walks $z < t$. Suppose we do such transformation ℓ times. Then the number of connected components is increased by ℓ the number of vertices in the cycle is increased by ℓ. Further the number of vertices in the path remains unchanged. Since there are $(1 \pm o(1))n^{\ell}$ ways of choosing the new vertices added in the transformations, each $H \in \mathrm{NBSAW}^*_{s,t,m,z,\ell_1,\ell_2}$ can be transformed into $(1 \pm o(1))n^{\ell}$ $H' \in \mathrm{NBSAW}^*_{s,t,m,z+\ell}$.

Next we show that for each $H' \in \mathrm{NBSAW}^*_{s,t,m,z+\ell}$, it can be generated by at most 2^t different $H \in \mathrm{NBSAW}^*_{s,t,m,z,\ell_1,\ell_2}$. Given H', if for the $z + \ell$ roots of multiplicity-2 trees which are contained in the multiplicity-1 cycle, we know whether they are also contained in the multiplicity-1 cycle of H, then we can identify the unique H from H'. Since there are at most 2^t choices for determining which of the $z + \ell$ roots are contained in the multiplicity-1 cycle of H, we can conclude that each $H' \in \mathrm{NBSAW}^*_{s,t,m,z+\ell}$ can be generated by at most 2^t different $H \in \mathrm{NBSAW}^*_{s,t,m,z,\ell_1,\ell_2}$

Combining the two facts, we have

- each $H' \in \mathrm{NBSAW}^*_{s,t,m,z+\ell}$ can be generated by at most 2^t different $H \in \mathrm{NBSAW}^*_{s,t,m,z,\ell_1,\ell_2}$
- each $H \in \mathrm{NBSAW}^*_{s,t,m,z,\ell_1,\ell_2}$ can be transformed into $(1 \pm o(1))n^{\ell}$ $H' \in \mathrm{NBSAW}^*_{s,t,m,z+\ell}$.

It follows that

$$\left|\mathrm{NBSAW}^*_{s,t,m,z,\ell_1,\ell_2}\right| \leqslant 2^t n^{-\ell} \left|\mathrm{NBSAW}^*_{s,t,m,z+\ell}\right|$$

which fulfills the proof. □

B.4.2.5 Counting nice block self-avoiding walks with short cycles

We prove the Lemma 4.71

Lemma B.97. *Consider the settings of Theorem 4.43. For $A = (1000s)^2$, Let $m, z, q > 1$ be integers such that $st - 2m + 2z + q = \frac{t}{10\sqrt{A}}$, then*

$$\left|NBSAW^*_{s,t,m,z}\right| \leqslant 2^{10t} s^t \cdot n^{-m+z} \cdot \left|NBSAW^*_{s,t,0,0}\right|.$$

Proof. We denote $\ell = st - 2m + 2z$ as the number of multiplicity-1 edges in the cycle. Then we have $z \leqslant \ell \leqslant \frac{t}{10\sqrt{A}}$. We first pick the number of vertices and leaves in each of the z trees. Since there are m vertices and t leaves to assign, there are at most $\binom{m}{z}\binom{t}{z}$ choices. Since $z \leqslant \frac{t}{10\sqrt{A}}$, this is bounded by $(100As)^{t/10\sqrt{A}} \leqslant 2^t$. Further we fix the staring point of these trees in the sequence. This take $\binom{\ell}{z} \leqslant 2^{t/10\sqrt{A}}$ choices.

We denote the number of vertices and leaves in the tree T_i as m_i and t_i, for $i \in [z]$. Then for each tree, we note that if the order of vertices and degree of each vertex is determined, then the subsequence of this tree in H is totally determined (For order of vertices, we mean order of the first appearence of vertices in the sequence). For determining the vertices and the order, there are at most n^{m_i} choices.

Next we determine the degrees of the vertices, there are at most $\binom{m_i}{t_i}$ ways to choose the leaves. We note that the leaves must be pivot vertices, and there are at most $m_i/s + 2$ pivots of self-avoiding walks in tree i. Therefore we have $t_i \leqslant m_i/s + 2$, and it follows that $\binom{2m_i/s+2}{t_i} \leqslant 2^{2m_i/s+2}$. Let $d_{T_i}(v)$ be the degree of vertice v in the tree T_i. Then by the degree constraint, we have

$$\sum_{v \in T_i} (d_{T_i}(v) - 2)\mathbb{1}_{[d_{T_i}(v) \geqslant 2]} = t_i$$

Since there are $m_i - t_i$ vertices with degree at least 2 in T_i, the total number of choices for the degrees of these vertices is upper bounded by

$$\binom{m_i - t_i + t_i}{t_i} \leqslant (es)^{t_i}$$

In all, we have the number of choices for the degrees of vertices upper bounded by $(es)^{t_i} 2^{2m_i/s+2}$.

Now we multiplying the number of choices for each tree T_i, and we have

$$\prod_{i=1}^{z} (es)^{t_i} 2^{2m_i/s+2} \leqslant 2^{10t} s^t$$

Therefore there are at most $2^{10t} s^t n^{m_i}$ choices for the multiplicity-2 tree. Finally we choose the $st - 2m + z$ vertices which are only incident to multiplicity-1 edges, which takes at most $n^{st-2m+z}$ choices. Multiplying together, we get the claim. □

B.4.2.6 Counting pairs of block self-avoiding walk

We prove lemma Lemma 4.93 here. The proof is very similar to lemma Lemma B.89.

Lemma B.98. *Let B_i be a connected subgraph of the underlying graph of a block self-avoiding walk pair $H \in BSAW_{s,t,u} \times BSAW_{s,t,v}$ with m_i vertices. Let $H(B_i)$ be the multigraph on $V(B_i)$ that is induced by $H(B_i)$. Suppose that:*

- *The cut $H(V(B_i), V(H) \setminus V(B_i))$ consists of ℓ_i edges of multiplicity 1.*
- *All edges in $H(B_i)$ of multiplicity $\geqslant 2$ are in B_i.*
- *The number of edges of multiplicity 1 in $H(B_i)$ is q_i.*
- *The number of edges of multiplicity 2 in $H(B_i)$ is h_i.*
- *The edges of multiplicity larger than 2 satisfy*

$$\sum_{\substack{e \in H(B_i) \\ m_H(e) \geqslant 3}} m_H(e) = p_i.$$

Then we have

$$\sum_{\substack{v \in V(B_i) \\ d^H(v) \geqslant 3}} \left(\frac{m_H(v)}{2} - 1 \right) \leqslant (8h_i + 4p_i)/s + 11q_i + 8\ell_i + 4p_i + 6(h_i - m_i + 1)\,,$$

where $d^H(v)$ is the degree of v in the underlying graph $G(H)$, i.e., without counting multiplicities, and

$$m_H(v) = \sum_{\substack{e \in E(H), \\ e \text{ is incident to } v}} m_H(e).$$

Proof. For the case where $u = v$, each block self-avoiding walk pair in $H \in \mathrm{BSAW}_{s,t,u} \times \mathrm{BSAW}_{s,t,v}$ is a block self-avoiding walk contained in $M_{s,2t}$. Further the B_i, q_i, h_i and p_i are defined in the same way as the Lemma B.88. Therefore, the bound follows by as a corollary of Lemma B.88.

It remains to bound for the case $u \neq v$, which also follows from the same proof. Let $H = H_1 \otimes H_2$ be the decomposition of H into two block self-avoiding walks $H_1 \in \mathrm{BSAW}_{s,t,u}$ and $H_2 \in \mathrm{BSAW}_{s,t,v}$. Let B^* be a spanning tree of B_i. We start by deriving an upper bound on the number of leaves of B^*. Notice that if a vertex $v \in V(B^*)$ is a leaf of B^*, then it must satisfy at least one of the following three conditions:

(a) v is incident to an edge in the cut $H(V(B_i), V(H) \setminus V(B_i))$.

(b) v is incident to an edge in $G(H(B_i)) \setminus B^*$.

(c) v is a pivot of H_1 or H_2, i.e., v is an end-vertex of one of the $2t$ blocks of self-avoiding walks forming H.

By the same argument in Lemma B.88, there are at most ℓ_i leaves satisfying (a), $2q_i + 2h_i + \frac{2}{3}p_i - 2(m_i - 1)$ satisfying (b), and $2(q_i + 2h_i + p_i)/s + 2\ell_i$ leaves satisfying condition (c). Therefore the number of leaves of B^* is at most

$$\ell_i + 2q_i + 2h_i + \frac{2}{3}p_i - 2(m_i - 1) + 2(q_i + 2h_i + p_i)/s + 2\ell_i \leqslant (4h_i + 2p_i)/s + 3\ell_i + 4q_i + p_i + 2(h_i - m_i + 1)$$

Let $d^{B^*}(v)$ be the degree of a vertex $v \in B^*$. By degree constraint the number of leaves in B^* is equal to

$$2 + \sum_{\substack{v \in V(B_i) \\ d^{B^*}(v) \geqslant 3}} (d^{B^*}(v) - 2).$$

Therefore, same as (B.4.2), we have

$$\sum_{\substack{v \in V(B_i) \\ d^{B^*}(v) \geqslant 3}} (d^{B^*}(v) - 2) \leqslant (4h_i + 2p_i)/s + 3\ell_i + 4q_i + p_i + 2(h_i - m_i + 1). \tag{B.4.4}$$

Now define

$$m_{H,2}(v) = \sum_{\substack{e \in E(H), \\ e \text{ is incident to } v, \\ m_H(e)=2}} m_H(e)\,,$$

and

$$m_{H,\neq 2}(v) = \sum_{\substack{e \in E(H), \\ e \text{ is incident to } v, \\ m_H(e)\neq 2}} m_H(e)\,.$$

Clearly, $m_H(v) = m_{H,2}(v) + m_{H,\neq 2}(v)$. Therefore, by the same argument in Lemma B.88

$$\begin{aligned}
\sum_{\substack{v \in V(B_i) \\ d^H(v) \geqslant 3}} \left(\frac{m_H(v)}{2} - 1\right) &= \sum_{\substack{v \in V(B_i) \\ d^H(v) \geqslant 3}} \left(\frac{m_{H,2}(v)}{2} - 1\right) + \sum_{\substack{v \in V(B_i) \\ d^H(v) \geqslant 3}} \frac{m_{H,\neq 2}(v)}{2} \\
&\overset{(a)}{\leqslant} \sum_{\substack{v \in V(B_i) \\ d^H(v) \geqslant 3}} \left(d^{B_i}(v) - 1\right) + \sum_{v \in V(B_i)} \frac{m_{H,\neq 2}(v)}{2} \\
&= \sum_{\substack{v \in V(B_i) \\ d^H(v) \geqslant 3}} \left(d^{B^*}(v) - 1\right) + 2\big(|E(B_i)| - |E(B^*)|\big) + \frac{\ell_i}{2} + q_i + p_i \\
&\leqslant \sum_{\substack{v \in V(B_i) \\ d^H(v) \geqslant 3}} (d^{B^*}(v) - 1) + 2\left(h_i + \frac{p_i}{3} - (m_i - 1)\right) + \frac{\ell_i}{2} + q_i + p_i
\end{aligned}$$

$$
\begin{aligned}
&\leqslant \sum_{\substack{v\in V(B_i)\\ d^{B^*}(v)\geqslant 3}} (d^{B^*}(v)-1) + \left(\sum_{\substack{v\in V(B_i)\\ d^{H}(v)\geqslant 3\\ d^{B^*}(v)=2}} 1\right) + 2(h_i-m_i+1)+\ell_i+q_i+2p_i \\
&\overset{(b)}{\leqslant} 2\cdot \sum_{\substack{v\in V(B_i)\\ d^{B^*}(v)\geqslant 3}} (d^{B^*}(v)-2) + (2q_i+\ell_i) + 2(h_i-m_i+1)+\ell_i+q_i+2p_i \\
&\overset{(c)}{\leqslant} (8h_i+4p_i)/s + 6\ell_i + 8q_i + 2p_i + 4(h_i-m_i+1) \\
&\qquad\qquad + 2(h_i-m_i+1) + 2\ell_i + 3q_i + 2p_i \\
&= (8h_i+4p_i)/s + 11q_i + 8\ell_i + 4p_i + 6(h_i-m_i+1)\,,
\end{aligned}
$$

where (a) follows from the fact that every edge of multiplicity 2 and incident to $v \in V(B_i)$ must be in B_i, hence $m_{H,2}(v) \leqslant 2d^{B_i}(v)$. (b) follows from the fact that if $v \in V(B_i)$ satisfies $d^H(v) \geqslant 3$ and $d^{B_i}(v) = 2$, then v must be incident to an edge of multiplicity 1, i.e., v must be incident to an edge in $E(H(B_i)) \setminus E(B_i)$, or an edge in the cut $H(V(B_i), V(H) \setminus V(B_i))$. There are q_i edges in $E(H(B_i)) \setminus E(B_i)$, each of which is incident to two vertices in $V(B_i)$, and there are ℓ_i edges in the cut $H(V(B_i), V(H) \setminus V(B_i))$, each of which is incident to one vertex in $V(B_i)$. (c) follows from (B.4.4). □

Now we prove lemma Lemma 4.93.

Lemma B.99. *Consider the settings of Theorem 4.44. Let $u, v \in [n]$. Let $\mathcal{B} = \{B_1, \dots, B_z\}$ be collections of disjoint connected graphs each with respectively $m_1, \dots, m_z \geqslant 2$ vertices. Let B^{uv} be a connected graph disjoint from any graph in $\mathcal{B}$ and with $m_{z+1} \geqslant 2$ vertices.Let $\{\mathcal{F}_k\}_{i=1}^{z+1}$ be a sequence of tuples of integers as in Definition 4.91. Let $f^*_{s,t}, g^*_{s,t}$ be the functions*

$$
\begin{aligned}
f^*_{s,t}(m, m', \mathcal{F}, \Psi) =& (\Psi)^{2h/s+10(q+\ell+p+1)+2h-2(m'-1)} \cdot (st)^{5\ell+5q+8p+4h+4-4(m'-1)}\,, \\
g^*_{s,t}(m', \mathcal{F}) =& n^{-p-\ell/2-q-2h+m'}\,.
\end{aligned}
$$

Let $m = \sum_{j\in[z+1]} m_j$. Then there are at most

$$
2n^{2st-2+\mathbb{1}_{[u=v]}\mathbb{1}_{[B_{uv}\neq\emptyset]}} \cdot \prod_{1\leqslant k\leqslant z'} f^*_{s,t}(m, m_k, \mathcal{F}_i, \Psi_i)\cdot g^*_{s,t}(m_i, \mathcal{F}_i, h_i)
$$

block self-avoiding walk pairs in the set $\mathcal{M}_{s,t,u,v\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+z'}}(\mathcal{B})$.

