Local Statistics, Semidefinite Programming, and Community Detection*

Jess Banks† Sidhanth Mohanty‡ Prasad Raghavendra§

Abstract

We propose a new, efficiently solvable 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 adversarial perturbation of a *linear* number of edges. 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.

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 into k equal-sized 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 $d < d_{IT}$

*Full version at https://arxiv.org/abs/1911.01960

†JB is supported by the NSF Graduate Research Fellowship Program under Grant DGE-1752814

‡Supported by NSF grant CCF-1718695..

§Supported by NSF grant CCF-1718695.

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, AS18]. 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. More recently, elaborate spectral methods such as those in [ABARS20, SM19, AR20] have been shown to be robust to adversarial perturbations effecting $O(n^\epsilon)$ or $O(\log^\epsilon n)$ vertices respectively. Other robust algorithms for recovery are known, but only when the edge-densities are significantly higher than Kesten-Stigum [GV16, MMV16, CSV17, SVC16]. Finally, Montanari and Sen in [MS15] study an 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 $o(|E|)$ adversarial edge perturbations. 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 $D_G, D_x \in \mathbb{N}$. In the setting of this paper, the $\text{LoSt}(D_G, D_x)$ SDP has size $O(n^{D_x})$, with a constant (in n) number of affine constraints dependent on D_x, D_G, and the number of coommunities k.[1] 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. Though natural in hindsight, the definition of Local Statistics SDP hierarchy is the main conceptual contribution of this paper.

We will demonstrate the power of the *Local Statistics* hierarchy through two test cases, namely 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

[1] In particular, there will be one affine constraint for each partially labelled graph with D_G edges and D_x $[k]$-labelled vertices (see Definition 4).

formally as follows.

Definition 1. (Detection and Robustness) *Let $\mathcal{P}_n$ and $\mathcal{N}_n$ denote two sequences of distributions on graphs. We say that an algorithm* $\mathsf{A} : Graphs \to \{\text{P}, \text{N}\}$ *solves the detection problem, or can distinguish $\mathcal{P}_n$ and $\mathcal{N}_n$. if*

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

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*

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

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 for some λ,

$$M_{i,j} = \begin{cases} 1 + (k-1)\lambda & i = j \\ 1 - \lambda & i \neq j. \end{cases} \tag{2.1}$$

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 reversible, 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 (2.1); 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 result is the following.

Theorem 2.1. *Let $\mathcal{N}_n = \mathcal{G}(n, d/n)$, and $\mathcal{P}_n$ denote the n-vertex SBM with parameters (d, k, M, π). For every $\epsilon > 0$, if $d > d_{\mathrm{KS}} + \epsilon$, then there exists constant $m \in \mathbb{N}$ and $\rho > 0$ such that local statistic SDP relaxation* LoSt$(2, m)$ *can ρ-robustly solve the detection problem.*

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 1. *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 case.*

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.

THEOREM 2.2. *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). For every $\epsilon > 0$, if $d > d_{\mathrm{KS}} + \epsilon$, then there exists a constant $m \in \mathbb{N}$ and $\rho > 0$ so that* $\mathrm{LoSt}(2, m)$ *can ρ-robustly solve the detection problem. Conversely, if $d < d_{\mathrm{KS}} - \epsilon$, then for every $m \in \mathbb{N}$,* $\mathrm{LoSt}(2, m)$ *fails to (0-robustly) solve the detection problem.*

Future Work There are two regrettable omissions from the above results: we lack a complementary lower bound in the Stochastic Block Model, and we do not solve the problem of recovery above Kesten-Stigum in either model.

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, $\langle X, Y \rangle = \mathrm{tr}\, X^T Y$ 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 $\mathrm{diag} : \mathbb{R}^{n\times n} \to \mathbb{R}$ be the function extracting the diagonal of a matrix, and $\mathrm{Diag} : \mathbb{R}^n \to \mathbb{R}^{n\times n}$ 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 $p_{\mathrm{out}} > p_{\mathrm{in}}$, then the partition induced by the communities should have an unusually large number $\frac{dn}{2} \cdot \frac{p_{\mathrm{out}}}{p_{\mathrm{out}}+p_{\mathrm{in}}}$ 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., $x = \mathrm{argmax}_x\, p(x|G)$.

