Local Statistics, Semidefinite Programming, and Community Detection

Jess Banks * Sidhanth Mohanty † Prasad Raghavendra‡

September 21, 2020

Abstract

We propose a new hierarchy of semidefinite programming relaxations for inference problems. As test cases, we consider the problem of community detection in block models. The vertices are partitioned into k communities, and a graph is sampled conditional on a prescribed number of inter- and intra-community edges. The problem of *detection*, where we are to decide with high probability whether a graph was drawn from this model or the uniform distribution on regular graphs, is conjectured to undergo a computational phase transition at a point called the Kesten-Stigum (KS) threshold.

In this work, we consider two models of random graphs namely the well-studied (irregular) stochastic block model and a distribution over random regular graphs we'll call the Degree Regular Block Model. For both these models, we show that sufficiently high constant levels of our hierarchy can perform detection arbitrarily close to the KS threshold and that our algorithm is robust to up to a linear number of adversarial edge perturbations. Furthermore, in the case of Degree Regular Block Model, we show that below the Kesten-Stigum threshold no constant level can do so.

In the case of the (irregular) Stochastic Block Model, it is known that efficient algorithms exist all the way down to this threshold, although none are robust to a linear number of adversarial perturbations of the graph when the average degree is small. More importantly, there is little complexity-theoretic evidence that detection is hard below the threshold. In the DRBM with more than two groups, it has not to our knowledge been proven that any algorithm succeeds down to the KS threshold, let alone that one can do so robustly, and there is a similar dearth of evidence for hardness below this point.

Our SDP hierarchy is highly general and applicable to a wide range of hypothesis testing problems.

*University of California, Berkeley
†University of California, Berkeley
‡University of California, Berkeley

Contents

1 Introduction

Community detection is a canonical example of a high-dimensional inference problem, one that is a test-bed to develop algorithmic and lower bound techniques. Much of the existing literature on community detection concerns the *stochastic block model (SBM)*. For now let us discuss the *symmetric* setting where we first partition n vertices in to k groups, and include each edge independently and with probability p_{in} or p_{out} depending on whether or not the labels of its endpoints coincide. Research in this area spans several decades, and it will not be fruitful to attempt a thorough review of the literature here; we refer the reader to [Abb17] for a survey. Most salient to us, however, is a rich theory of computational threshold phenomena which has emerged out of the past several years of collaboration between computer scientists, statisticians, and statistical physicists.

The key computational tasks associated with the SBM are *recovery* and *detection*: we attempt either to reconstruct the planted communities from the graph, or to decide whether a graph was drawn from the planted model or the Erdős-Rényi model with the same average degree. A set of fascinating conjectures were posed in Decelle et al. [DKMZ11b], regarding these tasks in the case of 'sparse' models where $p_{\text{in}}, p_{\text{out}} = O(1/n)$ and the average degree is $O(1)$ as the number of vertices diverges.

It is typical to parametrize the symmetric SBM in terms of k, the average degree

$$d = \frac{np_{\text{in}} + (k-1)np_{\text{out}}}{k},$$

and a 'signal-to-noise ratio'

$$\lambda \triangleq \frac{np_{\text{in}} - np_{\text{out}}}{kd}.$$

In this setup, it is believed that as we hold k and λ constant, then there is an *information-theoretic threshold* $d_{IT} \approx \frac{\log k}{k\lambda^2}$, in the sense that when $d < d_{IT}$ both detection and recovery are impossible for any algorithm. Moreover, Decelle et al. conjecture that efficient algorithms for both tasks exist only when the degree is larger than a point known as the *Kesten-Stigum threshold* $d_{KS} = \lambda^{-2}$. Much of this picture is now rigorous [MNS18, Mas14, BLM15, ABH16]. Still, fundamental questions remain unanswered. What evidence can we furnish that detection and recovery are indeed intractible in the so-called 'hard regime' $d_{IT} < d < d_{KS}$? How robust are these thresholds to adversarial noise or small deviations from the model?

Zooming out, this discrepancy between information-theoretic and computational thresholds is conjectured to be quite universal among planted problems, where we are to reconstruct or detect a structured, high-dimensional signal observed through a noisy channel. The purpose behind our work is to begin developing a framework capable of providing evidence for average case computational intractability in such settings. To illustrate this broader motivation, consider a different average-case problem also conjectured to be computationally intractable: refutation of random 3-SAT. A random instance of 3-SAT with n literals and, say $m = 1000n$ clauses is unsatisfiable with high probability. However, it is widely conjectured that the problem of *certifying* that a given random 3-SAT instance is unsatisfiable is computationally intractable (all the way

up to $n^{3/2}$ clauses) [Fei02]. While proving intractability remains out of reach, the complexity theoretic literature now contains ample evidence in support of this conjecture. Most prominently, exponential lower bounds are known for the problem in restricted computational models such as linear and semidefinite programs [Gri01] and resolution based proofs [BSW01]. Within the context of combinatorial optimization, the Sum-of-Squares (SoS) SDPs yield a hierarchy of successively more powerful and complex algorithms which capture and unify many other known approaches. A lower bound against the SoS SDP hierarchy such as [Gri01] provides strong evidence that this refutation problem is computationally intractable. This paper is a step towards developing a similar framework to reason about the computational complexity of detection and recovery in stochastic block models specifically, and planted problems generally.

A second motivation is the issue of robustness of computational thresholds under adversarial perturbations of the graph. Spectral algorithms based on non-backtracking walk matrix [BLM15] achieve weak-detection as soon as $d > d_{KS}$, but are not robust in this sense. Conversely, robust algorithms for recovery are known, but only when the edge-densities are significantly higher than Kesten-Stigum [GV16, MMV16, CSV17, SVC16]. The positive result that gets closest to robustly achieving the conjectured computational phase transition at d_{KS} is the work of Montanari and Sen [MS15] who observe that their SDP-based algorithm for testing whether the input graph comes from the Erdős-Rényi distribution or a Stochastic Block Model with $k = 2$ communities also works in presence of $o(|E|)$ edge outlier errors. On the negative side, Moitra et al. [Moi12] consider the problem of weak recovery in a SBM with two communities and $p_{\text{in}} > p_{\text{out}}$ in the presence of *monotone errors* that add edges within communities and delete edges between them. Their main result is a statistical lower bound indicating the phase transition for weak recovery changes in the presence of monotone errors. This still leaves open the question of whether there exist algorithms that weakly recover right at the threshold and are robust to $o(|E|)$ perturbations in the graph.

2 Main Results

We define a new hierarchy of semidefinite programming relaxations for inference problems that we refer to as the *Local Statistics* hierarchy, denoted $\text{LoSt}(D_G, D_x)$ and indexed by parameters $D_G, D_x \in \mathbb{N}$. This family of SDPs is inspired by the technique of pseudocalibration in proving lower bounds for sum-of-squares (SoS) relaxations, as well as subsequent work of Hopkins and Steurer [HS17] extending it to an SoS SDP based approach to inference problems. The LoSt hierarchy can be defined for a broad range of inference problems involving a joint distribution μ on an observation and hidden parameter.

As test cases, we apply our SDP relaxations to community detection in two families of random graphs with planted community structure: the sparse Stochastic Block Model (SBM) discussed above, and a degree-regular analogue that we term the *Degree Regular Block Model (DRBM)*. Our results will concern the problem of *detection*, defined formally as follows.

Definition 2.1 (Detection and Robustness)**.** Let $\mathcal{P}_n$ and $\mathcal{N}_n$ denote two sequences of distributions on graphs. We say that an algorithm $\mathsf{A} : \text{Graphs} \to \{\text{P}, \text{N}\}$ *solves the detection problem*, or *can distinguish*

$\mathcal{P}_n$ *and* $\mathcal{N}_n$. if

$$\mathcal{P}_n\big[\mathsf{A}(G) = \mathrm{P}\big] = 1 - o_n(1) \qquad \text{and} \qquad \mathcal{N}_n\big[\mathsf{A}(G) = \mathrm{N}\big] = 1 - o_n(1).$$

Fix $\epsilon > 0$, and write $G \approx_\epsilon \widetilde{G}$ to mean that two graphs on the same vertex set V differ at at most $\epsilon|V|$ edges. If A solves the detection problem, we say that it does so *ϵ-robustly* if

$$\mathcal{P}_n\big[\mathsf{A}(G) = \mathsf{A}(\widetilde{G}),\, \forall G \approx_\epsilon \widetilde{G}\big] = 1 - o_n(1) \qquad \text{and} \qquad \mathcal{N}_n\big[\mathsf{A}(G) = \mathsf{A}(\widetilde{G}),\, \forall G \approx_\epsilon \widetilde{G}\big] = 1 - o_n(1).$$

The Stochastic Block Model Adapting notation from [BLM15], we will parameterize the SBM by average degree d, number of communities k, group size distribution $\pi \in \mathbb{R}^k$, and symmetric, nonnegative edge probability matrix $M \in \mathbb{R}^{k\times k}$. To sample a graph $\boldsymbol{G} = (V(\boldsymbol{G}), E(\boldsymbol{G}))$, first choose the label $\sigma(u)$ of each vertex $u \in V(\boldsymbol{G})$ independently according to π, and then include each potential edge (u, v) with probability $M_{\sigma(u),\sigma(v)} \cdot d/n$. We adopt the natural requirement that the average degree of a vertex conditional on any group label is d, which is equivalent to the normalization condition $M\pi = e$, where the latter is the all-ones vector in $\mathbb{R}^k$. We will call the model *symmetric* if

$$M_{i,j} = \begin{cases} 1 + (k-1)\lambda & i = j \\ 1 - \lambda & i \neq j. \end{cases} \tag{1}$$

One can check that this recovers the setup in the previous section.

The general SBM, like this symmetric subcase, is conjectured to undergo a series of phase transitions as (k, M, π) are held fixed and the average degree is varied. These include an information-theoretic threshold and, most salient to this paper, a computational 'Kesten-Stigum' transition [DKMZ11a]. To describe the latter, it is necessary to introduce one further piece of notation, which will be of repeated use to us in the course of the paper. Write $T \triangleq M \operatorname{Diag} \pi$, noting that T is the transition matrix for a reversible Markov chain with stationary distribution π. For any vertex in group i, the label of a uniformly random neighbor is roughly distributed according to the ith row of T, and, more generally, the vertex labels encountered by a random non-backtracking random walk are approximately governed by the Markov process that T defines. As this process is stationary, the spectrum of T is real, and we will write its eigenvalues as $1 = \lambda_1 \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_k|$. The second eigenvalue λ_2 is a generalization of the signal-to-noise ratio λ from equation (1); in fact one can verify that in the symmetric SBM, $\lambda_2 = \cdots = \lambda_k = \lambda$. The Kesten-Stigum threshold is thus defined as $d_{\mathrm{KS}} \triangleq \lambda_2^{-2}$.

Our main theorem regarding the SBM is that, when $d > d_{\mathrm{KS}}$, the $\mathrm{LoSt}(2, D)$ SDP can robustly solve the detection problem for some $D = O(1)$ (albeit tending to infinity as $d \to d_{\mathrm{KS}}$).

Theorem 2.2. *Let $\mathcal{N}_n = \mathcal{G}(n, d/n)$, and $\mathcal{P}_n$ denote the n-vertex SBM with parameters (d, k, M, π). If $d > d_{\mathrm{KS}}$, then there exist $\delta > 0$, $D = O(1)$, and $\rho > 0$ (all dependent on d) for which the* $\mathrm{LoSt}(2, D)$ *SDP with error tolerance δ can ρ-robustly solve the detection problem.*

We additionally show that a simplified version of the $\mathrm{LoSt}(2, D)$ SDP (Definition 6.2) which is

powerful enough to solve the detection problem above the KS threshold, fails to do so below it at every constant level. This is the content of the forthcoming Theorem 6.3.

The Degree Regular Block Model We will parametrize the DRBM identically to the SBM, by a quadruple (d, k, M, π); this time we of course require that d is an integer. To sample a graph $\boldsymbol{G} = (V(\boldsymbol{G}), E(\boldsymbol{G}))$, first choose a uniformly random "π-balanced" partition $V(\boldsymbol{G}) = \bigsqcup_{i\in[k]} V_i(\boldsymbol{G})$, by which we mean that $|V_i(\boldsymbol{G})| = \pi(i)n$ for every i. Then, choose a uniformly random d-regular graph, conditioned on there being exactly $\pi(i)\pi(j)M(i,j)\cdot dn$ edges between each pair of distinct groups $i \neq j$, and $\pi(i)^2 M(i,j)\cdot dn/2$ edges internal to each group i. For simplicity, we will assume that the parameters are such that these group sizes and edge counts are integers. As with the SBM, we will call the model *symmetric* if the entries of M are constant on the diagonal and off-diagonal respectively. As a warm-up for the main technical arguments of the paper, we will study in Section 4 a simplified version of the Local Statistics SDP that can solve the detection problem on the symmetric DRBM.

Remark 2.3. The DRBM as we have defined it differs from the Regular Stochastic Block Model of [BDG+16], in which each vertex has a prescribed number of neighbors in every community. Although superficially similar, the behavior of this 'equitable' model (as it is known in the physics literature [NM14]) is quite different from ours. For instance, [BDG+16] show that whenever detection is possible in the two community case, one can *exactly* recover the planted labels. This is not true in our setting.

It is widely believed that the threshold behavior of the general DRBM is analogous to that of the SBM, including an information-theoretic threshold, and Kesten-Stigum threshold at $d_{\mathrm{KS}} \triangleq \lambda_2^{-2} + 1$. However, most formal treatment in the literature has been limited to random d-regular graphs conditional on having a planted k-coloring, a case not fully captured by our model. Characterization of the information-theoretic threshold, even in simple cases, remains largely folklore.

Our main result on the DRBM is analogous to Theorem 2.2 on the SBM.

Theorem 2.4. *Let $\mathcal{N}_n$ denote the uniform distribution on d-regular graphs with n-vertices, and $\mathcal{P}_n$ the DRBM with parameters (d, k, M, π). If $d > d_{\mathrm{KS}}$, then there exists a constant $m \in \mathbb{N}$, $\delta > 0$, and and $\rho > 0$ (all dependent on d) so that $\mathrm{LoSt}(2, m)$ with error tolerance δ can ρ-robustly solve the detection problem. Conversely, if $d < d_{\mathrm{KS}}$, then every constant level, no matter the error tolerance, fails to do so.*

Along the way we will inadvertently prove that standard spectral detection using the adjacency matrix succeeds above d_{KS}, but cannot have the same robustness guarantee. It is a now-classic result of Friedman that, with probability $1 - o_n(1)$, the spectrum of a uniformly random d-regular graph is within $o_n(1)$ of $(-2\sqrt{d-1}, 2\sqrt{d-1}) \cup \{d\}$. Conversely, we show:

Corollary 2.5. *Let $\boldsymbol{G}$ be drawn from the DRBM with parameters (d, k, M, π) satisfying $d > d_{\mathrm{KS}} + \epsilon$. There exists some $\eta = \eta(\epsilon)$ such that, for each eigenvalue λ of M satisfying $|\lambda| > 1/\sqrt{d-1} + \epsilon$, the adjacency matrix $A_{\boldsymbol{G}}$ is guaranteed one eigenvalue μ satisfying $|\mu| > 2\sqrt{d-1} + \eta$.*

Future Work Regrettably, we do not solve the problem of recovery above Kesten-Stigum in either model. However, we will in Appendix A reduce recovery in the DRBM to the following conjecture regarding the spectrum of A_G for $\boldsymbol{G}$ drawn from the planted model.

Conjecture 2.6. *Let $\mathcal{P}_{(d,k,M,\pi)}$ be any DRBM with $|\lambda_1|, ..., |\lambda_\ell| > (d-1)^{-1/2}$. Then, for any η, with high probability, A_G has only ℓ eigenvalues with modulus larger than $2\sqrt{d-1}+\eta$.*

Related Work. Semidefinite programming approaches have been most studied in the dense, irregular case, where exact recovery is possible (for instance [ABH16, AS15]), and it has been shown that an SDP relaxation can achieve the information-theoretically optimal threshold [HWX16]. However, in the sparse regime we consider, the power of SDP relaxations for weak recovery remains unclear. Guedon and Vershynin [GV16] show upper bounds on the estimation error of a standard SDP relaxation in the sparse, two-community case of the SBM, but only when the degree is roughly 10^4 times the information theoretic threshold. More recently, in a tour-de-force, Montanari and Sen [MS15] showed that for two communities, the SDP of Guedon and Vershynin achieves the information theoretically optimal threshold for large but constant degree, in the sense that the performance approaches the threshold if we send the number of vertices, and then the degree, to infinity. Semi-random graph models have been intensively studied in [BS95, FK00, FK01, CO04, KV06, CO07, MMV12, CJSX14, GV16] and we refer the reader to [MMV16] for a more detailed survey. In the logarithmic-degree regime, robust algorithms for community detection are developed in [CL+15, KK10, AS12]. Far less is known in the case of regular graphs.

3 Technical Overview

Notation. We will use bold face font for random objects sampled from these distributions. Because we care only about the case when the number of vertices is very large, we will use *with high probability (w.h.p)* to describe any sequence of events with probability $1 - o_n(1)$ in $\mathcal{N}$ or $\mathcal{P}$ as $n \to \infty$. We will write $[n] = \{1, ..., n\}$, and in general use the letters u, v, w to refer to elements of $[n]$ and i, j for elements of $[k]$. The identity matrix will be denoted by $\mathbb{1}$, and we will write X^T for the transpose of a matrix X, $\langle X, Y \rangle = \operatorname{tr} X^T Y$ for the standard matrix inner product, and $\|X\|_F$ for the associated Frobenius norm. Positive semidefiniteness will be indicated with the symbol $\succeq$. The standard basis vectors will be denoted $e_1, e_2, ...$, the all-ones vector written as e, and the all-ones matrix as $\mathbb{J} = ee^T$. Finally, let $\operatorname{diag} : \mathbb{R}^{n \times n} \to \mathbb{R}$ be the function extracting the diagonal of a matrix, and $\operatorname{Diag} : \mathbb{R}^n \to \mathbb{R}^{n \times n}$ be the one which populates the nonzero elements of a diagonal matrix with the vector it is given as input.

3.1 Optimization vs Inference

While it was suspected that a semidefinite programming relaxation could be used towards community detection in sparse stochastic block models, many earlier attempts at it [GV16, MS15] failed to detect communities right up to the KS threshold at a fixed degree. These works studied

the Goemans-Williamson SDP relaxation for MaxCut applied to the problem of detecting two communities ($k = 2$). The idea being that if we consider a two community SBM with $p_{out} > p_{in}$, then the partition induced by the communities should have an unusually large number ($\frac{dn}{2} \cdot \frac{p_{out}}{p_{out}+p_{in}}$) of crossing edges. Hence an SDP relaxation of MaxCut could be harnessed towards detecting and possibly recovering the communities. Indeed, in this special case, the maximum bisection in the graph is a Maximum Likelihood Estimate (MLE) for the communities x given the graph G, i.e., $x = \text{argmax}_x\, p(x|G)$.

This approach of casting inference as optimization has its limitations. In particular, as one approaches the KS threshold, the number of crossing edges between the two communities, namely $\frac{dn}{2} \cdot \frac{p_{out}}{p_{out}+p_{in}}$, is lower than the value of MaxCut in a random Erdos-Renyi graph! In other words, if we run an exponential-time algorithm that finds the maximum cut via a brute-force enumeration, then it will find a better MaxCut in a random Erdos-Renyi graph than the true communities in the planted model. It is therefore unclear whether an SDP relaxation of MaxCut can solve the problem.

In hindsight, the number of crossing edges is but one statistic associated with the partition and there is no canonical reason why optimizing this statistic would be the optimal way to distinguish the two models. For example, in the same setting one could minimize the number of paths of length two that go between the two sides of the partition, or maximize the number of paths of length three that cross the partition and so on. At a more basic level, if we are interested in estimating the moments of the distribution $x|G$, it is not clear that we should cast this problem as optimization.

The local statistics SDP hierarchy that we propose is a "feasibility SDP" that looks for candidate low-degree moments for the distribution $x|G$. The constraints of the SDP ensure that the value of local statistics such as number of crossing edges is roughly the same as we would expect in a graph drawn from the communities.

3.2 Detection, Refutation, and Sum-of-Squares

We will begin the discussion of the Local Statistics algorithm by briefly recalling Sum-of-Squares programming. Say we have a constraint satisfaction problem presented as a system of polynomial equations in variables $x = (x_1, ..., x_n)$ that we are to simultaneously satisfy. In other words, we are given a set

$$\mathcal{S} = \{x \in \mathbb{R}^n : f_1(x), ..., f_m(x) = 0\}$$

and we need to decide if it is non-empty. Whenever the problem is satisfiable, any probability distribution supported on $\mathcal{S}$ gives rise to an operator $\mathbb{E} : \mathbb{R}[x] \to \mathbb{R}$ mapping a polynomial x to its expectation. Trivially, $\mathbb{E}$ has the properties:

$$\textit{Normalized} \qquad \mathbb{E}\, 1 = 1 \tag{2}$$

$$\textit{Satisfies of } \mathcal{S} \qquad \mathbb{E}\, f_i(x) \cdot p(x) = 0 \qquad \forall i \in [m], \forall p \in \mathbb{R}[x] \tag{3}$$

$$\textit{Positive} \qquad \mathbb{E}\, p(x)^2 \geqslant 0 \qquad \forall p \in \mathbb{R}[x] \tag{4}$$

We will extend these definitions to any operator mapping some subset of $\mathbb{R}[x] \to \mathbb{R}$.

Refuting the constraint satisfaction problem, e.g. proving that $\mathcal{S} = \emptyset$, is equivalent to showing that no operator obeying (2)-(4) can exist. The key insight of SoS is that often one can do this by focusing only on polynomials of some bounded degree. Writing $\mathbb{R}[x]_{\leqslant D}$ for the polynomials of degree at most D, we call an operator $\widetilde{\mathbb{E}} : \mathbb{R}[x]_{\leqslant D} \to \mathbb{R}$ a *degree-D pseudoexpectation* if it is normalized, positive, and satisfies $\mathcal{S}$ for every polynomial in its domain. It is well-known that one can search for a degree D pseudoexpectation with a semidefinite program of size $O(n^D)$, and if this smaller, relaxed problem is infeasible, we've shown that $\mathcal{S}$ is empty. This is the *degree-D Sum-of-Squares relaxation* of our CSP.

3.3 The Local Statistics Hierarchy

Let $\mathcal{P}_n$ denote a sequence of distributions on graphs with a planted community structure, and $\mathcal{N}_n$ a corresponding 'null' distribution with no such prescribed structure. For us, $\mathcal{P}_n$ will always denote the DRBM or SBM, and $\mathcal{N}_n$ the Erdős-Rényi model with average degree d, or the uniform distribution on d-regular graphs. Our goal is to devise an algorithm that can discern, with high probability, which of these two distributions a graph was drawn from. In this setup, the details of the null and prior distribution are known to us; the main idea of this work is that it is only natural to grant an SDP hypothesis testing algorithm access to this information as well. Our strategy will be do devise an SDP that is satisfiable with high probability when a graph is drawn from $\mathcal{P}_n$, and unsatisfiable with high probability when it is drawn from $\mathcal{N}_n$.

The Local Statistics SDP will be assembled from components of the Sum-of-Squares algorithm, and as such we will need to carefully articulate the null and planted distribution, and their statistical properties, in the language of polynomials. Let us write $x = \{x_{u,i}\}$ for a collection of variables indexed by vertices $u \in [n]$ and group labels $i \in [k]$, and $G = \{G_{u,v}\}$ for a collection indexed by two-element subsets $\{u, v\} \subset [n]$. We will regard a random graph from the null model as a collection of random variables $\boldsymbol{G} = \{\boldsymbol{G}_{u,v}\}$ indexed in the same way, where $\boldsymbol{G}_{u,v}$ is the Boolean indicator for the edge (u, v). Similarly, the planted model is a joint distribution over pairs $(\boldsymbol{x}, \boldsymbol{G})$, where $\boldsymbol{G}$ is a graph, and $\boldsymbol{x}_{i,u}$ is the indicator that vertex u has label i. Thus for each polynomial $p \in \mathbb{R}[G, x]$, we can compute the *statistic* $\mathbb{E}\, p(\boldsymbol{G}, \boldsymbol{x})$. We will see below that one can easily construct such a polynomial that counts, for instance, the number of triangles in a graph, or the number of edges between vertices in the same group.

The random variables $\boldsymbol{G}$ and $\boldsymbol{x}$ take values in the zero locus of the following set of polynomials in $\mathbb{R}[G, x]$:

$$G_{u,v}^2 = G_{u,v} \qquad \forall u, v \in [n] \tag{5}$$

$$x_{u,i}^2 = x_{u,i} \qquad \forall u \in [n],\, i \in [k] \tag{6}$$

$$x_{u,1} + \cdots + x_{u,k} = 1 \qquad \forall u \in [n]. \tag{7}$$

For brevity, we will throughout the paper denote by $\mathcal{B}_k$ the set of polynomials constraints in the x variables appearing in (6) and (7). Moreover, in our case both the null and planted models have a natural symmetry: they are invariant under permutations of the vertices. To a first approximation,

the (D_G, D_x) level of the Local Statistics SDP, on input $G_0 \in \{0,1\}^{\binom{n}{2}}$, will endeavor to find a degree-D_x pseudoexpectation $\widetilde{\mathbb{E}} : \mathbb{R}[x]_{\leqslant D_x} \to \mathbb{R}$ that (i) satisfies $\mathcal{B}_k$, and (ii) obeys *moment constraints* of the form

$$\widetilde{\mathbb{E}}\, p(G_0, x) \approx \mathop{\mathbb{E}}_{(\boldsymbol{G},\boldsymbol{x})\sim\mathcal{P}_n} p(\boldsymbol{G}, \boldsymbol{x})$$

for symmetric polynomials $p \in \mathbb{R}[G, x]$ with degree D_G in the G variables. We ask that these moment constraints are only approximately satisfied to ensure that, when $(\boldsymbol{G}, \boldsymbol{x})$ is drawn from the planted model, the pseudoexpectation $\widetilde{\mathbb{E}}\, p(G, x) \triangleq p(\boldsymbol{G}, \boldsymbol{x})$ is with high probability a feasible solution. This formulation is inspired by the technique of pseudocalibration from the SOS lower bounds literature [BHK+19, HS17, HKP+17].

Each polynomial $p(G, x)$, when evaluated at a point in the zero locus described above, counts occurrences of a certain combinatorial structure in G, in which some of the vertices are restricted to have particular labels. For instance,

$$\sum_u \prod_{u \neq v} (1 - G_{u,v}) \qquad \text{and} \qquad \sum_{u \neq v} G_{u,v} x_{u,i} x_{v,j}$$

count the number of isolated vertices, and the number of edges between vertices in groups i and j, respectively. Note that since $\widetilde{\mathbb{E}}$ is required to satisfy the Boolean constraints on the G variables and the $\mathcal{B}_k$ constraints on the x variables, we are free to consider only polynomials that have been reduced modulo these constraints: for simplicity we will assume that they are multilinear in G and x, and furthermore that monomial containts $x_{u,i} x_{u,j}$ for $i \neq j$.

Remark 3.1. Although we have stated it in the specific context of the DRBM, the local statistics framework extends readily to any planted problem involving a joint distribution μ on pairs $(\boldsymbol{G}, \boldsymbol{x})$ of a hidden structure and observed signal, if we take appropriate account of the natural symmetries in μ. For a broad range of such problems, including spiked random matrix models [AKJ18, PWBM16], compressed sensing [ZK16, Ran11, KGR11] and generalized linear models [BKM+19] (to name only a few) there are conjectured computational thresholds where the underlying problem goes from being efficiently solvable to computationally intractable, and the algorithms which are proven or conjectured attain this threshold are often not robust. We hope that the local statistics hierarchy can be harnessed to design robust algorithms up to these computational thresholds, as well as to provide evidence for computational intractibility in the conjectured hard regime. The relation (if any) between the local statistics SDP hierarchy and iterative methods such as belief propagation or AMP is also worth investigating.

3.4 Analyzing the Local Statistics SDP

By design, the Local Statistics SDP is always feasible when given as input a graph drawn from the planted model. To show that LoSt$(2, m)$ can distinguish between the null and planted models, then, it suffices to show that it is with high probability infeasible when passed a graph from the null model.

For a matrix $C \in \mathbb{R}^{n\times n}$, let $C^{(t)}$ denote the t^{th} "non-backtracking power" of the matrix:

$$C^{(t)}_{i,j} \triangleq \sum_{\text{n.b. paths } p:i\to j} \prod_{(u,v)\in p} C_{u,v}$$

where the sum is over non-backtracking paths of length t from i to j. The local statistic that serves as a dual certificate to show infeasibility of $\mathrm{LoSt}(2,m)$ in the null model is given by,

$$p^{(m)}(G,x) = \langle \phi(x), (A-(d/n)\mathbb{J})^{(m)}\phi(x)\rangle$$

for an appropriately chosen $\phi : [k] \to \mathbb{R}$. In particular, we will see in the sections below that, if $\mathrm{LoSt}(2,m)$ SDP is feasible on input G, there is some matrix $X \succeq 0$ with unit trace and bounded entries on its diagonal for which

$$|\langle X, (A-(d/n)\mathbb{J})^{(m)}\rangle| \geqslant \omega(d^{m/2})n.$$

The use of this centered non-backtracking walk matrix $\overline{A}^{(m)}_G = (A-(d/n)\mathbb{J})^{(m)}$ was inspired by the work of Fan and Montanari [FM17], who use the centered non-backtracking matrix for $m = 2$. Thus, to show infeasibility it would be sufficient to bound the spectral norm of the matrix $\overline{A}^{(m)}_G = (A-(d/n)\mathbb{J})^{(m)}$ by $d^{m/2}$ for sufficiently large constant m.

In the d-regular case, the non-backtracking powers of the adjacency matrix A can be expressed as univariate polynomials in the matrix A. Thus spectral norm bounds on the adjacency matrix of a random d-regular graph [Fri03] can be translated into spectral norm bounds that we require. This is roughly the approach taken in the d-regular case.

Unfortunately, things are not so simple in the irregular case: the analogous bound fails for constant m due to the presence of high-degree vertices in $\boldsymbol{G}$. The main challenge in studying $\overline{A}^{(m)}_{\boldsymbol{G}}$, when $\boldsymbol{G}$ is a sparse Erdős-Rényi random graph, is the presence of of certain localized combinatorial structures which inflate the number of non-backtracking walks: high-degree vertices and small subgraphs with many cycles. Instead, we show the spectral norm bound after deleting these structures from the random graph $\boldsymbol{G}$ and that the deletion does not affect the global statistic significantly.

Let us make this precise. In any graph G, write $\mathsf{B}_t(v,G)$ for the set of vertices with distance at most t from v; call v (t,ε)-*heavy* if $|\mathsf{B}_t(v,G)| \geqslant (1+\epsilon)^t d^t$. We will call a vertex v (t,r,ε)-*vexing* if either it participates in a cycle of length less than r or it is (t,ε)-heavy.

Fix $r = \Theta(\frac{\log n}{(\log\log n)^2})$. Let $\boldsymbol{G}$ be an Erdős-Renyi $G(n,d/n)$ graph, let $\boldsymbol{S}$ its the set of (t,r,ε)-vexing vertices, and let $\boldsymbol{G}_{t,r,\varepsilon}$ be the (t,r,ε)-truncation obtained by deleting all the vertices in $\boldsymbol{S}$ from $\boldsymbol{G}$. Let $\boldsymbol{A}$ be the adjacency matrix of $\boldsymbol{G}_{t,\varepsilon,r}$. Define

$$\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus S}1^\top_{[n]\setminus S}\right)^{(\ell)}[u,v] = \sum_{\substack{W \text{ length-}\ell \text{ nonbacktracking walk} \\ \text{from } u \text{ to } v \text{ in complete graph } K_{[n]\setminus S}}} \prod_{ij\in W}\left(\boldsymbol{A} - \frac{d}{n}11^\top\right)[i,j]$$

We prove the following spectral norm bound via the trace method:

Theorem 3.2. *With probability* $1 - n^{-100}$, $\left\| \left(\boldsymbol{A} - \frac{d}{n} 1_{[n]\setminus S} 1_{[n]\setminus S}^{\mathsf{T}} \right)^{(\ell)} \right\| \leqslant \left((1+\varepsilon)^4 \sqrt{d} \right)^{\ell}$.

3.5 Proving the spectral norm bound

The proof of the above spectral norm bound is the most technical argument of the paper. As expected, the proof of the spectral norm bound via trace method reduces to the problem of computing the expected number of copies of combinatorial structures that we call linkages in the underlying graph $\boldsymbol{G}$.

Definition 3.3 (Linkages)**.** A closed walk W of length $k\ell$ is a $(k \times \ell)$-*linkage* if it can be split into k segments each of length-ℓ such that the walk W is nonbacktracking on each segment. Each ℓ-step non-backtracking segment is a "link".

We will bound the number of $(k \times \ell)$-linkages using an encoding argument.

It is instructive to consider the encoding argument in the case when the graph $\boldsymbol{G}$ is a $d+1$-regular tree and the walk W starts at the root. Let us encode a $(k \times \ell)$-linkage starting at the root, one link at a time. Each link which is a ℓ-step n.b.walk in a tree consists of t-steps towards the root followed by $\ell - t$ steps away from the root for some $t \in \{0, \ldots, r\}$. We refer to the steps towards the root as "up-steps" and steps away from the root as "down-steps". Encode each link by specifying:

- The number of up-steps t using $\log \ell$ bits.
- For each down-step, the index of the child as an integer from $\{1, \ldots, d\}$.

Since the walk begins and ends at the root, the number of up-steps is equal to the number of down-steps. Therefore the number of down-steps is precisely $k\ell/2$. Hence the above encoding uses precisely $k\ell/2 \cdot (\log d) + k \log \ell$ bits. As $\ell \to \infty$, this is approximately $\frac{1}{2} \log d$ bits on average per step. Therefore the number of $k \times \ell$-linkages starting at the root in a d-regular tree is at most $((1+\epsilon)\sqrt{d})^{k\ell}$ for sufficiently large constant ℓ.

In an Erdos-Renyi random graph $\boldsymbol{G}$, there will be cycles of length $< k\ell$ thus breaking the above encoding argument. In other words, if we consider the graph $G(W)$ formed by the edges in the $(k \times \ell)$-linkage W, then $G(W)$ can include cycles once we set $k = \Omega(\log n)$. However, since we deleted all (t, r, ϵ)-vexing vertices $G(W)$ has no cycles of length $< \Theta(\frac{\log n}{(\log \log n)^2}$.