Proof. We will prove by analyzing how we can construct a block self-avoiding walk pair $H \in \mathcal{M}_{s,t,u,v\{\mathcal{F}_i,\Psi_i\}_{i=1}^{z+z'}}(\mathcal{B})$ as a sequence of edges $\{v_0 = u, v_1, v_2, \dots, v_{2st-1}, v_{2st} = v\}$.We denote the respective copies of $\{B_1, \dots, B_z, B_{uv}\}$ in H as $\{B'_1, \dots, B'_z, B'_{z+1}\}$ The construction is divided into two steps

- We first choose each B'_i and the underlying graph of $H(V(B'_i))$ for $i \in [z+1]$.

- Second we choose the indice of the multiplicity-1 edges in the sequence of H. Furthermore we decide which of the vertices in the cuts are contained in $V(B'_i)$ for $i \in [z+1]$.
- Finally we construct a sequence of $2st$ vertices $\{v_0 = u, v_1, v_2, \ldots, v_{2st-1}, v\}$ as the block self-avoiding walk pair H, respecting the underlying graph of H and indices of cut edges fixed in the first two steps.

We denote the union of the z disjoint graphs $B'_1, B'_2, \ldots, B'_z, B'_{z+1}$ as

$$B' = \bigcup_{i=1}^{z+1} B'_i.$$

We recap the setting of the theorem. The size of multiway cut $H\left(V(B'_1), \ldots, V(B'_z), V(H) \setminus \left(\bigcup_{j\in[z]} V(B'_z)\right)\right)$ is bounded by ℓ, and all the edges in the cuts have multiplicity 1. Let $H(B')$ be the multi-graph that is induced by H on $V(B')$. There are q multiplicity-1 edges and h multiplicity-2 edges in $H(B')$. The edges of multiplicity at least 3 in $H(B')$ satisfy the following equation:

$$\sum_{\substack{e\in H(B')\\ m_H(e)\geqslant 3}} m_H(e) = p.$$

For convenience of notation, we denote the number of vertices u, v contained in $V(B'$ as $c_{u,v}$

$$c_{u,v} = \mathbb{1}_{[u\in V(B')]} + \mathbb{1}_{[u\neq v]}\mathbb{1}_{[v\in V(B')]}$$

For the first construction step, we first choose the vertices of the respective copies of $\{B_1, \ldots, B_z, B_{uv}\}$ in H. There are at most $n^{m-c_{uv}}$ choices.

Next, we choose the multi-graph $H(B')$ that is induced by H on $V(B')$, i.e., we have to choose multiplicities for the edges of B' (which must be at least 2), and we have to choose q edges of multiplicity 1 whose both end-vertices must lie in $V(B')$. Note that there are p multi-edges in $H(B')$ that correspond to edges of multiplicity at least 3, so specifying these multi-edges is equivalent to specifying the multiplicities of the edges of B', because once we specify the edges of B' of multiplicity at least 3, any remaining edge in B' must have multiplicity-2. Since both end-vertices of these multi-edges lie in $V(B')$ and since $|V(B')| = m$, there are at most m^{2p} ways to specify the p multi-edges. For choosing the q multiplicity-1 edges, there are at most m^{2q} ways of doing so.

Thus there are at most $m^{2q+2p}n^{m-c_{u,v}}$ choices for the first step.

In the second step of construction, we consider $H \setminus H(V(B'))$, which consists of edges of multiplicity-1. More precisely, since H is a block self-avoiding walk, the remaining edges of H can be partitioned into a number of disjoint walks of multiplicity 1. Each time the walk of multiplicity 1 starts

- either by exiting $V(B'_i)$ for any $i \in [z+1]$ through a cut edge
- or from u as the start of H_1 and $v_{st+1} = v$ as the start of H_2

and then ends by

- either entering $V(B'_j)$ through a cut edge for any $j \in [z+1]$
- or reaching $v_{st} = u$ as the end of H_1 or v_{2st} as the end of H_2.

. We call these walks as *outside walks*. We call the first edge of each outside walk as an *outgoing edge*, and we call the last edge of each outside walk as a *returning edge*. The remaining edges of the outside walks are called *outside edges*. Note that for $i \neq j$, then an outside walk between B'_i and B'_j can be of length 1, and in such case the outgoing edge and returning edge will be the same edge.

We decide the index of cutting edges in the block self-avoiding walk pair sequence H, and for each vertex incident to an cut edge. Since there are at most ℓ cut edges, the number of choices for the indices of cut edges in the sequence of H is bounded by $\binom{2st}{\ell}$. Further for deciding which of the vertices in the cuts are contained in each of B_i for $i \in [z+1]$, there are at most $(z+1)^\ell$ choices.

Finally there are at most $(st)^q$ choices for the indices of the q multiplicity-1 edges contained in $H(V(B_i))$ for $i \in [z+1]$. Thus the total number of choice for the second step is upper bounded by $\binom{2st}{\ell}(z+1)^\ell (st)^q$

In the third step of construction, for $k \in [0, 2st-1]$, given $v_1, v_2, \ldots, v_k$, we choose v_{k+1}. For $k = st$, we have v_{k+1} fixed as v. Since the second step of construction, we can decide whether $v_k, v_{k+1} \in V(B')$ for each k. Thus for $k \neq st$, we can divide them into 3 conditions:

- If $v_{k+1} \in V(B'_i)$ and $v_k \notin V(B'_i)$, then there are at most m choices for v_{k+1} given v_k.
- If $v_{k+1} \in V(B'_i)$ and $v_k \in V(B'_i)$, then there are at most $\Psi_i + q_i$ choices for v_{k+1} given v_k.
- If $v_{k+1} \notin V(B')$, then there are at most n choices for v_{k+1} given v_k.

The total number of possibilities is upper bounded by the product of choices at each step.

Since (v_k, v_{k+1}) must be a cut for satisfying the first condition, there are at most ℓ values of $k \in [0, 2st-1]$ satisfying the first condition. For each such k there are at most st choices for v_{k+1}.

For the third condition, v_{k+1} is not contained in $V(B')$ and $k \neq st$. The number of vertices not contained in $V(B')$ is upper bounded by the number of edges in the outside walks which are not returning edges. There are $2st - 2h - p - q$ edges in the outside walks, and there are at least $\ell/2$ returning edges, thus there are at most $2st - 2h - p - q - \ell/2$ values of k such that v_{k+1} is not contained in $V(B')$. We consider three cases

- When B_{uv} exists, u, v are contained in $V(B')$ then there are at least $\ell/2$ returning edges.

- When B_{uv} doesn't exist and $u = v$, by definition the vertex $u \notin V(B')$ (otherwise B_{uv} will exist as one of the $B_1, B_2, \ldots, B_z$, leading to contradiction). In this case, there are at least $\ell/2 + 2$ returning edges. Also for such case $c_{u,v} = 0$.
- When B_{uv} doesn't exist and $u \neq v$. In this case there are at least $\ell/2 + 2 - c_{u,v}$ returning edges,

Thus when B_{uv} exists, there are at most $2st - 2h - p - q - \ell/2$ values of $k \neq st$ satisfying the third condition. When B_{uv} doesn't exist, there are at most $2st - 2 + c_{u,v} - 2h - p - q - \ell/2$ values of $k \neq st$ satisfying the third condition.

For the second condition, we already decide whether (v_k, v_{k+1}) is a multiplicity-1 edge in the second step of construction. If (v_k, v_{k+1}) is of multiplicity 1, then v_{k+1} is already fixed in the second step of construction. Thus we only need to consider k such that (v_k, v_{k+1}) has multiplicity $\geqslant 2$. By product rule for counting, the number of possibilities is bounded by

$$\begin{aligned}\prod_{i=1}^{z} \prod_{v \in V(B_i')} \Psi_i^{\frac{m_H(v)}{2}-1} &= \prod_{i=1}^{z} \Psi_i^{\sum_{v \in V(B_i')} \frac{m_H(v)}{2}-1} \\ &\leqslant \prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)},\end{aligned}$$

where the last inequality follows from Lemma B.98.

Therefore multiplying together, when B_{uv} is empty, the number of choices for the third step of construction is bounded by

$$\prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)} n^{2st-2h-p-q-\ell/2-2+c(u,v)} (st)^{\ell}$$

when B_{uv} exists, the number of choices is bounded by

$$\prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)} n^{2st-2h-p-q-\ell/2} (st)^{\ell}$$

Combining all three construction steps, when B_{uv} exists, the number of choices is bounded by

$$\prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)} n^{2st-2h-p-q-\ell/2-c(u,v)} (st)^{4\ell+3q+2p}$$

which is equivalent to

$$\prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)} n^{2st-2h-p-q-\ell/2-2+\mathbb{1}_{[u=v]}} (st)^{4\ell+3q+2p}$$

When B_{uv} doesn't exist, the number of choices is bounded by

$$\prod_{i=1}^{z} \Psi_i^{(8h_i+4p_i)/s+11q_i+8\ell_i+4p_i+6(h_i-m_i+1)} n^{2st-2h-p-q-\ell/2-2} (st)^{4\ell+3q+2p}$$

which proves the claim. □

B.4.2.7 Counting nice block self-avoiding walk pairs

Here we prove lemma Lemma 4.99

Lemma B.100. *Consider the settings of Theorem 4.44. Let $u \in [n]$ and Then*

$$|NMULTIG_{s,t,u,u,m}| \leqslant \frac{m^{10}}{n^m} |NMULTIG_{s,t,u,u,0}|$$

For proving this, our plan is first to show that for every m, the size of the set $\text{NMULTIG}_{s,t,u,u,m}$ is dominated by two multiplity-1 cycles sharing m edges. Then we compare the number of such simple block self-avoiding pairs for different m.

Lemma B.101. *For every $m \geqslant 0$, we denote $CyclePair_{u,u,m}$ as the set of nice block self-avoiding walk pairs which are two simple cycles starting from u and only sharing a length-m path containing u. Then we have*

$$CyclePair_{u,u,m} \leqslant |NMULTIG_{s,t,u,u,m}| \leqslant 2CyclePair_{u,u,m}$$

Proof. The nice block self-avoiding walks in the set $\text{NMULTIG}_{s,t,u,u,m}$ can be represented as $\{v_0, v_1, \ldots, v_{2st-1}, v_0\}$, where $v_0, v_1, \ldots, v_{2st-1} \in [n]$ are vertices and $v_0 = v_{st} = u$.

First we note that for each $H \in \text{CyclePair}_{u,u,m}$, the underlying graph contains $2st - m - 2$ vertices excluding vertex u. Therefore the size of $\text{CyclePair}_{u,u,m}$ is at least $n^{2st-m-2}$. Further when $m = st$, the size of $\text{CyclePair}_{u,u,m}$ is at least $n^{2st-m-1}$.

Now for $H \in \text{NMULTIG}_{s,t,u,u,m}$, we denote

$$p = \sum_{e \in H} m_e(H) \mathbb{1}_{[m_e(H) \geqslant 2]}$$
$$\ell = \sum_{e \in H} \mathbb{1}_{[m_e(H) \geqslant 2]}$$

As an ordered sequence of edges, the block self-avoiding walk pairs H can be decomposed into alternating segments of multiplicity-$\geqslant 2$ edges(type I) and multiplicity-1 edges(type II). For convenience, we view the first segment and last segment of H as the same segment if they are of the same type. Then we denote the number of type I segments in H as z. and the number of connected components formed by multiplicity-$\geqslant 2$ edges in

H as k. Further we denote the number of type I segments containing any of the m shared edges in H as z'. Finally we denote the indicator variable of the event that v_0 and v_{st} are both contained in type I segments as $\mathbb{1}_{[E]}$.

We note the relation between k, z and z'. For each connected component containing any of the m shared edges in H, it must satisfy one of the following conditions

- it is visited in one segment of H_1 and one segment of H_2
- it is visited in a segment containing v_{st} or v_0.

Thus we must have $z - k \geqslant (z' - 2)/2$

We further define $\text{NMULTIG}_{s,t,u,u,m,k,z,\mathbb{1}_{[E]}}$ as the subset of $\text{NMULTIG}_{s,t,u,u,m}$ respecting $k, z, \mathbb{1}_{[E]}$ as defined above.

By the above definitions, the number of multiplicity 1 edges is equal to $2st - p$, and there are at most $2st - p - z$ vertices in the interior of type I segments. Further there are at most $\ell + k$ different vertices contained in the type II segments. Thus there are at most $2st - p + \ell + k - z$ different vertices in H.

If $\mathbb{1}_{[E]} = 0$, then at least one of v_0, v_{st} are contained in type II segments. Thus in this case there are at most $2st - p + \ell + k - z - 2$ different vertices other than u contained in $\text{NMULTIG}_{s,t,u,u,m,k,z,\mathbb{1}_{[E]}}$. The only case that $p - \ell - k + z = m$ is that $p = 2\ell = m$ and $k = z$. In such case H must be two multiplicity-1 cycles overlapping a length m path. The number of such $H \in \text{NMULTIG}_{s,t,u,u,m,k,z,0}$ is bounded by $n^{2st-2-m}$. Otherwise we have $2st - p + \ell + k - z - 2 \leqslant 2st - 2 - m - 1$

If $\mathbb{1}_{[E]} = 1$, then both of v_0, v_{st} are contained in type II segments, then there are at most $2st - p + \ell + k - z - 1$ different vertices other than u contained. Now because $2st - p + \ell + k - z - 1 = 2st - 2 - m - (p - \ell + k - z - m - 1)$, there are 3 cases

- $p - \ell + k - z - m = 0$, in this case $H^{(1)}$ and $H^{(2)}$ are two identical cycles. For such case, we must have $m = st$, and it corresponds exactly to the set $\text{CyclePair}_{u,u,m}$
- $p - \ell + k - z - m = 1$, in this case $H^{(1)}$ and $H^{(2)}$ are two multiplicity-1 cycles sharing a path containing u. Thus such case also corresponds to the set $\text{CyclePair}_{u,u,m}$
- Finally we have $p - \ell + k - z - m > 1$.