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 $\frac{dn}{2} \cdot \frac{p_{\mathrm{out}}}{p_{\mathrm{out}}+p_{\mathrm{in}}}$ 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:

(3.2) *Normalized* $\quad \mathbb{E}\,1 = 1$

(3.3) *Satisfies $\mathcal{S}$* $\quad \mathbb{E}\, f_i(x) \cdot p(x) = 0 \quad \forall i \in [m], \forall p \in \mathbb{R}[x]$

(3.4) *Positive* $\quad \mathbb{E}\, p(x)^2 \geqslant 0 \quad \forall p \in \mathbb{R}[x]$

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 (3.2)-(3.4) 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 $\mathbb{R}[G, x]$:

(3.5) $$G_{u,v}^2 = G_{u,v} \qquad \forall u, v \in [n]$$

(3.6) $$x_{u,i}^2 = x_{u,i} \qquad \forall u \in [n],\ i \in [k]$$

(3.7) $$x_{u,1} + \cdots + x_{u,k} = 1 \qquad \forall u \in [n].$$

For brevity, we will throughout the paper denote by $\mathcal{B}_k$ the set of polynomials constraints in the x variables appearing in (3.6) and (3.7). 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 2. *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 $\mathrm{LoSt}(2, m)$ 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 $\overline{A}^{(m)}_{\boldsymbol{G}} = (A - (d/n)\mathbb{J})^{(m)}$ 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 $\overline{A}^{(m)}_{\boldsymbol{G}} = (A - (d/n)\mathbb{J})^{(m)}$ 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 $r = \Theta(\frac{\log n}{(\log logn)^2})$. Let $\boldsymbol{G}$ be an Erdős-Renyi $G(n, d/n)$ graph, let $\boldsymbol{S}$ be 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

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

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

THEOREM 3.1. *With probability* $1 - n^{-100}$, $\left\| \left(\boldsymbol{A} - \frac{d}{n} 1_{[n]\setminus \boldsymbol{S}} 1_{[n]\setminus \boldsymbol{S}}^\mathsf{T}\right)^{(\ell)} \right\| \leqslant \left((1+\varepsilon)^4 \sqrt{d}\right)^\ell$.

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 2. (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 2.2 by way of a reduction to some key results from this simpler, symmetric case. Finally, in the full version of this paper we prove Theorem 2.1 regarding the SBM.

4 A Simplified SDP for the Symmetric DRBM

A few 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 $d_{\text{KS}} \triangleq \lambda^{-2} + 1$. 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

$$\boldsymbol{Y} \triangleq \frac{k}{k-1}\left(\boldsymbol{y}\boldsymbol{y}^* - \frac{1}{k}\mathbb{J}\right) \succeq 0. \tag{4.8}$$

This 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 $s \geqslant 1$,*

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

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

Definition 3. *Fix a small number $\delta > 0$ and write $a \simeq b$ to mean that $|a - b| \leqslant \delta n$ (one should treat δ as a small number which we will set at the end). For each $m \geqslant 1$, the level m symmetric path statistics SDP with error tolerance $\delta > 0$ is the feasibility problem*

$$\begin{aligned} &\textit{Find } Y \succeq 0 \textit{ s.t.} && Y_{u,u} = 1 && \forall u \in [n] \\ &&& \langle Y, \mathbb{J} \rangle = 0 && \\ &&& \langle Y, A_G^{(s)} \rangle \simeq \mathbb{E}\langle \boldsymbol{Y}, A_{\boldsymbol{G}}^{(s)} \rangle && \forall s \in [m]. \end{aligned} \tag{4.9}$$

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 G 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 4.1. *If $(d-1)\lambda^2 > \epsilon$, then there exists constant $m \in \mathbb{N}$ and $\rho > 0$ so that $SPS(m, \lambda)$ solves the detection problem ρ-robustly. Conversely if $(d-1)\lambda^2$ then every constant level fails to do so.*

Remark 3. *Throughout the proofs of Theorem 4.1 and our other main theorems, the reader should imagine that δ is allowed to change line-by-line; all expressions in the paper are bounded by some large constant, so δ will never become too small.*

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,

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

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*

$$\mathrm{d}\mu_{\text{KM}}(z) = \frac{1}{2\pi} \frac{d}{\sqrt{d-1}} \frac{\sqrt{4(d-1) - z^2}}{d^2 - z^2}$$

$$\mathrm{d}z\,\mathbf{1}\left[|z| < 2\sqrt{d-1}\right],$$

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

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

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

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

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

Lemma 4.2. *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 $s + t < \mathrm{girth}(G)$

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

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 4.3. *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 full version, 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 4.1, namely that if $(d-1)\lambda^2 > 1$, then the $SPS(m,\lambda)$ SDP, for 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 (4.8) is with high probability a feasible solution to SDP (4.9). 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 4.1. *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 $\boldsymbol{G} \sim \mathcal{N}$.