The starting point of our encoding argument is a decomposition of $G(W)$ into a spanning forest F and a few additional edges $E(W) \setminus F$, such that the non-forest edges $E(W) \setminus F$ are in total traversed $o(k\ell)$ times during the walk. We prove the existence of such a decomposition using a linear programming based argument.

Roughly speaking, this decomposition lets us encode the walk W by breaking it up into closed walks in trees, with the decomposition only introducing a negligible overhead in the encoding. Therefore, one recovers a bound analogous to the bound in a d-regular tree, which is approximately $\frac{1}{2}\log d$ bits per step in the walk.

The remainder of the paper will be laid out as follows. Before embarking on our investigation of the Local Statistics SDP in the DRBM and SBM in full generality, we will in Section 4 study a simplified SDP that can robustly solve the detection problem for the symmetric Degree Regular Block Model. Having done so, we will move on in Section 5 to the case of the general DRBM, proving Theorem 2.4 by way of a reduction to some key results from this simpler, symmetric case. Finally, in Section 6 we prove Theorem 2.2 regarding the SBM.

4 A Simplified SDP for the Symmetric DRBM

Many key ideas from the remainder of the paper are captured by the symmetric case of the Degree Regular Block Model, in which each group has size exactly n/k, and the edge probability matrix is

$$M = k\lambda\mathbb{1} + (1-\lambda)\mathbb{J}.$$

Since the communities have equal sizes, we have $T = k^{-1}M$, and the Kesten-Stigum threshold is $d_{\mathrm{KS}} \triangleq \lambda^{-2} + 1$. Throughout this section, let $\mathcal{P}$ denote this symmetric case of the DRBM, and $\mathcal{N}$ the uniform distribution on d-regular graphs. The purpose of this section is to show, in this symmetric case, that a simplified version of the Local Statistics SDP can robustly solve the detection problem.

To introduce this simpler SDP, let $G = (V, E)$ be any graph on n vertices, and write $A_G^{(s)}$ for the $n \times n$ matrix that counts non-backtracking random walks of length s; we will develop some further theory regarding these matrices in Section 4.1 below. Now, let $(\boldsymbol{G}, \boldsymbol{y}) \sim \mathcal{P}$ be drawn from the symmetric DRBM, and—thinking of $\boldsymbol{y}$ as an $n \times k$ matrix—write

$$\boldsymbol{Y} \triangleq \frac{k}{k-1}\left(\boldsymbol{y}\boldsymbol{y}^* - \frac{1}{k}\mathbb{J}\right) \succeq 0. \tag{8}$$

This is a rank-$(k-1)$ positive semidefinite matrix that is n/k times the projector onto the subspace spanned by the indicator vectors for the k groups and orthogonal to the all-ones vector. The inner product $\langle \boldsymbol{Y}, A_{\boldsymbol{G}}^{(s)} \rangle$ counts non-backtracking walks weighted according to the labels of their initial and terminal vertices.

Lemma 4.1. *Let $(\boldsymbol{G}, \boldsymbol{Y}) \sim \mathcal{P}$. Then for every $s \geqslant 1$,*

$$\mathbb{E}\langle \boldsymbol{Y}, A_{\boldsymbol{G}}^{(s)} \rangle = \lambda^s d(d-1)^{s-1}n + o(n)$$

and with high probability these quantities enjoy concentration of $o(n)$.

Definition 4.2. Fix a small number $\delta > 0$. The *level m symmetric path statistics SDP* with error

tolerance $\delta > 0$, on input G_0, is the feasibility problem

$$\begin{aligned} \text{Find } Y \succeq 0 \text{ s.t.} \qquad Y_{u,u} &= 1 && \forall u \in [n] \\ \langle Y, \mathbb{J} \rangle &= 0 \\ \left| \langle Y, A_G^{(s)} \rangle - \lambda^s d(d-1)^{s-1} n \right| &\leqslant \delta n && \forall s \in [m] \end{aligned} \tag{9}$$

We will refer to this as the $SPS(m, \lambda)$ SDP. To handle adversarial edge corruption, it is necessary to include the following contingency if the input G_0 is not d-regular: before running the above SDP, delete all edges incident to vertices with degree greater than d, and then greedily add edges between vertices with degree less than d to obtain a d-regular graph.

Theorem 4.3. *If $(d-1)\lambda^2 > 1$, then there exists constant $m \in \mathbb{N}$, $\delta > 0$, and $\rho > 0$ so that $SPS(m, \lambda)$ solves the detection problem ρ-robustly. Conversely if $(d-1)\lambda^2$ then no such m, δ, ρ exist.*

4.1 Non-backtracking Walks and Orthogonal Polynomials

The central tool in our proofs will be *non-backtracking walks*—these are walks which on every step are forbidden from visiting the vertex they were at two steps previously. We will collect here some known results on these walks specific to the case of d-regular graphs. Write $A_G^{(s)}$ for the $n \times n$ matrix whose (v, w) entry counts the number of length-s non-backtracking walks between verties v and w in a graph G. It is standard that the $A_G^{(s)}$ satisfy a two-term linear recurrence,

$$\begin{aligned} A_G^{(0)} &= \mathbb{1} \\ A_G^{(1)} &= A_G \\ A_G^{(2)} &= A_G^2 - d\mathbb{1} \\ A_G^{(s)} &= A A_G^{(s-1)} - (d-1) A_G^{(s-2)} \qquad s > 2, \end{aligned}$$

since to enumerate non-backtracking walks of length s, we can first extend each such walk of length $s-1$ in every possible way, and then remove those extensions that backtrack.

On d-regular graphs, the above recurrence immediately shows that $A_G^{(s)} = q_s(A_G)$ for a family of monic, scalar *non-backtracking polynomials* $\{q_s\}_{s \geqslant 0}$, where $\deg q_s = s$. To avoid a collision of symbols, we will use z as the variable in all univariate polynomials appearing in the paper. It is well known that these polynomials are an orthogonal polynomial sequence with respect to the *Kesten-McKay measure*

$$\mathrm{d}\mu_{\mathrm{KM}}(z) = \frac{1}{2\pi} \frac{d}{\sqrt{d-1}} \frac{\sqrt{4(d-1) - z^2}}{d^2 - z^2} \, \mathrm{d}z \, \mathbf{1}\left[|z| < 2\sqrt{d-1} \right],$$

with its associated inner product

$$\langle f, g \rangle_{\mathrm{KM}} \triangleq \int f(z) g(z) d\mu_{\mathrm{KM}}(z)$$

on the vector space of square integrable functions on $(-2\sqrt{d-1}, 2\sqrt{d-1})$. One quickly verifies that

$$\|q_s\|_{\mathrm{KM}}^2 \triangleq \int q_s(z)^2 \, \mathrm{d}\mu_{\mathrm{KM}} = q_s(d) = \begin{cases} 1 & s = 0 \\ d(d-1)^{s-1} & s \geqslant 1 \end{cases} = \frac{1}{n} \left(\#\text{ length-}s\text{ n.b. walks on } G\right)$$

in the normalization we have chosen [ABLS07]. Thus any function f in this vector space can be expanded as

$$f = \sum_{s \geqslant 0} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} q_s.$$

We will also need the following lemma of Alon et al. [ABLS07, Lemma 2.3] bounding the size of the polynomials q_s:

Lemma 4.4. *For any $\varepsilon > 0$, there exists an $\eta > 0$ such that for $z \in [-2\sqrt{d-1} - \eta, 2\sqrt{d-1} + \eta]$,*

$$|q_s(z)| \leqslant 2(s+1)\|q_s\|_{\mathrm{KM}} + \varepsilon.$$

The behavior of the non-backtracking polynomials with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathrm{KM}}$ idealizes that of the $A_G^{(s)} = q_s(A_G)$ under the trace inner product. In particular, if $s + t < r(G)$

$$\langle A_G^{(s)}, A_G^{(t)} \rangle = n \langle q_s, q_t \rangle_{\mathrm{KM}} = \begin{cases} n(\#\text{ length-}s\text{ n.b. walks on } G) & s = t \\ 0 & s \neq t \end{cases}.$$

This is because the diagonal entries of $A_G^{(s)} A_G^{(t)}$ count pairs of non-backtracking walks with length s and t respectively: if $s \neq t$ any such pair induces a cycle of length at most $s + t$, leaving only the degenerate case when $s = t$ and the two walks are identical. Above the girth, if we can control the number of cycles, we can quantify how far the $A_G^{(s)}$ are from orthogonal in the trace inner product.

Luckily for us, sparse random graphs have very few cycles. To make this precise, call a vertex *bad* if it is at most L steps from a cycle of length at most C. These are exactly the vertices for which the diagonal entries of $A_G^{(s)} A_G^{(t)}$ are nonzero, when $s + t < C + L$.

Lemma 4.5. *For any constant C and L, with high probability any graph $\boldsymbol{G} \sim \mathcal{P}$ has at most $O(\log n)$ bad vertices.*

We will defer the proof of this lemma to the appendix, but one can immediately observe the consequence that, with high probability,

$$\langle A_{\boldsymbol{G}}^{(s)}, A_{\boldsymbol{G}}^{(t)} \rangle = O(\log n)$$

for any $s, t = O(1)$.

4.2 Distinguishing

Let us now prove the first assertion in Theorem 4.3, namely that if $(d-1)\lambda^2 > 1$, then the $SPS(m,\lambda)$ SDP, for some $\delta > 0$ sufficiently large m, can distinguish the null and planted models. From Lemma 4.1, if $(\boldsymbol{G}, \boldsymbol{Y}) \sim \mathcal{P}$, then the matrix $\boldsymbol{Y}$ from equation (8) is with high probability a feasible solution to SDP (9). Thus, it remains only to show that with high probability over $\boldsymbol{G} \sim \mathcal{N}$, some round of the $SPS(m,\lambda)$ SDP is infeasible. Our strategy will be to first reduce this infeasibility to a univariate polynomial design problem, and then solve this with the machinery developed in the prior subsection.

Proposition 4.6. *If there exists a degree-m polynomial $f \in \mathbb{R}[z]$ which is (i) strictly nonnegative on the interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ and (ii) satisfies*

$$\langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle_{\mathrm{KM}} < 0,$$

then with high probability the $SPS(m,\lambda)$ SDP is infeasible for $\boldsymbol{G} \sim \mathcal{N}$, at any error tolerance

$$\delta < \frac{|\langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle_{km}|}{\sqrt{m}\|f\|_{\mathrm{KM}}}.$$

Proof. First note that, for any such polynomial f, our discussion in the previous section implies

$$f = \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} q_m. \tag{10}$$

Moreover, since f is strictly positive on $[-2\sqrt{d-1}, 2\sqrt{d-1}]$, it is nonnegative on some fattening I of this interval.

Now, let $\boldsymbol{G}$ be a uniformly random d-regular graph. By Friedman's Theorem [Fri08], the spectrum of $A_{\boldsymbol{G}}$ consists of a 'trivial' eigenvalue at d, plus $n-1$ eigenvalues whose magnitudes—with high probability—are at most $2\sqrt{d+1} + o_n(1)$. In particular, these remaining eigenvalues with high probability lie inside the fattening of $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ on which f is nonnegative. We can project away this trivial eigenvalue by passing to the centered adacency matrix $\overline{A}_{\boldsymbol{G}} = (\mathbb{1} - \mathbb{J}/n)A_{\boldsymbol{G}}(\mathbb{1} - \mathbb{J}/n) = A_{\boldsymbol{G}} - d\mathbb{J}/n$, and observe that $0 \preceq f(\overline{A}_{\boldsymbol{G}})$.

Assume, seeking contradiction, that $\boldsymbol{Y}$ is a feasible solution to the $SPS(m)$ SDP. We can compute that

$$\begin{aligned}
0 &\leqslant \langle \boldsymbol{Y}, f(\overline{A}_{\boldsymbol{G}}) \rangle \\
&= \langle \boldsymbol{Y}, \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} q_m(\overline{A}_{\boldsymbol{G}}) \rangle \\
&= \langle \boldsymbol{Y}, \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} \left(q_m(A_{\boldsymbol{G}}) - q_s(d)\mathbb{J}/n\right) \rangle
\end{aligned}$$

$$
\begin{aligned}
&= \langle \boldsymbol{Y}, \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} A_{\boldsymbol{G}}^{(s)} \rangle \\
&\leqslant \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} \cdot \lambda^s \|q_s\|_{\mathrm{KM}}^2 n + \delta \sum_{s=0}^{m} \frac{|\langle f, q_s \rangle_{\mathrm{KM}}|}{\|q_s\|_{\mathrm{KM}}^2} \\
&\leqslant \langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle + \delta \sqrt{m} \|f\|_{\mathrm{KM}} < 0
\end{aligned}
$$

□

The following proposition implies a proof of the first part of Theorem 4.3.

Proposition 4.7. *If $\lambda^2(d-1) > 1$, there exists a polynomial satisfying the hypotheses of Proposition 4.6.*

Proof. Call m' the largest even number less than or equal to m, let $\varepsilon > 0$ be a very small number, and take

$$f(z) = -q_{m'}(z) + 2m' \|q_{m'}\|_{\mathrm{KM}} + \varepsilon,$$

which by Lemma 4.4 has the desired positivity property. This choice of f satisfies

$$\langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle = -\|q_{m'}\|_{\mathrm{KM}}^2 |\lambda|^{m'} + 2m' \|q_{m'}\|_{\mathrm{KM}} + \varepsilon,$$

which is negative when

$$\lambda^2 > \left(\frac{2m'}{\|q_{m'}\|_{\mathrm{KM}}} + \frac{\varepsilon}{\|q_{m'}\|_{\mathrm{KM}}^2} \right)^{\frac{2}{m'}} = \left(\frac{2m'}{\sqrt{d(d-1)^{m'-1}}} + \frac{\varepsilon}{d(d-1)^{m-1}} \right)^{\frac{2}{m'}};$$

this tends to $\frac{1}{d-1}$ as $m \to \infty$. □

4.3 Lower Bound

We now turn to the complementary bound: when $(d-1)\lambda^2 < 1$, no constant level of the symmetric path statistics SDP can distinguish the null and planted distributions. It suffices to show that, for d in this regime, $SPS(m, \lambda)$ is feasible for every constant m. Once again, we will reduce to and solve a univariate polynomial design problem.

Proposition 4.8. *If there exists a polynomial $g \in \mathbb{R}[z]$ that is (i) strictly positive on $(-2\sqrt{d-1}, 2\sqrt{d-1})$, and (ii) satisfies*

$$\langle g, q_s \rangle_{\mathrm{KM}} = \lambda^s \|q_s\|_{\mathrm{KM}}^2 \qquad \text{For all } s = 0, \ldots, m,$$

then the $SPS(m, \lambda)$ SDP at any constant error tolerance $\delta > 0$ is with high probability feasible for a uniformly random d-regular graph.

Proof. Letting $\boldsymbol{G}$ be the random regular graph in question, and fixing arbitrary $\delta > 0$, we need to

produce $Y \succeq 0$ with ones on the diagonal, zero inner product with the matrix $\mathbb{J}$, and satisfying

$$\left| \langle Y, A_G^{(s)} \rangle - \lambda^s \|q_s\|_{\mathrm{KM}}^2 n \right| \leqslant \delta n.$$

Our strategy will be to modify the matrix $g(\overline{A}_G) = g(A_G) - g(d)\mathbb{J}/n$.

First, note that by expanding g in the non-backtracking basis and invoking Lemma 4.5, for any $0 \leqslant s \leqslant m$ we have

$$\langle g(\overline{A}_G), A_G^{(s)} \rangle = \langle g(A_G), A_G^{(s)} \rangle + g(d)\|q_s\|_{\mathrm{KM}}^2 = \lambda^s \|q_s\|_{\mathrm{KM}}^2 \cdot n + O(\log n),$$

since $g(d)\|q_s\|_{\mathrm{KM}}^2$ is a constant. Moreover, as g is strictly positive on $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ it is by continuity nonnegative on any constant size fattening of this interval, and by Friedman's theorem the spectrum of A_G other than the eigenvalue at d is contained w.h.p. in such a set. Thus $g(\overline{A}_G)$ is positive semidefinite, and as a polynomial in the centered adjacency matrix, is orthogonal to the all-ones matrix.

However, the diagonal of $g(\overline{A}_G)$ may not be equal to one, for two different reasons. The diagonal entries of $g(A_G) = g(\overline{A}_G) + g(d)\mathbb{J}/n$ different from one are exactly those corresponding to vertices within $\deg g$ steps of a constant length cycle; from Lemma 4.5 we know that there are at most $O(\log n)$ of these *bad* vertices (keeping the terminology from the aforementioned Lemma). However, when we subtract $g(d)\mathbb{J}/n$, even the $\Omega(n - \log n)$ diagonal entries equal to one—those corresponding to *good* vertices—are shifted. Let us therefore define

$$\tilde{Y} = \frac{1}{1 - g(d)/n} g(\overline{A}_G),$$

which restores the diagonal entries of the good vertices.

Now, $\tilde{Y}$ is PSD, and is accordingly the Gram matrix of some vectors $\alpha_1, \ldots, \alpha_n \in \mathbb{R}^n$. The scale factor we have applied ensures that for every good vertex u, $\|\alpha_u\| = 1$, and orthogonality to the all-ones matrix—which is preserved by this constant scaling—is equivalent to $\sum_u \alpha_u = 0$.

The remaining diagonal elements are at worst some constant C dependent on d and g, since the diagonal entries of each $A_G^{(s)}$ are all $O(1)$. Thus, writing Γ for the set of good vertices, we know

$$\left\| \sum_{u \in \Gamma} \alpha_u \right\| = \left\| \sum_{u \notin \Gamma} \alpha_u \right\| \leqslant C \log n$$

It is clear that by removing at most $C \log n$ vertices from Γ to create a new set Γ' we can choose a collection of unit vectors β_u for each $u \in U'$ so that

$$\sum_{u \notin \Gamma'} \beta_u = \sum_{u \in \Gamma'} \alpha_u.$$

Our final matrix Y will be the Gram matrix of these new β and remaining α vectors. We must finally check that the affine constraints against the $A_G^{(s)}$ matrices are still approximately satisfied. However,

even starting from a bad vertex, there are at most a constant number of vertices within s steps of it, and at most a constant number of non-backtracking walks to any such vertex. Thus

$$\begin{aligned}\left|\langle Y, A_G^{(s)}\rangle - \langle \tilde{Y}, A_G^{(s)}\rangle\right| &= \left|2\sum_{u\notin\Gamma', v\in\Gamma'}(A_G^{(s)})_{u,v}\alpha_u^T(\alpha_v-\beta_v) + \sum_{u,v\notin\Gamma'}(A_G^{(s)})_{u,u}\left(\|\alpha_u\|-\|\beta_u\|\right)\right| \\ &= O(\log n)\end{aligned}$$

where we have used that $\max_u \|\alpha_u\| = O(1)$ and broken up both summations by first enumerating the $O(\log n)$ vertices in U' and then the at most $O(1)$ vertices in its depth s neighborhood. Thus, for our fixed $\delta > 0$, we have

$$\left|\langle \tilde{Y}, A_G^{(s)}\rangle - \lambda^s \|q_s\|_{\mathrm{KM}}^2\right| = O(\log n) \leqslant \delta n$$

for n sufficiently large. □

The second part of Theorem 4.3 ensues from the following proposition.

Proposition 4.9. *Whenever $\lambda^2(d-1) < 1$, there exists a polynomial satisfying the conditions of Proposition 4.8.*

Proof. Such a polynomial y is exactly of the form

$$g = \sum_{s=0}^{m} \lambda^s q_s + \text{terms with larger } q_s\text{'s.}$$

We will use the extremely simple construction of letting the coefficients on the terms $q_{m+1}, q_{m+1}, \cdots$ *also* be powers of λ. The idea here is that, whenever $\lambda^2(d-1) < 1$, the series $\sum_{s\geqslant 0}\lambda^s q_s$ converges to a positive function on $(-2\sqrt{d-1}, 2\sqrt{d-1})$, so by taking a long enough initial segment, we can get a positive approximant.

In particular, let $p \gg m$ be even, and set

$$g = \sum_{s=0}^{p} \lambda^s q_s.$$

It is a standard calculation, employing the recurrence relation on the polynomials q_s, that

$$g(z) = \frac{1-\lambda^2+\lambda^{p+2}(d-1)q_p(z) - \lambda^{p+1}q_{p+1}(z)}{(d-1)\lambda^2 - \lambda z + 1}.$$

One an quickly verify that

$$\frac{1-\lambda^2}{(d-1)\lambda^2-\lambda z+1} > 0 \qquad \text{for all } |z| \leqslant 2\sqrt{d-1},$$

so we only need to check that $\lambda^2(d-1) < 1$ ensures $\lambda^{p+2}(d-1)q_p - \lambda^{p+1}q_{p+1} \to_p 0$. This follows immediately from Lemma 4.4, as $|q_p| \leqslant 2p\sqrt{d(d-1)^p}$. □

4.4 Robustness

We have shown already that if $(d-1)\lambda^2 > 1$, then for some constant $m(\lambda)$ and error tolerance $\delta(\lambda) > 0$, the level m symmetric path statistics SDP can solve the detection problem, and that otherwise no such δ and $m = O(1)$ can exist. In this section we show that this result is *robust*. To do so, we need to argue (i) that when $\boldsymbol{G} \sim \mathcal{P}$, or $\boldsymbol{G} \sim \mathcal{N}$ with $(d-1)\lambda^2 < 1$, the SDP with high probability remains feasible for any error tolerance δ, even after perturbing ρn edges, and (ii) that when $\boldsymbol{G} \sim \mathcal{N}$ and $(d-1)\lambda^2 > 1$, for some $\rho > 0$ and $\delta' < \delta(\lambda)$, the SDP remains infeasible at tolerance δ', even after perturbing ρn edges.

Assume that $\boldsymbol{G}$ was drawn from either the planted or null distribution, and that $\tilde{\boldsymbol{H}} \approx_\rho \boldsymbol{G}$. When we defined the $SPS(m, \lambda)$ SDP, we stipulated that in the event of an irregular input, we greedily remove edges until the maximum degree is d, and then greedily add edges among degree-deficient vertices until the minimum degree is d as well. Thus the actual input to the SDP is a graph $\boldsymbol{H}$, which one can verify satisfies $\boldsymbol{H} \approx_{\rho\xi} \boldsymbol{G}$ for some absolute constant ξ. Call a vertex $v \in [n]$ *corrupted* if its $(m+1)$-neighborhood in $\boldsymbol{H}$ differs from its $(m+1)$-neighborhood in $\boldsymbol{G}$. We begin by analyzing the difference $A_{\boldsymbol{G}}^{(s)} - A_{\boldsymbol{H}}^{(s)}$ for $s \in [m]$. Supposing v is not a corrupted vertex, then $A_{\boldsymbol{G}}^{(s)}$ and $A_{\boldsymbol{H}}^{(s)}$ agree on the vth row and column, which means $(A_{\boldsymbol{G}}^{(s)} - A_{\boldsymbol{H}}^{(s)})_{v,:} = 0$. On the other hand, if v is a corrupted vertex,

$$\left\| \left(A_{\boldsymbol{G}}^{(s)} - A_{\boldsymbol{H}}^{(s)} \right)_{v,-} \right\|_1 \leqslant \left\| A_{\boldsymbol{G}}^{(s)} \right\|_1 + \left\| A_{\boldsymbol{H}}^{(s)} \right\|_1 \leqslant 2d(d-1)^{s-1}$$

In particular, this means the entrywise 1-norm of $A_{\boldsymbol{G}}^{(s)} - A_{\boldsymbol{H}}^{(s)}$, is bounded by $2\xi\rho n \cdot 2d(d-1)^{s-1}$ since there are at most $2\xi\rho n$ corrupted vertices (i.e. if all corrupted edges had disjoint endpoints).

To prove (i), assume that the SDP is feasible at error tolerance δ on input $\boldsymbol{G}$, and write Y for a solution. Then

$$\left| \langle Y, A_{\boldsymbol{H}}^{(s)} \rangle - \langle Y, A_{\boldsymbol{G}}^{(s)} \rangle \right| \leqslant \left\| A_{\boldsymbol{H}}^{(s)} - A_{\boldsymbol{G}}^{(s)} \right\|_1 \leqslant 2\xi\rho d(d-1)^{s-1},$$

and thus Y is feasible on input $\boldsymbol{H}$ with error tolerance

$$\delta' = 2\xi\rho d(d-1)^{m-1} + \delta.$$

Since on $\boldsymbol{G} \sim \mathcal{P}$, or $\boldsymbol{G} \sim \mathcal{N}$ with $(d-1)\lambda^2 < 1$ the SDP is feasible for every $\delta > 0$, we can take $\delta \to 0$, and find we are free to choose ρ so long as we work at tolerance $2\xi\rho d(d-1)^m$.

To prove (ii), assume $\boldsymbol{G} \sim \mathcal{N}$ and $(d-1)\lambda^2 > 1$. Infeasibility of the SDP on input $\boldsymbol{G}$ is witnessed by the polynomial f from Proposition 4.7. So, let $\boldsymbol{Y}$ be a putative solution to the SDP on input $\boldsymbol{H}$, at tolerance δ', seeking a contradiction: recycling some computations from the proof of Proposition 4.6

$$0 \leqslant \langle \boldsymbol{Y}, f(\overline{A}_{\boldsymbol{G}}) \rangle$$

$$
\begin{aligned}
&= \langle \boldsymbol{Y}, \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} A_{\boldsymbol{G}}^{(s)} \rangle \\
&\leqslant \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} \left(\langle \boldsymbol{Y}, A_{\boldsymbol{H}}^{(s)} \rangle \pm 2\rho\xi d(d-1)^s \right) \\
&\leqslant \langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle_{\mathrm{KM}} + 2\rho\xi \sum_{s=0}^{m} |\langle f, q_s \rangle|_{\mathrm{KM}} + \delta' \sqrt{m} \|f\|_{\mathrm{KM}} \\
&\leqslant \langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle_{\mathrm{KM}} + (\delta' + 2\rho\xi) \sqrt{m} \|f\|_{\mathrm{KM}}.
\end{aligned}
$$

Thus we have a contradiction if

$$
\delta' < \frac{|\langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle_{\mathrm{KM}}|}{\sqrt{m}\|f\|_{\mathrm{KM}}} - 2\rho\xi.
$$

Here our choice of ρ must be constrained so that the right hand side of this expression is positive. This indicates a tradeoff between proximity to the KS threshold and robustness.

5 The Degree Regular Block Model

In this section we generalize the results from the previous section in two ways simultaneously: we study the fully general Degree Regular Block Model, and the full Local Statistics SDP. Both add some technical hurdles, but we will find that once these have been dealt with, the core arguments reduce to the symmetric results from Section 4. Throughout, assume that $\mathcal{N}$ is the uniform distribution on d-regular graphs, and $\mathcal{P}$ is the DRBM with fixed parameters (d, k, M, π). In this section we prove Theorem 2.4.

5.1 Local Statistics and Partially Labelled Subgraphs

As in the introduction let $x = \{x_{u,i}\}$ and $G = \{G_{u,v}\}$ be sets of variables indexed by $u \in [n]$ and $i \in [k]$. Our random graphs $\boldsymbol{G}$ and community labels $\boldsymbol{x}$ take values in the subset of $\{0,1\}^{\binom{n}{2}} \times \{0,1\}^{n \times k} \subset \mathbb{R}^{\binom{n}{2}} \times \mathbb{R}^{n \times k}$ defined by the polynomial equations

$$
\begin{aligned}
G_{u,v}^2 - G_{u,v} &= 0 \\
x_{u,i}^2 - x_{u,i} &= 0 \\
\sum_i x_{u,i} - 1 &= 0
\end{aligned} \tag{11}
$$

as in the introduction, we will write the ideal generated by the polynomials on the left of the second two equations as $\mathcal{B}_k$. Any point x in the vanishing locus of $\mathcal{B}_k$ corresponds to a map $\sigma_x : [n] \to [k]$. Write $\mathbb{S}[G, x] \subset \mathbb{R}[G, x]$ for the vector subspace of multilinear polynomials, fixed under the action of the symmetric group $\mathfrak{S}_n$ on the index set $[n]$, and for which no monomial contains $x_{u,i}x_{u,j}$ for $i \neq j$. This contains some polynomials that vanish modulo the equations above, but is convenient

to work with.

The local statistics SDP, given as input a graph $G_0 \in \{0,1\}^{\binom{n}{2}}$, attempts to find a pseudoexpectation $\widetilde{\mathbb{E}} : \mathbb{R}[x] \to \mathbb{R}$ that (i) evaluates to zero on any polynomial in $\mathcal{B}_k$, and (ii) assigns certain prescribed values to polynomials $p(G_0, x)$ obtained by evaluating a low-degree-polynomial $p \in \mathbb{S}[G, x]$ at the input graph. To state it fully, we will first construct a combinatorially meaningful vector space basis for $\mathbb{S}[G, x]$.

Definition 5.1 (Partially Labelled Subgraph)**.** A *partially labelled graph* (H, S, τ) consists of a graph H, distinguished subset of vertices $S \subset V(H)$, and a labelling $\tau : S \to [k]$. An *occurrence* of (H, S, τ) in a fully labelled graph (G, σ) is an injective homomorphism $\varphi : H \to G$ which respects the labelling. In other words, it is an injective map $\varphi : V(H) \to V(G)$ satisfying (i) $(\varphi(u), \varphi(v)) \in E(G)$ for every edge $(u, v) \in E$, and (ii) $\sigma(\varphi(v)) = \tau(v)$ for every $v \in S$.

Lemma 5.2 (Partially Labelled Subgraphs are a Basis)**.** *Let (H, S, τ) be a partially labelled subgraph. Then there is a symmetric polynomial $p_{H,S,\tau} \in \mathbb{R}[G, x]$ with degree $|S|$ in x and $|E(H)|$ in G that, for any $(\boldsymbol{G}, \boldsymbol{x})$ satisfying equations (11), counts occurrences of H in $(\boldsymbol{G}, \sigma_{\boldsymbol{x}})$. Furthermore, these polynomials form a basis for $\mathbb{S}[G, x]$.*

Proof. These polynomials are exactly the *monomial basis* obtained by considering the $\mathfrak{S}_n$ orbit of each multilinear monomial in G and x which does not conatin $x_{u,i}x_{u,j}$ for $i, j \in [k]$. Each such monomial is of the form

$$\prod_{(u,v)\in E} G_{u,v} \prod_{u \in S} x_{u,\tau(u)},$$

where $E \subset \binom{[n]}{2}$, $S \subset [n]$, and $\tau : S \to [k]$. Letting H be the graph whose vertices are those present either in S or in one of the pairs in E, when this monomial is evaluated at (G_0, x_0) satisfying the above equations, it is simply the indicator for one occurrence of (H, S, τ). By symmetrizing with respect to $\mathfrak{S}_n$, one obtains indicators for all possible such occurrences. □

The Local Statistics $L(2, m)$, on input G_0, contains constraints of the form

$$\widetilde{\mathbb{E}}\, p_{H,S,\tau}(G_0, x) \approx \mathop{\mathbb{E}}_{(\boldsymbol{G},\boldsymbol{x})\sim\mathcal{P}} p_{H,S,\tau}(\boldsymbol{G}, \boldsymbol{x}).$$

where $|S| \leqslant 2$ and $|E(H)| \leqslant m$. The following theorem computes the right hand side of the above equation in the planted, for this class of partially labelled subgraphs. We will discuss it briefly below and remit the proof to the appendix. Let (H, S, τ) be a partially labelled graph, and define

$$C_H(d) \triangleq \frac{\prod_{v\in V(H)} (d)_{\deg(v)}}{d^{|E(H)|}} \tag{12}$$

$$L_{(H,S,\tau)}(M, \pi) \triangleq \sum_{\hat{\tau}:\hat{\tau}|_S=\tau} \prod_{v\in v(H)} \pi(\hat{\tau}(v)) \prod_{(u,v)\in E(H)} M_{\hat{\tau}(u),\hat{\tau}(v)}. \tag{13}$$

Here $(d)_s = d(d-1)\cdots(d-s+1)$ is the falling factorial, and the sum in the second line is over all $\hat{\tau} : V(H) \to [k]$ which agree with τ on S. Define also $\chi(H) = |V(H)| - |E(H)|$ and $c(H) = \#$ connected components of H.

Theorem 5.3 (Local Statistics). *Let (H, S, τ) be a partially labelled graph with $O(1)$ edges. Then, with high probability over $(\boldsymbol{x}, \boldsymbol{G}) \sim \mathcal{P}$,*

$$p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G}) = n^{\chi(H)} L_{(H,S,\tau)}(M, \pi) \cdot C_H(d) \pm o(n^{c(H)}).$$

The proof may be found in Appendix B, but some comments are in order here. First, when H is a forest and $\chi(H) = c(H)$, we see that $p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G})$ concentrates, and that

$$n^{-c(H)} p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G}) \to L_{(H,S,\tau)}(M, \pi) C_H(d),$$

Conversely, it is well-known that (for instance) the number of cycles in $\boldsymbol{G}$ is Poisson distributed with constant mean, and thus all we can say with high probability is that there are $o(n)$ of them. This fact is reflected in greater generality in the discrepency between the $O(n^{\chi(H)})$ and $O(n^{c(H)})$ scales of $p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G})$ and its fluctuations, respectively, when H contains at least one cycle. Since we need to give the Local Statistics algorithm affine constraints that are satisfied with high probability in the planted model, we will include these two distinct scales in our full statement of the algorithm.

Second, the constants $L_{(H,S,\tau)}(M, \pi)$ and $C_H(d)$ have a pleasant interpretation in the case when H is a forest. If G is an unlabelled and locally treelike d-regular graph, in the sense that the shortest cycle is much larger than the longest path in H, then there are exactly $n^{c(H)} C_H(d)$ injective homomorphisms of H into G. On the other hand, $L_{(H,S,\tau)}(M, \pi)$ describes the probability of a certain outcome in a natural Markov process: start at some vertex $s \in S$, choose its label i according to π, and for each neighbor choose a label j with probability $T_{i,j} = M_{i,j}\pi(j)$. If one continues this until all of H is labelled, $L_{(H,S,\tau)}(M, \pi)$ gives the probability that every vertex $s \in S$ is given label $\tau(s)$.

We may finally define formally the Local Statistics algorithm.

Definition 5.4. The *degree (D_x, D_G) Local Statistics algorithm* with error tolerance $\delta > 0$, on input G_0, is the following SDP: find a pseudoexpectation $\widetilde{\mathbb{E}} : \mathbb{R}[x]_{\leqslant D_x} \to \mathbb{R}$ that is positive, normalized, satisfies $\mathcal{B}_k$, and for which

$$\widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, G_0) = n^{\chi(H)} L_{(H,S,\tau)}(M, \pi) C_H(d) \pm \delta n^{c(H)}$$

for every (H, S, τ) with $|S| \leqslant D_x$ and $|E(H)| \leqslant D_G$.