For proving $|\text{NMULTIG}_{s,t,u,u,m}| \leqslant 2\left|\text{CyclePair}_{u,u,m}\right|$, we only need to prove that for $m \neq st$

$$\left|\text{NMULTIG}_{s,t,u,u,m} \setminus \text{CyclePair}_{u,u,m}\right| \leqslant \left|\text{CyclePair}_{u,u,m}\right|$$

By the above analysis, for any $H \in \text{NMULTIG}_{s,t,u,u,m} \setminus \text{CyclePair}_{u,u,m}$, there are at most $2st - m - 2 - (p - \ell + k - z - m)/2$ different vertices other than u in H.

For fixed p, ℓ, k, z, we count the number of such $H \in \text{NMULTIG}_{s,t,u,u,m,k,z,\mathbb{1}_{[E]}}$. We divide the construction of H as a sequence of vertices into 3 steps

- In the first step, we put the multiplicity-$\geqslant 2$ edges which are not shared by $H^{(1)}$ and $H^{(2)}$ in the sequence.
- In the second step, we put the m shared edges which are shared by $H^{(1)}$ and $H^{(2)}$
- In the final step, we label the vertices in H

Then we denote

$$p_1 = \sum_{e \in H^{(1)} \cap H^{(2)}} m_e(H) \mathbb{1}_{[m_e(H) \geqslant 2]}$$
$$p_2 = \sum_{e \in H^{(1)} \Delta H^{(2)}} m_e(H) \mathbb{1}_{[m_e(H) \geqslant 2]}$$

It's easy to see $p = p_1 + p_2$.

For the first step, there are at most $(st)^{2p_2}$ choices.

For the second step, we first assign the multiplicities to the shared edges in H. Since each shared edge has multiplicity 2, 3 or 4, there are at most $(2st)^{p_1 - 2m}$ choices. For each of the shared edges with multiplicity 3 or 4, there are at most $(2st)^4$ choices for fixing all its locations in the sequence. Therefore there are at most $(2st)^{4(p_1 - 2m)}$ choices for the locations of all shared edges with multiplicity at least 3.

Next we decide the locations of multiplicity-2 shared edges in the sequence. We call the segments of multiplicity-2 shared edges in H as type III segments. Since the shared edges are contained in at most z' type I segments, and there are at most $p - 2m$ shared edges with multiplicity other than 2, the multiplicity-2 shared edges are split into at most $z' + p - 2m$ type III segments. For fixing the locations of these segments in the sequence, we only need to specify the starting indices and ending indices of these segments. Thus there are only $(st)^{z' + p - 2m}$ choices for the locations of these type III segments.

For each of the multiplicity-2 shared edge, they appear exactly once both in $H^{(1)}$ and $H^{(2)}$(where $H = H^{(1)} \otimes H^{(2)}$ is the decomposition of nice block self-avoiding walk H). For fixing two locations of each multiplicity-2 shared edge with multiplicity-2, we first choose the set of locations taken by multiplicity-2 shared edges in H, and then match the locations in $H^{(1)}$ with locations in $H^{(2)}$. For choosing the set of locations taken by multiplicity-2 shared edges, there are at most $(2st)^{p_2}$ ways.

Now for the number of ways of matching locations in $H^{(1)}$ and $H^{(2)}$, we make the following observation: once for the first and last edges in every type III segment, we fix all their locations in the sequence of H, then the locations of each multiplicity-2 shared edge are fixed. For proving this observation, we note that for $H \in \mathrm{NMULTIG}_{s,t,u,u}$, for each vertex incident to a shared multiplicity-2 edge, it must be incident to 2 multiplicity-1 edges in the subgraph of $H^{(1)}$ and subgraph of $H^{(2)}$ respectively. For any such vertex v, there are two possibilities for the two edges e_i, e_{i+1} incident to it in $H^{(1)}$

- e_i, e_{i+1} are two consecutive edges in the sequence

- each of e_i, e_{i+1} is the first or last edge of some type III segment

We denote the correspondance of e_i, e_{i+1} in $H^{(2)}$ as e'_i, e'_{i+1}. For the second case, e'_i, e'_{i+1} has been fixed under the setting. For the first case, the correspondance of e_i, e_{i+1} are either consecutive in the sequence of $H^{(2)}$, or has been fixed as the first or last edge of some type III segment. Therefore suppose e'_i is fixed, then e'_{i+1} is also fixed.

Since for fixing the first and last edge in each type III segment of $H^{(1)}(H^{(2)})$, , there are at most $(st)^{4(z'+p-2m)}$ choices. We conclude that there are at most $(2st)^{p_2}(st)^{4(z'+p-2m)}(st)^{z'+p-2m}$ choices for the shared multiplicity-2 edges in the sequence.

Finally we only need to label at most $2st - m - 2 - (p - \ell - m + z - k)/2$ vertices other than u. Thus there are at most $n^{2st-m-2-(p-\ell-m+z-k)/2}$ choices for labelling the vertices.

Putting together for fixed p, ℓ, k, z, z', the size of $\mathrm{NMULTIG}_{s,t,u,u,m} \setminus \mathrm{CyclePair}_{u,u,m}$ is bounded by

$$n^{2st-m-2-(p-\ell-m+z-k)/2}(st)^{5(z'+p-2m)}(st)^{z'+2p_2+4(p_1-2m)+z'}$$

Arranging the terms, this is upper bounded by

$$n^{2st-m-2}n^{-(p-\ell-m+z-k)/2}(st)^{7z'+9(p-2m)}$$

Now we note that $p - 2m \geqslant 2(\ell - m)$, thus $p - \ell - m \geqslant (p - 2m)/2$. Therefore this is upper bounded by

$$n^{2st-m-2}n^{-(p-2m)/4}n^{-(z-k)/2}(st)^{7z'+9(p-2m)}$$

Since we have proved above that $z - k \geqslant (z' - 2)/2$, we can sum this geometric series satisfying constraints that $z' \leqslant 2(z-k)+2, z-k \geqslant 0, p \geqslant 2m$ and $p - \ell + z - k > 0$.

$$\begin{aligned}
&\sum_{p\geqslant 2m}\sum_{\substack{z-k\geqslant 0\\ z-k+p-2m>0}}\sum_{z'\leqslant 2(z-k)+2} n^{2st-m-2}n^{-(p-2m)/4}n^{-(z-k)/2}(st)^{7z'+9(p-2m)}\\
\leqslant&\sum_{p\geqslant 2m}\sum_{\substack{z-k\geqslant 0\\ z-k\geqslant 2m+1-p}} n^{2st-m-2}n^{-(p-2m)/4}n^{-(z-k)/2}(st)^{14(z-k)+15+9(p-2m)}\\
\leqslant&\sum_{p\geqslant 2m} n^{2st-m-2}n^{-\max(1,p-2m)/4}(st)^{30+9(p-2m)}\\
\leqslant&n^{-0.24}n^{2st-m-2}
\end{aligned}$$

Therefore we obtain that

$$\mathrm{NMULTIG}_{s,t,u,u,m} \setminus \mathrm{CyclePair}_{u,u,m} \leqslant o(\left|\mathrm{CyclePair}_{u,u,m}\right|)$$

The claim thus follows. □

Now the Lemma 4.99 is easy to prove.

Proof. By the last lemma, we have

$$|\mathrm{NMULTIG}_{s,t,u,u,m}| \leqslant 2\left|\mathrm{CyclePair}_{u,u,m}\right|$$

Thus for $m \neq st$,

$$|\mathrm{NMULTIG}_{s,t,u,u,m}| \leqslant 4stn^{2st-m-2}$$

For $m = st$, we have

$$|\mathrm{NMULTIG}_{s,t,u,u,st}| \leqslant 4n^{st-1}$$

Further

$$|\mathrm{NMULTIG}_{s,t,u,u,0}| \geqslant (1-o(1))n^{2st-m-2}$$

Therefore it follows that for $m \neq st$

$$|\mathrm{NMULTIG}_{s,t,u,u,m}| \leqslant \frac{2st}{n^m}|\mathrm{NMULTIG}_{s,t,u,u,0}|$$

and for $m = st$

$$|\mathrm{NMULTIG}_{s,t,u,u,st}| \leqslant \frac{4}{n^{st-1}}|\mathrm{NMULTIG}_{s,t,u,u,0}|$$

□

B.5 Additional tools

We present here some generic tools that are used throughout Chapter 4. We start by stating the classical Paley-Zygmund inequality:

Lemma B.102 (Paley-Zygmund inequality)**.** *If $Z \geqslant 0$ is a random variable with finite variance, then for $0 \leqslant \theta \leqslant 1$, we have*

$$\mathrm{P}(Z > \theta \mathrm{E}[Z]) \geqslant (1-\theta)^2 \frac{\mathrm{E}[Z]^2}{\mathrm{E}[Z^2]}$$

The next lemma relates different norms in $\mathbb{R}^n$.

Lemma B.103. *Let $v \in \mathbb{R}^n$ be a vector and let $q < p \geqslant 1$. Then*

$$\|v\|_p \leqslant n^{1/p-1/q}\|v\|_q .$$

Proof. Let $r = q/p$. Applying Hölder's inequality

$$\|v\|_p^p \leqslant \left(\sum_{i\in[n]} v_i^q\right)^{1/r} \cdot \left(\sum_{i\in[n]} 1\right)^{1-1/r} = n^{1-p/q}\left(\sum_{i\in[n]} v_i^q\right)^{p/q} .$$

Taking the p-th root the lemma follows. □

Lemma B.104 (Courant-Fischer min-max theorem)**.** *Let Z be an $n \times n$ Hermitian matrix with eigenvalues $\lambda_1 \leqslant \lambda_2 \ldots \leqslant \lambda_k \leqslant \ldots \leqslant \lambda_n$, then we have*

$$\lambda_{k+1} = \max_{U} \left\{ \min_{x} \{ R_Z(x) \mid x \in U \text{ and } x \neq 0 \} \mid \dim(U) = n - k \right\}$$

where $R_Z(x) = \frac{\langle Z, xx^\top \rangle}{\|x\|^2}$

The fact below counts the number of non-isomorphic trees with bounded number of leaves.

Fact B.105. *Let $2 \leqslant s \leqslant t$. The number of non-isomorphic trees on st vertices with at most t leaves, denoted by $T(st, t)$ is at most*

$$T(st, t) \leqslant 2t \cdot (8e \cdot s)^{2t} .$$

Proof. Each non-isomorphic tree can be encoded with a length-$2st$ planar code. There is a one-to-one mapping between leaves in the graph and flips from 1 to 0 in the code. Hence there are at most $2t$ flips in the code. The number of length-$2st$ binary codes with at most $2t$ flips is bounded by

$$\binom{2st}{2t} \cdot 2^{2t} \leqslant (8e \cdot s)^{2t} .$$

□

Appendix C

Deferred proofs and addendum to Chapter 5

C.1 Push-out effect of basic SDP

In this section, we present Theorem C.1 that captures the push-out effect of the basic SDP value of uncorrupted stochastic block model. This theorem is based on Theorem 5 and Theorem 8 of [MS16] and is stated in a way that is easier for us to use in our analysis. It is intensively used in Section 5.3, where we prove weak recovery guarantees of our SoS algorithm.

Theorem C.1 (Restatement of Theorem 5 and Theorem 8 of [MS16])**.** *Given a graph $G \sim \mathsf{SBM}_n(d, \varepsilon)$, there exists $C = C(\delta)$ and d_δ that only depend on $\delta = \varepsilon^2 d - 1$ such that, with probability at least $1 - Ce^{-n/C}$, we have*

$$\mathrm{SDP}(\tilde{A}) \geqslant (2 + \Delta)n\sqrt{d}$$

and,

$$\mathrm{SDP}(\tilde{A} - \frac{\varepsilon d}{n} X^*) \leqslant (2 + \rho)n\sqrt{d}$$

where $\rho = \frac{C \log d}{d^{1/10}}$ and $\Delta = \Delta(\delta)$ for some value $\Delta(\delta)$ that only depends on δ.

C.2 Spectral bound of degree-pruned submatrix

In this section, we use the following result to show that we can prune out a small fraction of the high degree vertices to get a spectrally bounded submatrix of the centered adjancency matrix.

Theorem C.2 (Restatement of [FO05, CRV15, LM22])**.** *Suppose M is a random symmetric matrix with zero on the diagonal whose entries above the diagonal are independent with the following*

distribution

$$M_{ij} = \begin{cases} 1 - p_{ij} & w.p.\ p_{ij} \\ p_{ij} & w.p.\ 1 - p_{ij} \end{cases}$$

Let α be a quantity such that $p_{ij} \leqslant \frac{\alpha}{n}$ and M_1 be the matrix obtained from M by zeroing out all the rows and columns having more than 20α positive entries. Then with probability $1 - \frac{1}{n^2}$, we have

$$\|M_1\|_{\mathrm{op}} \leqslant \chi\sqrt{\alpha}$$

for some constant χ.

From Theorem C.2, we can get the following spectral bound for degree-pruned adjacency matrix.

Corollary C.3. *In the setting of Definition 5.1, with probability at least $1 - o(1)$, there exists a subset $T \subseteq [n]$ of size at least $(1 - \beta)n$ such that*

$$\left\|\tilde{A}_T\right\|_{\mathrm{op}} \leqslant C_s\sqrt{d}$$

where C_s is some constant and $\beta = \beta(\delta)$ is a value that only depends on δ.