Lemma 5.5. *For any $\delta > 0$, the $\mathrm{LoSt}(D_x, D_G)$ algorithm is with high probability feasible on input $\boldsymbol{G} \sim \mathcal{P}$.*

Proof. Let $\boldsymbol{x}$ be the hidden signal; we will set $\widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G}) = p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G})$. This is clearly positive, satisfies $\mathcal{B}_k$, and from Theorem 5.3 it satisfies the affine constraints in Definition 5.4. □

5.2 Distinguishing

Let us prove the first part of Theorem 2.4: when $(d-1)\lambda_2^2 > 1$, there exist constant m, ρ, and $\delta > 0$ for which the $\mathrm{LoSt}(2, m)$ SDP at error tolerance δ solves the detection problem ρ-robustly. Since the

SDP is with high probability feasible for any m and $\delta > 0$ when $\boldsymbol{G} \sim \mathcal{P}$, it remains only to show infeasibility for some m, ρ, δ when $\boldsymbol{G} \sim \mathcal{N}$.

Let $\boldsymbol{G} \sim \mathcal{N}$, and assume we have a viable pseudoexpectation $\widetilde{\mathbb{E}}$ for the LoSt$(2, m)$ SDP with some tolerance $\delta > 0$. Write $X \succeq 0$ for the $nk \times nk$ matrix whose $(u, i), (v, j)$ entry is $\widetilde{\mathbb{E}}\, x_{u,i} x_{v,j}$; it is routine that positivity of $\widetilde{\mathbb{E}}$ implies positive semidefiniteness of X. It will at times be useful to think of X as a $k \times k$ matrix of $n \times n$ blocks $X_{i,j}$, and at others as an $n \times n$ matrix of $k \times k$ blocks $X_{u,v}$. Let us also define matrices $A_{\boldsymbol{G}}^{\langle s \rangle}$ that count *self-avoiding* walks of length s, as opposed to the non-backtracking walks counted by the matrices $A_{\boldsymbol{G}}^{(s)}$ whose notation they echo. Our strategy will be to first write the moment matching constraints on $\widetilde{\mathbb{E}}$ as affine constraints of the form $\langle X_{i,j}, Y \rangle = C$, and then combine these to contradict feasibility of X.

Lemma 5.6. *For any i, j, and any $s = 0, \ldots, m$, recall that $A_G^{(s)}$ is the matrix counting non-backtracking walks of length s, and $\mathbb{J}$ is the all-ones matrix. For any $\delta' > \delta$,*

$$\langle X_{i,j}, A_{\boldsymbol{G}}^{(s)} \rangle = \pi(i) T_{i,j}^s \|q_s\|_{\mathrm{KM}}^2 n \pm \delta' n$$
$$\langle X_{i,j}, \mathbb{J} \rangle = \pi(i)\pi(j) n^2 \pm \delta' n^2$$

Proof. For the first assertion, let (H, S, τ) be the path of length s whose endpoints are labelled $i, j \in [k]$. In this case $C_H(d) = d(d-1)^{s-1} = \|q_s\|_{\mathrm{KM}}^2$, and one can quickly verify that $L_{(H,S,\tau)} = \pi(i) T_{i,j}^s$. Each *self-avoiding* walk of length s in G is an occurrence of H, so from Theorem 5.3

$$\langle X_{i,j}, A_{\boldsymbol{G}}^{\langle s \rangle} \rangle = \widetilde{\mathbb{E}}\, p_{H,S,\tau}(x, \boldsymbol{G}) = \pi(i) M_{i,j}^s \|q_s\|_{\mathrm{KM}}^2 n \pm \delta n$$

It is an easy consequence of Lemma 4.5 that for every constant s, $A_{\boldsymbol{G}}^{(s)}$ and $A_{\boldsymbol{G}}^{\langle s \rangle}$ differ only on $O(\log n)$ rows, and since each row has constant L_2 norm,

$$\left\| A_{\boldsymbol{G}}^{(s)} - A_{\boldsymbol{G}}^{\langle s \rangle} \right\|_F^2 = O(\log n).$$

The matrix X has diagonal elements $X_{(u,i),(u,i)} = \widetilde{\mathbb{E}}\, x_{u,i}^2 = \widetilde{\mathbb{E}}\, x_{i,u}$ by the Boolean constraint, and $\widetilde{\mathbb{E}}\,(x_{u,1} + \cdots + x_{u,k}) = 1$ by the Single Color constraint. By PSD-ness of X, every $\widetilde{\mathbb{E}}\, x_{u,i}^2 = \widetilde{\mathbb{E}}\, x_{u,i}$ is nonnegative, so each is between zero and one. It is a standard fact that the off-diagonal entries of such a PSD matrix have magnitude at most one, so from Lemma 4.4

$$\langle X_{i,j}, A_{\boldsymbol{G}}^{(s)} \rangle = \langle X_{i,j}, A_{\boldsymbol{G}}^{\langle s \rangle} \rangle + \langle X_{i,j} A_{\boldsymbol{G}}^{\langle s \rangle} - A_{\boldsymbol{G}}^{(s)} \rangle = \langle X_{i,j}, A_{\boldsymbol{G}}^{\langle s \rangle} \rangle \pm O(\log n) = \pi(i) M_{i,j}^s \|q_s\|_{\mathrm{KM}}^2 n \pm \delta' n$$

for $s = 0, \ldots, m$, any $\delta' > \delta$, and n sufficeintly large. For the second assertion, when $i \neq j$ take (H, S, τ) to be the partially labelled graph on two disconnected vertices, with labels i and j respectively. In this case $C_H(d) = 1$, and $L_{(H,S,\tau)}(M, \pi) = \pi(i)\pi(j)$. We then have

$$\langle X_{i,j}, \mathbb{J} \rangle = \widetilde{\mathbb{E}}\, p_{H,S,\tau}(x, \boldsymbol{G}) = \pi(i)\pi(j) n^2 \pm \delta n^2$$

For the case $i = j$, let (H', S', τ') be a single vertex labelled i, for which $C_H(d) = 1$ and $L_{(H',S',\tau')}(M, \pi) =$

$\pi(i)$. We can write

$$\langle X_{i,i}, \mathbb{J}\rangle = \widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G}) + \widetilde{\mathbb{E}}\, p_{(H',S',\tau')}(x, \boldsymbol{G}) = \pi(i)^2 n^2 + \pi(i) n \pm \delta n^2 = \pi(i)\pi(j) n^2 + \delta' n^2$$

for any $\delta' > \delta$ and n sufficiently large. □

We will now apply a fortuitous change of basis furnished to us by the transition matrix T. Let us write F for the matrix of right eigenvectors of T, normalized so that every column has unit norm, and sorted so that the first column is a multiple of the all-ones vector. Thus $TF = F\Lambda$, where Λ is a diagonal matrix containing the eigenvalues, sorted in decreasing order of magnitude. It is a standard fact from the theory of reversible Markov chains that $F^{-1}\operatorname{Diag}(\pi)F = \mathbb{1}$.

Now, define a matrix $\check{X} \triangleq (F^T \otimes \mathbb{1})X(F \otimes \mathbb{1})$, by which we mean that

$$\check{X} = \begin{pmatrix} F_{1,1}\mathbb{1} & \cdots & F_{1,k}\mathbb{1} \\ \vdots & \ddots & \vdots \\ F_{k,1}\mathbb{1} & \cdots & F_{k,k}\mathbb{1} \end{pmatrix} \begin{pmatrix} X_{1,1} & \cdots & X_{1,k} \\ \vdots & \ddots & \vdots \\ X_{k,1} & \cdots & X_{k,k} \end{pmatrix} \begin{pmatrix} F_{1,1}\mathbb{1} & \cdots & F_{1,k}\mathbb{1} \\ \vdots & \ddots & \vdots \\ F_{k,1}\mathbb{1} & \cdots & F_{k,k}\mathbb{1} \end{pmatrix}.$$

We will think of $\check{X}$, analogous to X, as a $k \times k$ matrix of $n \times n$ blocks $\check{X}_{i,j}$. Note that we can also think of this as as a change of basis $x \mapsto F^T x$ directly on the variables appearing in polynomials accepted by our pseudoexpectation.

Lemma 5.7. *For any $s = 0, \ldots, m$, and any $\delta'' > \|F\|^2\sqrt{k}\delta$, we have*

$$\langle \check{X}_{i,j} A_{\boldsymbol{G}}^{(s)}\rangle = \begin{cases} 0 & i \neq j \\ \lambda_i^s \|q_s\|_{\mathrm{KM}}^2 & i = j \end{cases} \pm \delta'' n$$

$$\langle \check{X}_{i,j}, \mathbb{J}\rangle = \begin{cases} n^2 & i = j = 1 \\ 0 & \textit{else} \end{cases} \pm \delta'' n^2$$

Proof. Our block-wise change of basis commutes with taking inner products between the blocks $X_{i,j}$ and the non-backtracking walk matrices. In other words, invoking Lemma 5.6 with $\delta' > \delta$ and keeping track of how the additive errors compound as we take linear combinations,

$$\begin{aligned} \begin{pmatrix} \langle \check{X}_{1,1}, A_{\boldsymbol{G}}^{(s)}\rangle & \cdots & \langle \check{X}_{1,k}, A_{\boldsymbol{G}}^{(s)}\rangle \\ \vdots & \ddots & \vdots \\ \langle \check{X}_{k,1}, A_{\boldsymbol{G}}^{(s)}\rangle & \cdots & \langle \check{X}_{k,k}, A_{\boldsymbol{G}}^{(s)}\rangle \end{pmatrix}_{i,j} &= \left(F^T \begin{pmatrix} \langle X_{1,1}, A_{\boldsymbol{G}}^{(s)}\rangle & \cdots & \langle X_{1,k}, A_{\boldsymbol{G}}^{(s)}\rangle \\ \vdots & \ddots & \vdots \\ \langle X_{k,1}, A_{\boldsymbol{G}}^{(s)}\rangle & \cdots & \langle X_{k,k}, A_{\boldsymbol{G}}^{(s)}\rangle \end{pmatrix} F \right)_{i,j} \\ &= \left(F^T \operatorname{Diag}(\pi) T^s F \right)_{i,j} \cdot \|q_s\|_{\mathrm{KM}}^2 n \pm \|F\|^2\sqrt{k}\delta' n \\ &= \left(F^T \operatorname{Diag}(\pi) F \Lambda^s \right)_{i,j} \cdot \|q_s\|_{\mathrm{KM}}^2 n \pm \|F\|^2\sqrt{k}\delta' n \\ &= \Lambda_{i,j}^s \cdot \|q_s\|_{\mathrm{KM}}^2 n \pm \|F\|^2\sqrt{k}\delta' n. \end{aligned}$$

A parallel calculation gives us

$$\begin{aligned}\begin{pmatrix} \langle \check{X}_{1,1}, \mathbb{J} \rangle & \cdots & \langle \check{X}_{1,k}, \mathbb{J} \rangle \\ \vdots & \ddots & \vdots \\ \langle \check{X}_{k,1}, \mathbb{J} \rangle & \cdots & \langle \check{X}_{k,k}, \mathbb{J} \rangle \end{pmatrix}_{i,j} &= \left(F^T \begin{pmatrix} \langle X_{1,1}, \mathbb{J} \rangle & \cdots & \langle X_{1,k}, \mathbb{J} \rangle \\ \vdots & \ddots & \vdots \\ \langle X_{k,1}, \mathbb{J} \rangle & \cdots & \langle X_{k,k}, \mathbb{J} \rangle \end{pmatrix} F \right)_{i,j} \\ &= \left(F^T \pi \pi^T F \right)_{i,j} \cdot n^2 \pm \|F\|^2 \sqrt{k} \delta' n^2 \\ &= \left(e_1 e_1^T \right)_{i,j} \cdot n^2 \pm \|F\|^2 \sqrt{k} \delta' n^2\end{aligned}$$

where e_1 is the first standard basis vector. The final line comes since π, being the left eigenvector associated to $\lambda_1 = 1$, is (up to scaling) the first row of F^{-1}. □

With Lemma 5.7 in hand, the remainder of the proof follows from Proposition 4.6 and Proposition 4.7 in the previous section. In particular, each block $\check{X}_{i,i}$ for $i = 2, ...k$ is a feasible solution to $SPS(m, \lambda_i)$ SDP with error tolerance $\delta'' > \|F\|^2 \sqrt{k}\delta$. We showed already that when $\lambda^2(d-1) > 1$, and for small enough error tolerance and large enough m, this SDP is w.h.p. infeasible on input $\boldsymbol{G} \sim \mathcal{N}$. Thus we need simply to make δ small enough so that δ'' is below the minimum tolerance in Proposition 4.6.

5.3 Spectral Distinguishing

Our argument in the previous section can be recast to prove Corollary 2.5, namely that above the Kesten-Stigum threshold the spectrum of the adjacency matrix can also be used to distinguish the null and planted distributions.

Let $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathcal{P}_{d,k,M,\pi}$, and write $\boldsymbol{X} \triangleq \boldsymbol{x}\boldsymbol{x}^T$, and

$$\check{\boldsymbol{X}} = (F^T \otimes \mathbb{1}) \boldsymbol{X} (F \otimes \mathbb{1}) = (F^T \boldsymbol{x})(F^T \boldsymbol{x})^T \triangleq \check{\boldsymbol{x}} \check{\boldsymbol{x}}^T.$$

Think of $\check{\boldsymbol{X}}$ as a block matrix $(\boldsymbol{X}_{i,j})_{i,j \in [k]}$, as we did X in the previous section, and $\check{\boldsymbol{x}}$ as a block vector $(\check{\boldsymbol{x}}_i)_{i \in [k]}$. Applying Theorem 5.3 and repeating the calculations in Lemma 5.6 and Lemma 5.7 *mutatis mutandis* with $\boldsymbol{X}$ instead of X, we can show that w.h.p.

$$\langle \check{\boldsymbol{X}}_{i,j}, A_{\boldsymbol{G}}^{(s)} \rangle = \lambda_i \|q_s\|_{\mathrm{KM}}^2 n + o(n) \qquad \text{if } i = j$$

and zero otherwise, for every $s = O(1)$ and

$$\langle \check{\boldsymbol{X}}_{1,1}, \mathbb{J} \rangle = \begin{cases} n^2 & i = j = 1 \\ 0 & \text{else} \end{cases},$$

with strict equality following from the rigidity of the group sizes in the planted model. Because $A_{\boldsymbol{G}}^{(s)} = \mathbb{1}$, we know

$$\check{\boldsymbol{x}}_i^T \check{\boldsymbol{x}}_j = \langle \check{\boldsymbol{X}}_{i,j}, \mathbb{1} \rangle = 0$$

when $i \neq j$. In other words, the k vectors $\check{x}_1, ..., \check{x}_k$ are orthogonal.

We can show that A_G has an eigenvalue with a separation $\eta > 0$ from the bulk spectrum by proving

$$\check{x}_i^T f(A_G)\check{x}_i = \langle \check{\mathbf{X}}_{i,i}, f(A_G)\rangle < 0$$

for some polynomial $f(x)$ positive on of $(-2\sqrt{d-1}-\eta, 2\sqrt{d-1}+\eta)$. As long as $(d-1)\lambda_i^2 > 1$, the same polynomial from Proposition 4.7 works here. As the $\check{x}_i$ are orthogonal, we get one distinct eigenvalue outside the bulk for each eigenvalue of T satisfying this property.

Remark 5.8. To distinguish the null model from the planted one using the spectrum of A_G, simply return PLANTED if A_G has a single eigenvalue other than d whose magnitude is bigger than $2\sqrt{d-1}+\delta$ for any error tolerance δ you choose, and NULL otherwise. Unfortunately, this distinguishing algorithm is not robust to adversarial edge insertions and deletions. For instance, given a graph $G \sim \mathcal{N}$, the adversary can create a disjoint copy of K_{d+1}, the complete graph on $d+1$ vertices, whose eigenvalues are all $\pm d$. The spectrum of the perturbed graph is the disjoint union of $\pm d$ and the eigenvalues of the other component(s), so the algorithm will be fooled. We will show in Section 5.5 that the Local Statistics SDP is robust to this kind of perturbation.

5.4 Lower Bounds

In this section, we prove the second half of Theorem 2.4, which gives a complementary lower bound: if every one of $\lambda_2, ..., \lambda_k$ has modulus at most $1/\sqrt{d-1}$ there exists some feasible solution to the Local Path Statistics SDP for every $m \geqslant 1$. We can specify a pseudoexpectation completely by way of an $(nk+1) \times (nk+1)$ positive semidefinite matrix

$$\begin{pmatrix} 1 & \widetilde{\mathbb{E}}\, x^T \\ \widetilde{\mathbb{E}}\, x & \widetilde{\mathbb{E}}\, x^T x \end{pmatrix} \triangleq \begin{pmatrix} 1 & l^T \\ l & X \end{pmatrix}.$$

After first writing down the general properties required of *any* quadratic pseudoexpectation satisfying $\mathcal{B}_k$, we'll show that in order for $\widetilde{\mathbb{E}}$ to match every moment asked of it by the LoSt$(2, m)$ SDP, it suffices for it to satisfy

$$\widetilde{\mathbb{E}}\, p_{H,S,\tau}(x, G) \approx \mathbb{E}\, p_{H,S,\tau}(\boldsymbol{G}, \boldsymbol{x})$$

when (H, S, τ) is a path of length $0, ..., m$ with labelled endpoints, or a pair of disjoint, labelled vertices. Finally, we'll construct a pseudoexpectation matching these path moments out of feasible solutions to the symmetric path statistics SDP from the previous section.

Lemma 5.9. *The set of $\mathcal{B}_k$-satisfying pseudoexpectations is parameterized by pairs $(X, l) \in \mathbb{R}^{nk\times nk} \times \mathbb{R}^{nk}$ for which*

$$\begin{pmatrix} 1 & l^T \\ l & X \end{pmatrix} \succeq 0 \tag{14}$$

$$\operatorname{diag}(X) = l \tag{15}$$

$$\operatorname{tr} X_{u,u} = e^T l = 1 \qquad \forall u \in [n] \tag{16}$$
$$X_{u,v} e = l_u \qquad \forall u, v \in [n] \tag{17}$$

Proof. Recall that the set $\mathcal{B}_k$ is defined by the polynomial equations

$$\begin{aligned} &\textit{Boolean} && x_{u,i}^2 = x_{u,i} && \forall u \in [n] \text{ and } i \in [k] \\ &\textit{Single Color} && \sum_i x_{u,i} = 1 && \forall u \in [n] \end{aligned}$$

That a degree-two pseudoexpectation *satisfies* these constraints means

$$\begin{aligned} \widetilde{\mathbb{E}}\, p(x) x_{u,i}^2 &= \widetilde{\mathbb{E}}\, p(x) x_{u,i} && \forall p \text{ s.t. } \deg p = 0 \\ \widetilde{\mathbb{E}}\, p(x) \sum_i x_{u,i} &= \widetilde{\mathbb{E}}\, p(x) && \forall p \text{ s.t. } \deg p \leqslant 1. \end{aligned}$$

Writing $X = \widetilde{\mathbb{E}}\, x^T x$ and $l = \widetilde{\mathbb{E}}\, x$ as above, the first constraint is equivalent to $l = \operatorname{diag}(X)$, since the degree-zero polynomials are just constants, and we can guarantee that the second holds for every polynomial of degree at most one by requiring it on $p = 1$ and $p =, x_{v,j}$ for all v and j. The Lemma is simply a concise packaging of these facts, using the block notation $X = (X_{u,v})_{u,v \in [n]}$ and $l = (l_u)_{u \in [n]}$. □

Proposition 5.10. *Let $\boldsymbol{G} \sim \mathcal{N}$, and let the pair $(X, l) \in \mathbb{R}^{nk \times nk} \times \mathbb{R}^{nk}$ satisfies (14)-(17) and*

$$\begin{aligned} \langle e, l_i \rangle &= \pi(i) n \pm \delta n \\ \langle X_{i,j}, A_{\boldsymbol{G}}^{(s)} \rangle &= \pi(i) T_{i,j}^s n \pm \delta n \\ \langle X_{i,j}, \mathbb{J} \rangle &= \pi(i)\pi(j) n^2 \pm \delta n^2, \end{aligned}$$

then with high probability the degree-two pseudoexpectation that they induce is a feasible solution to the $\mathrm{LoSt}(2, m)$ *SDP with any error tolerance $\delta' > \delta$.*

We will defer the proof of Proposition 5.10 to Appendix B. Its conclusion in hand, we can now set about constructing a pseudoexpectation. Since $(d-1)\lambda_2^2 < 1$, the $SPS(\lambda_i, m)$ SDP is feasible for every error tolerance $\delta' > 0$. Thus for each $i = 2, ..., k$ there exists a feasible solution in the form of a PSD matrix $Y(\lambda_i)$ satisfying

$$\begin{aligned} Y(\lambda_i)_{u,u} &= 1 && \forall u \in [n] \\ \langle Y(\lambda_i), A_{\boldsymbol{G}}^{(s)} \rangle &= \lambda_i^s \|q_s\|_{\mathrm{KM}}^2 n \pm \delta' n && \forall s \in [m] \\ \langle Y(\lambda_i), \mathbb{J} \rangle &= 0 \pm \delta' n^2. \end{aligned}$$

Now, define $\check{X}$ to be the $k \times k$ block diagonal matrix

$$\check{X} = \begin{pmatrix} \mathbb{J} & & & \\ & Y(\lambda_2) & & \\ & & \ddots & \\ & & & Y(\lambda_k), \end{pmatrix}$$

i.e. $\check{X}_{i,j} = 0$ when $i \neq j$, and the diagonal blocks are as above, and similarly let $\check{l} = (e, 0, ..., 0)^T$. We claim that the pair $X = (F^{-T} \otimes \mathbb{1})\check{X}(F^{-1} \otimes \mathbb{1})$ and $l = (F^{-1} \otimes \mathbb{1})\check{l}$ satisfies the conditions of Lemma 5.9 and Proposition 5.10.

First

$$\begin{pmatrix} 1 & l^T \\ l & X \end{pmatrix} = \begin{pmatrix} 1 & \\ & F^{-T} \otimes \mathbb{1} \end{pmatrix} \begin{pmatrix} 1 & \check{l}^T \\ \check{l} & \check{X} \end{pmatrix} \begin{pmatrix} 1 & \\ & F^{-1} \otimes \mathbb{1} \end{pmatrix} \succeq 0$$

by taking a Schur complement. Since π is the first row of F^{-1}, we know $l_i = \pi(i)e$ for each $i =\in [k]$. Moreover, since X is obtained by changing basis block-wise, the diagonal of X depends only on the diagonals of $\mathbb{J}$ and the $Y(\lambda_i)$, all of which are all ones, so

$$\begin{aligned} \operatorname{diag} X &= \operatorname{diag}\left((F^{-T} \otimes \mathbb{1}) \operatorname{Diag}(\operatorname{diag}(\check{X}))(F^{-1} \otimes \mathbb{1})\right) \\ &= \operatorname{diag}\left((F^{-T} \otimes \mathbb{1})\mathbb{1}(F^{-1} \otimes \mathbb{1})\right) \\ &= \operatorname{diag}\left(F^{-T}F^{-1} \otimes \mathbb{1}\right) \\ &= \operatorname{diag}\left(\operatorname{Diag}\pi \otimes \mathbb{1}\right) \\ &= (\pi(1)e, ..., \pi(k)e) = l \end{aligned}$$

as desired. Similarly, because $\check{X}$ is block diagonal when regarded as $k \times k$ matrix of $n \times n$ blocks, if we treat it instead as an $n \times n$ matrix of $k \times k$ blocks $\check{X}_{u,v}$, then $\check{X}_{u,u} = \mathbb{1}$ for every $u \in [n]$, and

$$\operatorname{tr} X_{u,u} = \operatorname{tr} F^{-T}\check{X}_{u,u}F^{-1} = \operatorname{tr} F^{-T}F^{-1} = \operatorname{tr}\operatorname{Diag}\pi = 1.$$

Finally, the top row of each $\check{X}_{u,v}$ is the vector e_1^T, so

$$X_{u,v}e = F^{-T}\check{X}_{u,v}F^{-1}e = F^{-T}\check{X}_{u,v}e_1 = F^{-T}e_1 = \pi = l_u.$$

It remains to verify the affine conditions in Proposition 5.10. As in the proof of Lemma 5.7, since each $Y(\lambda_i)$ is a feasible solution to the $SPS(\lambda_i, m)$ SDP with error tolerance δ',

$$\begin{aligned} \langle X_{i,j}, A_G^{(s)} \rangle &= \left(F^{-T} \begin{pmatrix} \langle \mathbb{J}, A_G^{(s)} \rangle & & & \\ & \langle Y(\lambda_2), A_G^{(s)} \rangle & & \\ & & \ddots & \\ & & & \langle Y(\lambda_k), A_G^{(s)} \rangle \end{pmatrix} F^{-1} \right)_{i,j} \\ &= \left(F^{-T}\Lambda^s F^{-1}\right)_{i,j} \cdot \|q_s\|_{\mathrm{KM}}^2 n \pm \|F^{-1}\|^2\sqrt{k}\delta' n \\ &= (\operatorname{Diag}(\pi)T^s)_{i,j} \cdot \|q_s\|_{\mathrm{KM}}^2 n \pm \|F^{-1}\|^2\sqrt{k}\delta' n \\ &= \pi(i)T^s_{i,j} \cdot \|q_s\|_{\mathrm{KM}}^2 n \pm \|F^{-1}\|^2\sqrt{k}\delta' n \end{aligned}$$

and

$$\begin{aligned}\langle X_{i,j}, \mathbb{J}\rangle &= \left(F^{-T}\begin{pmatrix}\langle \mathbb{J},\mathbb{J}\rangle & & & \\ & \langle Y(\lambda_2),\mathbb{J}\rangle & & \\ & & \ddots & \\ & & & \langle Y(\lambda_k),\mathbb{J}\rangle\end{pmatrix}F^{-1}\right)_{i,j}\\ &= \left(F^{-T}e_1e_1^TF^{-1}\right)_{i,j}\cdot n^2 \pm \|F^{-1}\|^2\sqrt{k}\delta' n^2\\ &= \left(\pi\pi^T\right)_{i,j}\cdot n^2 \pm \|F^{-1}\|^2\sqrt{k}\delta' n^2\\ &= \pi(i)\pi(j)\cdot n^2 \pm \|F^{-1}\|^2\sqrt{k}\delta' n^2,\end{aligned}$$

and by setting δ' sufficiently small, we can make each of these errors at most any $\delta > 0$ of our choosing.

5.5 Robustness

The proof of robustness largely reduces to the discussion in Section 4.4. Recall that we need to produce a $\rho > 0$ for which (i) when $\boldsymbol{G} \sim \mathcal{P}$, or $\boldsymbol{G} \sim \mathcal{N}$ with $(d-1)\lambda_2^2 < 1$, the SDP with high probability remains feasible for any error tolerance δ, even after perturbing ρn edges, and (ii) that when $\boldsymbol{G} \sim \mathcal{N}$ and $(d-1) > \lambda_2^2$, if the SDP is infeasible at tolerance δ, it remains so at some tolerance $\delta' < \delta$ even after perturbing ρn edges.

For (i), assume that the SDP is feasible at error tolerance δ on input $\boldsymbol{G}$. We build the SDP as a linear combination of solutions $Y(\lambda_i)$ to the $SPS(m, \lambda_i)$ SDP, which we argued in Section 4.4 is robust in the desired sense. For (ii), when $\boldsymbol{G} \sim \mathcal{N}$ and $(d-1)\lambda_2^2 > 1$, we reduced infeasibility of the LoSt$(2, m)$ SDP to that of the $SPS(m, \lambda_2)$ SDP, which we showed already is infeasible. Moreover, from Section 4.4, the latter remains infeasible after a sufficiently small perturbation.

6 The Stochastic Block Model

We turn, finally, to the proof of Theorem 2.2 concerning the local statistics algorithm and Stochastic Block Model. For the sake of exposition, as we did for the DRBM, we will first write down a simpler SDP that can robustly solve the detection problem above the KS threshold, and then show that feasibility of the full SDP implies feasibility of this simpler one.

Throughout this section, let $\mathcal{P}$ denote the SBM with fixed parameters (d, k, M, π), and $\mathcal{N} = \mathcal{G}(n, d/n)$. Recall that to sample a pair $(\boldsymbol{G}, \boldsymbol{x})$ from the planted model, we first choose a partition $V_1(\boldsymbol{G}) \sqcup \cdots \sqcup V_k(\boldsymbol{G}) = [n]$ by placing each vertex in group V_i with probability $\pi(i)$, setting $\boldsymbol{x}_{u,i}$ equal to 1 if $u \in V_i$; it will be convenient to write $\boldsymbol{\sigma} : [n] \to [k]$ for this random labelling map. Then, we include each edge $(u, v) \in E(\boldsymbol{G})$ with probability $M_{\boldsymbol{\sigma}(u),\boldsymbol{\sigma}(v)}d/n$, setting $\boldsymbol{G}_{u,v} = 1$ in this event.

Now, for any graph G, define the matrices $\overline{A}_G^{(s)}$ as follows. For each walk in the complete graph K_n, write $\gamma : u \to v$ if it begins with u and ends with v, let $w_G(\gamma) = \prod_{e\in\gamma}(G_e - d/n)$, and set

$$\left(\overline{A}_G^{(s)}\right)_{u,v} = \sum_{\gamma:u\to v,|\gamma|=s} w_G(\gamma). \tag{18}$$

When $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathcal{P}$, define as in Section 5.1 and Section 5.2 an $nk \times nk$ matrix $\boldsymbol{X} \triangleq \boldsymbol{x}\boldsymbol{x}^*$. As before, we will at times think of $\boldsymbol{X}$ as an $n \times n$ matrix of $k \times k$ blocks $\boldsymbol{X}_{u,v}$, and at others a $k \times k$ matrix of $n \times n$ blocks $\boldsymbol{X}_{i,j}$.

Claim 6.1. Let $\overline{T} = T - e^T\pi$. Then

$$\mathbb{E}\langle \boldsymbol{X}_{i,j}, \overline{A}_{\boldsymbol{G}}^{(s)}\rangle = \pi(i)\overline{T}^s_{i,j} \cdot d^s n + O(1),$$

and this inner product enjoys concentration of $O(\sqrt{n})$.

Proof. Let γ be some length-s self-avoiding walk in the complete graph; WLOG we can label its vertices with the set $[s+1]$. We need to calculate the expectation of $w_{\boldsymbol{G}}(\gamma)$ on the event that its endpoints are labelled i and j:

$$\begin{aligned}\mathbb{E}[w_{\boldsymbol{G}}(\gamma), \text{labels } i \text{ and } j] &= \sum_{\eta:[s+1],\eta(1)=i,\eta(s+1)=j} \pi(i)\prod_{t\in[s]}(M_{\eta(t),\eta(t+1)} - 1)(d/n)\pi(\eta(t+1)) \\ &= \pi(i)(T - e^T\pi)^s_{i,j} \cdot (d/n)^s.\end{aligned}$$

There are $n(n-1)(n-2)\cdots(n-s)$ length-s self avoiding walks in the complete graph, so already the total expectation of $w_{\boldsymbol{G}}(\gamma)$ among these walks accounts for the quantity in the claim. On the other hand, those γ's which contain a cycle contribute negligiblly: there are $O(n^v)$ γ's with v vertices, and for each one $\mathbb{E}\, w_{\boldsymbol{G}}(\gamma) = O(n^{-e})$, so the total contribution of γ's with $e \geqslant v$ is at best $O(1)$. □

This motivates the following SDP:

Definition 6.2. For each $m \geqslant 1$, the *level m path statistics SDP* with error tolerance $\delta > 0$ is the feasibility problem

$$\begin{aligned}\text{Find } X = (X_{i,j}) \succeq 0 \text{ s.t. } (X_{i,i})_{u,u} &\leqslant 1 && \forall u \in [n], i \in [k] && (19)\\ \operatorname{tr}(X_{u,u}) &= 1 && \forall u \in [n] && (20)\\ \langle X_{i,j}, \overline{A}_G^{(s)}\rangle &= \pi(i)\overline{T}^s_{i,j} \cdot d^s n \pm \delta n && \forall i,j \in [k], s = 0,...,m. && (21)\end{aligned}$$

Theorem 6.3. *When $\lambda_2^2 d > 1$, there exists $m = O(1)$ and $\delta > 0$ for which the level m path statistics SDP can solve the detection problem. Conversely, when $\lambda_2^2 d < 1$, no such m and δ exist.*

Recall that Λ is a $k \times k$ diagonal matrix containing the eigenvalues of T, sorted in descending order of modulus from the upper left corner. Since e and π^* are right and left eigenvectors, respectively, of T, $\overline{T}$ commutes with T and satisfies $\overline{T}F = F\overline{\Lambda}$, where $\overline{\Lambda}$ is obtained from Λ by

deleting the upper left entry (which in our setup is equal to 1). We will accordingly take the same change-of-basis approach as in the DRBM. For any feasible solution X to this SDP, we can form an analogous matrix $\check{X} \triangleq (F^T \otimes \mathbb{1})X(F \otimes \mathbb{1})$, with blocks $\check{X}_{i,j}$. Following Lemma 5.7, observe that

$$\begin{aligned}
\langle \check{X}_{i,j}, \overline{A}_{\boldsymbol{G}}^{(s)} \rangle = \begin{pmatrix} \langle \check{X}_{1,1}, \overline{A}_G^{(s)} \rangle & \cdots & \langle \check{X}_{1,k}, \overline{A}_G^{(s)} \rangle \\ \vdots & \ddots & \vdots \\ \langle \check{X}_{k,1}, \overline{A}_G^{(s)} \rangle & \cdots & \langle \check{X}_{k,k}, \overline{A}_G^{(s)} \rangle \end{pmatrix}_{i,j} &= \left(F^T \begin{pmatrix} \langle X_{1,1}, \overline{A}_G^{(s)} \rangle & \cdots & \langle X_{1,k}, \overline{A}_G^{(s)} \rangle \\ \vdots & \ddots & \vdots \\ \langle X_{k,1}, \overline{A}_G^{(s)} \rangle & \cdots & \langle X_{k,k}, \overline{A}_G^{(s)} \rangle \end{pmatrix} F \right)_{i,j} \\
&= \left(F^T \operatorname{Diag}(\pi)(T - e\pi^*)^s F \right)_{i,j} \cdot d^s n \pm \|F\|^2 \sqrt{k}\delta n \\
&= \left(F^T \operatorname{Diag}(\pi) F \overline{\Lambda}^s \right)_{i,j} \cdot d^s n \pm \|F\|^2 \sqrt{k}\delta n \\
&= \overline{\Lambda}^s_{i,j} \cdot d^s n \pm \|F\|^2 \sqrt{k}\delta n.
\end{aligned}$$

Moreover, the diagonal entries of $\check{X}$ are bounded by a constant dependant only on M and π, which we can check by considering the blocks $\check{X}_{u,u} = F^T X_{u,u} F$. Since $\operatorname{tr} X_{u,u} = 1$, the maximal diagonal entry of $\check{X}_{u,u}$ is at most $\operatorname{tr} F^T X_{u,u} F \leqslant \|F\|^2$.

Our observations about the matrices $\check{X}_{i,i}$ from Section 5 carry over here—namely each of these is PSD with ones on the diagonal. Thus we have shown that if the level-m Path Statistics SDP is feasible, then for some constant C,

$$\sup_{Y \succeq 0, \operatorname{tr} Y = n, Y_{i,i} \leqslant C} |\langle Y, \overline{A}_G^{(m)} \rangle| \geqslant |d\lambda_2|^s \cdot n - O(\delta)n,$$

(where the constant in the $O(\delta)$ may be taken as the quantity $\|F\|^2\sqrt{k}\delta$ above). In particular this is true when $\boldsymbol{G} \sim \mathcal{P}$. On the other hand, we will prove the following upper bound on this quantity when $\boldsymbol{G}$ is drawn from the null model.

Theorem 6.4. *Let $\boldsymbol{G} \sim \mathcal{N}$. Then for any $\epsilon, C > 0$ there exists $m \in \mathbb{N}$ so that with high probability*

$$\sup_{Y \succeq 0, \operatorname{tr} Y = n, Y_{i,i} \leqslant C} |\langle Y, \overline{A}_G^{(m)} \rangle| \leqslant ((1+\epsilon)d)^{m/2} n.$$

This, and the preceding dicsussion, prove one half of Theorem 6.3, namely that whenever $\lambda_2^2 d > 1$, there are some $m = O(1)$ and $\delta > 0$ for which the level m Path Statistics SDP is with high probability infeasible on inout $\boldsymbol{G} \sim$, but feasible on input $\boldsymbol{G} \sim \mathcal{P}$. We will prove the other half in Section 7. Theorem 2.2, the analogous statement to Theorem 6.3 for the full local statistics algorithm, follows from a final observation:

Observation 6.5. With high probability over $\boldsymbol{G} \sim \mathcal{N}$, if the level-$m$ Path Statistics SDP is infeasible at error tolerance δ, then the $\mathrm{LoSt}(2, m)$ SDP at some error tolerance $\delta'(\delta)$ is infeasible as well.

Proof. The quadratic block of the $\mathrm{LoSt}(2, m)$ SDP concerns $nk \times nk$ matrices, and includes all hard constraints—bounds on diagonal entries, trace of diagonal blocks—present in the Path Statistics

SDP. Moreover, it has access to affine constraints involving the counts of subgraphs with at most m edges. Since the entries of $\overline{A}_G^{(s)}$ for $s \leqslant m$ are simply linear combinations of such counts, LoSt$(2, m)$ has access to the affine constraints from the Local Path Statistics SDP as well. □

The promised robustness guarantee in Theorem 2.2 can be achieved by choosing B as per Theorem D.4, deleting edges incident to all vertices of degree $> B$, and inputting the resulting graph into the LoSt$(2, m)$ SDP.

6.1 Local Statistics in the SBM

We pause to compute the local statistics of the SBM; by setting $k = 1$ and $M = 1$, we recover analogous results for the Erdős-Rényi model. Recall that for a partially subgraph (H, S, τ),

$$L_{(H,S,\tau)}(M, \pi) \triangleq \sum_{\hat{\tau}:\hat{\tau}|_S=\tau} \prod_{v \in v(H)} \pi(\hat{\tau}(v)) \prod_{(u,v)\in E(H)} M_{\hat{\tau}(u),\hat{\tau}(v)}.$$

Theorem 6.6. *Let (H, S, τ) be a partially labelled graph with $O(1)$ edges and ℓ connected components. Then with high probability*

$$p_{(H,S,\tau)}(\boldsymbol{G}, \boldsymbol{x}) = n^{\chi(H)} L_{(H,S,\tau)}(M, \pi) \cdot d^{|E(H)|} n^{|V(H)|-|E(H)|} + o(n^{c(H)})$$

Proof. Fix (H, S, τ). There are

$$\binom{n}{|V(H)|} |V(H)|! = n^{|V(H)|} + O(n^{|V(H)|-1})$$

injective maps from $V(H) \hookrightarrow [n]$. The probability that each is an occurrence, once we condition on the labels σ of the relevant vertices, is given by $\prod_{(u,v)\in E(H)} M_{\sigma(u),\sigma(v)} d/n$. The probability of each labelling σ is $\prod_{u \in V(H)} \pi(\sigma(u))$, and we only consider labellings that agree with τ at the relevant vertices. Thus

$$\mathbb{E}\, p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G}) = n^{\chi(H)} L_{(H,S,\tau)}(M, \pi) + O(n^{\chi(H)-1}).$$

and one immediately sees that this expectation decomposes as a product of analogous expectations over the connected components of H.

To prove concentration, in the case when H has at least one cycle, $c(H) > \chi(H)$ and the assertion follows from Markov. Otherwise, let us consider $\mathbb{E}\, p_{H,S,\tau,U,W}(\boldsymbol{G}, \boldsymbol{x})^2$. This is a sum over pairs of maps $\phi, \psi : V(H) \to [n]$, and as H is acyclic, it is dominated by terms in which the images of these two maps are disjoint. Thus, to leading order, this variance is equal to the expected number of occurrences of two disjoint copies of (H, S, τ), which we just observed is $(\mathbb{E}\, p_{H,S,\tau,U,W}(\boldsymbol{G}, \boldsymbol{x}))^2$ to leading order. We finish by using Chebyshev and observing that $\chi(H) = c(H)$. □

6.2 Proof of Theorem 6.4

The main challenge in studying $\overline{A}_{\boldsymbol{G}}^{(m)}$, when $\boldsymbol{G}$ is a sparse Erdős-Rényi random graph, is the presence of of certain localized combinatorial structures which inflate the number of non-backtracking walks: high-degree vertices and small subgraphs with many cycles. Our strategy will be to decompose $\overline{A}_{\boldsymbol{G}}^{(s)}$ as a sum of two matrices, one of which "avoids" these structures and admits spectral norm bounds, and the other of which has a small entrywise L_1 norm. Let us make this precise. In any graph G, write $\mathsf{B}_t(v, G)$ for the set of vertices with distance at most t from v; call v (t,ε)-*heavy* if $|\mathsf{B}_t(v, G)| \geqslant (1+\epsilon)^t d^t$. We will call a vertex v (t, r, ε)-*vexing* if either it participates in a cycle of length less than r or it is (t,ε)-heavy. Let $\boldsymbol{H}$ be the subgraph obtained by deleting every vexing vertex, and write

$$\left(\overline{A}_{\boldsymbol{H}}^{(m)}\right)_{u,v} = \sum_{\gamma: u\to v, |\gamma|=s, \gamma\in V(\boldsymbol{H})} w_{\boldsymbol{G}}(\gamma).$$

We will also refer to $\boldsymbol{H}$ as the (t, r, ε)-*truncation of* G. In the sequel, we assume $r = \Theta\left(\frac{\log n}{(\log\log n)^2}\right)$. Then Theorem 6.4 is an immediate consequence of the following two results.

Theorem 6.7 (Truncated Spectral Norm Bound)**.** *For every $\epsilon > 0$, there exist t, m satisfying $m = t^3$ so that with high probability*

$$\|\overline{A}_{\boldsymbol{H}}^{(m)}\| \leqslant ((1+\epsilon)d)^{m/2}.$$

Proposition 6.8 (L_1 Bound)**.** *For every $\delta > 0$, and every $r = O(1)$, for any $t \geqslant \Omega(\frac{\log m - \log\delta}{\log(1+\epsilon)})$ so that with high probability*

$$\|\overline{A}_{\boldsymbol{G}}^{(m)} - \overline{A}_{\boldsymbol{H}}^{(m)}\|_1 \leqslant \delta n.$$

With these two results in hand, Theorem 6.4 quickly follows: any matrix $Y \succeq 0$ with unit diagonal satisfies $|Y_{u,v}| \leqslant 1$, so

$$\begin{aligned}
|\langle Y, \overline{A}_{\boldsymbol{G}}^{(m)}\rangle| &\leqslant |\langle Y, \overline{A}_{\boldsymbol{H}}^{(m)}\rangle| + |\langle Y, \overline{A}_{\boldsymbol{G}}^{(m)} - \overline{A}_{\boldsymbol{H}}^{(m)}\rangle| \\
&\leqslant n\|\overline{A}_{\boldsymbol{H}}^{(m)}\| + \|\overline{A}_{\boldsymbol{G}}^{(m)} - \overline{A}_{\boldsymbol{H}}^{(m)}\|_1 \\
&\leqslant \left(((1+\epsilon)d)^{m/2} + \delta\right) n.
\end{aligned}$$

Theorem 6.7 is the heavier technical lift, so we will warm up with the proof of Proposition 6.8. The proof of Theorem 6.7 is deferred to Section 6.3.

6.2.1 Proof of Proposition 6.8

For a non-backtracking walk γ on the complete graph, write $\mathcal{V}(\gamma)$ for the event that γ visits a vexing vertex. Then

$$\|\overline{A}_{\boldsymbol{G}}^{(m)} - \overline{A}_{\boldsymbol{H}}^{(m)}\|_1 \leqslant \sum_{\gamma\in K_n, |\gamma|=m} |w_{\boldsymbol{A}}(\gamma)| \mathbf{1}\left[\mathcal{V}(\gamma)\right] \leqslant \sum_{\gamma\in K_n, |\gamma|=m} (d/n)^{\#\text{ non-edges}} \mathbf{1}\left[\mathcal{V}(\gamma)\right].$$

Once we choose $\boldsymbol{G}$, γ alternates between segments of edges on $\boldsymbol{G}$ and segments of non-edges. We do not lose too much by relaxing slightly the condition that γ is non-backtracking, instead asking only that it is non-backtracking whenever it walks on $\boldsymbol{G}$.

Let us define an *m-scribble* $\mathfrak{s}$ with type $(p_1|q_1|\cdots|p_l|q_l)$ on the complete graph to be a path comprised of l non-backtracking segments of lengths $p_1, ..., p_l$ interspersed with l 'free' segments of lengths $q_1, ..., q_l$. We require that $\sum p_i + \sum q_i = m$, and all but perhaps p_1, q_l are strictly positive. Define $w(\mathfrak{s}) = (d/n)^{\sum q_i}$, and let us write $\mathfrak{s} \subset \boldsymbol{G}$ to mean that every non-backtracking segment of $\mathfrak{s}$ appears in $\boldsymbol{G}$. We will call a scribble *vexing* and write $\mathcal{V}(\mathfrak{s})$, if any of the vertices of $\mathfrak{s}$ is vexing. In view of the preceding paragraph, it suffices to bound

$$\sum_{\mathfrak{s} \in K_n, |\mathfrak{s}|=m} w(\mathfrak{s}) \mathbf{1}\left[\mathfrak{s} \subset \boldsymbol{G}\right] \mathbf{1}\left[\mathcal{V}(\mathfrak{s})\right].$$

We will divide the event $\mathcal{V}(\mathfrak{s})$ that $\mathfrak{s}$ is vexing into two subcases: write $\mathcal{H}(\mathfrak{s})$ if $\mathfrak{s}$ contains a heavy vertex, and $\mathcal{C}(\mathfrak{s})$ if it ever encounters a vertex on a cycle of length at most r.

We will need the following simplified version of the forthcoming Lemma C.8.

Lemma 6.9. *Let $\Gamma \subset K_n$, and write $\Gamma \subset \boldsymbol{G}$ to mean that every edge of Γ appears in $\boldsymbol{G}$. Then there exist universal C, c so that*

$$\mathbb{P}[\Gamma \subset \boldsymbol{G} \text{ and contains a } (t,\varepsilon)\text{-heavy vertex}] \leqslant \mathbb{P}[\Gamma \subset \boldsymbol{G}] \cdot |V(\Gamma)| \cdot C \exp\left(-\frac{c}{1+|V(\Gamma)|}(1+\varepsilon)^t\right)$$

Proof. Write $\boldsymbol{G}^c$ for the graph obtained by removing every one of Γ's edges, and write $\mathsf{B}_t(v, \boldsymbol{G}^c)$ for the t-neighborhood of a vertex in this modified graph. We claim that if $\Gamma \subset \boldsymbol{G}$ and one of its vertices is (t,ε)-heavy, then one of its vertices is (t,ε')-heavy in $\boldsymbol{G}^c$, where $\varepsilon' = (1+\varepsilon)(1+|V(\Gamma)|)^{-1/t} - 1$. Assume that v is the heavy vertex in $\boldsymbol{G}$, noting that

$$|B_t(v, G)| \leqslant |B_t(v, \boldsymbol{G}^c)| + |B_t(V(\Gamma), \boldsymbol{G}^c)|$$

by dividing the shortest paths of length t emanating from v according to whether they use edges from Γ or not. Since v is (t,ε)-heavy, the left hand side is at least $(1+\epsilon)^t d^t$. If for some ε' no other vertex in Γ is (t,ε')-heavy in $\boldsymbol{G}^c$, then $|B_t(V(\Gamma), \boldsymbol{G}^c)| \leqslant |V(\Gamma)|(1+\varepsilon')^t d^t$, and we conclude that $|B_t(v, \boldsymbol{G}^c)| \geqslant (1+\varepsilon)^t(1-\varepsilon^t|V(\Gamma)|)$, which is a contradiction if ε' is set as in the theorem statement.

Thus we have shown that the event we care about is contained in the intersection of two independent ones: that $\Gamma \subset \boldsymbol{G}$, and that there exists a vertex in Γ that is (t,ϵ')-heavy in $\boldsymbol{G}^c$. We can bound this second probability by taking a union bound over all vertices in Γ, and noting that the probability of being heavy in $\boldsymbol{G}^c$ is at most the probability of being heavy in $\boldsymbol{G}$. From Lemma C.8, the probability that a given vertex is (t,ϵ')-heavy in $\boldsymbol{G}$ is, for some universal C, c, at most $C\exp(-c(1+\epsilon')^t) \leqslant C\exp(-\frac{c}{|V(\Gamma)|+1}(1+\epsilon)^t)$. We then execute the union bound and assemble everything. □

With this lemma in hand, and using the fact that $\mathfrak{s}$ contains at most $m+1$ vertices,

$$\mathbb{P}[\mathfrak{s} \subset \boldsymbol{G}, \mathcal{H}_N(\mathfrak{s})] \leqslant \mathbb{P}[\mathfrak{s} \subset \boldsymbol{G}] C(m+1) \exp\left(-\frac{c}{m+2}(1+\epsilon)^t\right) \triangleq \mathbb{P}[\mathfrak{s} \subset \boldsymbol{G}] \Upsilon(m,\varepsilon).$$

Thus

$$\mathbb{E} \sum_{\mathfrak{s} \in K_n, |\mathfrak{s}|=m} w(\mathfrak{s}) \mathbf{1}\left[\mathfrak{s} \subset \boldsymbol{G}\right] \mathbf{1}\left[\mathcal{H}(\mathfrak{s})\right] \leqslant \Upsilon(m,\varepsilon)\, \mathbb{E} \sum_{\mathfrak{s} \in K_n |\mathfrak{s}|=m} w(\mathfrak{s}) \mathbf{1}\left[\mathfrak{s} \subset \boldsymbol{G}\right].$$

We need to perform a similar calculation for the scribbles which visit a vertex on a cycle. Fixing $\mathfrak{s}$, if $\mathfrak{s}$ itself contains a cycle, then $\mathbb{P}[\mathfrak{s} \subset \boldsymbol{G}, \mathcal{C}(\mathfrak{s})] = \boldsymbol{b}[\mathfrak{s} \subset \boldsymbol{G}]$. Otherwise, $\mathfrak{s}$ does not contain a cycle, and there must be a path of length at most r, using no edges in $\mathfrak{s}$, that connects two of its vertices. This event is independent of the event $\mathfrak{s} \subset \boldsymbol{G}$; for any fixed length s, there are at most n^{s-1} such paths, and each occurs with probability $O(n^{-s})$, meaning that the total probability is bounded by $O(r/n)$.

Combining all of this,

$$\begin{aligned}\mathbb{E} \sum_{\mathfrak{s} \in K_n, |\mathfrak{s}|=m} w(\mathfrak{s}) \mathbf{1}\left[\mathfrak{s} \subset \boldsymbol{G}\right] \mathbf{1}\left[\mathcal{V}(\mathfrak{s})\right] \leqslant (\Upsilon(m,\varepsilon) + O(r/n))\, \mathbb{E} \sum_{\mathfrak{s} \in K_n, |\mathfrak{s}|=m} w(\mathfrak{s}) \mathbf{1}\left[\mathfrak{s} \subset \boldsymbol{G}\right] \\ + \mathbb{E} \sum_{\mathfrak{s} \in K_n, |\mathfrak{s}|=m} w(\mathfrak{s}) \mathbf{1}\left[\mathfrak{s} \subset \boldsymbol{G}\right] \mathbf{1}\left[\mathfrak{s} \text{ contains a cycle}\right] \\ + \text{lower order terms.}\end{aligned}$$

To compute term in the second line, fix a scribble of type $(p_1|q_2|\cdots|q_l)$. To choose a scribble with this type in $\boldsymbol{G}$, one needs to select a subgraph in $\boldsymbol{G}$ with at most l connected components, at least one of which contains a cycle of length at most m. In expectation there are $o(n^{l-1})$ of these. For each of $q_1, ..., q_{l-1}$, there are $q_i - 1$ choices of a free vertex, and we pay a weight of $O(n^{-q_i})$; for q_l, if it is nonzero, there are q_l free vertices at a cost of $O(n^{-q_l})$. Thus the final term, the expected, weighted counts of scribbles that contain a cycle, contributes $o(n)$.

It therefore remains only to compute the expected weighted sum of all m-scribbles in $\boldsymbol{G}$. Analogous to the previous paragraph, to choose a scribble of type $(p_1|q_1|\cdots|p_l|q_l)$ in $\boldsymbol{G}$, one first selects a tuple of non-backtracking walks in $\boldsymbol{G}$ with lengths $p_1, ..., p_l$, and then connects them with free segments. In expectation there are $d^{p_1+\cdots p_l} n^l + O(n^{l-1})$ such tuples of walks in $\boldsymbol{G}$. For each of $q_1, ..., q_{l-1}$, there are $q_i - 1$ choices of a free vertex, and we pay a weight $(d/n)^{q_i}$; for q_l, if it exists, there are q_l free vertices at a cost of $(d/n)^{q_l} = d^{q_l}$. There are at most 2^{m+1} types, giving

$$\mathbb{E} \sum_{\mathfrak{s} \in K_n, |\mathfrak{s}|=m} w(\mathfrak{s}) \mathbf{1}\left[\mathfrak{s} \subset \boldsymbol{G}\right] \leqslant 2(2d)^m n + O(1).$$

Having computed its expectation, we now need to show that the number of vexing scribbles is concentrated. We begin by recalling the well-known Efron-Stein inequality.

Lemma 6.10. *Let $(Y, X_1, X_2, ..., X_T)$ be i.i.d. real random variables. Then for any function $f : \mathbb{R}^T \to \mathbb{R}$,*

$$\mathbf{Var} f(X_1, ..., X_T) \leqslant \frac{1}{2} \sum_{u \in [T]} \mathbb{E}\left[(f(X_1, ..., X_u, ..., X_T) - f(X_1, ..., Y, ..., X_T))^2 \right].$$

We will apply this to the function f that counts weighted, vexing scribbles. Let $\boldsymbol{G}$ be an ER random graph, and $\tilde{\boldsymbol{G}}$ be the same graph with some edge re-randomized. With probability $1 - 2d/n$, the graphs $\boldsymbol{G} = \tilde{\boldsymbol{G}}$ and the weighted scribble counts are the same. With the remaining probability, we are comparing the weighted, vexing scribble counts on two graphs that differ at an edge. Since the addition of an edge can only make more vertices vexing, the count can only increase; thus we can clumsily bound the difference by the total number of scribbles (vexing or not) that use the added edge.

Fact 6.11. *Let $\boldsymbol{G} \sim \mathcal{N}$. Then the probability that there is a vertex with degree larger than Δd is at most $n(e/\Delta)^{d\Delta}$. In particular, setting $\Delta = 2\log n$, this probability is $o(n^{-c})$ for every c.*

In a graph with maximum degree Δ, the weighted sum of m-scribbles with the property that at least one of the non-backtracking segments uses a given edge is upper bounded by $m(2\Delta)^m$, so let us split into the events that the maximum degree in $\boldsymbol{G}$ is less than vs. greater than $\log n$. On the first event, whose probability we will upper bound by 1, we get $m(2\log n)^m$ weighted scribbles. The second event gives us at most $m(2n)^m$ weighted scribbles, has probability better than any inverse polynomial in n. Thus by Efron-Stein inequality the variance of the number of weighted scribbles is at most

$$2(d/n) \cdot \binom{n}{2} \left(m^2 (2\log n)^{2m} + o(1) \right) = O(n \log^{2m} n),$$

so we get concentration of $O(\sqrt{n}\log^m n)$.

All told, then, we have that with high probability

$$\|\overline{A}_{\boldsymbol{G}}^{(m)} - \overline{A}_{\boldsymbol{H}}^{(m)}\|_1 \leqslant (m+1)C \exp\left(-\frac{c}{m+2}(1+\epsilon)^t \right) \cdot 2(2d)^m n + O(\sqrt{n}\log^{m+1} n).$$

To make this smaller than $\delta n + o(n)$, it suffices to set $t = \Omega(\frac{\log m - \log \delta}{\log(1+\epsilon)})$.

6.3 Spectral norm bounds

In this section, we prove Theorem 6.7.

6.3.1 Setup

Choosing parameters. Let ε and d be constants given to us. With the privilege of hindsight, we choose a small constant $\delta < \frac{1}{100\sqrt{(1+\varepsilon)d}}$; t to be a large enough integer (depending on ε, d and δ) so that:

1. the hypothesis of Theorem C.1 holds on parameters $d' := (1+\varepsilon)d, d$ and δ,
2. $((1+\varepsilon)d)^{1/t^2} t^{30/t^3} < 1+\varepsilon$,
3. $\delta^{-24/t} < 1+\varepsilon$,

$\ell := t^3$; k is any even integer in $\left[\frac{\log n \log\log n}{2\ell}, \frac{4\log n \log\log n}{\ell}\right]$; and $r := \frac{k\ell}{\ln^3(k\ell)}$. Observe that since t is constant, $r = O\left(\frac{\log n}{(\log\log n)^2}\right)$.

Let $\boldsymbol{G}$ be an Erdős-Renyi $G(n, d/n)$ graph, let $\boldsymbol{S}$ its the set of (t,r,ε)-vexing vertices, and let $\boldsymbol{G}_{t,r,\varepsilon}$ be the (t,r,ε)-truncation of $\boldsymbol{G}$. Let $\boldsymbol{A}$ be the adjacency matrix of $\boldsymbol{G}_{t,\varepsilon,r}$. Define

$$\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus S}1_{[n]\setminus S}^\top\right)^{(\ell)}[u,v] = \sum_{\substack{W \text{ length-}\ell \text{ nonbacktracking walk} \\ \text{from } u \text{ to } v \text{ in } K_{[n]\setminus S}}} \prod_{ij\in W}\left(\boldsymbol{A} - \frac{d}{n}11^\top\right)[i,j]$$

We are interested in obtaining bounds on the spectral norm of $\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus S}1_{[n]\setminus S}^\mathsf{T}\right)^{(\ell)}$, and towards doing so we employ the trace method. In particular, we prove:

Theorem 6.12. *With probability* $1 - n^{-100}$, $\left\|\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus S}1_{[n]\setminus S}^\mathsf{T}\right)^{(\ell)}\right\| \leqslant \left((1+\varepsilon)^4\sqrt{d}\right)^\ell$.

We will obtain spectral norm bounds on $\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus S}1_{[n]\setminus S}^\mathsf{T}\right)^{(\ell)}$ that hold with high probability by achieving high probability bounds on

$$\boldsymbol{F} := \operatorname{tr}\left(\left(\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus S}1_{[n]\setminus S}^\mathsf{T}\right)^{(\ell)}\right)^{2k}\right).$$

When $\boldsymbol{F}$ is bounded by R,

$$\left\|\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus S}1_{[n]\setminus S}^\mathsf{T}\right)^{(\ell)}\right\| \leqslant R^{\frac{1}{2k}}.$$

6.3.2 From High Trace to Counting

We borrow some more terminology from [MOP19b]:

Definition 6.13 (Linkages)**.** We call a closed walk of length $k\ell$ on K_n a $(k\times\ell)$-*linkage* if it can be split into k segments each of length-ℓ such that W is nonbacktracking on each segment. We refer to each such length-ℓ nonbacktracking segment as a *link*. We use $V(W)$ to denote the vertices visited by W and $E(W)$ to denote the (undirected) edges visited by W.

Within a linkage W, we use $a_{ij}(W)$ to denote the number of times the *undirected* edge $\{i,j\}$ is walked on (which in this exposition we will simply abbreviate to a_{ij}), $S(W)$ to denote the set of

singleton edges in $E(W)$, i.e. all edges $\{i,j\}$ such that $a_{ij} = 1$, and $D(W)$ to denote all the remaining edges (each of which has $a_{ij} \geqslant 2$), which we call *duplicative edges*. We use $e(W)$ to denote the "excess" number of edges in W, i.e., $e(W) = |E(W)| - |V(W)| - 1$ Finally, let $\mathcal{E}(W)$ denote the event that $V(W) \cap \boldsymbol{S}$ is empty. We will call a subset of edges E' *good* if there are no (t, r, ε)-vertices in the graph induced by E'. We have,

$$\begin{aligned}
\boldsymbol{F} &= \sum_{W \text{ is } (k \times \ell)\text{-linkage of } K_n} \prod_{ij \in W} \left(\boldsymbol{A}[i,j] - \frac{d}{n} \right) \cdot \mathbf{1}[\mathcal{E}(W)] \\
&= \sum_{W \text{ is } (k \times \ell)\text{-linkage of } K_n} \prod_{ij \in S(W)} \left(\boldsymbol{A}[i,j] - \frac{d}{n} \right) \prod_{ij \in D(W)} \left(\boldsymbol{A}[i,j] - \frac{d}{n} \right)^{a_{ij}(W)} \mathbf{1}[\mathcal{E}(W)] \qquad (22)
\end{aligned}$$

We write one of the terms in the above expression in a more convenient form:

$$\begin{aligned}
\left(\boldsymbol{A}[i,j] - \frac{d}{n} \right)^{a_{ij}(W)} &= \sum_{t=0}^{a_{ij}(W)} \boldsymbol{A}[i,j]^t \left(-\frac{d}{n} \right)^{a_{ij}(W)-t} \cdot \binom{a_{ij}(W)}{t} \\
&= \boldsymbol{A}[i,j] \sum_{i=1}^{a_{ij}(W)} \left(-\frac{d}{n} \right)^{a_{ij}(W)-t} \cdot \binom{a_{ij}(W)}{t} + \left(-\frac{d}{n} \right)^{a_{ij}(W)} \\
&= \boldsymbol{A}[i,j] \left(\left(1 - \frac{d}{n} \right)^{a_{ij}(W)} - \left(-\frac{d}{n} \right)^{a_{ij}(W)} \right) + \left(-\frac{d}{n} \right)^{a_{ij}(W)}.
\end{aligned}$$

Writing $\gamma_{ij} = \left(1 - \frac{d}{n}\right)^{a_{ij}(W)} - \left(-\frac{d}{n}\right)^{a_{ij}(W)}$, we can rewrite (22) as

$$\sum_{W \text{ is } (k \times \ell)\text{-linkage of } K_n} \prod_{ij \in S(W)} \left(\boldsymbol{A}[i,j] - \frac{d}{n} \right) \prod_{ij \in D(W)} \left(\boldsymbol{A}[i,j] \gamma_{ij} + \left(-\frac{d}{n} \right)^{a_{ij}(W)} \right) \mathbf{1}[\mathcal{E}(W)]$$

In the subsequent steps we will use $\sum_W$ as short for $\sum_{W \text{ is } (k \times \ell)\text{-linkage of } K_n}$. Thus, we have:

$$\boldsymbol{F} = \sum_W \prod_{ij \in S(W)} \left(A[i,j] - \frac{d}{n} \right) \sum_{L \subseteq D(W)} \prod_{ij \in L} A[i,j] \gamma_{ij} \prod_{ij \notin L} \left(-\frac{d}{n} \right)^{a_{ij}(W)} \mathbf{1}[\mathcal{E}(W)]$$

where $ij \notin L$ actually means $ij \in D(W) \setminus L$. We are interested in bounding $|\,\mathbb{E}[\boldsymbol{F}]|$. We first point out that $\gamma_{ij} \leqslant 1$ for large enough n. Then:

$$\begin{aligned}
|\,\mathbb{E}[\boldsymbol{F}]| &= \left| \mathbb{E} \left[\sum_W \prod_{ij \in S(W)} \left(A[i,j] - \frac{d}{n} \right) \sum_{L \subseteq D(W)} \prod_{ij \in L} A[i,j] \gamma_{ij} \prod_{ij \notin L} \left(-\frac{d}{n} \right)^{a_{ij}(W)} \mathbf{1}[\mathcal{E}(W)] \right] \right| \\
&= \left| \sum_W \sum_{L \subseteq D(W)} \prod_{ij \in L} \gamma_{ij} \prod_{ij \notin L} \left(-\frac{d}{n} \right)^{a_{ij}(W)} \mathbb{E} \left[\prod_{ij \in S(W)} \left(A[i,j] - \frac{d}{n} \right) \prod_{ij \in L} A[i,j] \mathbf{1}[\mathcal{E}(W)] \right] \right| \\
&\leqslant \sum_W \sum_{L \subseteq D(W)} \prod_{ij \in L} \gamma_{ij} \prod_{ij \notin L} \left(\frac{d}{n} \right)^{a_{ij}(W)} \left| \mathbb{E} \left[\prod_{ij \in S(W)} \left(A[i,j] - \frac{d}{n} \right) \prod_{ij \in L} A[i,j] \mathbf{1}[\mathcal{E}(W)] \right] \right|
\end{aligned}$$

By Theorem C.1:

$$\begin{aligned}
&\leqslant \sum_{W} \sum_{\substack{L \subseteq D(W) \\ L \text{ good}}} \prod_{ij \notin L} \left(\frac{d}{n}\right)^{a_{ij}(W)} \cdot C \log^2 n \cdot \left(\frac{d}{n}\right)^{|S(W) \cup L|} \cdot n^{8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt} \\
&\leqslant \sum_{W} \sum_{\substack{L \subseteq D(W) \\ L \text{ good}}} \prod_{ij \notin L} \left(\frac{d}{n}\right)^{a_{ij}(W)-1} \cdot C \log^2 n \cdot \left(\frac{d}{n}\right)^{|S(W) \cup L|} \cdot \left(\frac{d}{n}\right)^{|D(W)|-L} \cdot \\
&\qquad\qquad n^{8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt} \\
&\leqslant C(n) \sum_{W} \left(\sum_{\substack{L \subseteq D(W) \\ L \text{ good}}} \prod_{ij \notin L} \left(\frac{d}{n}\right)^{a_{ij}(W)-1} \right) \cdot \left(\frac{d}{n}\right)^{|S(W)|+|D(W)|} \cdot n^{8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt}
\end{aligned} \tag{23}$$

where $C(n) = C \log^2 n$. Now we analyze

$$\sum_{\substack{L \subseteq D(W) \\ L \text{ good}}} \prod_{ij \notin L} \left(\frac{d}{n}\right)^{a_{ij}(W)-1} .$$

Call the *weight* of a subset L of $D(W)$ as $w(L) \coloneqq \sum_{ij \in L} (a_{ij}(W) - 1)$. Let $D^*(W)$ be a maximum weight good subset of $D(W)$, and define $\Delta(W)$ as $w(D(W)) - w(D^*(W))$. We say $\Delta(W)$ is the number of *profligate steps* in the graph. Then:

$$\sum_{\substack{L \subseteq D(W) \\ L \text{ good}}} \prod_{ij \notin L} \left(\frac{d}{n}\right)^{a_{ij}(W)-1} = \sum_{\substack{L \subseteq W \\ L \text{ good}}} \left(\frac{d}{n}\right)^{w(D(W))-w(L)}$$

Since $a_{ij}(W)$ for every edge is at least 2, we can bound the above by:

$$\begin{aligned}
&\leqslant \sum_{L \subseteq W} \left(\frac{d}{n}\right)^{\max\{|D(W)|-|L|,\, \Delta(W)\}} \\
&= \sum_{\eta \leqslant \Delta(W)} \left(\frac{d}{n}\right)^{\Delta(W)} \cdot \binom{|D(W)|}{\eta} + \sum_{\eta > \Delta(W)} \left(\frac{d}{n}\right)^{\eta} \cdot \binom{|D(W)|}{\eta} \\
&\leqslant (\Delta(W)+1) \left(\frac{d|D(W)|}{n}\right)^{\Delta(W)} + \sum_{\eta > \Delta(W)} \left(\frac{d|D(W)|}{n}\right)^{\eta} \\
&\leqslant (\Delta(W)+2) \left(\frac{d|D(W)|}{n}\right)^{\Delta(W)} \\
&\leqslant 2 \left(\frac{2d|D(W)|}{n}\right)^{\Delta(W)} .
\end{aligned}$$

Plugging the above back into (23) and "absorbing" a factor of 2 into $C(n)$ tells us:

$$(23) \leqslant C(n) \sum_W \left(\frac{2d|D(W)|}{n}\right)^{\Delta(W)} \cdot \left(\frac{d}{n}\right)^{|S(W)|+|D(W)|} \cdot n^{8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt}. \tag{24}$$

We will split the above sum based on properties of the walk (such as $|S(W)|, e(W), \Delta(W), |V(W)|$) and count the number of terms in each split part using an encoding argument. Before we get into the counting argument, we make a key definition:

Definition 6.14. We say a step from u to v in a linkage W is *fresh* if v was never visited earlier in W. We will use $f(W)$ to denote the number of fresh steps in W.

Remark 6.15. For a linkage W, $|V(W)| = f(W) + 1$.