Proof. For simplicity, let us set a to be $a = (1 + \varepsilon)d$ and set $t = \beta n$. We apply Theorem C.2 by setting $\alpha > a$ to be a large enough constant. The probability that there exists more than βn vertices with degree at least 20α is at most

$$\binom{n}{t}\binom{tn}{10\alpha t}\left(\frac{a}{n}\right)^{10\alpha t} \leqslant \left(\frac{en}{t}\right)^{t}\left(\frac{en}{10\alpha}\right)^{10\alpha t}\left(\frac{a}{n}\right)^{10\alpha t} = \left(\frac{en}{t}\right)^{t}\left(\frac{ea}{10\alpha}\right)^{10\alpha t}$$

Since $\alpha > a$, we have

$$\left(\frac{en}{t}\right)^{t}\left(\frac{ea}{10\alpha}\right)^{10\alpha t} \leqslant \left(\frac{en}{t}\right)^{t}\left(\frac{e}{10}\right)^{10\alpha t} \leqslant e^{-10\alpha t + t(\log(n/t)+1)}$$

Plug in $t = \beta n$, we get the failure probability is $e^{-10\alpha\beta n + \beta n(\log(1/\beta)+1)}$. As long as $-10\alpha + \log(1/\beta) + 1 < 0$ for some α and β, the failure probability is $o(1)$. Take union bound with failure probability of Theorem C.2, we get that, with probability $1 - o(1)$, we have

$$\left\|\left(\tilde{A} - \frac{\varepsilon d}{n}X^*\right)_T\right\|_{\mathrm{op}} \leqslant \chi\sqrt{\alpha}$$

Since $\alpha > a > d$, we have

$$\left\|\left(\tilde{A} - \frac{\varepsilon d}{n}X^*\right)_T\right\|_{\mathrm{op}} \leqslant C_s'\sqrt{d}$$

for some constant C_s'. Notice that $\left\|(\frac{\varepsilon d}{n}X^*)_T\right\|_{\mathrm{op}} \leqslant \varepsilon d$. Apply triangle inequality, we get

$$\left\|\tilde{A}_T\right\|_{\mathrm{op}} \leqslant C_s'\sqrt{d} + \varepsilon d$$

When $\varepsilon\sqrt{d} = O(1)$, we get $\varepsilon d = O(\sqrt{d})$. Hence, with probability $1 - o(1)$, we have

$$\left\|\tilde{A}_T\right\|_{\mathrm{op}} \leqslant C_s\sqrt{d}$$

for some constant C_s. □

C.3 Deferred proofs

We present here deferred proofs of Chapter 5.

Claim C.4 (Restatement of Claim 5.9). Given matrix M, we have $\mathrm{SDP}(M) \leqslant \|M\|_{Gr}$.

Proof. If we look at the second definition of the basic SDP in Eq. (5.2.2) and the second definition of Grothendieck norm in Eq. (5.2.4), it is easy to check that the optimizer of Eq. (5.2.2) is a solution to Eq. (5.2.4) if we take $\delta_i = \sigma_i$. Hence, we have

$$\mathrm{SDP}(M) \leqslant \|M\|_{Gr}$$

□

Claim C.5 (Restatement of Claim 5.10). Let M be an $n \times n$ matrix whose diagonal entries are 0 and $S \subseteq [n]$ be a subset of indices, we have

$$\mathrm{SDP}(M_S) \leqslant \mathrm{SDP}(M)$$

Proof. Let X be the optimizer of $\mathrm{SDP}(M_S)$ and $Z = X_S + \mathrm{Id}_{[n]\setminus s}$. We have

$$\begin{aligned}\mathrm{SDP}(M_S) =& \langle X, M_S\rangle \\ =& \langle X_S, M\rangle\end{aligned}$$

Since M has zero on diagonals, we have $\langle X_S, M\rangle = \langle Z, M\rangle$. Notice that $Z \succeq 0$ and $Z_{ii} = 1$ for all $i \in [n]$. Therefore, Z is a solution to the basic SDP, which implies that

$$\langle Z, M\rangle \leqslant \mathrm{SDP}(M)$$

Thus, we have

$$\begin{aligned}\mathrm{SDP}(M_S) =& \langle X_S, M\rangle \\ =& \langle Z, M\rangle \\ \leqslant& \mathrm{SDP}(M)\end{aligned}$$

□

Lemma C.6 (Formal statement of Lemma 5.6)**.** *Let $\tilde{A} \in \mathbb{R}^{n\times n}$ and $S \subset [n]$ be a set of size $(1-\mu)n$. Suppose $\left\|\tilde{A}_S\right\|_{\mathrm{op}} \leqslant C_s\sqrt{d}$ for some constant C_s, then for all $S' \subseteq S$ with size at most $(1-2\mu)n$, there is a deg-4 SoS proof that*

$$\mathrm{SDP}(\tilde{A}_S - \tilde{A}_{S'}) \leqslant 3K_G C_s\sqrt{\mu}\cdot n\cdot\sqrt{d}$$

where K_G is the Grothendieck constant.

Proof. Consider arbitrary matrix $X \in \mathbb{R}^{n\times n}$ such that $X_{ii} = 1$ for $i \in [n]$ and $X \succeq 0$. To bound the value of $\langle X, \tilde{A}_S - \tilde{A}_{S'}\rangle$, we consider the $\infty \to 1$ norm of $\tilde{A}_S - \tilde{A}_{S'}$. Once we have its $\infty \to 1$ norm bound, we can apply the Grothendieck inequality and get

$$\langle X, \tilde{A}_S - \tilde{A}_{S'}\rangle \leqslant \left\|\tilde{A}_S - \tilde{A}_{S'}\right\|_{Gr} \leqslant K_G \left\|\tilde{A}_S - \tilde{A}_{S'}\right\|_{\infty\to 1} \tag{C.3.1}$$

where K_G is the Grothendieck constant. For simplicity, let us write $\tilde{A}_{S'} = \tilde{A}_S \odot (\mathbf{1} - z)(\mathbf{1} - z)^\top$ where z is the indicator vector such that $z_i = 1$ if and only if i is in S but not in S'. Consider $x, y \in \{\pm 1\}^n$, we have

$$\begin{aligned}
\langle x, (\tilde{A}_S - \tilde{A}_{S'})y\rangle =& \langle x, (\tilde{A}_S - \tilde{A}_S \odot (\mathbf{1} - z)(\mathbf{1} - z)^\top)y\rangle \\
=& 2\langle x, (\tilde{A}_S \odot z\,\mathbf{1}^\top)y\rangle - \langle x, (\tilde{A}_S \odot zz^\top)y\rangle \\
=& 2\langle x \odot z, \tilde{A}_S y\rangle - \langle x \odot z, \tilde{A}_S(y \odot z)\rangle \\
\leqslant& 2\|x \odot z\|\left\|\tilde{A}_S y\right\| + \|x \odot z\|\left\|\tilde{A}_S(y \odot z)\right\|
\end{aligned}$$

where we applied Cauchy Schwarz in the last step. Since $\left\|\tilde{A}_S\right\|_{\mathrm{op}} \leqslant C_s\sqrt{d}$ by our constraint and there can be at most μn vertices that is in S but not in S', we have

$$\begin{aligned}
\langle x, (\tilde{A}_S - \tilde{A}_{S'})y\rangle \leqslant& 2\|x \odot z\|\left\|\tilde{A}_S y\right\| + \|x \odot z\|\left\|\tilde{A}_S(y \odot z)\right\| \\
\leqslant& 2C_s\sqrt{d}\|x \odot z\|\|y\| + C_s\sqrt{d}\|x \odot z\|\|y \odot z\| \\
=& 2C_s\sqrt{d}\sqrt{\mu}n + C_s\sqrt{d}\mu n \\
\leqslant& 3C_s\sqrt{\mu}n\sqrt{d}
\end{aligned}$$

Therefore, $\left\|\tilde{A}_S - \tilde{A}_{S'}\right\|_{\infty\to 1}$ is bounded by $3C_s\sqrt{\mu}n\sqrt{d}$. Plug this into Eq. (C.3.1), we get

$$\langle X, \tilde{A}_S - \tilde{A}_{S'}\rangle \leqslant 3K_G C_s\sqrt{\mu}n\sqrt{d}$$

for any $X \in \mathbb{R}^{n\times n}$ such that $X_{ii} = 1$ for $i \in [n]$ and $X \succeq 0$. Thus, we have

$$\mathrm{SDP}(\tilde{A}_S - \tilde{A}_{S'}) \leqslant 3K_G C_s\sqrt{\mu} \cdot n \cdot \sqrt{d}$$

where K_G is the Grothendieck constant.

Notice that, every step of the proof can be made to be deg-4 SoS. Hence, the proof is deg-4 SoS. □

Lemma C.7 (Restatement of Lemma 5.16)**.** *The SoS program in Eq. (5.3.1) is feasible with probability* $1 - o(1)$.

Proof. From Corollary C.3, we know that, with probability $1 - o(1)$, there exists a submatrix $\tilde{A}_T$ of size $(1 - \beta)n$ whose spectral norm is bounded by $C_s\sqrt{d}$. By monotonicity of spectral norm, the spectral norm of all submatrices of size $(1 - \mu - \beta)n$ of $\tilde{A}_T$ are bounded by $C_s\sqrt{d}$. Therefore, if we consider the set $S = T \cap S^*$, it satisfies the spectral constraint with probability $1 - o(1)$.

Now, we need to show that, with probability $1 - o(1)$, the matrix $\tilde{A}_S$ with $S = T \cap S^*$ has large enough basic SDP value. Apply Theorem C.1, we get that with probability at least $1 - Ce^{-(1-\mu-\beta)n/C}$, a stochastic block model of size $(1-\mu-\beta)n$ has basic SDP value larger than or equal to $(2+\Delta)(1-\mu-\beta)n\sqrt{d}$. Consider all submatrices of size $(1-\mu-\beta)n$ of $\tilde{A}$ and take union bound, the failure probability is

$$\begin{aligned}\binom{n}{(1-\mu-\beta)n} Ce^{-(1-\mu-\beta)n/C} \leqslant & C\left(\frac{en}{(\mu+\beta)n}\right)^{(\mu+\beta)n} e^{-(1-\mu-\beta)n/C} \\ = & Ce^{(\mu+\beta)n(\log(1/(\mu+\beta))+1)-(1-\mu-\beta)n/C}\end{aligned}$$

When $\mu \leqslant \mu_\delta$ for some value μ_δ that only depends on δ and $\beta \ll 1/C$, the failure probability is $o(1)$. Therefore, with probability $1 - o(1)$, for the uncorrputed stochastic block model, the basic SDP value of all its submatrices of size $(1-\mu-\beta)n$ is larger than or equal to $(2+\Delta)(1-\mu-\beta)n\sqrt{d}$, which include the submatrix defined by the set $S = T \cap S^*$.

Hence, with probability $1 - o(1)$, there exists a subset $S = T \cap S^*$ of size $(1-\mu-\beta)n$ such that $\tilde{A}_S$ has basic SDP value larger than or equal to $(2+\Delta)(1-\mu-\beta)n\sqrt{d}$ and has spectral norm less than or equal to $C_s\sqrt{d}$.

For the value of X, we can simply take the optimizer of the basic SDP for $\tilde{A}_S$. This concludes the feasibility analysis of the program. □

Lemma C.8 (Restatement of Lemma 5.18)**.** *For X and w that satisfy the SoS program in Eq. (5.3.1), we have*

$$\mathcal{A}\left|\frac{X,w}{4}\right\rangle \langle X, X^*\rangle \geqslant \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2$$

where β is the small constant fraction of high degree nodes we need to prune to get bounded spectral norm according to Corollary C.3 and $\Delta' = \Delta'(\delta)$ for some value $\Delta'(\delta)$ that only depends on δ.

Proof. We decompose $\langle X, X^*\rangle$ into $\langle X, X^*\rangle = \langle X_{S'}, X^*_{S'}\rangle + \langle X - X_{S'}, X^*\rangle$. For $\langle X_{S'}, X^*_{S'}\rangle$, we can apply Lemma 5.17 and get

$$\mathcal{A}\left|\frac{X,w}{4}\right\rangle \langle X_{S'}, X^*_{S'}\rangle \geqslant \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{n^2}{\varepsilon\sqrt{d}})$$

Now, we consider $\langle X - X_{S'}, X^*\rangle$. Notice that, since X is positive semidefinite whose diagonals are 1's, all its entries are within $[-1, 1]$. This is because all principle submatrices of a positive semidefinite matrix are positive semidefinite. If we consider the principle submatrix formed by X_{ii}, X_{ij}, X_{ji} and X_{jj}, its determinant is non-negative. Hence, $X_{ij}^2 \leqslant X_{ii}X_{jj} = 1$. Since there can be at most $(2\mu+\beta)n$ vertices that are not in S', $\langle X - X_{S'}, X^*\rangle$ is a summation of at most $2(2\mu+\beta)n^2$ entries whose absolute values are less than or equal to 1. Therefore, we have

$$|\langle X - X_{S'}, X^*\rangle| \leqslant 2(2\mu+\beta)n^2$$

Combine the bounds on $\langle X_{S'}, X^*_{S'} \rangle$ and $\langle X - X_{S'}, X^* \rangle$, we have

$$\begin{aligned}
\mathcal{A}\left|_{\frac{1}{4}}^{X,w}\right. \langle X, X^* \rangle =& \langle X_{S'}, X^*_{S'} \rangle + \langle X - X_{S'}, X^* \rangle \\
\geqslant& \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{n^2}{\varepsilon\sqrt{d}}) - 2(2\mu+\beta)n^2 \\
\geqslant& \frac{\Delta'(1-\beta)n^2}{\varepsilon\sqrt{d}} - O(\sqrt{\mu}\frac{n^2}{\varepsilon\sqrt{d}}) - 2\beta n^2
\end{aligned}$$

which finishes the proof. □

Appendix D

Deferred proof and addendum to Chapter 6

D.1 Deferred proofs

This section contains proofs deferred troughout Chapter 6.

Deferred proofs of Section 6.4

We upper bound the number of $100(\log\log n)$ tangle free canonical paths for sparse graphs.

Lemma D.1 (Enumeration of canonical paths, restatement of Lemma 6.20)**.** *Let $\mathcal{W}^{2q,z}(v,e)$ be the set of canonical paths with v vertices and e distinct edges. We have*

$$\left|\mathcal{W}^{2q,z}(v,e)\right| \leqslant \left(2^{2t}z\right)^{2qt} \cdot (2zq)^{6tq\cdot(e-v+1)}\,.$$

Proof. Our proof is closely related to Lemma 17 in [BLM15], thus we only specify where it differs. Using similar notation, we may represent each walk $W \in \mathcal{W}^{2q,z}(v,e)$ as a sequence $A_{1,1},\ldots,A_{1,z},\ldots,A_{2q,z}$. We explore the sequence in lexicographic order and think of index (i,ℓ) as a time. We say $A_{i,\ell}$ is a first time if the target endpoint of the edge did not appear in the sequence before. The set of first time edges form a tree with vertex set $\{1,\ldots,v\}$. The distinct edges in W not in the tree are thus $\varepsilon = e - v + 1$. We use the same encoding of [BLM15]. In particular we encode long cycling times in the same way. For short cycling times we use them same encoding, however as we may have up to 2^t cycles in each subsequence $A_{i,1},\ldots,A_{i,z}$, we further need to specify which among the possible paths we are going to take for each short cycling time, there are 2^{2t} possible such paths. Finally, we may have up to t short cycling times in each walk and up to $\varepsilon\cdot t$ long cycling times. The result follows. □

The next lemma shows that graphs sampled from a distribution in $\mathcal{D}_{d,\gamma}$ for $\gamma \geqslant O(\log^2 n)$ have small average degree.