A consequence of Remark 6.15 along with the fact that $|S(W)| + |D(W)| = |E(W)|$, we get that $|E(W)| = f(W) + e(W)$. Thus, (24) is bounded by

$$C(n) \sum_W \left(\frac{2d|D(W)|}{n}\right)^{\Delta(W)} \cdot \left(\frac{d}{n}\right)^{e(W)+f(W)} \cdot n^{8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt}. \tag{25}$$

To complete the proof, we need the following, which is proved in Section 6.4:

Theorem 6.16. *The total number of (k, ℓ)-linkages with f fresh edges, e excess edges, s singleton edges and Δ profligate steps is at most:*

$$n^{f+1} \cdot (4\lambda(W))^{7\lambda(W)+1} \cdot (k\ell)^{3\lambda(W)+1} \cdot (\ell+1)^{6k} \cdot ((1+\varepsilon)d)^{tk+k\ell/2-|D(W)|-s/2}$$

where $\lambda(W) \leqslant 3e + \frac{12k\ell \ln(k\ell)}{r} + 3\Delta$.

Now recall that we wished to obtain bounds on the following from (25):

$$Q := C(n) \sum_W \left(\frac{2d|D(W)|}{n}\right)^{\Delta(W)} \cdot \left(\frac{d}{n}\right)^{e(W)+f(W)} \cdot n^{8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt}.$$

From Theorem 6.16:

$$\begin{aligned}
Q &\leqslant C(n) \sum_{f,e,\Delta,s\geqslant 0} \left(\frac{2d|D(W)|}{n}\right)^{\Delta} \cdot \left(\frac{d}{n}\right)^{f} \cdot \left(\frac{d}{n^2}\right)^{e} \cdot 4^s \cdot \delta^{s-24kt} \cdot \\
&\quad n^{f+1} \cdot (4\lambda(W))^{7\lambda(W)+1} \cdot (k\ell)^{3\lambda(W)+1} \cdot (\ell+1)^{6k} \cdot ((1+\varepsilon)d)^{tk+k\ell/2-|D(W)|-s/2} \\
&\leqslant C(n) \cdot n \sum_{f,e,\Delta,s\geqslant 0} \left(\frac{2d|D(W)|}{n}\right)^{\Delta} \cdot ((1+\varepsilon)d)^{f+k\ell/2-|D(W)|-s/2} \cdot \left(\frac{d}{n^2}\right)^{e} \cdot (4\delta)^s \cdot \delta^{-24kt} \cdot \\
&\quad (4\lambda(W)k\ell)^{7\lambda(W)+1} \cdot ((\ell+1)^6((1+\varepsilon)d)^t)^k
\end{aligned}$$

Defining $C'(n) := C(n) \cdot n$ and the fact $f = s + |D(W)| - e$, we get:

$$\leqslant C'(n) \sum_{f,e,\Delta,s\geqslant 0} \left(\frac{2d|D(W)|}{n}\right)^{\Delta} \cdot ((1+\varepsilon)d)^{s/2-e+k\ell/2} \cdot \left(\frac{d}{n^{\cdot 2}}\right)^{e} \cdot (4\delta)^{s} \cdot \delta^{-24kt} \cdot$$
$$(4\lambda(W)k\ell)^{7\lambda(W)+1} \cdot ((\ell+1)^6((1+\varepsilon)d)^t)^k$$
$$\leqslant C'(n) \cdot ((1+\varepsilon)d)^{k\ell/2} \sum_{f,e,\Delta\geqslant 0} \left(\frac{2d|D(W)|}{n}\right)^{\Delta} \cdot \left(\frac{1}{n^{\cdot 2}}\right)^{e} \cdot$$
$$\delta^{-24kt} \cdot (4\lambda(W)k\ell)^{7\lambda(W)+1} \cdot ((\ell+1)^6((1+\varepsilon)d)^t)^k \sum_{s\geqslant 0} \left(4\delta\sqrt{(1+\varepsilon)d}\right)^{s}$$

where the inequality above is true since no other term depends on s. By our choice of δ, the summation over s is bounded by 2.

$$\leqslant 2C'(n) \cdot ((1+\varepsilon)d)^{k\ell/2} \cdot ((\ell+1)^6((1+\varepsilon)d)^t)^k \cdot \delta^{-24kt}$$
$$\sum_{f,e,\Delta\geqslant 0} \left(\frac{2d|D(W)|}{n}\right)^{\Delta} \cdot \left(\frac{1}{n^{\cdot 2}}\right)^{e} \cdot (4\lambda(W)k\ell)^{7\lambda(W)+1}$$

By noting that $4\lambda(W)k\ell \leqslant \mathrm{poly}(k,\ell)$:

$$\leqslant 2C'(n) \cdot ((1+\varepsilon)d)^{k\ell/2} \cdot ((\ell+1)^6((1+\varepsilon)d)^t)^k \cdot \delta^{-24kt}$$
$$\sum_{f,e,\Delta} \left(\frac{2d|D(W)| \cdot \mathrm{poly}(k,\ell)}{n}\right)^{\Delta} \cdot \left(\frac{\mathrm{poly}(k,\ell)}{n^{\cdot 2}}\right)^{e} \cdot \mathrm{poly}(k,\ell)^{84k\ell\ln(k\ell)/r}$$
$$\leqslant 8C'(n) \cdot ((1+\varepsilon)d)^{k\ell/2} \cdot ((\ell+1)^6((1+\varepsilon)d)^t)^k \cdot \delta^{-24kt} \cdot \mathrm{poly}(k,\ell)^{84\ln^6(k\ell)} \cdot (k\ell)$$

We know that $\boldsymbol{F}$ is a nonnegative random variable since it is the trace of an even power of a Hermitian matrix, i.e., it is the trace of a positive semidefinite matrix. Hence, by Markov's inequality, we know that except with probability n^{-100}, the random variable $\boldsymbol{F}$ defined in (22) is bounded by

$$n^{100} \cdot 8C'(n) \cdot ((1+\varepsilon)d)^{k\ell/2} \cdot ((\ell+1)^6((1+\varepsilon)d)^t)^k \cdot \delta^{-24kt} \cdot \mathrm{poly}(k,\ell)^{84\ln^6(k\ell)} \cdot (k\ell) \tag{26}$$

By our choice of parameters, the $k\ell$-th root of the above is bounded by:

$$(1+\varepsilon)^4\sqrt{d}.$$

In particular, this means the k-th root of (26) is bounded by $\left((1+\varepsilon)^4\sqrt{d}\right)^{\ell}$ with probability $1 - n^{-100}$.

In summary, we have shown that whp:

$$\mathrm{tr}\left(\left(\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus s}1_{[n]\setminus s}^{\mathsf{T}}\right)^{(\ell)}\right)^{k}\right)^{1/k} \leqslant \left((1+\varepsilon)^4\sqrt{d}\right)^{\ell} \tag{27}$$

thereby establishing:

Theorem 6.17 (Restatement of Theorem 6.12)**.** *With probability* $1 - n^{-100}$*:*

$$\left\| \left(A - \frac{d}{n} 1_{[n]\setminus S} 1_{[n]\setminus S}^{\mathsf{T}} \right)^{(\ell)} \right\| \leqslant \left((1+\varepsilon)^4 \sqrt{d} \right)^{\ell}.$$

6.4 Counting Walks

This section is dedicated to proving Theorem 6.16. Let W be a $(k \times \ell)$-linkage with f fresh steps, e excess edges, s singleton edges, Δ profligate steps. In this section, we will give an efficient encoding of W which will help us upper bound the number of such linkages.

We use $G(W)$, defined to have vertex set $V(W)$ and edge set $E(W)$, to denote the graph of the linkage W. Further, each edge $\{i, j\}$ has a weight a_{ij}, which is the number of times edge $\{i, j\}$ is walked on in W. We can write $E(W)$ as the disjoint union $S(W) \cup D(W)$ where $S(W)$ is the set of singleton edges and $D(W)$ is the set of duplicative edges. Let $D^*(W)$ denote a maximum weight subset of $D(W)$ such that no vertices in the graph induced by those edges on vertices in W are (t, r, ε)-vexing within W.

Remark 6.18. For any set of edges E', we will use $V(E')$ to denote the set of endpoints of edges in E'.

Remark 6.19. $|D(W)| = |D^*(W)| + \Delta$ and $|E(W)| = |D^*(W)| + \Delta + s$.

Remark 6.20. For any subset of edges $E' \subseteq E(W)$, the total number of times W walks on an edge in E' is given by

$$\sum_{ij \in E'} a_{ij}.$$

A special case of the above is that when $E' = E(W)$, the sum above is equal to $k\ell$.

Our next step is to prove:

Claim 6.21. There is a spanning forest F of $D^*(W) \cup S(W)$ such that:

1. $\displaystyle\sum_{ij \in D^*(W)\setminus F} a_{ij} \leqslant \frac{4k\ell \ln(k\ell)}{r}$.
2. $|S(W) \setminus F| \leqslant e$.

In service of proving the above claim we will need the following standard fact which can be found in [KV12, Theorem 13.21]:

Fact 6.22. *Let P be the polytope in $\mathbb{R}^{D^*(W)}$ given by the convex hull of indicator vectors of spanning forests of $D^*(W)$. P is also the feasible region of the following linear program:*

$$x \in \mathbb{R}^{D^*(W)}$$

$$\begin{aligned} x &\geqslant 0 \\ \sum_{ij \in R} x_{ij} &\leqslant |V(R)| - 1 \qquad \forall R \subseteq D^*(W). \end{aligned} \tag{28}$$

We additionally also state the following from [MOP19a, Corollary 2.18] which is a consequence of the "irregular Moore bound" of [AHL02]:

Fact 6.23. *Let H be a graph with $v \geqslant 3$ vertices and girth $g \geqslant 20 \ln v$. Then $|E(H)| - v \leqslant \frac{2v \ln v}{g}$.*

Proof of Claim 6.21. Consider the following assignment to the variables of LP (28):

$$\widetilde{x}_{ij} = 1 - \frac{4 \ln(k\ell)}{r}.$$

We will first show that this assignment is indeed feasible for the LP. Recall that we need to show that for every $R \subseteq D^*(W)$, $\sum_{ij \in R} \widetilde{x}_{ij} \leqslant |V(R)| - 1$. The LHS of this expression is simply $|R| \left(1 - \frac{4 \ln(k\ell)}{r}\right)$ so it suffices to prove:

$$|R| \left(1 - \frac{4 \ln(k\ell)}{r}\right) \leqslant |V(R)| - 1 \qquad \forall R \subseteq D^*(W).$$

Case 1: $|V(R)| < r$**.** In this case R is a forest as there R has no cycles of length smaller than r. Since R is a forest $|R| \leqslant |V(R)| - 1$ and hence the above inequality we wish to prove is definitely true.

Case 2: $|V(R)| \geqslant r$**:** Since the girth of $(V(R), R)$ is at least r, by Fact 6.23,

$$\begin{aligned} |R| &\leqslant |V(R)| \left(1 + \frac{2 \ln |V(R)|}{r}\right) \\ \frac{|R|}{1 + \frac{2 \ln |V(R)|}{r}} &\leqslant |V(R)| \\ |R| - |R| \frac{2 \ln |V(R)|}{r} - 1 &\leqslant |V(R)| - 1 \\ |R| - |R| \frac{4 \ln(k\ell)}{r} &\leqslant |V(R)| - 1. \end{aligned}$$

If we augment the linear program (28) with the objective function

$$\max \sum_{ij \in D^*(W)} a_{ij} x_{ij},$$

by Fact 6.22 we know that there is a spanning forest $\widetilde{F}$ such that maximum of the above objective is achieved at the indicator vector of $\widetilde{F}$. Since we showed $\widetilde{x}$ is feasible,

$$\sum_{ij \in \widetilde{F}} a_{ij} \geqslant \sum_{ij \in D^*(W)} a_{ij} \widetilde{x}_{ij}.$$

Subtracting both sides of the above inequality from $\sum_{ij \in D^*(W)} a_{ij}$ yields

$$\sum_{ij \in D^*(W) \setminus \widetilde{F}} a_{ij} \leqslant \sum_{ij \in D^*(W)} a_{ij}(1 - \widetilde{x}_{ij}) \leqslant \frac{4k\ell \ln(k\ell)}{r}.$$

Now we extend $\widetilde{F}$ to a spanning forest F of $D^*(W) \cup S(W)$. Initially, we set $F = \widetilde{F}$ and process edges of $S(W)$ sequentially in an arbitrary order $i_1 j_1, \ldots, i_s j_s$. We add $i_t j_t$ to F if its addition does not create a cycle and "reject" it otherwise. The number of rejected edges is bounded by e and hence F must contain at least $s - e$ edges of S. Thus, $|S(W) \setminus F| \leqslant e$. Furthermore, since $D^*(W) \setminus F = D^*(W) \setminus \widetilde{F}$,

$$\sum_{ij \in D^*(W) \setminus F} a_{ij} \leqslant \frac{4k\ell \ln(k\ell)}{r}.$$

□

Claim 6.21 lends itself to a natural decomposition of the edges of $G(W)$ into *forest edges*, which we denote $F(W)$, and *crossing edges*, which we denote $C(W)$.

Remark 6.24. $C(W)$ can be written as the following natural disjoint union of sets:

$$C(W) = (S(W) \setminus F(W)) \cup (D^*(W) \setminus F(W)) \cup (D(W) \setminus D^*(W)).$$

At a high level, the linkage W breaks into stretches of steps on $F(W)$ between steps on $C(W)$; a large chunk of this section is dedicated to showing how to encode the portions of W on forest edges highly efficiently.

Let's now express the linkage W in terms of the sequence of vertices walked on: in particular $W = w_0 w_1 w_2 \ldots w_{k\ell}$.

Definition 6.25. We call each consecutive pair $w_i w_{i+1}$ a *step*. If the edge $\{w_i, w_{i+1}\}$ is a crossing edge, we call the step $w_i w_{i+1}$ a *crossing step*, and a *forest step* otherwise. We call a maximal contiguous sequence of forest steps a *cruise*.

Remark 6.26. Any $(k \times \ell)$-linkage W can be expressed as

$$W = C_1 s_1 C_2 s_2 \ldots C_{\gamma(W)} s_{\gamma(W)} C_{\gamma(W)+1}$$

where each C_i is a (possibly empty) cruise, each s_i is a crossing step, and $\gamma(W)$ is the number of crossing steps in W.

Next, we wish to bound $\gamma(W)$.

Claim 6.27. $\gamma(W) \leqslant e + \frac{4k\ell \ln(k\ell)}{r} + \Delta$.

Proof. By Remark 6.24:

$$\begin{aligned}
\gamma(W) &= \sum_{ij \in S(W) \setminus F(W)} a_{ij} + \sum_{ij \in D^*(W) \setminus F(W)} a_{ij} + \sum_{ij \in D(W) \setminus D^*(W)} a_{ij} \\
&\leqslant \sum_{ij \in S(W) \setminus F(W)} 1 + \frac{4k\ell \ln(k\ell)}{r} + \Delta && \text{(by Claim 6.21)} \\
&\leqslant e + \frac{4k\ell \ln(k\ell)}{r} + \Delta. && \text{(by Claim 6.21)}
\end{aligned}$$

□

Definition 6.28. We refer to endpoints of edges in $C(W)$ as well as the start/end vertex of W[1] as *terminal vertices*. We use $T(W)$ to refer to the set of terminal vertices of $G(W)$.

Remark 6.29. $|T(W)| \leqslant 2|C(W)| + 1 \leqslant 2\gamma(W) + 1$. We will use $\lambda(W)$ to refer to $2\gamma(W) + 1$.

Remark 6.30. Each cruise starts and ends at terminal vertices.

Definition 6.31 (Skeleton forest)**.** We use $\mathsf{Skel}(F(W))$ to refer to the subforest of $F(W)$ given by the union of paths in $F(W)$ connecting terminal vertices. Formally,

$$\mathsf{Skel}(F(W)) := \bigcup_{\substack{P \text{ path in } F(W) \\ \text{endpoints of } P \text{ in } T(W)}} P.$$

Observation 6.32. Every leaf in $\mathsf{Skel}(F(W))$ is a terminal vertex and hence by Remark 6.29 the number of leaves in $\mathsf{Skel}(F(W))$ is at most $\lambda(W)$.

Goal 1: Encoding $\mathsf{Skel}(F(W))$. Our first goal is to find an efficient encoding of $\mathsf{Skel}(F(W))$. Towards this goal, we first prove the following.

Lemma 6.33. *Let $\Gamma_{L,v}$ be the set of forests with at most L leaves on vertex set $\{1, \ldots, v\}$. There is a subset $Q \subseteq \Gamma_{L,v}$ of at most $(4Lv)^{2L+1}$ forests such that any forest in $\Gamma_{L,v}$ is isomorphic to a forest in Q.*

We will need the following classical graph theory fact called Cayley's formula; a reader can find multiple proofs in [Cas06]:

Fact 6.34. *The number of labeled spanning trees on v vertices is v^{v-2}.*

We will also need the following fact about trees:

Fact 6.35. *Let T be a tree where v_i is the number of vertices of degree i. Then the following are true:*

$$3(v_1 - 2) \geqslant \sum_{i \geqslant 3} i v_i$$

[1] which are the same since W is closed

$$v_1 - 2 \geqslant \sum_{i \geqslant 3} v_i$$

Proof. Using the fact that the sum of degrees in a tree is $2|V(T)| - 2$ we have:

$$\begin{aligned} 2\sum_{i \geqslant 1} v_i - 2 &= \sum_{i \geqslant 1} i v_i \\ v_1 + \sum_{i \geqslant 3} v_i - 2 &= \sum_{i \geqslant 3} i v_i \\ v_1 - 2 &= \sum_{i \geqslant 3} (i-2) v_i \end{aligned} \tag{29}$$

Lower bounding $i - 2$ by 1 in the RHS of (29), it follows that:

$$v_1 - 2 \geqslant \sum_{i \geqslant 3} v_i \tag{30}$$

Adding 2·(30) and (29) gives us:

$$3(v_1 - 2) \geqslant \sum_{i \geqslant 3} i v_i.$$

□

Proof of Lemma 6.33. Let Ξ be any tree in $\Gamma_{L,v}$. If we split $V(\Xi)$ into leaves $V_1(\Xi)$, degree-2 vertices $V_2(\Xi)$, and degree-$\geqslant 3$ vertices $V_{\geqslant 3}(\Xi)$, we have the following from Fact 6.35:

$$|V_1(\Xi)| - 2 \geqslant |V_{\geqslant 3}(\Xi)|.$$

Thus, $|V_{\geqslant 3}(\Xi)| \leqslant L - 2$. Let $\widetilde{\Xi}$ be the weighted tree described in the following way:

> Its vertex set is $[|V_1(\Xi) \cup V_{\geqslant 3}(\Xi)|]$. Let π be an arbitrary bijection from $[|V_1(\Xi) \cup V_{\geqslant 3}(\Xi)|]$ to $V_1(\Xi) \cup V_{\geqslant 3}(\Xi)$. Place an edge between vertices i and j if there is a path between $\pi(i)$ and $\pi(j)$ such that all vertices in between are in $V_2(\Xi)$. The weight of an edge ij in $\widetilde{\Xi}$ is the distance between i and j in Ξ.

Observe that $\widetilde{\Xi}$ has $\widetilde{v} \leqslant 2L$ vertices and the weight of an edge is an integer between 1 and $|V(\Xi)|$. By Cayley's formula (Fact 6.34) the number of labeled spanning trees on $\widetilde{v}$ vertices is at most $(\widetilde{v})^{\widetilde{v}-2}$. Consequently the number of spanning forests on $\widetilde{v}$ vertices is at most $(2\widetilde{v})^{\widetilde{v}}$ (since every spanning tree on $\widetilde{v}$ vertices has $2^{\widetilde{v}-1}$ subforests). Since $\widetilde{v} \leqslant 2L$ and there are at most $2L$ possibilities for $\widetilde{v}$, each labeled spanning forest on vertex set $[\widetilde{v}]$ that can be encoded by a number in $\left[(4L)^{2L+1}\right]$. In particular this gives us a way to encode the edge set of any $\widetilde{\Xi}$ by a number in $\left[(4L)^{2L+1}\right]$.

All the weights of the edges can be encoded by a number in $\left[|V(\Xi)|^{2L}\right]$, and consequently we can encode $\widetilde{\Xi}$ by a number in $\left[(4L|V(\Xi)|)^{2L+1}\right]$. It is possible to reconstruct a forest isomorphic to Ξ from $\widetilde{\Xi}$ and hence our proof is complete. □

Lemma 6.36. $\mathsf{Skel}(F(W))$ *can be encoded by a number in* $\left[(4\lambda(W)k\ell)^{2\lambda(W)+1} \cdot n^{|V(\mathsf{Skel}(F(W)))|}\right]$.

Proof. At a high level, our proof uses Lemma 6.33 to encode an unlabeled version of $\mathsf{Skel}(F(W))$ in $\left[(4\lambda(W)k\ell)^{2\lambda(W)+1}\right]$ bits and encodes labels using a number in $\left[n^{|V(\mathsf{Skel}(F(W)))|}\right]$.

Encoding "unlabeled" version of $\mathsf{Skel}(F(W))$. Let π be an arbitrary function that maps $V(F(W))$ to $\{1,\ldots,|V(F(W))|\}$. Note that the graph $\pi(\mathsf{Skel}(F(W)))$ is isomorphic to $\mathsf{Skel}(F(W))$. By 6.32, Lemma 6.33, and bounding $|V(\mathsf{Skel}(F(W)))|$ by $k\ell$, $\pi(\mathsf{Skel}(F(W)))$ can be encoded (up to isomorphism) by a number in $\left[(4\lambda(W)|V(\mathsf{Skel}(F(W)))|)^{2\lambda(W)+1}\right]$.

Encoding labels of $\mathsf{Skel}(F(W))$. From the encoding of $\pi(\mathsf{Skel}(F(W)))$, we can recover a graph on vertex set $\{1,\ldots,|V(\mathsf{Skel}(F))|\}$ isomorphic to $\mathsf{Skel}(F(W))$, which we call $\phi(\mathsf{Skel}(F(W)))$. We thus encode the map ϕ^{-1} as it is possible to reconstruct $\mathsf{Skel}(F(W))$ from $\phi(\mathsf{Skel}(F(W)))$ and ϕ^{-1}; such a map can be encoded using a number in $\left[n^{|V(\mathsf{Skel}(F(W)))|}\right]$.

Combining the above two encodings proves the lemma. □

Goal 2. Our next goal is to give an encoding of the collection of start and end points of each cruise.

Lemma 6.37. *Given the encoding of $\mathsf{Skel}(F(W))$ from Lemma 6.36, the collection of start and end points of each cruise*

$$\mathcal{C} = (C_1[\text{start}], C_1[\text{end}]), \ldots, (C_{\gamma+1}[\text{start}], C_{\gamma+1}[\text{end}])$$

can be encoded by a number in $\left[(k\ell)^{\lambda(W)}\right]$.

Proof. Let ϕ be the function from the proof of Lemma 6.36. The sequence

$$\Phi = \phi(C_1[\text{start}]), \phi(C_1[\text{end}]), \ldots \phi(C_{\gamma+1}[\text{start}])^2$$

is a sequence of length $\lambda(W)$ of elements in $\{1,\ldots,|V(\mathsf{Skel}(F))|\}$ and $|V(\mathsf{Skel}(F))| \leqslant k\ell$, and hence can be encoded by a number in $\left[(k\ell)^{\lambda(W)}\right]$. $\mathcal{C}$ can be recovered from Φ and ϕ^{-1}, and since the encoding of $\mathsf{Skel}(F(W))$ gives us ϕ^{-1}, so we are done. □

Goal 3: Encoding cruises. Now we move on to encoding cruises. Let C_i be a cruise that starts at terminal vertex t_{start} and ends at terminal vertex t_{end}.

Remark 6.38. There is a unique path between t_{start} and t_{end} in F as follows:

$$v_0 v_1 v_2 \ldots v_p v_{p+1}.$$

where $v_0 = t_{\mathsf{start}}$ and $v_{p+1} = t_{\mathsf{end}}$.

[2] We skip out on $C_{\gamma+1}[\text{end}]$ since it is equal to $C_1[\text{start}]$.

Definition 6.39. Let C_i be a cruise. We say a contiguous subwalk of C_i is a *detour* if it starts and ends at the same vertex.

Claim 6.40. Cruise C_i can be constructed by taking the path from t_{start} to t_{end} as described in Remark 6.38 and inserting at most one detour after each vertex in the path. In particular, C_i can be written in the form

$$C_i = v_0 \dots v_{j_1} \mathsf{Detour}_{i,j_1} \dots v_{j_2} \mathsf{Detour}_{i,j_2} \dots\dots v_{j_b} \mathsf{Detour}_{i,j_b} \dots v_{p+1}$$

where $0 \leqslant j_1 \leqslant \ldots \leqslant j_b \leqslant p+1$.

Proof. We can express C_i in the desired form using the following recursive procedure:

> If every vertex is visited once, the path from Remark 6.38 is the cruise. If there exists a vertex that occurs more than once, find the first such visited vertex v_{j_1}, and define Detour_{i,j_1} as the subwalk of C_i between the first and last occurrence of v_{j_1}; now repeat this procedure on the walk starting at the last occurrence of v_{j_1} and ending at the end of the cruise.

□

Goal 3.1: Encoding locations of detours. Recall that W is composed of k links of length-ℓ each. We utilize this structure of W to encode the locations as well as the length of all detours in W.

Definition 6.41. Given a detour Detour in W, we say the *timestamp* of Detour is the tuple (a, b) where a is the position of the start step of Detour in W and b is the position of the end step of Detour in W.

Lemma 6.42. *There is an encoding of the timestamps of all detours in W in $\left[(\ell+1)^{2k}\right]$.*

Proof. Let $L_1, \dots, L_k$ denote the k links that compose W. Due to the nonbacktracking nature of links, each link can have at most one "start step" of a detour and at most one "end step" of a detour. We associate a tuple (a_i, b_i) to link L_i where a_i is 0 if there is no start step of a detour in L_i and the position of that step (which is a number in $[\ell]$) if there is such a step. Likewise, b_i is 0 if L_i contains no end step, and is the position of the end step otherwise. It is possible to reconstruct timestamps of all detours from the m tuples (a_i, b_i), and since each tuple can be encoded by a number in $\left[(\ell+1)^2\right]$, this list of tuples can be encoded by a number in $\left[(\ell+1)^{2k}\right]$. □

Goal 3.2: Encoding detours. Before describing how we encode detours we make some structural observations about detours.

Claim 6.43. All the edges visited by any detour Detour are in $D^*(W) \cap F(W)$.

Proof. Since Detour is contained inside a cruise, all its edges are in $F(W)$. Hence, all edges of Detour are in $D^*(W) \cup S(W)$ because $F(W)$ is a spanning forest of $D^*(W) \cup S(W)$. Since Detour is a closed walk in a tree, it must visits each edge an even number of times; in particular, Detour does not contain any singleton edges and hence is completely contained in $D^*(W)$. □

Corollary 6.44. *For any detour* Detour*, the graph* $G(\mathsf{Detour})$ *has no* (t, r, ε)*-vexing vertices.*

Observation 6.45. Any detour Detour can be decomposed into a sequence of links of length exactly ℓ, with the exception of the first and last link, which can both have any length between 1 and ℓ.

Definition 6.46. Any detour Detour starts and ends at some vertex v. We call v the *root* of Detour and denote it with $\mathsf{Root}(\mathsf{Detour})$.

Remark 6.47. One should think of a detour as a closed walk on a tree rooted at a distinguished vertex.

Definition 6.48. We call a step from u to v in Detour an *up-step* if v is closer to $\mathsf{Root}(\mathsf{Detour})$ than u. In similar spirit, we call that step a *down-step* if v is further from $\mathsf{Root}(\mathsf{Detour})$ than u.

Definition 6.49. We further classify down-steps in a detour Detour into three types:

1. We call a down-step from u to v a *fresh skeleton step* if the edge $\{u, v\}$ is part of $\mathsf{Skel}(F(W))$ and has not been traversed by any detour so far.
2. We call a down-step from u to v a *fresh intrepid step* if the edge $\{u, v\}$ is *not* part of $\mathsf{Skel}(F(W))$ and has not been traversed so far. We use f_i to denote the total number of fresh intrepid steps across all detours in the walk.
3. We call a down-step from u to v a *stale step* if it is not a fresh skeleton step or a fresh intrepid step.

Claim 6.50. Suppose there is a stale step from u to v at time T. Then there is an occurrence of a step from u to v as well as from v to u in a detour at an earlier time.

Proof. Since the step at time T between u and v occurs in a detour, the edge $\{u, v\}$ must be part of $D^* \cap F(W)$. If $\{u, v\}$ is part of $\mathsf{Skel}(F(W))$, then it must have been traversed in a detour at a time before T, since otherwise this step would be classified as a fresh skeleton step. If $\{u, v\}$ is not part of $\mathsf{Skel}(F(W))$, then it must be part of $F(W) \setminus \mathsf{Skel}(F(W))$ and these edges are only traversed in detours; and if $\{u, v\}$ was not traversed in an earlier detour, it would have been classified as a fresh intrepid step.

Thus, we have established that the edge $\{u, v\}$ is traversed by a detour. Now, if $\{u, v\}$ was traversed in a detour, there must have been both a step from u to v and a step from v to u since if a directed edge is traversed in a detour then so is its reversal; in particular, a step between u and v occurs in a detour before time T. □

Definition 6.51. We call a (possibly empty) contiguous sequence of steps a *stretch*.

Observation 6.52. Due to nonbacktracking nature of links and the tree structure of detours, every link in a detour can be broken into 4 phases:

- Phase 1: an up-stretch,
- Phase 2: a stale stretch,
- Phase 3: a fresh skeleton stretch
- Phase 4: a fresh intrepid stretch.

Lemma 6.53. *Given the encoding of* $\mathsf{Skel}(F(W))$ *from Lemma 6.36, the encoding of endpoints of cruises from Lemma 6.37, and the encoding of timestamps of detours from Lemma 6.42, it is possible to encode all detours in* W *using a number in*

$$\left[\ell^{4k} \cdot ((1+\varepsilon)d)^{tm} \cdot ((1+\varepsilon)d)^{\frac{1}{2}(k\ell - 2|D(W)| - |S(W)|)} \cdot (3\lambda(W)+1)^{5\lambda(W)} \cdot n^{f_i}\right].$$

Proof. Let $\mathsf{Detour}_1, \ldots, \mathsf{Detour}_b$ be the sequence of detours of W in order of time. We first specify how we encode detours, and then prove that the encoding is valid, i.e., recovery of all Detour_a from the given encoding is possible. As pointed out in 6.45 each Detour_a can be broken into a sequence of links $L_1, \ldots, L_\tau$. We now describe how to encode each L_j.

Encoding metadata. For each link L_j, we first specify four numbers in $[\ell]$ denoting the lengths of the up-stretch, stale stretch, fresh skeleton stretch, and fresh intrepid stretch in the detour. Now we zoom in and encode each phase carefully.

Encoding up-stretches. We don't specify any extra information about the up-stretch.

Encoding a stale stretch. Given a stale stretch ζ, let E_ζ denote the set of edges visited before ζ starts. From Claim 6.43 ζ is completely contained in $D^*(W)$. Since ζ is a stale stretch, it must be contained in $E_\zeta \cap D^*(W)$. We first break ζ into $\left\lceil \frac{|\zeta|}{t} \right\rceil$ substretches $\zeta_1, \ldots, \zeta_{\left\lceil \frac{|\zeta|}{t} \right\rceil}$ each of length at most t and encode each substretch. Let v_i be the vertex at the start of ζ_i and v_i' be the end of ζ_i. Since $E_\zeta \cap D^*$ has no (t, r, ε)-vexing vertices, there are at most $((1+\varepsilon)d)^t$ vertices within distance t of v_i; in particular, there are at most $((1+\varepsilon)d)^t$ possible candidates for v_i'. We sort these candidates in increasing order of time first visited in a detour, and encode ζ_i with the index of v_i' in this list of candidates. Note that this index is a number in $\left[((1+\varepsilon)d)^t\right]$. To encode ζ, we specify $\left\lceil \frac{|\zeta|}{t} \right\rceil$ such numbers, one corresponding to each ζ_i.

Encoding a fresh skeleton stretch. For each step $u \to v$ of the fresh skeleton stretch, we don't specify any information if the degree of u within $\mathsf{Skel}(F(W))$ is $\leqslant 2$ and u is not a terminal. If the degree of u is at least 3 or if u is a terminal, we create a list of neighbors of u sorted in increasing

order of their identities in K_n, and specify the index of v in this list. Note that this index is at most the degree of u within $\mathsf{Skel}(F(W))$, which from Fact 6.35 is at most $3 \times$ (# leaves in $\mathsf{Skel}(F(W))$), which in turn from 6.32 is bounded by $3\lambda(W)$.

Encoding a fresh intrepid stretch. For every fresh intrepid step uv, we specify the identity of v in K_n, so each fresh intrepid step is encoded by a number in $[n]$.

Recovery of detours. We now show how to recover the detours from the given encodings. First, it is possible to recover the root of every detour from the encodings given by Lemma 6.36, Lemma 6.37 and Lemma 6.42. We now show how to recover the detours in order

$$\mathsf{Detour}_1, \mathsf{Detour}_2, \ldots, \mathsf{Detour}_b.$$

Suppose $\mathsf{Detour}_1, \ldots, \mathsf{Detour}_i$ have been recovered, we show how to recover Detour_{i+1}. Let $L_1, \ldots, L_\tau$ be the links in Detour_{i+1}. We show how to sequentially recover the links. Suppose $L_1, \ldots, L_j$ have been recovered. We now describe how to recover L_{j+1}.

Recovering the up-stretch in L_{j+1}. The length of the up-stretch, which is part of the "metadata encoding" is sufficient to reconstruct the up-stretch of L_{j+1}.

Recovering the stale stretch in L_{j+1}. By Claim 6.50 every step in the stale stretch of L_{j+1} has been taken in a detour before. Since we know the the length of the stale stretch in L_{j+1} from the metadata encoding, and we have recovered all steps before the stale stretch in L_{j+1} that are part of a detour, we can infer a list of candidate endpoints of the stale stretch. Further, we also know the order in which these candidates were visited in detours, and hence we can recover the stale stretch in L_{j+1} from the encoding of stale stretches we described.

Recovering the fresh skeleton stretch in L_{j+1}. Now we describe how to recover the fresh skeleton stretch of L_{j+1}. Once the stale stretch of L_{j+1} has been recovered, we know the start vertex of this stretch, v. We also can infer the length of the fresh skeleton stretch $\mathsf{LenSkel}$ from the metadata encoding. We recover this full stretch by performing the following walk, which traces the same steps as the fresh skeleton stretch of L_{j+1}:

- Let x be a counter that is initially 0.
- Let v' be initially set to v (v' denotes the "current vertex" in our walk).
- While $x \leqslant \mathsf{LenSkel}$:
  - If the degree of v' within $\mathsf{Skel}(F(W))$ is $\leqslant 2$ and v' is not a terminal, then step along the unique unvisited edge incident to v' (called $v'w$) and update v' to w. *Note that if the first x*

steps of this walk and those in the fresh skeleton stretch coincide, then $v'w$ must be the $(x+1)$-th step in the fresh skeleton stretch.