Lemma D.2. *Let n be an integer, $d > 0$, $\gamma \geqslant O(d + \log n)$ and consider a distribution $P_{d,\gamma} \in \mathcal{D}_{d,\gamma}$. Then for $\mathbf{A} \sim \mathcal{P}_{d,\gamma}$, with probability 0.999*

$$\operatorname{Tr} D(\mathbf{A}) \leqslant O(nd)\,.$$

Proof. The expected average degree of the graph $\mathbf{G}$ associated with $\mathbf{A}$ is d. By linearity of expectation we thus have $\mathbb{E} \operatorname{Tr} D(\mathbf{A}) = nd$. By Markov's inequality the result follows. □

Deferred proofs of Section 6.5

We start by proving the rough bound on $\mathbf{A}''$ of Lemma 6.25.

Lemma D.3 (Restatement of Lemma 6.25)**.** *Consider the settings of Lemma 6.24. Let $\mathbf{A}'$ as defined in Eq. (6.5.3) and let $\mathbf{A}'' = \mathbf{A} - \mathbf{A}'$. Then with probability $1 - o(1)$*

$$\|\mathbf{A}''\|_{\infty\to 1} \leqslant n^{k-1-\Omega(1)} \cdot O(p \cdot n^{k/2})\,.$$

Proof. Since each entry in $\mathbf{A}''$ is either zero or at least $\Omega(1)$ in absolute value, we have $\|\mathbf{A}''\|_{\infty\to 1} \leqslant O\left(\|\mathbf{A}''\|_F^2\right)$. We may decompose $\|\mathbf{A}''\|_{\mathrm{F}}^2$ as

$$\|\mathbf{A}''\|_{\mathrm{F}}^2 = \sum_{\substack{\alpha_1,\alpha_2,\beta_1,\beta_2 \in [n]^{(k-1)/2} \\ \text{s. t. } S(\alpha_1,\alpha_2)=S(\beta_1,\beta_2)}} \mathbf{A}''_{(\alpha_1,\beta_1),(\alpha_2,\beta_2)} + \sum_{\substack{\alpha_1,\alpha_2,\beta_1,\beta_2 \in [n]^{(k-1)/2} \\ \text{s. t. } k-1 > \left|S(\alpha_1,\alpha_2)\cap S(\beta_1,\beta_2)\right| > 1}} \mathbf{A}''_{(\alpha_1,\beta_1),(\alpha_2,\beta_2)} \tag{D.1.1}$$

We have less than $n^{k-1}(k-1)!$ elements in the first sum and less than $n^{2k-3}(k-1)!$ elements in the second one. To bound the first sum, notice that for any $t \geqslant 1$ and and $\alpha, \beta \in [n]^{(k-1)/2}$

$$\begin{aligned}
\mathbb{P}\left(\left(\mathbf{A}''_{(\alpha,\alpha),(\beta,\beta)}\right)^2 = t^2\right) &= \mathbb{P}\left(\left(\sum_{\ell\in[n]} T^2_{(\alpha,\beta,\ell)}\right)^2 = t^2\right) = \\
&= \mathbb{P}\left(\left|\sum_{\ell\in[n]} T^2_{(\alpha,\beta,\ell)}\right| = t\right) \\
&\leqslant \binom{n}{t}(1-p)^{n-t}p^t \\
&\leqslant \binom{n}{t}p^t \\
&\leqslant \left(\frac{e\cdot n\cdot p}{t}\right)^t.
\end{aligned}$$

We split the contribution of entries based on their magnitude and bound their number using Markov's inequality. For any $t \geqslant 1$, denote by $\mathcal{E}_t$ the event that

$$\sum_{\substack{\alpha_1,\alpha_2,\beta_1,\beta_2 \in [n]^{(k-1)/2} \\ \text{s. t. } S(\alpha_1,\alpha_2)=S(\beta_1,\beta_2)}} \left(\mathbf{A}''_{(\alpha_1,\alpha_2),(\beta_1,\beta_2)}\right)^2 [\![\mathbf{A}''_{(\alpha_1,\alpha_2),(\beta_1,\beta_2)} = t]\!] \leqslant 100 \cdot n^{k-1} \cdot (k-1)! \cdot p^{1/10} \cdot 2^{-t}\,,$$

where the Iverson brackets denote the indicator function. Let $\bar{\mathcal{E}}_t$ be its complement event. Notice that if $\mathcal{E}_t$ is verified for all $t \geqslant 1$, then

$$\sum_{\substack{\alpha_1,\alpha_2,\beta_1,\beta_2\in[n]^{(k-1)/2} \\ \text{s. t. } S(\alpha_1,\alpha_2)=S(\beta_1,\beta_2)}} \left(\mathbf{A}''_{(\alpha_1,\beta_1),(\alpha_2,\beta_2)}\right)^2 \leqslant 10^3 \cdot n^{k-1} \cdot (k-1)! \cdot p^{1/10}$$
$$\leqslant n^{(k-1)/2-\Omega(1)}(k-1)! \cdot O(p \cdot n^{k/2}),$$

where in the last step we used the assumption that $p \geqslant n^{-k/2}$. We verify that with high probability this intersection event happens. By Markov's inequality

$$\mathbb{P}(\bar{\mathcal{E}}_t) \leqslant (2e \cdot n \cdot p)^t \cdot \frac{100}{p^{1/10}}.$$

Thus

$$\mathbb{P}\left(\bigcup_{t\geqslant 1}\bar{\mathcal{E}}_t\right) \leqslant O(n \cdot p^{9/10}) \leqslant n^{-\Omega(1)}.$$

We focus next onto the second sum in Eq. (D.1.1). Let $\alpha_1, \alpha_2, \beta_1, \beta_2 \in [n]^{(k-1)/2}$ chosen accordingly, then

$$\begin{aligned}\mathbb{E}\left[\left(\mathbf{A}''_{(\alpha_1,\beta_1),(,\alpha_2,\beta_2)}\right)^2\right] &= \sum_{\ell,\ell'\in[n]} \mathbb{E}\,\mathbf{T}_{\alpha_1\alpha_2\ell}\mathbf{T}_{\beta_1\beta_2\ell}\mathbf{T}_{\alpha_1\alpha_2\ell'}\mathbf{T}_{\beta_1\beta_2\ell'} \\ &= \sum_{\ell\in[n]} \mathbb{E}\,\mathbf{T}^2_{\alpha_1\alpha_2\ell}\mathbf{T}^2_{\beta_1\beta_2\ell} \\ &\leqslant np^2.\end{aligned}$$

where in the second step we used independence of the entries of $\mathbf{T}$. As before, by Markov's inequality the result follows. □

Next we show that the trace of $D(\mathbf{A})$, for $\mathbf{A}'$ as in defined in Eq. (6.5.3) and the associated D as defined in Section 6.3, concentrates around its expectation.

Lemma D.4. *Consider the settings of Lemma 6.26. Let $\mathbf{A}'$ as defined in Eq. (6.5.3) and let D be the associated matrix as defined in Section 6.3. Then with probability at least $1 - 10^4$*

$$\operatorname{Tr} D(\mathbf{A}') \leqslant O\left(p^2 \cdot n^{2k-1}\right).$$

Proof. By linearity

$$\mathbb{E}\operatorname{Tr} D(\mathbf{A}') = \operatorname{Tr}\mathbb{E}\, D(\mathbf{A}') = \sum_{(\alpha_1,\beta_1)\in[n]^{k-1}} \sum_{(\alpha_2,\beta_2)\in[n]^{k-1}} \mathbb{E}\left|\mathbf{A}'_{(\alpha_1,\beta_1)(\alpha_2,\beta_2)}\right|.$$

As $\mathbb{E}\left|\mathbf{A}'_{(\alpha_1,\beta_1)(\alpha_2,\beta_2)}\right| = O(p^2 n)$, by Markov's inequality the result folllows. □

We restate and prove Lemma 6.31.

Lemma D.5 (Restatement of Lemma 6.31)**.** *Consider the settings of Theorem 6.27. Let $W \in BNBW^{2q,z}$. Then for any term in $\overline{AN}(W)$*

$$\mathbb{E}\prod_{i=1}^{2q}\left[\left|\sum_{\ell_1^i}\mathbf{T}_{(\alpha_1^i\alpha_2^i\ell_1^i)}\mathbf{T}_{(\beta_1^i\beta_2^i\ell_1^i)}\right|\left(\prod_{s=2}^{z-1}\mathbf{T}_{(\alpha_s^i\alpha_{s+1}^i\ell_s^i)}\mathbf{T}_{(\beta_s^i\beta_{s+1}^i\ell_s^i)}\right)\right]$$
$$\leqslant C^{2q}\cdot\mathbb{E}\prod_{i=1}^{2q}\left[\left(\sum_{\ell_1^i}\mathbf{T}_{(\alpha_1^i\alpha_2^i\ell_1^i)}\mathbf{T}_{(\beta_1^i\beta_2^i\ell_1^i)}\right)^2\left(\prod_{s=2}^{z-1}\mathbf{T}_{(\alpha_s^i\alpha_{s+1}^i\ell_s^i)}\mathbf{T}_{(\beta_s^i\beta_{s+1}^i\ell_s^i)}\right)\right],$$

for some constant $C > 0$.

Proof. We may rewrite the left hand side

$$\mathbb{E}\prod_{i=1}^{2q}\underbrace{\left|\sum_{\ell_1^i\in[n]}\mathbf{T}_{(\alpha_1^i\alpha_2^i\ell_1^i)}\mathbf{T}_{(\beta_1^i\beta_2^i\ell_1^i)}\right|}_{=:\mathbf{L}_1^i}\cdot\underbrace{\left(\prod_{s=2}^{z-1}\mathbf{T}_{(\alpha_s^i\alpha_{s+1}^i\ell_s^i)}\mathbf{T}_{(\beta_s^i\beta_{s+1}^i\ell_s^i)}\right)}_{=:\mathbf{L}_s^i}$$

and the right hand side

$$\mathbb{E}\prod_{i=1}^{2q}\underbrace{\left(\sum_{\ell_1^i\in[n]}\mathbf{T}_{(\alpha_1^i\alpha_2^i\ell_1^i)}\mathbf{T}_{(\beta_1^i\beta_2^i\ell_1^i)}\right)^2}_{=:\mathbf{R}_1^i}\cdot\underbrace{\left(\prod_{s=2}^{z-1}\mathbf{T}_{(\alpha_s^i\alpha_{s+1}^i\ell_s^i)}\mathbf{T}_{(\beta_s^i\beta_{s+1}^i\ell_s^i)}\right)}_{=:\mathbf{L}_s^i}.$$

Opening up the sums in $\mathbf{L}_2^1,\dots,\mathbf{L}_{z-1}^{2q}$ we get sums of terms of the form

$$\text{on the LHS:}\quad \mathbf{L}_1^1\cdots\mathbf{L}_1^{2q}\cdot\underbrace{\mathbf{T}_{(\alpha\alpha'\ell)}\cdots}_{=:\mathbf{S}}$$
$$\text{on the RHS:}\quad \mathbf{R}_1^1\cdots\mathbf{R}_1^{2q}\cdot\underbrace{\mathbf{T}_{(\alpha\alpha'\ell)}\cdots}_{=\mathbf{S}}.$$

By linearity of expectation we may consider each such element independently. By independence of the entries of $\mathbf{T}$ if there is a term in $\mathbf{S}$ of odd degree that does not appear in $\mathbf{L}_1^1\cdots\mathbf{L}_1^{2q}$ we have

$$\mathbb{E}\,\mathbf{L}_1^1\cdots\mathbf{L}_1^{2q}\cdot\mathbf{S}\leqslant 0\leqslant\mathbb{E}\,\mathbf{R}_1^1\cdots\mathbf{R}_1^{2q}\cdot\mathbf{S}.$$

It remains to consider the case in which each term in $\mathbf{S}$ has even degree. But then for any possible realization T of $\mathbf{T}$, as all entries are non-zero entries are bounded away from 0, we have

$$\mathbb{E}\left[\mathbf{L}_1^1 \cdots \mathbf{L}_1^{2q} \cdot \mathbf{S} \;\middle|\; \mathbf{T} = T\right] \leqslant C^{2q}\, \mathbb{E}\left[\mathbf{R}_1^1 \cdots \mathbf{R}_1^{2q} \cdot \mathbf{S} \;\middle|\; \mathbf{T} = T\right],$$

for some constant $C > 0$. This concludes the proof. □

D.2 Additional tools

Fact D.6. *Let $T \in (\mathbb{R}^n)^{\otimes k}$ be a tensor. Let $\tilde{T} \in (\mathbb{R}^n)^{\otimes k}$ be its symmetrization, that is for any multi-index $\alpha \in [n]^k$*

$$\tilde{T}_\alpha = \frac{1}{k!} \sum_{\alpha' \in \Pi(\alpha)} T_{\alpha'},$$

where $\Pi(\alpha)$ is the set of permutations of α. Then for any $x \in \mathbb{R}^n$

$$\langle T, x^{\otimes k} \rangle = \langle \tilde{T}, x^{\otimes k} \rangle.$$

Proof. Fix a mapping $\pi : [k] \to [k]$, then we have

$$\langle T, x^{\otimes k} \rangle = \sum_{\alpha \in [n]^k} T_\alpha x^\alpha = \sum_{\alpha \in [n]^k} T_{\pi(\alpha)} x^\alpha.$$

Repeating the reasoning for all $k!$ possible permutations the result follows. □

Next we provide some matrix concentration inequalities.