- If the degree of v' within $\mathsf{Skel}(F(W))$ is $\geqslant 3$ or v' is a terminal vertex: then assuming the first x steps of the current walk match those of the fresh skeleton stretch, we can recover the next step $v'w$ of the fresh skeleton stretch from the encoding of $\mathsf{Skel}(F(W))$ in Lemma 6.36 combined the encoding of fresh skeleton stretches described earlier in this proof. Thus, we update v' to w.
- Increment x by 1.

Recovering the fresh intrepid stretch in L_{j+1}. We can straightforwardly recover this stretch step-by-step since the identity of each vertex within K_n is given in the encoding.

Recovery wrapup. Thus, we have established how we recover link L_{j+1} from the given encoding and all links in all detours that occurred before. Inductively, this gives us a method to recover all detours in W.

Counting. Now we finally turn our attention to bounding the number of encodings of all detours. We will bound the number of metadata encodings, the number of stale stretch encodings, the number of fresh skeleton stretch encodings and finally the number of fresh intrepid stretch encodings.

Bounding the number of metadata encodings. Since there are at most k links in detours and the metadata of each link contains 4 numbers in $[\ell]$, there are at most ℓ^{4k} possible metadata encodings.

Bounding the number of stale stretch encodings. Let us call the stale stretch corresponding to a link L as $\zeta(L)$. Each stale stretch ζ is encoded using $\left\lceil \frac{|\zeta|}{t} \right\rceil$ numbers in $\left[((1+\varepsilon)d)^t\right]$. The total number of stale stretch encodings is then bounded by

$$\prod_{L \in \text{Links}(W)} \left(((1+\varepsilon)d)^t\right)^{\left\lceil \frac{|\zeta(L)|}{t} \right\rceil} \leqslant \left(((1+\varepsilon)d)^t\right)^{\sum_{L \in \text{Links}(W)} \left(\frac{|\zeta(L)|}{t}+1\right)}. \tag{31}$$

We turn our attention to bounding $\sum_{L \in \text{Links}(W)} \left(\frac{|\zeta(L)|}{t} + 1\right)$.

$$\sum_{L \in \text{Links}(W)} \left(\frac{|\zeta(L)|}{t} + 1\right) = k + \frac{1}{t} \sum_{L \in \text{Links}(W)} |\zeta(L)| \tag{32}$$

Note that $\sum_{L \in \text{Links}(W)} |\zeta(L)|$ is the total number of stale steps across all detours. From Claim 6.50 the (undirected) edge that a stale step is taken on is being traversed for *at least* the third time. Further, since the stale step is a down-step, there must be a corresponding up-step that is the reversal of

the down-step in the detour. Thus, an edge $\{i,j\}$ is traversed by a stale step at most $\frac{a_{ij}-2}{2}$ times. Further, since there are multiple steps that traverse the same edge that a given stale step traverses, every stale step must traverse an edge in $D(W)$. Thus, we can bound (32) by:

$$k + \frac{1}{t}\sum_{\{i,j\}\in D(W)} \frac{1}{2}(a_{ij}-2) = k + \frac{1}{2t}\left(\sum_{ij:a_{ij}\geqslant 2}(a_{ij}-2) + \sum_{ij:a_{ij}=1}(a_{ij}-1)\right)$$
$$= k + \frac{1}{2t}\left(k\ell - 2|D(W)| - |S(W)|\right)$$

Plugging in the above into (31) gives us a bound of:

$$((1+\varepsilon)d)^{tk} \cdot ((1+\varepsilon)d)^{\frac{1}{2}(k\ell - 2|D(W)| - |S(W)|)}.$$

Bounding the number of fresh skeleton stretch encodings. Let P be the set of vertices that either are terminal vertices or have degree-$\geqslant 3$ in $\mathsf{Skel}(F(W))$. We can extract our encoding of fresh skeleton stretches from the following map H.

> For every $v \in P$, $H(v)$ is equal to the list of numbers in $[\deg_{\mathsf{Skel}(F(V))}(v)]$ such that number i is in this list if vw_i is a fresh skeleton step, where w_i is the ith neighbor of v in lexicographic order of names in K_n; further, this list is sorted in order of time the corresponding steps are taken.

There are at most $(\deg_{\mathsf{Skel}(F(V))}(v)+1)^{\deg_{\mathsf{Skel}(F(V))}(v)}$ possibilities for $H(v)$ since every edge in the skeleton can occur at most once in a fresh skeleton stretch. Since the number of possible encodings is upper bounded by the number of candidates for H, we have a bound of

$$\prod_{v\in P}(\deg_{\mathsf{Skel}(F(V))}(v)+1)^{\deg_{\mathsf{Skel}(F(V))}(v)} \leqslant (3\lambda(W)+1)^{\sum_{v\in P}\deg_{\mathsf{Skel}(F(V))}(v)} \tag{33}$$

Now we focus on bounding $\sum_{v\in P}\deg_{\mathsf{Skel}(F(V))}(v)$.

$$\sum_{v\in P}\deg_{\mathsf{Skel}(F(V))}(v) = \sum_{v:\deg_{\mathsf{Skel}(F(V))}(v)\geqslant 3}\deg_{\mathsf{Skel}(F(V))}(v) + \sum_{\substack{v:\deg_{\mathsf{Skel}(F(V))}(v)\leqslant 2\\ v\in T(W)}}\deg_{\mathsf{Skel}(F(V))}(v)$$

From Fact 6.35 the first term is bounded by $3\times$ # leaves in $\mathsf{Skel}(F(W))$, which from 6.32 is bounded by $3\lambda(W)$. The second term is bounded by $2|T(W)|$, which from Remark 6.29 is at most $2\lambda(W)$. As an upshot we have:

$$\sum_{v\in P}\deg_{\mathsf{Skel}(F(V))}(v) \leqslant 5\lambda(W).$$

Plugging this into (33) gives us a bound on the number of possible skeleton fresh stretch encodings of:

$$(3\lambda(W)+1)^{5\lambda(W)}.$$

Bounding the number of fresh intrepid stretch encodings: The encoding of fresh intrepid stretches comprises of f_i identities of vertices in K_n, each of which is represented by a number in $[n]$. Hence there are at most n^{f_i} fresh intrepid stretch encodings.

Combining all the above bounds, we get a bound on the total number of possible encodings of all the detours of

$$\ell^{4k} \cdot ((1+\varepsilon)d)^{tk} \cdot ((1+\varepsilon)d)^{\frac{1}{2}(k\ell - 2|D(W)| - |S(W)|)} \cdot (3\lambda(W)+1)^{5\lambda(W)} \cdot n^{f_i}$$

□

Since it is possible to recover a linkage W from $\mathsf{Skel}(W)$, the endpoints of its cruises and the order in which the cruises occur, the timestamps of the detours, and the detours, by a combination of Lemma 6.36, Lemma 6.37, Lemma 6.42 and Lemma 6.53 along with a bound on $\lambda(W)$ from Claim 6.27 we have the following bound:

Theorem 6.54 (Restatement of Theorem 6.16)**.** *The total number of (k,ℓ)-linkages with f fresh edges, e excess edges, s singleton edges and Δ profligate steps is at most:*

$$n^{f+1} \cdot (4\lambda(W))^{7\lambda(W)+1} \cdot (k\ell)^{3\lambda(W)+1} \cdot (\ell+1)^{6k} \cdot ((1+\varepsilon)d)^{tk + k\ell/2 - |D(W)| - s/2}$$

where $\lambda(W) \leqslant 3e + \frac{12k\ell \ln(k\ell)}{r} + 3\Delta$.

7 Lower Bounds in the Stochastic Block Model

In this section, we finish the proof of Theorem 6.3 by proving lower bounds for the level-M path statistics SDP (as described by Definition 6.2) for every constant M for detection in the stochastic block model under the Kesten-Stigum threshold.

An ingredient we will need is an Ihara–Bass formula for weighted graphs, which appears in [WF11, FM17] as well as a related power series identity, which to our knowledge is novel. We give a proof for the sake of being self-contained.

7.1 Weighted Ihara-Bass and a Power Series Identity

Let $G = (V, E)$ be any graph. For any edge weights $c : E \to \mathbb{R}$, write A_c for the weighted adjacency matrix of G, and D_c for the diagonal matrix of c-weighted vertex degrees. More generally let $A_c^{(\ell)}$ count c-weighted non-backtracking walks on G, $C \in \mathbb{R}^{2|E| \times 2|E|}$ be the diagonal matrix with $C_{i\to j, i\to j} = C(i \to j)$, and write $B_c = CB$ where B is the nonbacktracking matrix of the complete graph.

Theorem 7.1 (Weighted Ihara-Bass). *For any weights $c : E \to \mathbb{R}$, let $\hat{c} = c(1-c^2)^{-1}$. Then*

$$\det(1 - B_c) = \prod_{(i,j)\in E} (1 - c(i,j)^2) \det(1 - A_{\hat{c}} + D_{c\hat{c}}),$$

and

$$(1 - A_{\hat{c}} + D_{c\hat{c}})^{-1} = \sum_{\ell \geqslant 0} A_c^{(\ell)}$$

whenever this series converges.

Proof. Regard each edge as a pair of directed edges in opposite directions. Write $S \in \mathbb{R}^{|V|\times 2|E|}$ and $T \in \mathbb{R}^{2|E|\times|V|}$ for the *start* and *terminal* matrices (i.e. if $(u,v) \in E$ the former has $S_{u,u\to v} = 1$ and the latter has $T_{u\to v,v} = 1$) and $\Pi \in \mathbb{R}^{2|E|\times 2|E|}$ for the involution that reverses directed edges. Let's adopt the convention that $B = TS - \Pi$, and note for later that $C\Pi = \Pi C$, since the weights c are a function of undirected edges. Moreover $S\Pi CT = D_c$ and $S(CB)^\ell CT = A_c^{(\ell+1)}$ for every $\ell \geqslant 0$; indeed analogous identities hold for any diagonal weight matrix commuting with Π.

Now consider the matrix

$$\mathfrak{B}_c \triangleq \begin{pmatrix} 1 & S \\ CT & 1 + C\Pi \end{pmatrix}.$$

We can compute the determinant of $\mathfrak{B}_c$ using two different Schur complements:

$$\det \mathfrak{B}_c = \det(1 - CB) = \det(1 + C\Pi) \det(1 - S(1 + C\Pi)^{-1} CT).$$

It remains now to understand the matrix $1 - S(1 + C\Pi)^{-1}CT$. Since C and Π commute,

$$(1 + C\Pi)^{-1} = (1 - C^2)^{-1}(1 - C\Pi)$$

making

$$\begin{aligned} 1 - S(1 + C\Pi)^{-1}CT &= 1 - S\left((1 - C^2)^{-1} - C(1 - C^2)^{-1}\Pi\right) CT \\ &= 1 - A_{\hat{c}} + D_{c\hat{c}}; \end{aligned}$$

the second line follows from our initial discussion and the definition $\hat{c} = c(1 - c^2)^{-1}$.

To prove the power series identity, let invert $\mathfrak{B}_c(z)$ with the Schur complement formula:

$$\begin{aligned} 1 &= \mathfrak{B}_c \mathfrak{B}_c^{-1} \\ &= \begin{pmatrix} 1 & S \\ CT & 1 + C\Pi \end{pmatrix} \begin{pmatrix} (1 - A_{\hat{c}} + D_{c\hat{c}})^{-1} & -S(1 - CB)^{-1} \\ -(1 - CB)^{-1}CT & (1 - CB)^{-1} \end{pmatrix}. \end{aligned}$$

Considering the upper left block, we see

$$(1 - A_{\hat{c}} + D_{c\hat{c}})^{-1} = 1 + S(1 - CB)^{-1}CT$$

$$= 1 + \sum_{\ell \geqslant 0} S(CB)^\ell CT$$
$$= \sum_{\ell \geqslant 0} A_c^{(\ell)}$$

□

7.2 Construction of SDP solution

Let $\boldsymbol{G}$ be a $G(n, d/n)$ graph. Our goal is to construct a solution to the SDP given in Definition 6.2 when $\boldsymbol{G} \sim G(n, d/n)$, and d is under the KS threshold. We instead construct a solution to the following simpler SDP, and obtain a solution for the SDP in Definition 6.2 via an identical procedure to the one described after the statement of Proposition 5.10. Given parameters λ, M, δ and graph G:

$$\begin{aligned} &\text{Find } n \times n \text{ matrix } Y \succeq 0 \text{ s.t.} \\ &\qquad Y_{i,i} = 1 && \forall i \in [n] \\ &\left\langle Y, \left(A_G - \frac{d}{n} 1 1^\top\right)^{(\ell)} \right\rangle = d^\ell \lambda^\ell n \pm O(\delta n) && \forall \ell \leqslant M. \end{aligned} \tag{34}$$

Our main technical result in this section is:

Theorem 7.2. *For $\boldsymbol{G} \sim G(n, d/n)$, for $|\lambda| < \frac{1}{\sqrt{d}}$, and for any $\delta, M > 0$, the SDP (34) is feasible with high probability.*

Let $\varepsilon > 0$ be an arbitrary constant, $\ell_0 \in [\lceil \log n \log\log n \rceil, 2\lceil \log n \log\log n \rceil]$, $t = \ell_0^{1/3}$, $r = \frac{2\ell_0}{\ln^3(2\ell_0)}$; let $\boldsymbol{G}_{t,r,\varepsilon}$ be its (t, r, ε)-truncation and let $\boldsymbol{A}_{t,r,\varepsilon}$ denote the adjacency matrix of $\boldsymbol{G}_{t,r,\varepsilon}$. Now, let S be the set of vertices deleted in truncating $\boldsymbol{G}$, and define edge weights $c : E \to \mathbb{R}$ so that

$$A_c = \boldsymbol{A}_{t,r,\varepsilon} - \frac{d}{n} 1_{[n]\setminus S} 1_{[n]\setminus S}^\top$$

Define $A_c^{(m)}$ as $\mathbb{1}$ when $m = 0$ and akin to how $\overline{A}^{(m)}$ was defined in Section 6.2 when $m \geqslant 1$. And finally define B_c the way it is defined in Section 7.1. Our next ingredient is establishing an operator norm bound on $B_c^{\ell_0}$. Indeed:

$$\|B_c^{\ell_0}\| \leqslant \sqrt{\operatorname{tr}\left(B_c^{\ell_0} (B_c^*)^{\ell_0}\right)}.$$

The above quantity can be seen to be upper bounded by:

$$\sqrt{n^2 \operatorname{tr}\left(\left(A_c^{(\ell_0 - 1)}\right)^2\right)}$$

which from (27) is bounded by:

$$n \cdot \left((1+\varepsilon)^4 \sqrt{d}\right)^{\ell_0 - 1},$$

which by our choice of ℓ_0 is at most

$$\left((1+\varepsilon)^5\sqrt{d}\right)^{\ell_0}.$$

Note that the following is true for *any* $\ell_0 \in I := [\lceil \log n \log\log n \rceil, 2\lceil \log n \log\log n \rceil]$:

$$\|B_c^{\ell_0}\| \leqslant \left((1+\varepsilon)^5\sqrt{d}\right)^{\ell_0}. \tag{35}$$

Since any $\ell \geqslant 2\lceil \log\log n \rceil$ can be expressed as

$$\ell := \ell_1 + \cdots + \ell_s$$

for $\ell_i \in I$, we can conclude from a combination of submultiplicativity of operator norm and (35) that

$$\|B_c^{\ell}\| \leqslant \|B_c^{\ell_1}\| \cdots \|B_c^{\ell_s}\| \leqslant \left((1+\varepsilon)^5\sqrt{d}\right)^{\ell}. \tag{36}$$

Via the expression $A_c^{(\ell)} = SB_c^{\ell-1}CT$ in the proof of Theorem 7.1 and the fact that $\|S\| \leqslant n$ and $\|CT\| \leqslant n$, we know:

$$\|A_c^{(\ell)}\| \leqslant \left((1+\varepsilon)^6\sqrt{d}\right)^{\ell} \tag{37}$$

for all $\ell \geqslant \ell_0$. Another consequence of (36) is

$$\rho(B_c) \leqslant \|B_c^{\ell}\|^{1/\ell} \leqslant (1+\varepsilon)^5\sqrt{d}. \tag{38}$$

Now, let

$$M_s(z) := \sum_{0 \leqslant \ell \leqslant s} A_c^{(\ell)} z^{\ell}.$$

Define $\hat{c}$ in terms of c identically to how it is defined in the statement of Theorem 7.1. From Theorem 7.1,

$$M_\infty(z) = (\mathbb{1} - A_{\hat{cz}} + D_{cz\hat{cz}})^{-1}.$$

Next, we use a proposition that is similar to (and whose proof follows) a similar statement in [WF11, FM17]:

Proposition 7.3. *Suppose $z \in \mathbb{R}$ and $|z| < \min\{1/\rho(B_c), 1\}$, then $M_\infty(z) \succeq 0$.*

Proof. $M_\infty(0)$ is the identity matrix and hence is certainly positive definite, which means all its eigenvalues are positive. Additionally, by the fact that all edge weights $c(i,j)$ are bounded by 1 and the weighted Ihara–Bass formula (Theorem 7.1), we can deduce that for all real z such that $|z| < \min\{1/\rho(B_c), 1\}$, $\det(M_\infty(z)) > 0$. Since the determinant (which is the product of eigenvalues) is strictly positive on a continuous interval, the eigenvalues of $M_\infty(z)$ are a continuous function of z on this interval, and the eigenvalues of M_∞ are strictly positive at one point in this interval, all eigenvalues of M_∞ must be positive for all real z where $|z| < \min\{1/\rho(B_c), 1\}$. Thus, the proposition follows. □

Our next goal will be to lower bound the minimum eigenvalue of $M_{r/2-1}(z)$, i.e. prove that the

minimum eigenvalue is not too negative when z is in an appropriate range.

Proposition 7.4. *Suppose* $|z| < \frac{1}{(1+2\varepsilon)^6\sqrt{d}}$, *then* $\lambda_{\min}(M_{r/2-1}(z)) \geqslant -\delta(n)$ *where* $\delta(n) = o_n(1)$.

Proof. $\lambda_{\min}(M_{r/2-1}(z)) = \lambda_{\min}\left(M_\infty(z) - \sum_{\ell \geqslant \lfloor r/2 \rfloor} A_c^{(\ell)} z^\ell\right)$, which by positive semidefiniteness of $M_\infty(z)$ is lower bounded by

$$-\left\|\sum_{\ell \geqslant r/2-1} A_c^{(\ell)} z^\ell\right\| \geqslant -\sum_{\ell \geqslant r/2-1} \left\|A_c^{(\ell)}\right\| |z|^\ell.$$

By a combination of Theorem 6.7 and (37), along with the assumption on $|z|$ we know the above is lower bounded by

$$-\sum_{\ell \geqslant r/2-1} \left(\frac{1+\varepsilon}{1+2\varepsilon}\right)^{6\ell} \geqslant -\alpha \left(\frac{1+\varepsilon}{1+2\varepsilon}\right)^{3r-6}.$$

where $\alpha := \sum_{\ell \geqslant 0} \left(\frac{1+\varepsilon}{1+2\varepsilon}\right)^{6\ell}$ is an absolute constant depending only on ε. The proposition follows from the choice of r. □

Let $\delta(n)$ be the function in the statement of Proposition 7.4, and define

$$X(z) := (1 - \delta(n)) \cdot M_{r/2-1}(z) + \delta(n) \cdot \mathbb{1}.$$

By Proposition 7.4, $X(z)$ is positive semidefinite when $|z| < \frac{1}{(1+2\varepsilon)^6\sqrt{d}}$. At this point, we state a fact that will be used later.

Fact 7.5. *With probability* $1 - o_n(1)$, *the maximum degree of a vertex in* $\boldsymbol{G}$ *is bounded by* $\log^2 n$.

Our next goal is to argue that the diagonal entries of $X(z)$ are $1 \pm o_n(1)$. Towards this, let us try to understand the contribution of $A_c^{(\ell)}$ to the diagonal. In particular:

Proposition 7.6. *Let* $\ell < r$. *The diagonal entries of* $A_c^{(\ell)}$ *are all bounded in magnitude by* $\frac{\left(2\log^2 n\right)^\ell}{n}$ *with probability* $1 - o_n(1)$.

Proposition 7.6 is a direct consequence of Fact 7.5 and the forthcoming Proposition 7.8 which we prove and use later.

Proposition 7.7. *The diagonal entries of* $X(z)$ *are all* $1 \pm \delta'(n)$ *with probability* $1 - o_n(1)$ *as long as* $|z| < 1$ *where* $\delta'(n)$ *is some function which is* $o_n(1)$.

Proof. As long as the maximum degree of G is bounded by $\log^2 n$, which happens with probability $1 - o_n(1)$, by Proposition 7.6 the diagonal entries of $X(z)$ are bounded by

$$\sum_{0 \leqslant \ell \leqslant r/2-1} \text{contribution of } A_c^{(\ell)} \text{ to diagonal} \leqslant r \cdot \frac{\left(2\log^2 n\right)^r}{n},$$

which is $o_n(1)$. □

Finally for $|z| < \frac{1}{(1+2\varepsilon)^6\sqrt{d}}$ we define $Y(z) \coloneqq (1-\delta'(n))X(z) + \Gamma$ where Γ is a diagonal matrix chosen so that the diagonal of $Y(z)$ is all-ones. By Proposition 7.7, Γ is positive semidefinite and combined with the fact that $X(z)$ is positive semidefinite, we can conclude that $Y(z)$ is positive semidefinite as well.

7.3 Matching path statistics

In this section, we are interested in understanding the value of $\left\langle \left(A - \frac{d}{n}11^\top\right)^{(\ell)}, Y(z) \right\rangle$ where A is the adjacency matrix of $\boldsymbol{G}$. It suffices to understand the value of each $\left\langle \left(A - \frac{d}{n}11^\top\right)^{(\ell)}, A_c^{(m)} \right\rangle$ where $\ell, m \leqslant \frac{r}{2} - 1$. To lighten the notation, in this subsection we will use A' to denote $\boldsymbol{A}_{t,r,\varepsilon}$.

To set up Proposition 7.8, we introduce some notation — let $\mathrm{NB}_{i,j,\ell}$ for $i, j \in [n]$ and $\ell \in \mathbb{N}$ denote the set of all nonbacktracking walks that start at i, end at j, and are of length-ℓ. Given a nonbacktracking walk W, we will interpret W as a set of tuples (i, ab) for $i \in [\ell]$ where ab is the edge walked on in the ith step of W. For any $S \subseteq W$, we will overload notation and use S to also denote the subset of edges with the timestep-indices removed.

Proposition 7.8. *Suppose G is a n-vertex graph with maximum degree bounded by $\Delta \geqslant d$, then for all $i, j \in [n]$:*

$$\left| \sum_{\substack{(W,S):S\subseteq W, S\neq W \\ W\in \mathrm{NB}_{i,j,\ell} \\ S\subseteq E(G)}} \left(-\frac{d}{n}\right)^{\ell-|S|} \right| \leqslant \frac{\ell\cdot(2\Delta)^\ell}{n}.$$

Proof.

$$\begin{aligned}\left| \sum_{\substack{(W,S):S\subseteq W, S\neq W \\ W\in \mathrm{NB}_{i,j,\ell} \\ S\subseteq E(G)}} \left(-\frac{d}{n}\right)^{\ell-|S|} \right| &\leqslant \sum_{k=1}^{\ell-1} \sum_{\substack{(W,S):S\subseteq W, |S|=k \\ W\in \mathrm{NB}_{i,j,\ell}}} \left(\frac{d}{n}\right)^{\ell-k} \\ &= \sum_{k=1}^{\ell-1} \left(\frac{d}{n}\right)^{\ell-k} \cdot |\{(W,S) : S\subseteq W, |S| = k, W \in \mathrm{NB}_{i,j,\ell}, S\subseteq E(G)\}|. \end{aligned} \tag{39}$$

For each k, we bound the above summand via an encoding argument. Given a walk $W = u_0u_1u_2\dots u_{\ell-1}u_\ell$ where $u_0 = i$ and $u_\ell = j$ along with a proper subset of steps S which are all contained in $E(G)$, we first find the last $u_{t-1}u_t$ on the segment which is not in S (which always

exists since S is always a proper subset). Therefore the segment $u_t \dots u_\ell$ must be composed only of edges in G. In our encoding we specify:

- The set S via timestamps (for which there are $2^\ell - 1$ choices),
- The value of t (for which there are ℓ choices),
- For $t \leqslant s \leqslant \ell - 1$, the index $a \in [\Delta]$ such that u_s is the a-th neighbor of u_{s+1} in G.
- For $1 \leqslant s \leqslant t-1$, if $u_{s-1}u_s$ is in S, we specify index $a \in [\Delta]$ such that u_s is the a-th neighbor of u_{s-1} in G; and if $u_{s-1}u_s$ is *not* in S, we specify the identity of u_s (of which there are n choices).

The total number of possible encodings is bounded by

$$\ell \cdot 2^\ell \cdot (\Delta)^k \, n^{\ell-k-1}.$$

Thus, the k-th summand is bounded by

$$\frac{\ell \cdot 2^\ell \cdot \left(\frac{\Delta}{d}\right)^k d^\ell}{n}.$$

Plugging in this bound into (39) implies the desired statement. □

As an upshot of Proposition 7.8 and Fact 7.5 we have:

$$\left(A - \frac{d}{n}11^\top\right)^{(\ell)} = A^{(\ell)} + R_\ell$$
$$A_c^{(\ell)} = A'^{(\ell)} + R'_\ell$$

where R and R' are entrywise bounded by $\frac{\ell \cdot (2\log^2 n)^\ell}{n}$ with high probability. Thus,

$$\left\langle \left(A - \frac{d}{n}11^\top\right)^{(\ell)}, A_c^{(m)} \right\rangle = \langle A^{(\ell)}, A'^{(m)} \rangle + \langle A^{(\ell)}, R'_m \rangle + \langle A'^{(m)}, R_\ell \rangle + \langle R_\ell, R'_m \rangle. \tag{40}$$

The entrywise ℓ_1 norms of $A^{(\ell)}$ and $A'^{(m)}$ are the total number of nonbacktracking walks of length-ℓ and m in $\boldsymbol{G}$ and $\boldsymbol{G}_{t,r,\varepsilon}$ respectively, which by the degree-bound of $\log^2 n$ and the fact that $\ell, m \leqslant r$, are each at most $n(\log^2 n)^r$. Since the entries of R_ℓ and R'_m are bounded by $\frac{\ell \cdot (2\log^2 n)^r}{n}$, the second and third terms of (40) are bounded by $r \cdot (2\log^2 n)^{2r}$. It is easy to see that the fourth term is bounded by $r^2(2\log^2 n)^{2r}$. It remains to understand the first term.

By Lemma 6.9 and our choice of t, there is no (t, ε)-heavy vertex in the graph with high probability. Thus, with high probability the only vertices in $\boldsymbol{G}_{t,r,\varepsilon}$ which were truncated are the ones that are part of a cycle of length at most r. To get a high probability bound on the number of such vertices, we observe that the number of cycles of length exactly k passing through a vertex v in the complete graph is at most n^{k-1}, whereas the probability of a fixed cycle occuring in $\boldsymbol{G}$ is $\left(\frac{d}{n}\right)^k$, which implies

via a union bound that the probability that v is part of a k-cycle is at most $\frac{d^k}{n}$. Consequently, the probability that v is part of a length-$\leqslant r$ cycle is at most $\frac{rd^r}{n}$, which means the expected number of vertices that are part of a length-$\leqslant r$ cycle is at most rd^r. By Markov's inequality, with high probability the number of of such vertices is bounded by, say, $rd^r \cdot \log n$.

We now use S to denote the set of vertices that are within distance $r-1$ of a truncated. From the high probability bound on the maximum degree in G of $\log^2 n$, the size of S is, with high probability, at most $r\left(d\log^2 n\right)^r$. For a matrix L and $T, U \subseteq [n]$, let's use $L_{T,U}$ to denote the principal submatrix obtained from the rows indexed by T and columns indexed by U. For all $t \leqslant r/2-1$:

$$A^{(t)}_{[n]\setminus S,[n]} = A'^{(t)}_{[n]\setminus S,[n]}$$
$$A^{(t)}_{[n],[n]\setminus S} = A'^{(t)}_{[n],[n]\setminus S}.$$

Thus,

$$\langle A^{(\ell)}, A'^{(m)}\rangle = \langle A^{(\ell)}_{S,S}, A'^{(m)}_{S,S}\rangle + \langle A^{(\ell)}_{S,[n]\setminus S}, A'^{(m)}_{S,[n]\setminus S}\rangle + \langle A^{(\ell)}_{[n]\setminus S,S}, A'^{(m)}_{[n]\setminus S,S}\rangle + \langle A^{(\ell)}_{[n]\setminus S,[n]\setminus S}, A'^{(m)}_{[n]\setminus S,[n]\setminus S}\rangle.$$

By Fact 7.5 and the bound on $|S|$, the first term is bounded by $r\left(d\log^2 n\right)^{2r}$ with high probability. If $\ell \neq m$, then each of the second to fourth terms is equal to 0. If $\ell = m$, the sum of the second to fourth terms is sandwiched between the total number of length-ℓ self-avoiding walks in $\boldsymbol{G}$ that also avoid S and the total number of length-ℓ self-avoiding walks in $\boldsymbol{G}$. By Fact 7.5 and the bound on $|S|$ with high probability the two quantity differ by at most $r\left(d\log^2 n\right)^{2r}$, and by Theorem 6.6 the latter quantity is $(1\pm o_n(1))d^\ell n$ with high probability. Since M from the statement of Theorem 7.2 is a constant, by a union bound, the latter quantity is $(1\pm o_n(1))d^\ell n$ with high probability simultaneously for all $\ell \leqslant M$. Additionally observe that when $\ell \neq m$, the inner product quantity we wish to bound remains bounded as long as the desired bounds on $|S|$ and the maximum degree in the graph hold. In conclusion, with high probability the following holds simultaneously for *all* $\ell \leqslant M$ and $m \leqslant r/2-1$:

$$\left\langle \left(A - \frac{d}{n}11^\top\right)^{(\ell)}, A_c^{(m)}\right\rangle = \begin{cases} (1\pm o_n(1))\cdot d^\ell n & \text{if } \ell = m \\ O(2r^2(2\log^2 n)^{2r}) & \text{if } \ell \neq m. \end{cases}$$

As a consequence:

$$\left\langle \left(A - \frac{d}{n}11^\top\right)^{(\ell)}, Y(z)\right\rangle = (1\pm o_n(1))d^\ell z^\ell n.$$

Since $Y(z)$ is PSD with high probability for all $|z| < \frac{1}{(1+\varepsilon)^6\sqrt{d}}$ and the choice of ε was arbitrary, Theorem 7.2 follows.

References

[Abb17] Emmanuel Abbe, *Community detection and stochastic block models: recent developments*, The Journal of Machine Learning Research **18** (2017), no. 1, 6446–6531. 1

[ABH16] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall, *Exact recovery in the stochastic block model*, IEEE Transactions on Information Theory **62** (2016), no. 1, 471–487. 1, 5

[ABLS07] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin, *Non-backtracking random walks mix faster*, Communications in Contemporary Mathematics **9** (2007), no. 04, 585–603. 13

[AHL02] Noga Alon, Shlomo Hoory, and Nathan Linial, *The moore bound for irregular graphs*, Graphs and Combinatorics **18** (2002), no. 1, 53–57. 42

[AKJ18] Ahmed El Alaoui, Florent Krzakala, and Michael I Jordan, *Fundamental limits of detection in the spiked wigner model*, arXiv preprint arXiv:1806.09588 (2018). 8

[AS12] Pranjal Awasthi and Or Sheffet, *Improved spectral-norm bounds for clustering*, Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, Springer, 2012, pp. 37–49. 5

[AS15] Emmanuel Abbe and Colin Sandon, *Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery*, 2015 IEEE 56th Annual Symposium on Foundations of Computer Science, IEEE, 2015, pp. 670–688. 5

[BBK+20] Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Cristopher Moore, and Alexander S Wein, *Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs*, arXiv preprint arXiv:2008.12237 (2020). 68, 69

[BDG+16] Gerandy Brito, Ioana Dumitriu, Shirshendu Ganguly, Christopher Hoffman, and Linh V Tran, *Recovery and rigidity in a regular stochastic block model*, Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms, Society for Industrial and Applied Mathematics, 2016, pp. 1589–1601. 4

[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. 8

[BKM+19] Jean Barbier, Florent Krzakala, Nicolas Macris, Léo Miolane, and Lenka Zdeborová, *Optimal errors and phase transitions in high-dimensional generalized linear models*, Proceedings of the National Academy of Sciences **116** (2019), no. 12, 5451–5460. 8

[BKW19] Afonso S Bandeira, Dmitriy Kunisky, and Alexander S Wein, *Computational hardness of certifying bounds on constrained pca problems*, arXiv preprint arXiv:1902.07324 (2019). 65

[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. 1, 2, 3

[BS95] Avrim Blum and Joel Spencer, *Coloring random and semi-random k-colorable graphs*, Journal of Algorithms **19** (1995), no. 2, 204–234. 5

[BSW01] Eli Ben-Sasson and Avi Wigderson, *Short proofs are narrowresolution made simple*, Journal of the ACM (JACM) **48** (2001), no. 2, 149–169. 2

[Cas06] Chad Casarotto, *Graph theory and Cayleys formula*. 44

[CJSX14] Yudong Chen, Ali Jalali, Sujay Sanghavi, and Huan Xu, *Clustering partially observed graphs via convex optimization*, The Journal of Machine Learning Research **15** (2014), no. 1, 2213–2238. 5

[CL$^+$15] T Tony Cai, Xiaodong Li, et al., *Robust and computationally feasible community detection in the presence of arbitrary outlier nodes*, The Annals of Statistics **43** (2015), no. 3, 1027–1059. 5

[CO04] Amin Coja-Oghlan, *Coloring semirandom graphs optimally*, International Colloquium on Automata, Languages, and Programming, Springer, 2004, pp. 383–395. 5

[CO07] _____, *Solving np-hard semirandom graph problems in polynomial expected time*, Journal of Algorithms **62** (2007), no. 1, 19–46. 5

[CSV17] Moses Charikar, Jacob Steinhardt, and Gregory Valiant, *Learning from untrusted data*, Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, ACM, 2017, pp. 47–60. 2

[DKMZ11a] 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. 3

[DKMZ11b] _____, *Inference and phase transitions in the detection of modules in sparse networks*, Physical Review Letters **107** (2011), no. 6, 065701. 1

[Fei02] Uriel Feige, *Relations between average case complexity and approximation complexity*, Proceedings of the thiry-fourth annual ACM symposium on Theory of computing, ACM, 2002, pp. 534–543. 2

[FK00] Uriel Feige and Robert Krauthgamer, *Finding and certifying a large hidden clique in a semirandom graph*, Random Structures & Algorithms **16** (2000), no. 2, 195–208. 5

[FK01] Uriel Feige and Joe Kilian, *Heuristics for semirandom graph problems*, Journal of Computer and System Sciences **63** (2001), no. 4, 639–671. 5

[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. 9, 53, 56, 74, 79

[Fri03] Joel Friedman, *A proof of Alon's second eigenvalue conjecture*, Proceedings of the thirty-fifth annual ACM symposium on Theory of computing, ACM, 2003, pp. 720–724. 9

[Fri08] ———, *A proof of Alon's second eigenvalue conjecture and related problems*. 14

[Gri01] Dima Grigoriev, *Linear lower bound on degrees of positivstellensatz calculus proofs for the parity*, Theoretical Computer Science **259** (2001), no. 1, 613–622. 2

[GV16] Olivier Guédon and Roman Vershynin, *Community detection in sparse networks via grothendiecks inequality*, Probability Theory and Related Fields **165** (2016), no. 3-4, 1025–1049. 2, 5

[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. 8

[HS17] Samuel B Hopkins and David Steurer, *Efficient bayesian estimation from few samples: community detection and related problems*, Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on, IEEE, 2017, pp. 379–390. 2, 8

[HWX16] Bruce Hajek, Yihong Wu, and Jiaming Xu, *Achieving exact cluster recovery threshold via semidefinite programming*, IEEE Transactions on Information Theory **62** (2016), no. 5, 2788–2797. 5

[KGR11] Ulugbek Kamilov, Vivek K Goyal, and Sundeep Rangan, *Optimal quantization for compressive sensing under message passing reconstruction*, 2011 IEEE International Symposium on Information Theory Proceedings, IEEE, 2011, pp. 459–463. 8

[KK10] Amit Kumar and Ravindran Kannan, *Clustering with spectral norm and the k-means algorithm*, 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, IEEE, 2010, pp. 299–308. 5

[KV06] Michael Krivelevich and Dan Vilenchik, *Semirandom models as benchmarks for coloring algorithms*, 2006 Proceedings of the Third Workshop on Analytic Algorithmics and Combinatorics (ANALCO), SIAM, 2006, pp. 211–221. 5

[KV12] Bernhard Korte and Jens Vygen, *Combinatorial optimization*, vol. 2, Springer, 2012. 41

[Mas14] Laurent Massoulié, *Community detection thresholds and the weak ramanujan property*, Proceedings of the forty-sixth annual ACM symposium on Theory of computing, ACM, 2014, pp. 694–703. 1

[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, ACM, 2012, pp. 367–384. 5

[MMV16] ———, *Learning communities in the presence of errors*, Conference on Learning Theory, 2016, pp. 1258–1291. 2, 5

[MNS18] Elchanan Mossel, Joe Neeman, and Allan Sly, *A proof of the block model threshold conjecture*, Combinatorica **38** (2018), no. 3, 665–708. 1

[Moi12] Ankur Moitra, *A singly-exponential time algorithm for computing nonnegative rank*, arXiv preprint arXiv:1205.0044 (2012). 2

[MOP19a] Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes, *Explicit near-ramanujan graphs of every degree*, arXiv preprint arXiv:1909.06988 (2019). 42

[MOP19b] _____, *The sdp value for random two-eigenvalue csps*, arXiv preprint arXiv:1906.06732 (2019). 36

[MS15] Andrea Montanari and Subhabrata Sen, *Semidefinite programs on sparse random graphs and their application to community detection*, arXiv preprint arXiv:1504.05910 (2015). 2, 5

[NM14] MEJ Newman and Travis Martin, *Equitable random graphs*, Physical Review E **90** (2014), no. 5, 052824. 4

[PWBM16] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra, *Optimality and sub-optimality of pca for spiked random matrices and synchronization*, arXiv preprint arXiv:1609.05573 (2016). 8

[Ran11] Sundeep Rangan, *Generalized approximate message passing for estimation with random linear mixing*, 2011 IEEE International Symposium on Information Theory Proceedings, IEEE, 2011, pp. 2168–2172. 8

[SVC16] Jacob Steinhardt, Gregory Valiant, and Moses Charikar, *Avoiding imposters and delinquents: Adversarial crowdsourcing and peer prediction*, Advances in Neural Information Processing Systems, 2016, pp. 4439–4447. 2

[WF11] Yusuke Watanabe and Kenji Fukumizu, *Loopy belief propagation, bethe free energy and graph zeta function*, arXiv preprint arXiv:1103.0605 (2011). 53, 56

[ZK16] Lenka Zdeborová and Florent Krzakala, *Statistical physics of inference: Thresholds and algorithms*, Advances in Physics **65** (2016), no. 5, 453–552. 8

A Conjectural recovery in the DRBM

As discussed in the introduction, this paper will not settle fully the question of recovering the planted communities. However, we can at least reduce some key aspects of this problem to Conjecture 2.6 regarding the spectrum of A_G when $\boldsymbol{G} \sim \mathcal{P}_{(d,k,M,\pi)}$.

There are numerous ways to pose the recovery task, and as many metrics of success, but let us set ourselves the modest goal of, given $\boldsymbol{G}$ drawn from a planted model with $\lambda_1^2, ..., \lambda_\ell^2 > (d-1)^{-1}$ and knowledge of the parameters (d, k, M, π), recovering a vector in $\mathbb{R}^n$ with constant correlation to each of the vectors $\check{\boldsymbol{x}}_1, ..., \check{\boldsymbol{x}}_\ell$ from the Section 5. If $\ell = k$, we can use this and our knowledge of M to apply the change-of-basis F^{-1} and recover vectors correlated to the indicators $\boldsymbol{x}_1, ..., \boldsymbol{x}_k$ for each of the k communities.

Our first claim is that, assuming Conjecture 2.6, the eigenvectors of A_G can be used to approximate the $\check{x}_i$'s. In Section 4 we showed that there exists a polynomial f strictly positive on $(-2\sqrt{d-1}, 2\sqrt{d-1}) \cup \{d\}$ with the property that

$$\check{x}_i^T f(A) \check{x}_i < -\delta n$$

for some constant δ. Writing $\mu_1, ..., \mu_n$ for the eigenvalues of A_G and $\Pi_1, \cdots \Pi_n$ for the orthogonal projectors onto their associated eigenspaces, we can expand this as

$$\begin{aligned}
-\delta n &> \sum_{u \in [n]} f(\mu_u) \check{x}_i^T \Pi_u \check{x}_i \\
&= \sum_{|\mu_u| < 2\sqrt{d-1}} f(\mu_u) \check{x}_i^T \Pi_u \check{x}_i + \sum_{|\mu_u| \geqslant 2\sqrt{d-1}} f(\mu_u) \check{x}_i^T \Pi_u \check{x}_i \\
&\geqslant \sum_{|\mu_u| \geqslant 2\sqrt{d-1}} f(\mu_u) \check{x}_i^T \Pi_u \check{x}_i \qquad\qquad f(x) \text{ positive on } (-2\sqrt{d-1}, 2\sqrt{d-1}) \\
&\geqslant \inf_{|x| \leqslant d} f(x) \cdot \check{x}_i^T \left(\sum_{|\mu_u| \geqslant 2\sqrt{d-1}} \Pi_u \right) \check{x}_i.
\end{aligned}$$

Thus, even if there are only constantly many eigenvectors outside the bulk, a (for instance) random vector in their span will have $O(n)$ correlation with each of the $\check{x}_i$'s.

In order to recover *robustly* we will lean on the results of Section 5.5. If we begin with $\boldsymbol{G}$ from the planted model, perform ϵn adversarial edge insertion or deletions, and then run the SDP again, we showed that the old SDP solution will *still* be feasible. Thus, if we take $\check{X}$ from the SDP run on the corrupted graph, we will still have

$$-\delta n > \langle f(A_G), \check{X}_{i,i} \rangle \geqslant \inf_{|x| \leqslant d} f(x) \cdot \langle \sum_{|\mu_u| \geqslant 2\sqrt{d-1}} \Pi_u, \check{X}_{i,i} \rangle,$$

so a, say, Gaussian vector with covariance $\check{X}_{i,i}$ will have constant correlation with the subspace spanned by the outside-the-bulk eigenvectors of A_G, the adjacency matrix of the *unperturbed* graph, which we showed above have the same correlation guarantee with the $\check{x}_i$'s.

B Local Statistics in the DRBM

In this section we will prove Theorem 5.3 and Proposition 5.10. Since the first posting of this paper, we have updated and streamlined the arguments using the framework developed by one of the authors in [BKW19]. Several lemmas below have analogues in that work with similar proofs, and we will point these out along the way.

We first prove Theorem 5.3 by computing the quantities $\mathbb{E}\, p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G})$ in the planted model. Fix parameters d, k, M, π, recalling that M is symmetric, and $M\,\mathrm{Diag}(\pi)$ is stochastic. For any partially labelled graph (H, S, τ), let $\chi(H) = |V(H)| - |E(H)|$, $c(H)$ denote its number of connected

components, and recall

$$C_H(d) \triangleq \frac{\prod_{v\in V(H)}(d)_{\deg(v)}}{d^{|E(H)|}}$$
$$L_{(H,S,\tau)}(M,\pi) \triangleq \sum_{\hat{\tau}:\hat{\tau}|_S=\tau}\prod_{v\in V(H)}\pi(\tau(v))\cdot\prod_{(u,v)\in E(H)}M_{\tau(u),\tau(v)},$$

where the latter sum is over extensions $\hat{\tau} : V(H) \to [k]$ of τ. Note that both quantities are well-defined, by the symmetry relation of M and π, and that both are multiplicative on disjoint unions. We are aiming to show if (H, S, τ) has $O(1)$ edges, then with high probability over $(\boldsymbol{x}, \boldsymbol{G}) \sim \mathcal{P}$,

$$p_{(H,S,\tau)}(\boldsymbol{x},\boldsymbol{G}) = n^{\chi(H)}L_{(H,S,\tau)}(M,\pi)\cdot C_H(d) \pm o(n^{c(H)}).$$

To do so, we will work in the *configuration model* $\hat{\mathcal{P}}$, a distribution over d-regular multigraphs which, (i) outputs a simple graph with probability bounded away from zero in n, and (ii) conditioned on simple output agrees with $\mathcal{P}$. It is routine that high probability statements in the configuration model, like the conclusion of this proposition, therefore transfer to $\mathcal{P}$.

To sample a multigraph $\hat{\boldsymbol{G}}$ from $\hat{\mathcal{P}}$, first choose a random π balanced labelling σ, and adorn each vertex $v \in [n]$ with d half-edges $v_1, ..., v_d$. Then, for each $i \in [k]$ randomly label the half-edges on the ith group $i \to 1, ..., i \to k$ so that a $\pi(i)\pi(j)M_{i,j}dn$ have label $i \to j$ for each $j \in [k]$. Finally, choose random perfect matchings joining the $i \to j$ and $j \to i$ edges for each $i, j \in [k]$.

Lemma B.1. *Condition on a π-balanced labelling $\sigma : [n] \to [k]$, and let $\boldsymbol{P}$ be the random perfect matching on half-edges output by $\hat{\mathcal{P}}$. Let R be a simple partial matching involving constantly many half-edges, and for short write $(u, v) \in R$ if some pair (u_a, v_b) appears in the matching. Then*

$$\hat{\mathcal{P}}[R \subset \boldsymbol{P}] = (1 \pm o_n(1))\prod_{(u,v)\in R}\frac{M_{\sigma(u),\sigma(v)}}{dn}.$$

Proof. Throughout this proof, we will call a pair (u_a, v_b) appearing in R an *edge*. Write S_i for the collection of half-edges that adorn the vertices in $\sigma^{-1}(i)$, U_i for the number of half-edges in R that belong to S_i, and $U_{i,j}$ the number of edges in R with one half-edge each in S_i and S_j, respectively. We have

$$\hat{\mathcal{P}}[R \subset \boldsymbol{P}] = \frac{\prod_{i<j}\frac{(\pi(i)\pi(j)M_{i,j}dn)!}{(\pi(i)\pi(j)M_{i,j}dn-U_{i,j})!}\prod_i\frac{(\pi(i)^2M_{i,i}dn)!}{(\pi(i)^2M_{i,i}dn-2U_{i,i})!}\frac{\pi(i)^2M_{i,i}dn-2U_{i,i}-1)!!}{(\pi(i)^2M_{i,i}dn-1)!!}}{\prod_i\frac{(\pi(i)dn)!}{(\pi(i)dn-U_i)!}}.$$

Up to $o_n(1)$ fluctuations, this is equal to

$$\frac{\prod_{i<j}(\pi(i)\pi(j)M_{i,j}dn)^{U_{i,j}}\prod_i(\pi(i)^2M_{i,i}dn)^{U_{i,i}}}{\prod_i(\pi(i)dn)^{U_i}}.$$

For each edge $(u_a, v_b) \in R$, the numerator has a term $\pi(\sigma(u))M_{\sigma(u),\sigma(v)}$, and the denominator has two terms $\pi(\sigma(u))dn$ and $\pi(\sigma(v))dn$; the dropped subscripts are intentional here. Since R is

simple, we can alternatively account for these terms by looking at pairs $(u, v) \in R$. Thus, again suppressing $o_n(1)$ fluctuations, we can rewrite as

$$\prod_{(u,v)\in R} \frac{\pi(\sigma(u))\pi(\sigma(v))M_{\sigma(u),\sigma(v)}dn}{\pi(\sigma(u))dn \cdot \pi(\sigma(v))dn} = \prod_{(u,v)\in R} \frac{M_{\sigma(u),\sigma(v)}}{dn}.$$

□

Proof of Theorem 5.3. Let's begin by computing the expectation of $p_{(H,S,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}})$ over $(\boldsymbol{x}, \hat{\boldsymbol{G}})$ sampled from the configuration model. This necessitates that we extend the quantity $p_{(HS,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}})$ to the case when $\hat{\boldsymbol{G}}$ is a multigraph—we will simply take it to mean the evaluation of $p_{(H,S,\tau)}$ on the simple graph obtained by removing all self-loops and merging all multi-edges.

Choose an extension $\hat{\tau} : V(H) \to [k]$ of τ, and an injection $\phi : V(H) \to [n]$ that agrees on labels. The image of each vertex in $V(H)$ has d half-edges, so there are

$$\prod_{v\in V(H)} (d)_{\deg(v)}$$

matchings that "collapse" to H. For each, Lemma B.1 tells us the probability of inclusion in $\hat{\boldsymbol{G}} \sim \hat{\mathcal{P}}$. Finally, there are $\prod_{v\in(H)}(\pi(\tau(v))n)$ such injective maps ϕ. Putting this all together, and summing over all extensions $\hat{\tau}$,

$$\mathbb{E}\, p_{(H,S,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}}) = n^{\chi(H)} L_{(H,S,\tau)}(dM, \pi) \cdot C_H(d) + O(n^{\chi(H)-1}).$$

If H has at least one cycle, then $c(H) > \chi(H)$, and an application of Markov's inequality finishes the proof. If instead H is a forest, then the assertion will follow from an application of Chebyshev's inequality. In particular, $\mathbb{E}\, p_{(H,S,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}})^2$ is a sum over pairs of injective maps ϕ_1, ϕ_2 of the probability that both are occurrences. We can think of each pair as a single injective map $\phi' : V(H') \to [n]$, where (H', S', τ') is the image of the union of the two copies of (H, S, τ) under ϕ_1, ϕ_2 respectively. In other words,

$$\mathbb{E}\, p_{(H,S,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}})^2 = \sum_{H'} p_{(H',S',\tau')}(\boldsymbol{x}, \hat{\boldsymbol{G}}),$$

where the sum is over all (H', S', τ') that can arise by identifying some pairs of vertices in two copies of (H, S, τ). Since H has no cycles, $\chi(H') \leqslant 2\chi(H)$, with equality only if $H' = H \sqcup H$. Thus, as $L_{(H,S,\tau)}$ and C_H are multiplicative on disjoint unions,

$$\mathbb{E}\, p_{(H,S,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}})^2 = \left(\mathbb{E}\, p_{(H,S,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}})\right)^2 + O(n^{2\chi(H)-1}).$$

We finally apply Chebyshev and note that $c(H) = \chi(H)$ for forests. □

It remains now to prove Proposition 5.10, the content of which is that in constructing a feasible pseudoexpectation in the planted model, it suffices to check only certain moment constraints. We

first show that the moment constraints involving partially labelled subgraphs which contain a cycle are automatically satisfied. The argument below is essentially identical to [BBK+20, Lemma 5.19]

Lemma B.2. *Let $\boldsymbol{G} \sim \mathcal{N}$, and assume that $\widetilde{\mathbb{E}}$ is a degree-D_x pseudoexpectation—perhaps dependent on $\boldsymbol{G}$—satisfying $\mathcal{B}_k$. For every $\delta > 0$, with high probability*

$$\widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G}) = \mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{G})\sim\mathcal{P}} p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G}) \pm \delta n^{c(H)}$$

for every (H, S, τ) with constantly many edges and containing a cycle.

Proof. Using Cauchy-Schwartz for pseudoexpectations, for every multilinear monomial $m(x)$ we have $(\widetilde{\mathbb{E}}\, m(x))^2 \leqslant \widetilde{\mathbb{E}}\, m(x)^2 = \widetilde{\mathbb{E}}\, m(x)$, since $\widetilde{\mathbb{E}}\, x_{u,i}^2 = \widetilde{\mathbb{E}}\, x_{u,i}$; thus $\widetilde{\mathbb{E}}\, m(x) \in [0, 1]$. In other words, for any (H, S, τ), we have

$$\left|\widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G})\right| = \left|\sum_{\phi:V(H)\hookrightarrow[n]} \prod_{(u,v)\in E(H)} \boldsymbol{G}_{\phi(u),\phi(v)} \prod_{u\in S} x_{\phi(u),\tau(u)}\right| \leqslant \left|\sum_{\phi:V(H)\hookrightarrow[n]} \prod_{(u,v)\in E(H)} \boldsymbol{G}_{\phi(u),\phi(v)}\right|.$$

The latter is the number of occurrences of H in $\boldsymbol{G}$, with both regarded as unlabelled graphs; from the proof of Theorem 5.3 above, if H has a cycle, then this is $o(n^{c(H)})$. Thus with high probability

$$\widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G}) = o(n^{c(H)}) = n^{\chi(H)} L_{(H,S,\tau)}(M, \pi) C_H(d) \pm \delta n^{(c(H)}$$

for any $\delta > 0$, as $\chi(H) < c(H)$. □

This lemma leaves us to check only the partially labelled trees, and we next show that in fact it suffices to verify only a subset of these. The following appeared as Definition 5.20 in [BBK+20].

Definition B.3. The *pruning* of a partially labelled tree (H, S, τ) is the unique maximal subtree with the property that all leaves are distinguished vertices; if (H, S, τ) is unlabelled, its pruning is empty. The pruning of a forest is obtained by taking the pruning of each tree.

Lemma B.4. *Let (H, S, τ) be a partially labelled, tree, $(\tilde{H}, S, \tau)$ its pruning. Then*

$$L_{(H,S,\tau)}(M, \pi) = L_{(\tilde{H},S,\tau)}(M, \pi)$$

Proof. We'll argue inductively that one can delete any unlabelled leaf without affecting $L_{(H,S,\tau)}$. Let v be such a leaf, w its parent, and (H', S, τ) be obtained by deleting v. Then

$$L_{(H,S,\tau)}(M, \pi) = L_{(H',S,\tau)}(M, \pi) \sum_{\ell\in[k]} M_{\tau(w),\ell}\pi(\ell) = L_{(H',S,\tau)}(M, \pi),$$

as $M\,\mathrm{Diag}(\pi)$ is Stochastic. □

Lemma B.5. *Let $\boldsymbol{G} \sim \mathcal{N}$, and let (H, S, τ) and $(\tilde{H}, S, \tilde{\tau})$ be a partially labelled forest and its pruning,*

respectively. Then, with high probability,

$$\left\| p_{(H,S,\tau)}(x, \boldsymbol{G}) - n^{c(H)-c(\tilde{H})} \frac{C_H(d)}{C_{\tilde{H}}(d)} p_{(\tilde{H},S,\tilde{\tau})}(x, \boldsymbol{G}) \right\|_1 = o(n^{c(H)}),$$

where by $\|\cdot\|_1$ we mean the L_1 norm of the coefficients.

Proof. This argument adapted with minor elaboration from [BBK+20, Lemma 5.21]. For each occurrence $\tilde{\phi} : V(\tilde{H}) \hookrightarrow [n]$ of $(\tilde{H}, S, \tilde{\tau})$, call an occurrence $\phi : V(H) \hookrightarrow [n]$ of H an extension of $\tilde{\phi}$ if they agree on $V(\tilde{H}) \subset V(H)$. Write $\tilde{\Phi}$ for the set of occurrences of the pruning, and for each $\tilde{\phi} \in \tilde{\phi}$, write $\Phi(\tilde{\phi})$ for its set of extensions.

Again using the fact that each multilinear monomial has $\widetilde{\mathbb{E}}\, m(x) \in [0, 1]$, we may write

$$\left\| p_{(H,S,\tau)}(x, \boldsymbol{G}) - n^{c(H)-c(\tilde{H})} \frac{C_H(d)}{C_{\tilde{H}}(d)} p_{(\tilde{H},S,\tilde{\tau})}(x, \boldsymbol{G}) \right\|_1 \leqslant \sum_{\tilde{\phi} \in \tilde{\Phi}} \left| |\Phi(\tilde{\phi})| - n^{c(H)-c(\tilde{H})} \frac{C_H(d)}{C_{\tilde{H}}(d)} \right|$$

Only $o(n^{c(\tilde{H})})$ occurrences in this sum have the property that their $|E(H)|$ neighborhoods in $\boldsymbol{G}$ contain a cycle, so to prove the assertion in the lemma we are free to ignore these terms entirely. For each remaining occurrence $\tilde{\phi}$, and each connected component $\tilde{J}$ of $\tilde{H}$ containing a distinguished vertex, and the corresponding component J of H, there are precisely

$$\prod_{v \in V(H)} \prod_{q=\deg_{\tilde{J}}(v)}^{\deg_J(v)-1} (d - q) = \frac{C_J(d)}{C_{\tilde{J}}(d)}$$

ways to extend it to an occurrence of J. To finish counting the number of extensions of $\tilde{\phi}$, we need to choose an occurrence of K, the disjoint union of every connected component of H which contains no distinguished vertex, that does not interact with $\tilde{\phi}(\tilde{H})$ or the already-chosen extensions of the $\tilde{J}$'s. But, there are

$$n^{c(K)} \prod_{v \in V(K)} \prod_{q=0}^{\deg_K(v)-1} (d - q) + o(n^{c(K)}) = n^{c(H)-c(\tilde{H})} C_K(d) + o(n^{c(H)-c(\tilde{H})})$$

ways to do this. Thus, using multiplicativity of $C_H(d)$ on disjoint unions,

$$|\Phi(\tilde{\phi})| = n^{c(H)-c(\tilde{H})} C_H(d) + o(n^{c(H)-c(\tilde{H})}),$$

there are $O(n^{c(\tilde{H})}$ possible occurrences of $\tilde{H}$, and we are done. □

Taking a union bound and applying the above lemma, we immediately obtain:

Lemma B.6. *Let $\boldsymbol{G} \sim \mathcal{N}$, and assume that $\widetilde{\mathbb{E}}$ is is a degree-D_x pseudoexpectation, which may depend on $\boldsymbol{G}$. If $\widetilde{\mathbb{E}}$ satisfies the affine moment constraints for every pruned, partially labelled forest with at most D_G*

edges and D_x distinguished vertices, then with high probability

$$\left|\widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G}) - \mathop{\mathbb{E}}_{(\boldsymbol{x},\boldsymbol{G})\sim\mathcal{P}} p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G})\right| = o(n^{\chi(H)})$$

for every partially labelled forest (H, S, τ) with at most D_G edges and D_x distinguished vertices.

Proof. For each (H, S, τ), let $(\tilde{H}, S, \tilde{\tau})$ be its pruning. Recalling again that each monomial has pseudoexpectation in the interval $[0, 1]$, we have, with high probability,

$$\widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G}) = n^{c(H)-c(\tilde{H})}\frac{C_H(d)}{C_{\tilde{H}(d)}} p_{(\tilde{H},S,\tilde{\tau})}(x, \boldsymbol{G}) + o(n^{c(H)}) = n^{c(H)} L_{(H,S,\tau)}(M, \pi) C_H(d) + o(n^{\chi(H)}).$$

Taking a union bound over the finitely many extensions of the finitely many pruned partially labelled forests, we are done. □

To prove Proposition 5.10, we need to specialize this result to the case of pruned partially labelled forests with at most two distinguished vertices. These are exactly the paths with labelled endpoints, a pair of labelled vertices, and a single labelled vertex. Recalling the notation of $X \in \mathbb{R}^{nk\times nk}$ as the matrix $X_{(u,i),(v,j)} = \widetilde{\mathbb{E}}\, x_{u,i}x_{v,j}$ and $l \in \mathbb{R}^{nk}$ as the vector with $l_{(u,i)} = \widetilde{\mathbb{E}}\, x_{u,i}$, our argument in Lemma 5.6 may be rephrased to say that the moment constraints on $\widetilde{\mathbb{E}}$ for the first two cases at any error tolerance $\delta' > \delta$ are implied by

$$\langle X_{i,j}, A_{\boldsymbol{G}}^{(s)}\rangle = \pi(i)T_{i,j}^s\|q_s\|_{\mathrm{KM}}^2 n \pm \delta n$$
$$\langle X_{i,j}, \mathbb{J}\rangle = \pi(i)\pi(j)n^2 \pm \delta n^2.$$

The third case, of a single labelled vertex, is implied by

$$\langle l_i, e\rangle = \pi(i)n \pm \delta n$$

Proposition 5.10 now follows from Lemma B.2 and Lemma B.6.

C Bounding Singleton Expectation

Let

$$\zeta_W(A) \triangleq \prod_{i\in V(W)} \beta_i(A)\cdot\gamma_i(A)$$

where γ_i, β_i are boolean functions given by,

$$\beta_i(A) \triangleq 1[i \text{ is } (t, d') \text{ bounded in } A]$$
$$\gamma_i(A) \triangleq 1[i \text{ is NOT in a cycle of length} \leqslant r \text{ in } A]$$

This section is devoted to showing the following bound on the expectation of products involving singleton edges $S(W)$.

Theorem C.1. *For every $d' > d > 1$ and $\delta \in (0,1)$, the following holds for all sufficiently large t. Suppose $S(W)$ is the singleton edges and $J \subseteq D(W)$ a set of duplicative edges in a (k,ℓ)-linkage W, and $g \leqslant \frac{\log n}{\log\log n}$ we have,*

$$\left|\mathbb{E}\left[\prod_{ij\in S(W)}\left(A_{ij}-\frac{d}{n}\right)\cdot\prod_{ij\in J}A_{ij}\cdot\zeta_W(A)\right]\right| \leqslant C\log^2 n\cdot\left(\frac{d}{n}\right)^{|S(W)\cup J|}\cdot n^{0.8e(W)}\cdot 4^{|S(W)|}\delta^{|S(W)|-24kt} \tag{41}$$

for some absolute constant C. Here $e(W)$ is the excess edges in the walk defined as $e(W) = |E(W)| - |V(W)| + 1$.

We wish to emphasize that the key aspect of (41) is the term $\delta^{|S(W)|}$, showing that the expectation decays exponentially in $|S(W)|$.

Proof. Henceforth in this section, We will use S to denote $S(W)$. We begin the proof of the theorem by expanding out the expectation in (41).

$$\begin{aligned}&\mathbb{E}\left[\prod_{ij\in S}\left(A_{ij}-\frac{d}{n}\right)\cdot\prod_{ij\in J}A_{ij}\cdot\zeta_W(A)\right]\\&=\sum_{\alpha\in\{0,1\}^S}\mathbf{Pr}[A_S=\alpha]\mathbb{E}\left[\left(1-\frac{d}{n}\right)^{|\alpha|}\left(-\frac{d}{n}\right)^{|S|-|\alpha|}\cdot\prod_{ij\in J}A_{ij}\cdot\zeta_W(A)\right].\end{aligned}$$

Using $\mathbf{Pr}[A_S=\alpha] = \left(\frac{d}{n}\right)^{|\alpha|}\cdot\left(1-\frac{d}{n}\right)^{|S|-|\alpha|}$, we can simplify the above expression to,

$$=\left(1-\frac{d}{n}\right)^{|S|}\cdot\left(\frac{d}{n}\right)^{S\cup J}\cdot\sum_{\alpha\in\{0,1\}^S}(-1)^{\alpha}\mathop{\mathbb{E}}_{A^c}\left[\zeta_W\left(A^c, A_S=\alpha, A_J=1\right)\right],$$

where A^c denotes the random variables $\{A_{ij} | ij \in \overline{S\cup J}\}$, each of which is an independent Bernoulli random variable with expectation $\frac{d}{n}$.

We will now select a subset of edges $Q \subseteq S(W)$ such that the following two conditions hold:

1. Edges in Q are far from each other in the graph $G(W)$. Formally, for all $ij, i'j' \in Q$,
$$\mathrm{dist}_{G(W)}(i,i') \geqslant 4t\,.$$
2. Neighborhoods of each of the edges in Q have a small number of vertices. Specifically, for all $ij \in Q$, $|\mathsf{B}_{2t}(i, G_0)| \leqslant 2t+2$.

We will show in Lemma C.2 that there exists such a set Q with $|Q| \geqslant |S(W)|/8t - 3k - 6e(W)$.

Let $R \triangleq S(W) \setminus Q$. Let $\alpha = \alpha_Q \cup \alpha_R$ where $\alpha_Q \in \{0,1\}^Q$ and $\alpha_R \in \{0,1\}^R$. We can upper bound the above term by,

$$\leqslant \left(\frac{d}{n}\right)^{S \cup J} 2^{|R|} \cdot \max_{\alpha_R \in \{0,1\}^R} \left| \sum_{\alpha_Q \in \{0,1\}^Q} (-1)^{|\alpha_Q|} \mathop{\mathbb{E}}_{A^c} \left[\zeta_W \left(A^c, A_Q = \alpha_Q, A_R = \alpha_R, A_J = 1\right)\right] \right| \tag{42}$$

For any fixed choice of A^c, let $\zeta_{A^c,\alpha_R} : \{0,1\}^Q \to \{0,1\}$ denote the function,

$$\zeta_{A^c,\alpha_R}(z) \triangleq \zeta_W \left(A^c, A_Q = z, A_R = \alpha_R, A_J = 1\right) .$$

Rewriting the LHS of (42) in terms of ζ_{A^c,α_R},

$$\leqslant \left(\frac{d}{n}\right)^{S \cup J} 2^{|R|} \cdot \max_{\alpha_R \in \{0,1\}^R} \left| \mathop{\mathbb{E}}_{A^c} \left[\sum_{\alpha_Q \in \{0,1\}^Q} (-1)^{|\alpha_Q|} \cdot \zeta_{A^c,\alpha_R}(\alpha_Q) \right] \right|$$

Observe that for any function $\psi : \{0,1\}^Q \to \{0,1\}$, $\sum_{z \in \{0,1\}^Q} (-1)^{|z|} \psi(z) = 0$ if ψ is independent of any bit in z. Otherwise, the sum is upper bounded by $2^{|Q|}$. Therefore, we can rewrite the above bound as,

$$\leqslant \left(\frac{d}{n}\right)^{S \cup J} 2^{|Q \cup R|} \cdot \max_{\alpha_R \in \{0,1\}^R} \left(\mathbf{Pr}_{A^c} \left[\zeta_{A^c,\alpha_R} \text{ depends on all bits in } Q\right]\right) \tag{43}$$

Recall that $\zeta_W(A) = \beta_W(A) \cdot \gamma_W(A)$ where $\beta_W(A) = \prod_{i \in W} \beta_i(A)$ and $\gamma_W(A) = \prod_{i \in W} \gamma_i(A)$.

Analogous to the definition of ζ_{A^C,α_R}, define corresponding boolean functions β_{A^C,α_R} and γ_{A^C,α_R} over $\{0,1\}^Q$, i.e.,

$$\beta_{A^c,\alpha_R}(z) \triangleq \beta_W \left(A^c, A_Q = z, A_R = \alpha_R, A_J = 1\right) .$$
$$\gamma_{A^c,\alpha_R}(z) \triangleq \gamma_W \left(A^c, A_Q = z, A_R = \alpha_R, A_J = 1\right) .$$

By a simple union bound, we can write

$$\begin{aligned}
&\mathbf{Pr}_{A^c} \left[\zeta_{A^c,\alpha_R} \text{ depends on all bits in } Q\right] \\
&\leqslant \sum_{Q' \subset Q} \mathbf{Pr}_{A^c} \left[\beta_{A^c,\alpha_R} \text{ depends on all bits in } Q' \bigwedge \gamma_{A^c,\alpha_R} \text{ depends on all bits in } Q \setminus Q'\right] \\
&\leqslant \sum_{Q' \subset Q} \min \left(\mathbf{Pr}_{A^c} \left[\beta_{A^c,\alpha_R} \text{ depends on all bits in } Q'\right], \mathbf{Pr}_{A^c} \left[\gamma_{A^c,\alpha_R} \text{ depends on all bits in } Q \setminus Q'\right]\right)
\end{aligned} \tag{44}$$

We will the probabilities in the above sum in Claim C.4 and Claim C.3 respectively. Substituting these bound on probabilities into (44)

$$\leqslant \sum_{Q' \subset Q} \min \left(\delta^{16t|Q'|}, C(\log^2 n) n^{-0.7(|Q \setminus Q'|/r - \#_c(Q \cup R \cup J)}\right) \tag{45}$$
$$\leqslant C(\log^2 n) n^{0.7\#_c(Q \cup R \cup J)} \sum_{Q' \subset Q} \cdot \min((\delta^{16t})^{|Q'|}, (n^{-0.7/g})^{|Q \setminus Q'|}) \tag{46}$$

$$\leqslant C(\log^2 n)n^{0.7e(W)} \cdot 2^{|Q|} \cdot (\delta^{16t})^{|Q|/2} \tag{47}$$
$$\leqslant C(\log^2 n)n^{0.7e(W)} \cdot 2^{|S(W)|} \cdot (\delta)^{|S(W)|-24kt-48e(W)} \tag{48}$$
$$\leqslant C\log^2 n \cdot n^{0.8e(W)} \cdot 2^{|S(W)|} \cdot \delta^{|S(W)|-24kt} \tag{49}$$

Substituting back in (43) we get the bound in the theorem.