Theorem D.7 (Matrix Bernstein, [Tro12])**.** *Consider a finite sequence $\{\mathbf{M}_\ell\}$ of independent, random, matrices with dimensions $n_1 \times n_2$. Assume that each random matrix satisfies*

$$\mathbb{E}\,\mathbf{M}_\ell = \mathbf{0} \quad \text{and} \quad \|\mathbf{M}_\ell\| \leqslant R \text{ almost surely.}$$

Define

$$\sigma^2 := \max\left\{ \left\| \sum_\ell \mathbb{E}\,\mathbf{M}_\ell \mathbf{M}_\ell^\mathsf{T} \right\|, \left\| \sum_\ell \mathbb{E}\,\mathbf{M}_\ell^\mathsf{T} \mathbf{M}_\ell \right\| \right\}.$$

Then, for all $t \geqslant 0$,

$$\mathbb{P}\left[\left\| \sum_\ell \mathbf{M}_\ell \right\| \geqslant t \right] \leqslant (n_1 + n_2) \cdot \exp\left(\frac{-t^2/2}{\sigma^2 + Rt/3} \right).$$

Appendix E

Deferred proofs and addendum to Chapter 7

E.1 Deferred proofs for stochastic block models

We prove Lemma 7.33 restated below.

Lemma E.1 (Restatement of Lemma 7.33)**.** *Consider the settings of Lemma 7.32. With probability* $1 - \exp(-\Omega(n))$ *over* $\mathbf{G} \sim \mathsf{SBM}_n(\gamma, d, x)$*,*

$$\left\| \hat{X}(Y(\mathbf{G})) - \frac{1}{n} x x^\top \right\|_F^2 \leqslant \frac{800}{\gamma \sqrt{d}}.$$

Proof. Recall $\mathcal{K} = \{X \in \mathbb{R}^{n \times n} : X \succeq 0, X_{ii} = 1/n \ \forall i\}$. Let $X^* := \frac{1}{n} x x^\top$. Since $\hat{X} = \hat{X}(Y(\mathbf{G}))$ is a minimizer of $\min_{X \in \mathcal{K}} \|Y(\mathbf{G}) - X\|_F^2$ and $X^* \in \mathcal{K}$, we have

$$\left\| \hat{X} - Y(\mathbf{G}) \right\|_F^2 \leqslant \|X^* - Y(\mathbf{G})\|_F^2 \iff \left\| \hat{X} - X^* \right\|_F^2 \leqslant 2 \left\langle \hat{X} - X^*, Y(\mathbf{G}) - X^* \right\rangle.$$

The infinity-to-one norm of a matrix $M \in \mathbb{R}^{m \times n}$ is defined as

$$\|M\|_{\infty \to 1} := \max\{\langle u, Mv \rangle : u \in \{\pm 1\}^m, v \in \{\pm 1\}^n\}.$$

By [GV16, Fact 3.2], every $Z \in \mathcal{K}$ satisfies

$$|\langle Z, Y(\mathbf{G}) - X^* \rangle| \leqslant \frac{K_G}{n} \cdot \|Y(\mathbf{G}) - X^*\|_{\infty \to 1},$$

where $K_G \leqslant 1.783$ is Grothendieck's constant. Similar to the proof of [GV16, Lemma 4.1], using Bernstein's inequality and union bound, we can show (cf. Fact E.2)

$$\|Y(\mathbf{G}) - X^*\|_{\infty \to 1} \leqslant \frac{100n}{\gamma \sqrt{d}}$$

with probability $1 - \exp(-\Omega(n))$. Putting things together, we have

$$\left\| \hat{X}(Y(\mathbf{G})) - \frac{1}{n} x x^\top \right\|_F^2 \leqslant \frac{400 \cdot K_G}{\gamma \sqrt{d}},$$

with probability $1 - \exp(-\Omega(n))$. □

Fact E.2. *Let $\gamma > 0, d \in \mathbb{N}, x^* \in \{\pm 1\}^n$, and $\mathbf{G} \sim \mathsf{SBM}(\gamma, d, x^*)$. Let $Y(\mathbf{G}) = \frac{1}{\gamma d}\big(A(\mathbf{G}) - \frac{d}{n} J\big)$, where $A((G))$ is the adjacency matrix of (G) with entries d/n on the diagonal. Then*

$$\max_{x \in \{\pm 1\}^n} \left| x^\top \big(Y(\mathbf{G}) - \tfrac{1}{n} x^* (x^*)^\top \big) x \right| \leqslant \frac{100n}{\gamma \sqrt{d}}$$

with probability at least $1 - e^{-10n}$.

Proof. The result will follow using Bernstein's Inequality and a union bound. Define $\boldsymbol{E} \coloneqq Y(\mathbf{G}) - \frac{1}{n} x^* (x^*)^\top$. Fix $x \in \{\pm 1\}^n$ and for $1 \leqslant i < j \leqslant n$, let $\boldsymbol{Z}_{i,j} \coloneqq \boldsymbol{E}_{i,j} x_i x_j$. Then $x^\top \boldsymbol{E} x = 2 \sum_{1 \leqslant i < j \leqslant n} \boldsymbol{Z}_{i,j}$. Note that

$$\begin{aligned}
&\mathbb{E}\, \boldsymbol{Z}_{i,j} = 0\,, \\
&\big|\boldsymbol{Z}_{i,j}\big| \leqslant \frac{1}{\gamma n} \cdot \left(\frac{n}{d} - 1\right) + \frac{1}{\gamma d n} \leqslant \frac{1}{\gamma d}\,, \\
&\mathbb{E}\, \boldsymbol{Z}_{i,j}^2 = \operatorname{Var}\big[\boldsymbol{Y}(\mathbf{G})_{i,j}\big] \leqslant \mathbb{E}\, \boldsymbol{Y}(\mathbf{G})_{i,j}^2 \leqslant (1+\gamma)\frac{d}{n} \cdot \frac{1}{\gamma^2 n^2} \left[\left(\frac{n}{d} - 1\right)^2 - \frac{1}{\gamma^2 n^2} \right] + \frac{1}{\gamma^2 n^2} \\
&\qquad \leqslant (1+\gamma)\frac{1}{d\gamma^2 n} + \frac{1}{\gamma^2 n^2} \leqslant \frac{3}{\gamma^2 d n}\,.
\end{aligned}$$

By Bernstein's Inequality (cf. [Wai19, Proposition 2.14]) it follows that

$$\begin{aligned}
\mathbb{P}\left(\sum_{i<j} \boldsymbol{Z}_{i,j} \geqslant \frac{50n}{\gamma\sqrt{d}} \right) &\leqslant \mathbb{P}\left(\sum_{i<j} \boldsymbol{Z}_{i,j} \geqslant \frac{n^2}{2} \cdot \frac{100n}{\gamma\sqrt{d}} \right) \leqslant 2\exp\left(- \frac{\frac{10^4}{\gamma^2 d}}{\frac{3}{\gamma^2 d n} + \frac{100}{3\gamma^2 d^{3/2} n}} \right) \\
&= 2\exp\left(-\frac{10^4 n}{3 + \frac{100}{\sqrt{d}}} \right) \leqslant \exp(-50n)\,.
\end{aligned}$$

Hence, by a union bound over all $x \in \{\pm 1\}^n$ it follows that

$$\max_{x \in \{\pm 1\}^n} \left| x^\top \big(Y(\mathbf{G}) - \tfrac{1}{n} x^* (x^*)^\top \big) x \right| \leqslant \frac{100n}{\gamma \sqrt{d}}$$

with probability at least $1 - e^{-10n}$. □

E.2 Deferred proofs for clustering mixtures of Gaussians

In this section, we will prove Lemma 7.65 restated below.

Lemma (Restatement of Lemma 7.65)**.** *Consider the settings of Theorem 7.53. Suppose* $\mathbf{Y}$ *is a good set as per Definition 7.56. Let* $W(\mathbf{Y}) \in \mathcal{W}(\mathbf{Y})$ *be the matrix computed by Algorithm 7.55. Suppose the algorithm does not reject. Then*

$$\left\|\phi(W(\mathbf{Y})) - \mathbf{W}^*\right\|_1 \leqslant \frac{n^2}{k} \cdot \frac{3}{k^{98}} .$$

We will need the following fact about our clustering program. Similar facts where used, e.g., in [HL18, FKP+19]. One difference for us is that we don't have a constraint on the lower bound on the cluster size indicated by our SoS variables. However, since we maximize a variant of the ℓ_1 norm of the second moment matrix of the pseudo-distribution this will make up for this.

Fact E.3. *Consider the same setting as in Lemma 7.65. Let* $0 < \delta \leqslant \frac{1}{1.5\cdot 10^{10}} \cdot \frac{1}{k^{201}}$ *and denote by* $\mathbf{C}_1, \dots, \mathbf{C}_k \subseteq [n]$ *the indices belonging to each true cluster. Then* $W(\mathbf{Y})$ *satisfies the following three properties:*

1. *For all* $i, j \in [n]$ *it holds that* $0 \leqslant \mathbf{W}_{i,j} \leqslant 1$,
2. *for all* $i \in [n]$ *it holds that* $\sum_{j=1}^n \mathbf{W}_{i,j} \leqslant \frac{n}{k}$ *and for at least* $(1 - \frac{1}{1000k^{100}})n$ *indices* $i \in [n]$ *it holds that* $\sum_{j=1}^n \mathbf{W}_{i,j} \geqslant (1 - \frac{1}{(10)^6 k^{200}}) \cdot \frac{n}{k}$,
3. *for all* $r \in [k]$ *it holds that* $\sum_{i\in\mathbf{C}_r, j\notin\mathbf{C}_r} \mathbf{W}_{i,j} \leqslant \delta \cdot \frac{n^2}{k}$.

We will prove Fact E.3 at the end of this section. With this in hand, we can proof Lemma 7.65.

Proof of Lemma 7.65. For brevity, we write $\mathbf{W} = W(\mathbf{Y})$. Since $\phi(\mathbf{W}^*) = \mathbf{W}^*$ and ϕ is 10-Lipschitz we can also bound

$$\left\|\phi(\mathbf{W}) - \mathbf{W}^*\right\|_1 \leqslant 10 \cdot \|\mathbf{W} - \mathbf{W}^*\|_1 .$$

Let $\delta \leqslant \frac{1}{1.5\cdot 10^{10}} \cdot \frac{1}{k^{201}}$ and again let $\mathbf{C}_1, \dots, \mathbf{C}_k \subseteq [n]$ denote the indices belonging to each true cluster. Note that by assumption that $\mathbf{Y}$ is a good sample it holds for each $r \in [k]$ that $\frac{n}{k} - n^{0.6} \leqslant |\mathbf{C}_r| \leqslant \frac{n}{k} + n^{0.6}$.

Let $r, r' \in [k]$. We can write

$$\|\mathbf{W} - \mathbf{W}^*\|_1 = \sum_{r=1}^k \sum_{i,j\in\mathbf{C}_r} \left|\mathbf{W}_{i,j} - 1\right| + \sum_{r=1}^k \sum_{i\in\mathbf{C}_r, j\notin\mathbf{C}_r} \left|\mathbf{W}_{i,j} - 0\right| \tag{E.2.1}$$

Note that we can bound the second sum by $k \cdot \delta \frac{n^2}{k}$ using Item 3. Further, in what follows consider only indices i such that $\sum_{j=1}^{n} \mathbf{W}_{i,j} \geqslant (1 - \frac{1}{(10)^6 k^{200}}) \cdot \frac{n}{k}$. By Item 2 we can bound the contribution of the other indices by

$$\frac{1}{1000k^{100}} n \cdot \left(\frac{n}{k} + n^{0.6}\right) \leqslant \frac{2}{1000k^{100}} \cdot \frac{n^2}{k} .$$

Focusing only on such indices, for the first sum in Eq. (E.2.1), fix $r \in [k]$. We will aim to show that most entries of $\mathbf{W}$ are large if and only if the corresponding entry of $\mathbf{W}^*$ is 1. By Item 3 and Markov's Inequality, it follows that for at least a $(1 - \frac{1}{1000k^{100}})$-fraction of the indices $i \in \mathbf{C}_r$ it holds that

$$\sum_{j \notin \mathbf{C}_r} \mathbf{W}_{i,j} \leqslant 1000k^{100} \cdot \delta \tfrac{n^2}{k \cdot |\mathbf{C}_r|} \leqslant 1000k^{100}\delta \cdot \tfrac{n}{1 - k \cdot n^{-0.4}} \leqslant 2000k^{101}\delta \cdot \tfrac{n}{k} ,$$

where we used that $|\mathbf{C}_r| \geqslant \frac{n}{k} - n^{0.6}$. Call such indices *good*. Notice that for good indices it follows using Item 2 that

$$\sum_{j \in \mathbf{C}_r} \mathbf{W}_{i,j} \geqslant \tfrac{n}{k} \cdot (1 - \frac{1}{(10)^6 k^{200}} - 2000k^{101}\delta) .$$

Denote by G the number of $j \in \mathbf{C}_r$ such that $\mathbf{W}_{i,j} \geqslant 1 - \frac{1}{1000k^{100}}$. Using the previous display and that $\mathbf{W}_{i,j} \leqslant 1$ we obtain

$$\begin{aligned}
\tfrac{n}{k} \cdot \left(1 - \frac{1}{(10)^6 k^{200}} - 2000k^{101}\delta\right) \leqslant \sum_{j \in \mathbf{C}_r} \mathbf{W}_{i,j} &\leqslant G \cdot 1 + (|\mathbf{C}_r| - G) \cdot (1 - \tfrac{1}{1000k^{100}}) \\
&\leqslant G \cdot \tfrac{1}{1000k^{100}} + \tfrac{n}{k} \cdot (1 + \tfrac{1}{kn^{0.4}}) \cdot (1 - \tfrac{1}{1000k^{100}}) \\
&\leqslant G \cdot \tfrac{1}{1000k^{100}} + \tfrac{n}{k} \cdot (1 + \tfrac{1}{kn^{0.4}}) ,
\end{aligned}$$

where we also used $|\mathbf{C}_r| \leqslant \frac{n}{k} + n^{0.6}$. Rearranging now yields

$$G \geqslant \frac{n}{k} \cdot \left(1 - \frac{1}{1000k^{100}} - \frac{10^3 k^{99}}{n^{0.4}} - 2 \cdot 10^6 k^{101}\delta\right) \geqslant \frac{n}{k} \cdot \left(1 - \frac{2}{1000k^{100}} - 2 \cdot 10^6 k^{101}\delta\right) .$$

We can now bound

$$\begin{aligned}
\sum_{i,j \in \mathbf{C}_r} \left|\mathbf{W}_{i,j} - 1\right| &= \sum_{i,j \in \mathbf{C}_r, i \text{ is good}} \left|\mathbf{W}_{i,j} - 1\right| + \sum_{i,j \in \mathbf{C}_r, i \text{ is not good}} \left|\mathbf{W}_{i,j} - 1\right| \\
&\leqslant |\mathbf{C}_r| \cdot \left((|\mathbf{C}_r| - G) \cdot 1 + |\mathbf{C}_r| \cdot \tfrac{1}{1000k^{100}}\right) + \tfrac{1}{1000k^{100}} \cdot |\mathbf{C}_r|^2 \\
&\leqslant |\mathbf{C}_r|^2 (1 + \tfrac{1}{500k^{100}}) - G \cdot |\mathbf{C}_r| \\
&\leqslant \tfrac{n^2}{k^2}(1 + \tfrac{k}{n^{0.4}})^2 (1 + \tfrac{1}{500k^{100}}) - \tfrac{n^2}{k^2}(1 - \tfrac{2}{1000k^{100}} - 2 \cdot 10^6 k^{101}\delta)(1 - \tfrac{k}{n^{0.4}}) \\
&\leqslant \tfrac{n^2}{k^2} \cdot (30 \cdot 10^6 k^{101}\delta + \tfrac{11}{500k^{100}}) \leqslant \tfrac{n^2}{k} \cdot (30 \cdot 10^6 k^{100}\delta + \tfrac{11}{500k^{101}})
\end{aligned}$$

$$\leqslant \tfrac{n^2}{k} \cdot \tfrac{3}{125k^{101}} \,.$$

Putting everything together, it follows that

$$\left\|\phi(\mathbf{W}) - \mathbf{W}^*\right\|_{\mathrm{F}}^2 \leqslant \left\|\phi(\mathbf{W}) - \mathbf{W}^*\right\|_1 \leqslant 10 \cdot \tfrac{n^2}{k}\left(\delta k + \frac{2}{1000k^{100}} + \tfrac{3}{125k^{100}}\right) \leqslant \tfrac{n^2}{k} \cdot \tfrac{4}{k^{100}} \leqslant \tfrac{n^2}{k} \cdot \tfrac{3}{k^{98}} \,.$$

□

It remains to verify Fact E.3.