□

Lemma C.2. *For all $t < \ell$, in a $k \times \ell$-linkage there exists $Q \subset S(W)$ with $|Q| \geqslant \frac{|S(W)|}{8t} - 3\ell - 6e(W)$ such that,*

1. *For all $ij, i'j' \in Q$, $\mathrm{dist}_{G(W)}(i, i') \geqslant 4t$.*
2. *For all $ij \in Q$, $|\mathsf{B}_{2t}(i, G_0)| \leqslant 2t + 2$.*

Proof. All the steps of the walk are divided into consecutive segments of singleton edges ("singleton stretches") and duplicative edges ("duplicative stretch").

The walk can step from a singleton stretch into a duplicative stretch, either by a turn-around or at an edge that creates a cycle. The number of such transitions is therefore at most $\ell + 2e(W)$ where $2e(W)$ is the number of excess edges.

Hence there are $|S(W)|$ singletons split into $\ell + 2e(W)$ disjoint path segments. Given a path of length Δ, delete segments of length $8t$ from both end, and then pick edges at a regular intervals of length $8t$ from the each other in the remaining. This yields $\lfloor \frac{(\Delta-16t)}{8t} \rfloor$ edges which are pairwise $8t$ away, and the $2t$ neighborhood around each of them is a path and thus has only $2t + 2$ edges. Perform this operation on each of the singleton segments to select a subset Q of singleton edges. By construction edges in Q satisfy the conditions of the Lemma above. It remains to lower bound the size of Q. If $\Delta_1, \ldots, \Delta_q$ are the lengths of the singleton stretches, we can write

$$\begin{aligned} |Q| \geqslant \sum_{i=1}^{q} \left\lfloor \frac{(\Delta_i - 16t)}{8t} \right\rfloor &\geqslant \sum_{i=1}^{q} \frac{(\Delta_i - 16t)}{8t} - 1 \\ &\geqslant \frac{|S(W)|}{8t} - 3q \geqslant \frac{|S(W)|}{8t} - 3\ell - 6e(W) \end{aligned}$$

edges. □

C.1 Away from short cycles

Claim C.3. For any subset $Q^* \subset Q$,

$$\mathbf{Pr}_{A^c}\left[\gamma_{A^c,\alpha_R} \text{ depends on all bits in } Q^*\right] \leqslant C(\log^2 n)n^{-0.7(|Q^*|/r - \mathrm{NumCycles}(Q \cup R \cup J)}$$

Proof. The function $\gamma_{A^c,\alpha_R}(z) = \gamma_W(A^c, A_Q = z, A_R = \alpha_R, A_J = 1)$ is a anti-monotone function of z.

For every pair $ij \in Q^*$, since γ_{A^c,α_R} depends on z_{ij} there is some setting of $z_{Q\setminus\{ij\}}$ such that $\gamma_{A^c,\alpha_R}(z_{ij} = 0, z_{Q\setminus\{ij\}}) = 1$ but $\gamma_{A^c,\alpha_R}(z_{ij} = 1, z_{Q\setminus\{ij\}}) = 0$. By definition of γ_W, this implies that addition of edge ij creates a cycle of length at most r.

Therefore, in the graph given by $A' = (A^c, A_R = \alpha_R, A_J = 1, A_Q = 1)$ every edge $ij \in Q^*$ is in a cycle of length at most r. There are at least $|Q^*|/r$ cycles in the graph A', and at least $|Q^*|/r - \text{NumCycles}(Q \cup R \cup J)$ involve edges of the random graph A^c.

Now we appeal to Lemma A.3 in [FM17] to conclude the claim. □

C.2 Heavy Vertices

The goal of this section is to prove the following claim, a component in the proof of Theorem C.1.

Claim C.4. Given $d' > d > 1$ and $\delta > 0$, for all sufficiently large value of t the following holds for every subset $Q^* \subset Q$,

$$\mathbf{Pr}_{A^c}\left[\beta_{A^c,\alpha_R} \text{ depends on all bits in } Q^*\right] \leqslant (\delta)^{t|Q^*|}$$

First, let us setup some notation. For a graph G and a set of vertices S, we make the following definitions.

$$\mathsf{B}_r(S, G) \triangleq \{i | \operatorname{dist}_G(i, S) \leqslant r\}$$

$$\mathsf{N}_r(S, G) \triangleq \{i | \operatorname{dist}_G(i, S) = r\}$$

Here dist_G refers to the shortest path distance on the graph G. We borrow the following tail bound on the sizes of neighborhoods in $\mathcal{G}(n, \frac{d}{n})$ from [FM17].

Lemma C.5. *Fix $d > 1$ and consider the Erdos-Renyi graph $G \sim \mathcal{G}(n, \frac{d}{n})$. Then there exists $C, c > 0$ such that for any $s \geqslant 0$, $t \geqslant 1$ and $v \in [n]$*

$$\mathbf{Pr}\left[|\mathsf{B}_t(v; G)| \geqslant sd^t\right] \leqslant Ce^{-cs}$$

Critical to the proof of Claim C.4 is the notion of being heavy vertex, and close-to heavy vertices. A heavy vertex is any vertex with $\mathsf{B}_t(v, G) \geqslant (d')^t$. A vertex is marked as close-to-heavy if it is within distance t of a heavy vertex. Formally, we have the following definition

Definition C.6. A vertex v in a graph $G = (V, E)$ is (t, d')-*close to heavy* if there exists v' such that $\operatorname{dist}_G(v, v') \leqslant t$ such that $|\mathsf{B}_t(v'; G)| > (d')^t$.

First, we will bound the probability that a vertex in an Erdős-Rényi graph is (t, d')-close to heavy.

Lemma C.7. *There exists absolute constant C, c such that for all t and $d' > d$, for a graph $G \sim \mathcal{G}(n, \frac{d}{n})$ and a vertex v,*

$$\mathbf{Pr}_G\left[v \text{ is } (t,d')\text{-close to heavy}\right] \leqslant Cd^t e^{-c(d'/d)^t}$$

Proof. Let X be the random variable denoting the number of vertices that are (t,d')-close to heavy in a graph $G \sim \mathcal{G}(n, \frac{d}{n})$. Clearly the above probability is given by $\frac{1}{n}\mathbb{E}[X]$. Suppose a vertex v has $|\mathsf{B}_t(v;G)| = \gamma(d')^t$ for some $\gamma > 1$. Then every vertex $u \in \mathsf{B}_t(v;G)$ is (t,d')-close to heavy. Therefore, we can upper bound the expected number of vertices that are (t,d')-close to heavy in a graph $G \sim \mathcal{G}(n, \frac{d}{n})$ by,

$$\begin{aligned}
\frac{1}{n}\mathbb{E}[X] &\leqslant \int_{\gamma=1}^{\infty} \mathbf{Pr}[|\mathsf{B}_t(v;G)| = \gamma(d')^t] \cdot (\gamma(d')^t) d\gamma \\
&\leqslant \int_{s=(d'/d)^t}^{\infty} \mathbf{Pr}[|\mathsf{B}_t(v;G)| = s(d)^t] \cdot (sd^t) \cdot \left(\frac{d^t}{(d')^t} ds\right) \\
&\leqslant \frac{d^{2t}}{(d')^t} \int_{s=(d'/d)^t}^{\infty} Ce^{-cs} s ds \leqslant \frac{C}{c^2} \cdot \frac{d^{2t}}{(d')^t} \cdot \left[-e^{-z}z - e^{-z}\right]_{c(d'/d)^t}^{\infty} < C' d^t e^{-c(d'/d)^t}
\end{aligned}$$

where the last inequality holds whenever $(d'/d) > 1$ and $C' > \frac{C(1+c)}{c^2}$. □

The following Lemma upper bounds the probability of a vertex v being close to heavy in a more complicated setup. Here a subgraph $G' = (V', E')$ is chosen to be included in the graph, and a set of vertices $\mathcal{F}$ are forbidden in the neighborhood of v.

Lemma C.8. *There exists absolute constants C, c such that the following holds for all t and $d' > d > 1$.*

Suppose $G' = (V', E')$ be a subgraph of the complete graph and let $v \in V'$ be a vertex in G'.

Let $\mathcal{F} \subseteq [n]$ be a set of vertices disjoint from V', i.e., $\mathcal{F} \cap V' = \emptyset$. Suppose we draw $G^c \sim \mathcal{G}(n, \frac{d}{n})$ and set $G = G' \cup G^c$ then,

$$\mathbf{Pr}_G\left[v \text{ is } (t,d')\text{-close to heavy in } G | \mathsf{B}_{2t}(v,G) \cap \mathcal{F} = \emptyset\right] \leqslant C|V'|d^t e^{-c\frac{1}{|V'|+1}\cdot\left(\frac{d'}{d}\right)^t}$$

Proof. Notice that the indicator of the event

$$\mathcal{E}_1 = 1[v \text{ is } (t,d')\text{-close to heavy in } G]$$

is a monotone function of the edges G^c. On the other hand, the event $\mathcal{E}_2 = 1[\mathsf{B}_{4t}(v,G) \cap \mathcal{F} = \emptyset]$ is an anti-monotone function.

By FKG inequality, the two events are negatively correlated and therefore conditioning on $\mathcal{E}_2$ reduces the chance of $\mathcal{E}_1$, i.e.,

$$\mathbf{Pr}_G\left[v \text{ is } (t,d')\text{-close to heavy in } G | \mathsf{B}_{4t}(v,G) \cap \mathcal{F} = \emptyset\right] \leqslant \mathbf{Pr}_G\left[v \text{ is } (t,d')\text{-close to heavy in } G\right].$$

Now we make the following claim which will prove subsequently.

Claim C.9. Let $\rho \triangleq \left(\frac{1}{|V'|+1}\right)^{1/t}$. If no vertex $w \in V'$ is $(t, \rho d')$-close to heavy in G^c, then v is not (t, d')-close to heavy in G.

Assuming the above claim, we can use the union bound to argue

$$\begin{aligned}\mathbf{Pr}_G\left[v \text{ is } (t,d')\text{-close to heavy in } G\right] &\leqslant \mathbf{Pr}_{G^c}\left[\exists u \in V' \text{ which is } (t,\rho d')\text{-close to heavy in } G^c\right] \\ &\leqslant \sum_{u \in V'} \mathbf{Pr}_{G^c}\left[u \text{ is } (t,\rho d')\text{-close to heavy in } G^c\right] \\ &\leqslant C|V'|d^t e^{-c\frac{1}{|V'|+1}\cdot\left(\frac{d'}{d}\right)^t}\end{aligned}$$

where the last inequality follows from Lemma C.7 □

Now we return to proving Claim C.9.

Proof. (Proof of Claim C.9) Suppose $v \in V'$ is (t, d')-close to heavy in G, and let $u \in [n]$ be the heavy vertex with $\mathrm{dist}_G(u, v) \leqslant t$.

Now we will lower bound $|\mathsf{B}_t(u, G^c)|$. To this end, consider any $u' \in [n]$ with $\mathrm{dist}_G(u, u') \leqslant t$. The path from $u \to u'$ is either completely contained in G^c in which case $\mathrm{dist}_{G^c}(u, u') \leqslant t$ or the path from $u \to u'$ uses edges in G' which implies that $u' \in \mathsf{B}_t(V', G^c)$. Therefore, we can write

$$|\mathsf{B}_t(u, G)| \leqslant |\mathsf{B}_t(u, G^c)| + |\mathsf{B}_t(V', G^c)|.$$

Since u is (t, d')-heavy, $|\mathsf{B}_t(u, G)| \geqslant (d')^t$. If no vertex $w \in V'$ is $(t, \rho d')$-close to heavy in G^c, then

$$|\mathsf{B}_t(V', G^c)| \leqslant \sum_{w \in V'} |\mathsf{B}_t(w, G^c)| \leqslant |V'| \cdot \left(\rho d'\right)^t$$

One can thus conclude that,

$$|\mathsf{B}_t(u, G^c)| \geqslant (d')^t \left(1 - \rho^t |V'|\right) \geqslant (\rho d')^t .$$

Finally since $\mathrm{dist}_G(v, u) \leqslant t$, there exists some vertex $w \in V'$ such that $\mathrm{dist}_{G^c}(w, u) \leqslant t$. Thus w is $(t, \rho d')$-close to heavy in G^c. □

Lemma C.10. *For every $d' > d$ and $\delta > 0$, there exists t such that the following holds. Fix a subset $V_0 \subset [n]$ of vertices and a graph $G_0 = (V_0, E_0)$ with at most $|E_0| < \log^2 n$ edges. Suppose $V^* \subset V_0$ be such that,*

1. *For every vertex $i \in V^*$, $|\mathsf{B}_{2t}(i; G_0)| < t^2$.*
2. *$\mathrm{dist}_{G_0}(i, j) \geqslant 4t$ for all $i, j \in V^*$.*

Then if we sample a graph G by including each of the remaining edges $\binom{n}{2} - E_0$ independently with probability $\frac{d}{n}$,

$$\mathbf{Pr}\left[\forall v \in V^*, v \text{ is } (t, d')\text{-close to heavy in } G\right] \leqslant \delta^{t|V^*|}$$

Proof. Let $G^c = ([n], E^c)$ denote the graph consisting of edges in $\binom{n}{2} - E_0$ each of which is included independently with probability $\frac{d}{n}$.

Consider the neighborhood $\mathsf{B}_{2t}(v; G)$ around a vertex $v \in V^*$. Clearly, the neighborhood contains the sub-graph $\mathsf{B}_{2t}(v, G_0)$ since $G_0 \subset G$. All the additional vertices (and edges) in $\mathsf{B}_{2t}(v; G)$ are those reachable by taking the newly sampled edges in G^c.

Intuitively, up to constant distances, the graph G^c will be "tree-like". More specifically, for a typical sample, one would expect that the neighborhood can be decomposed as,

$$\mathsf{B}_{2t}(v, G) = \mathsf{B}_{2t}(v, G_0) \cup \bigcup_{w \in \mathsf{B}_{2t}(v, G_0)} \mathsf{T}_w$$

where T_w is a tree with vertex w as root, and no other vertices in V_0. Call a vertex $v \in V^*$ to be *typical* if the above assumptions hold.

We will first show that there is a significant fraction of vertices in $|V^*|$ are *typical* with all but negligible probability.

To this end, consider the graph $\mathcal{H}$ formed by the edges in

$$E[\mathsf{B}_{2t}(V^*; G)] - E[\mathsf{B}_{2t}(V^*; G_0)]\,,$$

where $E[\mathcal{S}]$ denotes the set of edges contained in a set of vertices $\mathcal{S}$.

Consider a vertex $v \in V^*$. For every vertex $w \in \mathsf{B}_{2t}(v; G_0)$ and $\mathrm{dist}(w, v) = d$, the graph $\mathcal{H}$ contains the subgraph $\mathsf{B}_{2t-d}(w, G) - \mathsf{B}_{2t-d}(w, G_0)$. In fact, in a typical vertex $v \in V^*$, this would be a tree of depth $2t - d$ with vertex w as root.

Claim C.11. The number of typical vertices is at least $|V^*| - 2s$ where $s \triangleq \#_c(\mathsf{B}_{2t}(V_0; G)) - \#_c(G_0)$.

Proof. Consider the execution of a depth-first-traversal on the graph $\mathcal{H}$. More precisely, consider the execution of the following algorithm:

- ExploreGraph()
  - Set $visited[w] = false$ for all $w \in \mathcal{H} \cup V_0$
  - For each vertex $w \in \mathsf{B}_{2t}(V^*; G_0)$
    - If $visited[w] = false$ then Mark w as *isolated* and $Explore(w)$
- Explore(v)

- for each edge $(v, w) \in \mathcal{H}$ do
  - If $w \in V_0$, mark (v, w) as *stale edge* and set $visited[w] = true$.
  - If $w \notin V_0$ and $visited[w] = true$, mark the edge (v, w) as *back edge*
  - If $w \notin V_0$ and $visited[w] = false$ set visited[w] = true and call Explore(w)

Execution of ExploreGraph will consist of a sequence of DFS traversals each producing a connected component of $\mathcal{H}$. Each traversal starts at some node $w \in \mathsf{B}_{2t}(V^*, G_0)$ that has not been visited yet. The traversal goes through edges in $\mathcal{H}$, visiting new nodes, marking some edges as back and stale.

Observe that every stale edge or a back-edge increases the cycle number of $\mathsf{B}_2t(V_0; G)$ by adding an edge, but no new vertex. Therefore, the total number of stale/back edges is at most $\#_c(\mathsf{B}_{2t}(V_0; G)) - \#_c(G_0)$. For brevity, let us denote $s \triangleq \#_c(\mathsf{B}_{2t}(V_0; G)) - \#_c(G_0)$.

A vertex $v \in V^*$ is typical if the following hold:

1. Every vertex $w \in \mathsf{B}_{2t}(v; G_0)$ is marked *isolated* (never visited via a stale edge).
2. For every vertex $w \in \mathsf{B}_{2t}(v; G_0)$, the corresponding call $Explore(w)$ did not produce a *stale* or *back* edge in one of its descendants.

   Since there

As there are at most s-stale edges, at most s vertices $v \in V^*$ have some vertex $w \in \mathsf{B}_{2t}(v, G_0)$ visited by a stale edge. Furthermore, at most s vertices $v \in V^*$ have a vertex $w \in \mathsf{B}_{2t}(v, G_0)$ that produced a *stale* or *back* edge. Hence at least $|V^*| - 2s$ vertices are typical. □

Returning to the proof of Lemma C.10, let $V^*_{\mathsf{typ}} = \{i_1, \ldots, i_R\} \subseteq V^*$ denote the set of *typical* vertices in V^*. Now we will describe how to sample a graph $G = G^c \cup G_0$ from the conditional distribution: $\big(G \mid V^*_{\mathsf{typ}} \text{are typical}\big)$.

- For $j = 1$ to R
  - Sample $\mathsf{B}_{2t}(i_j; G)$ conditioned on i_j being *typical* or equivalently, $\mathsf{B}_{2t}(i_j; G) \cap G_{j-1} = \mathsf{B}_{2t}(i_j, G_0)$.

    For sake of concreteness, we will outline how to sample $\mathsf{B}_{2t}(i_j; G)$. For each vertex $w \in \mathsf{B}_{2t}(i_j, G_0)$, with distance $\mathrm{dist}(w, i_j) = D$, sample the local neighborhood tree T_w of depth D in a breadth first manner, while avoiding vertices in G_{j-1}.
  - $G_j = G_{j-1} \bigcup \mathsf{B}_{2t}(i_j; G)$.
- Sample all the remaining unrevealed edges by including them independently with probability $\frac{d}{n}$, conditioned on V^*_{typ} being typical.

By virtue of the above order of sampling, we can write

$$\mathbf{Pr}\left[\forall v \in V^*_{\mathsf{typ}}, v \text{ is } (t,d')\text{-close to heavy in } G | V^*_{\mathsf{typ}} \text{ is typical}\right] \tag{50}$$

$$= \prod_{j=1}^{R} \mathbf{Pr}\left[i_j \text{ is } (t,d')\text{-close to heavy in } G_j \middle| G_{j-1}\right] \tag{51}$$

where recall that G_j is sampled conditioned on i_j being typical.

By virtue of being typical, the neighborhood of i_j, $\mathsf{B}_{2t}(i_j, G_{j-1})$ is same as its original neighborhood $\mathsf{B}_{2t}(i_j, G_0)$ in G_0. Let $\mathcal{F}_j$ be the remaining vertices in G_{j-1} namely,

$$\mathcal{F}_j = G_{j-1} - \mathsf{B}_{2t}(i_j, G_0)$$

Since i_j conditioned on being typical, vertices in $\mathcal{F}_j$ are forbidden to be chosen in the neighborhood $\mathsf{B}_{2t}(i_j; G_j)$. Therefore deleting all edges among $\mathcal{F}_j$ has no effect on whether i_j is (t,d')-close to heavy. That is,

$$i_j \text{ is } (t,d')\text{-close to heavy in } G_j \iff i_j \text{ is } (t,d')\text{-close to heavy in } \mathsf{B}_{2t}(v, G_0) \cup G^c$$

We will bound the probability of the latter event using Lemma C.8. Specifically, apply Lemma C.8 with $G' = \mathsf{B}_t(i_j, G_0)$ and $\mathcal{F}_j$ we get that,

$$\mathbf{Pr}\left[i_j \text{ is } (t,d')\text{-close to heavy in } |\mathsf{B}_{2t}(v,G) \cap \mathcal{F} = \emptyset\right] \leqslant Ct^2 \cdot d^t e^{-c\left(\frac{d'}{d}\right)^t \cdot \frac{1}{1+t^2}} \triangleq \Delta(d,d',t)$$

Substituting back in (50),

$$\mathbf{Pr}\left[\forall v \in V^*_{\mathsf{typ}}, v \text{ is } (t,d')\text{-close to heavy in } G | V^*_{\mathsf{typ}} \text{ is typical}\right] \leqslant \Delta(d,d',t)^{|V^*_{\mathsf{typ}}|} \leqslant \Delta(d,d',t)^{|V^*|-2s} \tag{52}$$

where recall that $s \triangleq \#_c(\mathsf{B}_{2t}(V_0, G)) - \#_c(G_0)$. By Lemma A.3 in [FM17], we know that or all $k < \log^2 n$,

$$\mathbf{Pr}[\#_c(\mathsf{B}_{2t}(V_0,G)) - \#_c(G_0) \geqslant k] \leqslant C(\log^2 n) n^{-0.7k}$$

Along with (52), this implies that

$$\mathbf{Pr}\left[\forall v \in V^*, v \text{ is } (t,d')\text{-close to heavy in } G\right]$$
$$\leqslant \mathbf{Pr}\left[s > \frac{|V^*|}{4}\right] + \Delta(d,d',t)^{|V^*|-2(\frac{|V^*|}{4})} \leqslant 2\Delta(d,d',t)^{|V^*|/2}$$

for large enough n. Finally, the lemma follows by observing that for all fixed d, d', δ, we can make $\Delta(d,d',t) \leqslant \delta^{4t}$ for sufficiently large t.

□

We now have all the pieces need to prove Claim C.4.

Proof. (Proof of Claim C.4) The function $\beta_{A^c,\alpha_R}(z) = \beta_W(A^c, A_Q = z, A_R = \alpha_R, A_J = 1)$ is a anti-monotone function of z.

For every pair $ij \in Q^*$, since β_{A^c,α_R} depends on z_{ij}, there is some setting of $z_{Q\setminus\{ij\}}$ such that $\beta_{A^c,\alpha_R}(z_{ij} = 0, z_{Q\setminus\{ij\}}) = 1$ but $\beta_{A^c,\alpha_R}(z_{ij} = 1, z_{Q\setminus\{ij\}}) = 0$. This implies that addition of edge ij creates a vertex v that is not (t, d')-bounded. The vertex v is within distance t of both endpoints i and j, since otherwise the addition of the edge ij has no effect on the (t, d')-boundedness of vertex v.

Therefore we can upper bound,

$$\begin{aligned}&\mathbf{Pr}_{A^c}\left[\beta_{A^c,\alpha_R} \text{ depends on all bits in } Q^*\right]\\ &\leqslant \mathbf{Pr}_{A^c}\left[\forall ij \in Q^*, i \text{ is } (t,d')\text{-close to heavy in graph } A^c \cup J \cup Q \cup R\right]\end{aligned}$$

By construction of the set of edges Q, the set $V^* = \{i | ij \in Q^*\}$ satisfy the conditions in the hypothesis of Lemma C.10 in the graph formed by $Q \cup R \cup J$. Hence the claim follows by appealing to Lemma C.10. □

D Robustness in the Stochastic Block Model

In this section, we will show that the local statistic SDP relaxation yields a robust algorithm. Throughout this section, let $\boldsymbol{G}$ be drawn from either the Erdős-Rényi or Stochastic Block Model on n vertices, with average degree d. We will prove that an adversarial modification of ϵn edges, for sufficiently small ϵ, cannot meaningfully later subgraph occurrences, except by creating vertices of high degree. Therefore, if we run the Local Statistics SDP after deletion of sufficiently high-degree vertices, the resulting algorithm is robust to adversarial edge meddling.

Let us make this intuition precise. In a similar vein to the partially labelled graph formalism from the main body of the paper, let us define now a *pinned graph* to be a pair (H, R) where $R \subset V(H)$ contains exactly one vertex from each connected component of H. Write $\ell(H) = |R|$ for the number of such components. Given a graph G and a subset T of $\ell(H)$ vertices, an *occurrence* of (H, R) in (G, T) is an (injective?) homomorphism that maps the pinned vertices R to the target set T. Let's write $\Gamma_{H,R}(G, T)$ for the set of such occurrences.

Claim D.1. For every pinned graph (H, R) and any then for any $T \subset [n]$,

$$\lim_{n\to\infty} \mathop{\mathbb{E}}_{G}[|\Gamma_{H,W(H)}(G, T)|^2] = c_H$$

for a constant c_H dependent only on H.

Proof. First let us consider the following expectation:

$$\mathbb{E}[|\Gamma_{H,R}(\boldsymbol{G}, S)|] = \sum_{\phi: V(H)\to[n]} \mathbb{P}\left[\phi \text{ is an occurrence}\right]$$

The number of nonzero terms in the summation is $\binom{n}{|V(H)|-\ell(H)}(|V(H)| - \ell(H))! = n^{|V(H)|-\ell(H)} + O(n^{|V(H)|-\ell(H)})$. For each term, the probability that ϕ is an occurrence is $O(n^{-|E(H)|})$ Since $|E(H)| \geqslant |V(H)| - \ell(H)$ in a graph with at most ℓ connected components, the above expectation is a constant depending on graph H.

Now we turn our attention to $\mathbb{E}[|\Gamma_{H,R}(\boldsymbol{G}, T)|^2]$, which we can expand as

$$\mathbb{E}[|\Gamma_{H,R}(\boldsymbol{G}, T)|^2] = \sum_{\phi,\psi: V(H)\to[n]} \mathbb{P}[\phi, \psi \text{ are occurrences}]$$

Each nonzero term gives rise to a graph H^* obtained by taking the union of the images of $\phi(H)$ and $\psi(H)$; this union is a graph with ℓ connected components, each of which contains one of the target vertices T. There are only finitely many graphs on at most $2|V(H)|$ vertices that have this form, so we can write the expectation of concern to us as a sum of expected occurrence counts of these types, and apply our initial observation. □

Lemma D.2. *Fix $d > 0, \epsilon \in (0,1)$, and a finite pinned graph (H, R). There exists $\Delta(H, R, d, \epsilon) > 0$ such that the following holds: with probability $1 - \epsilon$ for all $Q \subset [n]$ with $|Q| \leqslant \Delta(H, R, \epsilon)n$,*

$$\left|\bigcup_{T\cap Q\neq\emptyset} \Gamma_{H,R}(\boldsymbol{G}, T)\right| \leqslant \epsilon n^{\ell(H)}$$

Proof. We can expand the size of this union as a sum over subsets $T \in \binom{[n]}{\ell(H)}$:

$$\begin{aligned}\left|\bigcup_{T\cap Q\neq\emptyset} \Gamma_{H,R}(\boldsymbol{G}, T)\right| &= \sum_{T\in\binom{[n]}{\ell(H)}} |\Gamma_{H,R}(\boldsymbol{G}, T)| \cdot \mathbf{1}\left[T \cap Q \neq \emptyset\right] \\ &\leqslant \left(\sum_{T\in\binom{[n]}{\ell}} |\Gamma_{H,R}(\boldsymbol{G}, T)|^2\right)^{1/2} \cdot \left(\sum_{T\in\binom{[n]}{\ell}} \mathbf{1}\left[T \cap Q \neq \emptyset\right]^2\right)^{1/2} \qquad (53)\end{aligned}$$

With probability at least $1 - \epsilon$, we have

$$\sum_{T\in\binom{[]n]}{\ell(H)}} |\Gamma_{H,R}(\boldsymbol{G}, T)|^2 \leqslant \frac{1}{\epsilon}\mathbb{E}\left[\sum_{T\in\binom{[n]}{\ell(H)}} |\Gamma_{H,R}(\boldsymbol{G}, T)|^2\right] \leqslant \frac{c_H}{\epsilon} n^{\ell(H)}$$

where c_H is the constant depending on H from Claim D.1. Set $\Delta(H, d, \epsilon) = \frac{\epsilon^3}{\ell c_T}$. Notice that for a set Q smaller than $\Delta(H, d, \epsilon)$, the number of $T \subset \binom{[n]}{\ell(H)}$ is at most $\ell \cdot \frac{\epsilon^3}{\ell(H)c_H} n^{\ell(H)} = \frac{\epsilon^3}{c_H} n^{\ell(H)}$.

Conditioned on this event of probability $1-\epsilon$, we can use (53) to conclude that,

$$\left|\bigcup_{T\cap Q\neq\emptyset}\Gamma_{H,R}(\boldsymbol{G},T)\right|\leqslant \epsilon n^{\ell(H)}$$

whenever $|Q|\leqslant \Delta(H,d,\epsilon)n$. □

By taking a union bound over all trees of size k and all choices of designated vertices, we have the following corollary.

Corollary D.3. *For every $d,k>0$ and $\epsilon\in(0,1)$, there exists η such that following holds. Denoting by $\mathcal{H}$ the set of all graphs with at most m edges, then with probability $1-\epsilon$, for all $Q\subset[n], |Q|\leqslant\eta n$ and $H\in\mathcal{H}$ we have*

$$|\Gamma_H(\boldsymbol{G},Q)|\leqslant\epsilon n^{\ell(H)}.$$

Now we are ready to prove the main theorem of this section, namely robustness of local statistics SDP relaxation.

Theorem D.4. *(Robustness of Local Statistics SDP) For every d,ϵ,k, there exist B and γ such that, with probability at least $1-\epsilon$ over $\boldsymbol{G}=(\boldsymbol{G},\boldsymbol{E})$, the following holds:*

Let $\tilde{G}=([n],\tilde{E})$ be an arbitrary graph such that $|\boldsymbol{E}\triangle\tilde{E}|\leqslant\gamma n$; write $G^*=([n],E^*)$ for the graph obtained by deleting edges incident to all vertices of degree $>B$ in $\tilde{G}$. Then for every graph H with at most m edges,

$$|\Gamma_H(G)\triangle\Gamma_H(G^*)|\leqslant\epsilon n^{\ell(H)}$$

Consequently, if $\widetilde{\mathbf{E}}:\mathbb{R}[x]_{\leqslant 2}\to\mathbb{R}$ is a pseudoexpectation that is a feasible solution to the level $(2,m)$ local statistics SDP on G^* (or G) with tolerance δ, then $\widetilde{\mathbf{E}}$ is a feasible solution on level $(2,m)$ local statistics SDP with tolerance $\delta+\epsilon$ on G (or G^*). Further, if $\widetilde{\mathbf{E}}$ is infeasible for the level $(2,m)$ local statistics SDP on G^* (or G) by a margin of δ, then $\widetilde{\mathbf{E}}$ remains infeasible on the level $(2,m)$ local statistics SDP by margin of $\delta-\epsilon$ on G (or G^*).

Proof. Let $\eta>0$ be the choice for which Corollary D.3 holds given $d,k,\epsilon/4$. Set $B\triangleq\lceil\frac{2d}{\eta}\rceil$ and $\gamma=\frac{\epsilon}{4m2^mB^{m^3}}$. We will express $\Gamma_H(\boldsymbol{G})\triangle\Gamma_H(G^*)=\Gamma_{del}\cup\Gamma_{trunc}\cup\Gamma_{add}$ and bound the size of each of the three sets.

- $\Gamma_{del}=\Gamma_H(\boldsymbol{G})-\Gamma_H(\tilde{G})$ are the occurrences of H in $\boldsymbol{G}$ that were deleted by the adversarial corruption of edges.

  Since the corruption deletes at most γn edges, which are incident on at most $2\gamma n<\eta n$ vertices, we can use Corollary D.3 to conclude that this set is at most $\epsilon n^{\ell(H)}/4$

- $\Gamma_{trunc}=(\Gamma_H(\boldsymbol{G})\cap\Gamma_H(\tilde{G}))\setminus\Gamma_H(G^*)$ are the occurrences of H that were deleted due to the removal of edges incident to high-degree vertices while constructing G^*.

  The average degree of the graph $\boldsymbol{G}$ is $d+o(1)$ with $1-o_n(1)$. Therefore, the average degree of $\tilde{G}$ is at most $d+2\gamma<2d$. Hence, the number of vertices of degree $>B$ is at most $(2d/B)\cdot n<\eta n$. Again by Corollary D.3, $|\Gamma_{trunc}|\leqslant\epsilon n/4$.

- $\Gamma_{add} = \Gamma_H(G^*) \setminus \Gamma_H(\boldsymbol{G})$ are the occurrences of H in $\boldsymbol{G}$ that were added by the adversarial corruption, and survived the truncation of high-degree vertices.

  Every occurrence in Γ_{add} includes one of the γn edges in $\tilde{E} - E$.

  Since the degree of each vertex of G^* is at most B, there are at most B^m vertices in their neighborhood of radius m around every vertex v. Hence, for any given connected component $\mathcal{C} \subseteq H$, the number of occurrences of $\mathcal{C}$ that contain a vertex $i \in [n]$ is at most $|\mathcal{C}| \cdot (B^m)^{|\mathcal{C}|}$.

  For every edge $e = (u, v) \in \tilde{E} - E$, there are at most $2B^m$ vertices in their neighborhood of radius m. The number of occurrences of any connected component $\mathcal{C}$ in this neighborhood is thus at most $(2B^m)^{|\mathcal{C}|}$.

  Hence the number of occurrences that use at least one edge in $|\tilde{E} - E|$ is at most

  $$\sum_{\mathcal{C} \subset H} \left(n^{\ell(H)-1} \cdot (B^m)^{|V(H)|-|\mathcal{C}|} \right) \cdot (|(\tilde{E} - E) \cap E^*| \cdot (2B^m)^{|\mathcal{C}|} \leqslant \ell(H) \cdot 2^m B^{m^2 \ell(H)} \gamma n^{\ell(H)}$$

  By the choice of γ, the desired bound follows.

Conditioned on the event that assertion in Corollary D.3 holds, for every choice of corruptions, we have that

$$\Gamma_H(\boldsymbol{G}) \triangle \Gamma_H(G^*) \leqslant \epsilon n/4 + \epsilon n/4 + \epsilon n/4 < \epsilon n$$

The claim about the solution to the level $(2, m)$-local statistics SDP is immediate by observing that for any partially labelled subgraph (H, S, τ),

$$\widetilde{\mathbf{E}}[|p_{H,S,\tau}(G^*, x) - p_{H,S,\tau}(G, x)]| \leqslant |\Gamma_H(G) \triangle \Gamma_H(G^*)|$$

for any $\widetilde{\mathbf{E}}$ that satisfies $\mathcal{B}_k$. □