Proof of Fact E.3. Let $\mathcal{P} = \mathcal{P}_{n,k,t}(\mathbf{Y})$ be the system of Eq. ($\mathcal{P}_{n,k,t}(Y)$). Recall that $\mathbf{W}_{i,j} = \tilde{\mathbb{E}} \sum_{l\in[k]} z_{i,l} z_{j,l}$. Since

$$\mathcal{P} \vdash_4 \left\{ 0 \leqslant \sum_{l\in[k]} z_{i,l} z_{j,l} \leqslant \sum_{l\in[k]} z_{i,l} \leqslant 1 \right\},$$

it follows that $0 \leqslant \mathbf{W}_{i,j} \leqslant 1$. Further, for each $i \in [n]$ it holds that

$$\mathcal{P} \vdash_4 \left\{ \sum_{j\in[n], l\in[k]} z_{j,l} z_{i,l} \leqslant \tfrac{n}{k} \sum_{l\in[k]} z_{i,l} \leqslant \tfrac{n}{k} \right\}$$

implying that $\sum_{j\in[n]} \mathbf{W}_{i,j} \leqslant \frac{n}{k}$. Further, by Lemma 7.64

$$\|\mathbf{W}\|_1 \geqslant \frac{n^2}{k} \cdot \left(1 - n^{-0.4} - \frac{1}{(10)^{10}k^{300}}\right) \geqslant \frac{n^2}{k} \cdot \left(1 - \frac{1}{(10)^9 k^{300}}\right).$$

Denote by $\mathbf{W}_i$ the i-th row of $\mathbf{W}$ and by L the number of rows which have ℓ_1 norm at least $(1 - \frac{1}{(10)^6 k^{200}}) \cdot \frac{n}{k}$. Since for all i it holds that $\|\mathbf{W}_i\|_1 \leqslant \frac{n}{k}$ it follows that

$$\begin{aligned}\frac{n^2}{k} \cdot \left(1 - \frac{1}{(10)^9 k^{300}}\right) \leqslant \sum_{i\in[n]} \|\mathbf{W}_i\|_1 &\leqslant L \cdot \frac{n}{k} + (n - L) \cdot \left(1 - \tfrac{1}{(10)^6 k^{200}}\right) \cdot \frac{n}{k} \\ &= L \cdot \tfrac{1}{(10)^6 k^{200}} \cdot \frac{n}{k} + \frac{n^2}{k} \cdot \left(1 - \tfrac{1}{(10)^6 k^{200}}\right)\end{aligned}$$

Rearranging then yields $L \geqslant (1 - \frac{1}{1000k^{100}}) \cdot n$ which proofs Item 2.

It remains to verify Item 3. Fix $r, l \in [k]$ and define $z_l(\mathbf{C}_r) = \frac{k}{n} \sum_{i\in\mathbf{C}_r} z_{i,l}$. Let $t > 0$ be an integer. We aim to show that for all unit vectors v it holds that

$$\mathcal{P} \vdash_{10t} \left\{ z_l(\mathbf{C}_r) \cdot \frac{1}{\Delta^{2t}} \sum_{r'\neq r} z_l(\mathbf{C}_{r'}) \langle \mu_r - \mu_{r'}, v\rangle^{2t} \leqslant \frac{\delta}{k} \right\}, \tag{E.2.2}$$

where Δ is the minimal separation between the true means. Before proving this, let us examine how we can use this fact to prove Item 3. Note, that for all $r \neq r'$ it holds that

$$\sum_{s,u\in[k]} \left\langle \mu_r - \mu_{r'}, \tfrac{\mu_s - \mu_u}{\|\mu_s - \mu_u\|} \right\rangle^{2t} \geqslant \Delta^{2t} \,.$$

Hence, if the above SoS proof indeed exists, we obtain

$$\begin{aligned}\sum_{i\in\mathbf{C}_r, j\notin\mathbf{C}_r}\mathbf{W}_{i,j} &= \sum_{l=1}^{k}\tilde{\mathbb{E}}\sum_{i\in\mathbf{C}_r, j\notin\mathbf{C}_r} z_{i,l}z_{j,l} = \frac{n^2}{k^2}\tilde{\mathbb{E}}z_l(\mathbf{C}_r)\cdot\sum_{r'\neq r} z_l(\mathbf{C}_{r'})\\ &\leqslant \frac{n^2}{\Delta^{2t}k^2}\sum_{s,u\in[k]}\tilde{\mathbb{E}}z_l(\mathbf{C}_r)\cdot\sum_{r'\neq r} z_l(\mathbf{C}_r)\left\langle \mu_r-\mu_{r'}, \tfrac{\mu_s-\mu_u}{\|\mu_s-\mu_u\|}\right\rangle^{2t}\\ &\leqslant \frac{\delta}{k}k^2\cdot\frac{n^2}{k^2} = \delta\cdot\frac{n^2}{k}\,.\end{aligned}$$

In the remainder of this proof we will prove Eq. (E.2.2). We will use the following SoS version of the triangle Inequality (cf. Fact E.14)

$$\left|\frac{x,y}{2t}\right. (x+y)^{2t} \leqslant 2^{2t-1}(x^{2t}+y^{2t})\,.$$

Recall that $\mu'_l = \frac{k}{n}\sum_{i=1}^{n} z_{i,l}y_i$ and denote by $\mu_{\pi(i)}$ the true mean corresponding to the i-th sample. Let v be an arbitrary unit vector, it follows that

$$\begin{aligned}&\mathcal{P}\left|\frac{}{10t}\right. \{z_l(\mathbf{C}_r)\cdot\frac{1}{\Delta^{2t}}\sum_{r'\neq r} z_l(\mathbf{C}_{r'})\langle\mu_r-\mu_{r'},v\rangle^{2t}\\ &\quad\leqslant z_l(\mathbf{C}_r)\cdot\frac{2^{2t-1}}{\Delta^{2t}}\sum_{r'\neq r} z_l(\mathbf{C}_{r'})\big(\langle\mu_r-\mu'_l,v\rangle^{2t}+\langle\mu_{r'}-\mu'_l,v\rangle^{2t}\big)\\ &\quad\leqslant \frac{2^{2t-1}}{\Delta^{2t}}\sum_{r=1}^{k} z_l(\mathbf{C}_r)\langle\mu_r-\mu'_l,v\rangle^{2t} = \frac{2^{2t-1}}{\Delta^{2t}}\cdot\frac{k}{n}\sum_{i=1}^{n} z_{i,l}\langle\mu_{\pi(i)}-\mu'_l,v\rangle^{2t}\}\,,\end{aligned}$$

where we used that $\mathcal{P}\left|\frac{}{1}\right. \sum_{r=1}^{k} z_l(\mathbf{C}_r)\leqslant 1$. Using the SoS triangle inequality again and that $\mathcal{P}\left|\frac{}{2}\right. z_{i,l}\leqslant 1$ we obtain

$$\begin{aligned}&\mathcal{P}\left|\frac{10t}{}\right. \{z_l(\mathbf{C}_r)\cdot\frac{1}{\Delta^{2t}}\sum_{r'\neq r} z_l(\mathbf{C}_{r'})\langle\mu_r-\mu_{r'},v\rangle^{2t}\\ &\quad\leqslant \frac{2^{4t-1}}{\Delta^{2t}}\cdot\left(k\cdot\frac{1}{n}\sum_{i=1}^{n}\langle\mathbf{y}_i-\mu_{\pi(i)},v\rangle^{2t}+\frac{k}{n}\sum_{i=1}^{n} z_{i,l}\langle\mathbf{y}_i-\mu'_l,v\rangle^{2t}\right)\}\,.\end{aligned}$$

We start by bounding the first sum. Recall that by assumption the uniform distribution over each true cluster is $2t$-explicitly 2-bounded. It follows that

$$\left|\frac{}{2t}\right. \{\frac{1}{n}\sum_{i=1}^{n}\langle\mathbf{y}_i-\mu_{\pi(i)},v\rangle^{2t} = \frac{1}{k}\sum_{r=1}^{k}\frac{k}{n}\sum_{i\in\mathbf{C}_r}\langle\mathbf{y}_i-\mu_r,v\rangle^{2t} \leqslant \frac{1}{k}\sum_{r=1}^{k}\frac{k}{n}\cdot|\mathbf{C}_r|\cdot(2t)^t\cdot\|v\|_2^{2t} \tag{E.2.3}$$

$$\leqslant \left(1+\frac{k}{n^{0.4}}\right)\cdot(2t)^t \leqslant 2(2t)^t\}\,, \tag{E.2.4}$$

where we used that $|\mathbf{C}_r| \leqslant \frac{n}{k} + n^{0.6}$. To bound the second sum, we will use the moment bound constraints. In particular, we know that

$$\mathcal{P} \mid\!\!\!\frac{}{10t} \left\{ \frac{k}{n} \sum_{i=1}^{n} z_{i,l} \langle \mathbf{y}_i - \mu_l', v \rangle^{2t} \leqslant (2t)^t \right\}. \tag{E.2.5}$$

Combining Eq. (E.2.4) and Eq. (E.2.5) now yields

$$\mathcal{P} \mid\!\!\!\frac{}{10t} \left\{ z_l(\mathbf{C}_r) \cdot \frac{1}{\Delta^{2t}} \sum_{r' \neq r} z_l(\mathbf{C}_{r'}) \langle \mu_r - \mu_{r'}, v \rangle^{2t} \leqslant k \frac{2^{2t+1}(2t)^t}{\Delta^{2t}} \leqslant k \left(\frac{8t}{\Delta^2} \right)^t \right\}.$$

Note that by assumption $\Delta \geqslant O(\sqrt{t k^{1/t}})$. Overloading notation, we can choose the t parameter in the SoS proof to be 202 times the t parameter in the lower bound in the separation to obtain[1]

$$\sum_{i \in \mathbf{C}_r, j \notin \mathbf{C}_r} \mathbf{W}_{i,j} \leqslant \delta \cdot \frac{n^2}{k}.$$

□

E.2.1 Privatizing input using the Gaussian Mechanism

In this section, we will proof the following helpful lemma used in the privacy analysis of our clustering algorithm (Algorithm 7.55). In summary, it says that when restricted to some set our input has small ℓ_2 sensitivity, we can first add Gaussian noise proportional to this sensitivity and afterwards treat this part of the input as "privatized". In particular, for the remainder of the privacy analysis we can treat this part as the same on adjacent inputs. Note that we phrase the lemma in terms of matrix inputs since this is what we use in our application. Of course, it also holds for more general inputs.

Lemma E.4. *Let $V, V' \in \mathbb{R}^{n \times d}$, $m \in [n]$ and $\Delta > 0$ be such that there exists a set S of size at least $n - m$ satisfying*

$$\forall i \in S. \left\| V_i - V_i' \right\|_2^2 \leqslant \Delta^2,$$

where V_i, V_i' denote the rows of V, V', respectively. Let $\mathcal{A}_2 \colon \mathbb{R}^{n \times d} \to \mathcal{O}$ be an algorithm that is $(\varepsilon_2, \delta_2)$-differentially private in the standard sense, i.e,., for all sets $\mathcal{S} \subseteq \mathcal{O}$ and datasets $X, X' \in \colon \mathbb{R}^{n \times d}$ differing only in a single row it holds that

$$\mathbb{P}(\mathcal{A}_2(X) \in S) \leqslant e^{\varepsilon_2} \mathbb{P}(\mathcal{A}_2(X') \in S) + \delta_2.$$

Further, let $\mathcal{A}_1 \colon \mathbb{R}^{n \times d} \to \mathbb{R}^{n \times d}$ be the Gaussian Mechanism with parameters $\Delta, \varepsilon_1, \delta_1$. I.e., on input M it samples $\mathbf{W} \sim N\left(0, 2\Delta^2 \cdot \frac{\log(2/\delta_1)}{\varepsilon_1^2}\right)^{n \times d}$ and outputs $M + \mathbf{W}$.

[1] Note that this influences the exponent in the running time and sample complexity only by a constant factor and hence doesn't violate the assumptions of Theorem 7.53.

Then for

$$\varepsilon' \coloneqq \varepsilon_1 + m\varepsilon_2\,,$$
$$\delta' \coloneqq e^{\varepsilon_1} m e^{(m-1)\varepsilon_2}\delta_2 + \delta_1\,.$$

$\mathcal{A}_2 \circ \mathcal{A}_1$ is (ε', δ')-differentially private with respect to V and V', i.e., for all sets $\mathcal{S} \subseteq \mathcal{O}$ it holds that

$$\mathbb{P}((\mathcal{A}_2 \circ \mathcal{A}_1)(V) \in S) \leqslant e^{\varepsilon'}\mathbb{P}((\mathcal{A}_2 \circ \mathcal{A}_1)(V') \in S) + \delta'\,.$$

Proof. Without loss of generality, assume that $S = \{1, \ldots, m\}$. Denote by V_1, V_2 the first m and last $n - m$ rows of V respectively. Analogously for V_1', V_2'. We will later partitin the noise $\mathbf{W}$ of the Gaussian mechanism in the same way. Further, for a subset A of $\mathbb{R}^{n\times n}$ and $Y \in \mathbb{R}^{m\times n}$ define

$$T_{A,Y} = \left\{ X \in \mathbb{R}^{(n-m)\times n} \;\middle|\; \begin{pmatrix} X \\ Y \end{pmatrix} \in A \right\} \subseteq \mathbb{R}^{(n-m)\times n}\,.$$

Note that $\begin{pmatrix} X \\ Y \end{pmatrix} \in A$ if and only if $X \in T_{A,Y}$.

Let $\mathcal{S} \subseteq \mathcal{O}$. It now follows that

$$\begin{aligned}
\mathbb{P}_{\mathcal{A}_2,\mathbf{W}}[(\mathcal{A}_2 \circ \mathcal{A}_1)(V) \in S] &= \mathop{\mathbb{E}}_{\mathcal{A}_2,\mathbf{W}}\left[\mathbb{1}\left\{V + \mathbf{W} \in \mathcal{A}_2^{-1}(S)\right\}\right] \\
&= \mathop{\mathbb{E}}_{\mathcal{A}_2,\mathbf{W}_2}\left[\mathop{\mathbb{E}}_{\mathbf{W}_1}\left[\mathbb{1}\left\{\begin{pmatrix} V_1 + \mathbf{W}_1 \\ V_2 + \mathbf{W}_2 \end{pmatrix} \in \mathcal{A}_2^{-1}(S)\right\}\right] \;\middle|\; \mathbf{W}_2\right] \\
&= \mathop{\mathbb{E}}_{\mathcal{A}_2,\mathbf{W}_2}\left[\mathop{\mathbb{E}}_{\mathbf{W}_1}\left[\mathbb{1}\left\{V_1 + \mathbf{W}_1 \in T_{\mathcal{A}_2^{-1}(S),V_2+\mathbf{W}_2}\right\}\right] \;\middle|\; \mathbf{W}_2\right] \\
&\leqslant e^{\varepsilon_1} \cdot \mathop{\mathbb{E}}_{\mathcal{A}_2,\mathbf{W}_2}\left[\mathop{\mathbb{E}}_{\mathbf{W}_1}\left[\mathbb{1}\left\{V_1' + \mathbf{W}_1 \in T_{\mathcal{A}_2^{-1}(S),V_2+\mathbf{W}_2}\right\}\right] \;\middle|\; \mathbf{W}_2\right] + \delta_1 \\
&= e^{\varepsilon_1} \cdot \mathop{\mathbb{E}}_{\mathcal{A}_2,\mathbf{W}}\left[\mathbb{1}\left\{\begin{pmatrix} V_1' + \mathbf{W}_1 \\ V_2 + \mathbf{W}_2 \end{pmatrix} \in \mathcal{A}_2^{-1}(S)\right\}\right] + \delta_1\,,
\end{aligned}$$

where the inequality follows by the guarantees of the Gaussian Mechanism. Further, we can bound

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathcal{A}_2,\mathbf{W}}\left[\mathbb{1}\left\{\begin{pmatrix} V_1' + \mathbf{W}_1 \\ V_2 + \mathbf{W}_2 \end{pmatrix} \in \mathcal{A}_2^{-1}(S)\right\}\right] &= \mathop{\mathbb{E}}_{\mathbf{W}}\left[\mathop{\mathbb{E}}_{\mathcal{A}_2}\left[\mathbb{1}\left\{\mathcal{A}_2\begin{pmatrix} V_1' + \mathbf{W}_1 \\ V_2 + \mathbf{W}_2 \end{pmatrix} \in S\right\} \;\middle|\; \mathbf{W}\right]\right] \\
&\leqslant e^{m\varepsilon_2} \cdot \mathop{\mathbb{E}}_{\mathbf{W}}\left[\mathop{\mathbb{E}}_{\mathcal{A}_2}\left[\mathbb{1}\left\{\mathcal{A}_2\begin{pmatrix} V_1' + \mathbf{W}_1 \\ V_2' + \mathbf{W}_2 \end{pmatrix} \in S\right\} \;\middle|\; \mathbf{W}\right]\right] + me^{(m-1)\varepsilon_2}\delta_2 \\
&= e^{m\varepsilon_2} \cdot \mathop{\mathbb{E}}_{\mathcal{A}_2,\mathbf{W}}\left[\mathbb{1}\left\{\begin{pmatrix} V_1' + \mathbf{W}_1 \\ V_2' + \mathbf{W}_2 \end{pmatrix} \in \mathcal{A}_2^{-1}(S)\right\}\right] + me^{(m-1)\varepsilon_2}\delta_2\,,
\end{aligned}$$

where the inequality follows by the privacy guarantees of $\mathcal{A}_2$ combined with standard group privacy arguments.

Putting the above two displays together and plugging in the definition of ε', δ' we finally obtain

$$\mathbb{P}_{\mathcal{A}_2,\mathbf{W}}[(\mathcal{A}_2 \circ \mathcal{A}_1)(V) \in S] \leqslant e^{\varepsilon'} \mathbb{P}_{\mathcal{A}_2,\mathbf{W}}[(\mathcal{A}_2 \circ \mathcal{A}_1)(V') \in S] + \delta' .$$

□

E.3 Additional tools

Concentration of measure

We introduce here several useful and standard concentration inequalities.

Fact E.5 (Concentration of spectral norm of Gaussian matrices)**.** *Let* $\mathbf{W} \sim \mathcal{N}(0,1)^{m \times n}$*. Then for any* t*, we have*

$$\mathbb{P}\left(\sqrt{m} - \sqrt{n} - t \leqslant \sigma_{\min}(\mathbf{W}) \leqslant \sigma_{\max}(\mathbf{W}) \leqslant \sqrt{m} + \sqrt{n} + t\right) \geqslant 1 - 2\exp\left(-\frac{t^2}{2}\right),$$

where $\sigma_{\min}(\cdot)$ *and* $\sigma_{\max}(\cdot)$ *denote the minimum and the maximum singular values of a matrix, respectively.*

Let $\mathbf{W}'$ *be an* n*-by-*n *symmetric matrix with independent entries sampled from* $N(0, \sigma^2)$*. Then* $\|\mathbf{W}'\| \leqslant 3\sigma\sqrt{n}$ *with probability at least* $1 - \exp(-\Omega(n))$*.*

Fact E.6 (Maximum degree of Erdős-Rényi graphs)**.** *Let* G *be an Erdős-Rényi graph on* n *vertices with edge probability* p*. Then with probability at least* $1 - n\exp(-np/3)$*, any vertex in* G *has degree at most* $2np$*.*

Fact E.7 (Gaussian concentration bounds)**.** *Let* $\mathbf{X} \sim \mathcal{N}(0, \sigma^2)$*. Then for any* $t \geqslant 0$*,*

$$\max\{\mathbb{P}(\mathbf{X} \geqslant t), \mathbb{P}(\mathbf{X} \leqslant -t)\} \leqslant \exp\left(-\frac{t^2}{2\sigma^2}\right).$$

Fact E.8 (Chernoff bound)**.** *Let* $\mathbf{X_1}, \ldots, \mathbf{X_n}$ *be independent random variables taking values in* $\{0, 1\}$*. Let* $\mathbf{X} := \sum_{i=1}^{n} \mathbf{X_i}$ *and let* $\mu := \mathbb{E}\,\mathbf{X}$*. Then for any* $\delta > 0$*,*

$$\mathbb{P}\big(\mathbf{X} \leqslant (1-\delta)\mu\big) \leqslant \exp\left(-\frac{\delta^2\mu}{2}\right),$$

$$\mathbb{P}\big(\mathbf{X} \geqslant (1+\delta)\mu\big) \leqslant \exp\left(-\frac{\delta^2\mu}{2+\delta}\right).$$

Lemma E.9 (Restatement of Lemma 8.9)**.** *Let* Φ *be a* d*-by-*n *Gaussian matrix, with each entry independently chosen from* $N(0, 1/d)$*. Then, for every vector* $u \in \mathbb{R}^n$ *and every* $\alpha \in (0, 1)$

$$\mathbb{P}(\|\Phi u\| = (1 \pm \alpha)\|u\|) \geqslant 1 - e^{-\Omega(\alpha^2 d)} .$$

Linear algebra

Lemma E.10 (Weyl's inequality)**.** *Let A and B be symmetric matrices. Let $R = A - B$. Let $\alpha_1 \geqslant \cdots \geqslant \alpha_n$ be the eigenvalues of A. Let $\beta_1 \geqslant \cdots \geqslant \beta_n$ be the eigenvalues of B. Then for each $i \in [n]$,*

$$\left|\alpha_i - \beta_i\right| \leqslant \|R\|.$$

Lemma E.11 (Davis-Kahan's theorem)**.** *Let A and B be symmetric matrices. Let $R = A - B$. Let $\alpha_1 \geqslant \cdots \geqslant \alpha_n$ be the eigenvalues of A with corresponding eigenvectors $v_1, \ldots, v_n$. Let $\beta_1 \geqslant \cdots \geqslant \beta_n$ be the eigenvalues of B with corresponding eigenvectors $u_1, \ldots, u_n$. Let θ_i be the angle between $\pm v_i$ and $\pm u_i$. Then for each $i \in [n]$,*

$$\sin(2\theta_i) \leqslant \frac{2\|R\|}{\min_{j \neq i}\left|\alpha_i - \alpha_j\right|}.$$

Convex optimization

Proposition E.12. *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a convex function. Let $\mathcal{K} \subseteq \mathbb{R}^m$ be a convex set. Then $y^* \in \mathcal{K}$ is a minimizer of f over $\mathcal{K}$ if and only if there exists a subgradient $g \in \partial f(y^*)$ such that*

$$\left\langle y - y^*, g \right\rangle \geqslant 0 \quad \forall y \in \mathcal{K}.$$

Proof. Define indicator function

$$I_{\mathcal{K}}(y) = \begin{cases} 0, & y \in \mathcal{K}, \\ \infty, & y \notin \mathcal{K}. \end{cases}$$

Then for $y \in \mathcal{K}$, one has

$$\partial I_{\mathcal{K}}(y) = \left\{ g \in \mathbb{R}^m : \left\langle g, y - y' \right\rangle \geqslant 0 \ \forall y' \in \mathcal{K} \right\}.$$

Note y^* is a minimizer of f over $\mathcal{K}$, if and only if y^* is a minimizer of $f + I_{\mathcal{K}}$ over $\mathbb{R}^m$, if and only if $\mathbf{0}_m \in \partial(f + I_{\mathcal{K}})(y^*) = \partial f(y^*) + \partial I_{\mathcal{K}}(y^*)$, if and only if there exists $g \in \partial f(y^*)$ such that $\langle g, y - y^* \rangle \geqslant 0$ for any $y \in \mathcal{K}$. □

Proposition E.13 (Pythagorean theorem from strong convexity)**.** *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a convex function. Let $\mathcal{K} \subseteq \mathbb{R}^m$ be a convex set. Suppose f is κ-strongly convex over $\mathcal{K}$. Let $x^* \in \mathcal{K}$ be a minimizer of f over $\mathcal{K}$. Then for any $x \in \mathcal{K}$, one has*

$$\|x - x^*\|^2 \leqslant \frac{2}{\kappa}(f(x) - f(x^*)).$$

Proof. By strong convexity, for any subgradient $g \in \partial f(x^*)$ one has

$$f(x) \geqslant f(x^*) + \left\langle x - x^*, g \right\rangle + \frac{\kappa}{2}\|x - x^*\|^2.$$

By Proposition E.12, $\langle x - x^*, g \rangle \geqslant 0$ for some $g \in \partial f(x^*)$. Then the result follows. □

Minor Lemmas

Fact E.14 (Lemma A.2 in [KSS18])**.** *For all integers $t > 0$ it holds that*

$$\left|\tfrac{x,y}{2t}\right. (x+y)^{2t} \leqslant 2^{2t-1}(x^{2t}+y^{2t})\,.$$

Fact E.15. *Let $\varepsilon, \delta > 0$. Let $\mathcal{M}\colon \mathcal{Y} \to \mathcal{O}$ be a randomized algorithm that, for every pair of adjacent inputs, with probability at least $1-\gamma \geqslant 1/2$ over the internal randomness of $\mathcal{Y}$[2] satisfies (ε, δ)-privacy. Then $\mathcal{M}$ is $(\varepsilon + 2\gamma, \delta + \gamma)$-private.*

Proof. Let X, X' be adjacent input and let B be the event under which $\mathcal{M}$ is (ε, δ)-private. By assumption, we know that $\mathbb{P}(B) \geqslant 1-\gamma$. Let $S \in \mathcal{O}$, it follows that

$$\begin{aligned}
\mathbb{P}(\mathcal{M}(X) \in S) &= \mathbb{P}(B)\cdot\mathbb{P}(\mathcal{M}(X)\in S \mid B) + \mathbb{P}(B^c)\cdot\mathbb{P}(\mathcal{M}(X)\in S \mid B^c)\\
&\leqslant \mathbb{P}(\mathcal{M}(X)\in S \mid B) + \gamma\\
&\leqslant e^{\varepsilon}\mathbb{P}(\mathcal{M}(X)\in S \mid B) + \delta + \gamma\\
&\leqslant \frac{e^{\varepsilon}}{\mathbb{P}(B)}\cdot\mathbb{P}(\mathcal{M}(X)\in S) + \delta + \gamma\\
&\leqslant e^{\varepsilon+\log\left(\frac{1}{1-\gamma}\right)}\cdot\mathbb{P}(\mathcal{M}(X)\in S) + (\delta+\gamma)\\
&\leqslant e^{\varepsilon+2\gamma}\cdot\mathbb{P}(\mathcal{M}(X)\in S) + (\delta+\gamma)\,,
\end{aligned}$$

where we used that $\log(1-\gamma) \geqslant -2\gamma$ for $\gamma \in [0, 1/2]$. □

[2] In particular, this randomness is independent of the input