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

$$f = \sum_{s=0}^{m} \frac{\langle f, q_s\rangle_{\mathrm{KM}}}{\|q_s\|^2_{\mathrm{KM}}} q_m. \qquad (4.10)$$

Moreover, since f is strictly positive on $[-2\sqrt{d-1}, 2\sqrt{d-1}]$, it is nonnegative on some fattening $[-2\sqrt{d-1} - \iota, 2\sqrt{d-1} + \iota]$ 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

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

□

The following proposition implies a proof of the first part of Theorem 4.1.

Proposition 4.2. *If $\lambda^2(d-1) > 1$, there exists a polynomial satisfying the hypotheses of Proposition 4.1.*

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

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

which by Lemma 4.2 has the desired positivity property. This choice of f satisfies

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

which is negative when

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

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 4.3. *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*

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

then the $SPS(m,\lambda)$ SDP is with high probability feasible for a uniformly random d-regular graph.

Proof. Letting $\boldsymbol{G}$ be the random regular graph in question, we need to produce $Y \succeq 0$ with ones on the diagonal, zero inner product with the matrix $\mathbb{J}$, and satisfying

$$\langle Y, A_{\boldsymbol{G}}^{(s)}\rangle \simeq \lambda^s \|q_s\|_{\text{KM}}^2 n;$$

as above $\simeq$ denotes equality up to $\pm\delta n$ for our small, fixed error tolerance $\delta > 0$. Our strategy will be to modify the matrix $g(\overline{A}_{\boldsymbol{G}}) = g(A_{\boldsymbol{G}}) - g(d)\mathbb{J}/n$.

First, note that by expanding g in the non-backtracking basis and invoking Lemma 4.3, for any $0 \leqslant s \leqslant m$ we have

$$\begin{aligned} \langle g(\overline{A}_{\boldsymbol{G}}), A_{\boldsymbol{G}}^{(s)}\rangle &= \langle g(A_{\boldsymbol{G}}), A_{\boldsymbol{G}}^{(s)}\rangle + g(d)\|q_s\|_{\text{KM}}^2 \\ &= \lambda^s \|q_s\|_{\text{KM}}^2 \cdot n + O(\log n) \simeq \lambda^s \|q_s\|_{\text{KM}}^2 \cdot n. \end{aligned}$$

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 $A_{\boldsymbol{G}}$ other than the eigenvalue at d is contained w.h.p. in such a set. Thus $g(\overline{A}_{\boldsymbol{G}})$ is positive semidefinite, and as a polynomial in the centered adjacency matrix, is orthogonal to the all-ones matrix.

However, the diagonal of $g(\overline{A}_{\boldsymbol{G}})$ may not be equal to one, for two different reasons. The diagonal entries of $g(A_{\boldsymbol{G}}) = g(\overline{A}_{\boldsymbol{G}}) + g(d)\mathbb{J}/n$ different from one are exactly those corresponding to vertices within $\deg g$ steps of a constant length cycle; from Lemma 4.3 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

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

which restores the diagonal entries of the good vertices.

Now, $\tilde{Y}$ is PSD, and is accordingly the Gram matrix of some vectors $\alpha_1, ..., \alpha_n \in \mathbb{R}^n$. 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 $A^{(s)}_{\boldsymbol{G}}$ 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 $A^{(s)}_{\boldsymbol{G}}$ 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

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

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. ∎

The second part of Theorem 4.1 ensues from the following proposition.

PROPOSITION 4.4. *Whenever $\lambda^2(d-1) < 1$, there exists a polynomial satisfying the conditions of Proposition 4.3.*

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 4.2, as $|q_p| \leqslant 2p\sqrt{d(d-1)^p}$. ∎

4.4 Robustness We have shown already that if $(d-1)\lambda^2$, then the level m symmetric path statistics can solve the detection problem—for some constant m. In this section we show that it can do so robustly. In other words, we need to show that (i) with high probability in $\mathcal{P}$, if $SPS(m,\lambda)$ is feasible, then it will remain feasible after $\rho_m n$ edge corruptions, and (ii) with high probability in $\mathcal{N}$, if $SPS(m,\lambda)$ is infeasible, then altering ρ_m edges cannot fix it. Let us maintain the notation from Definition 1, that $G \approx_\rho H$ means that these two differ on at most 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,

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

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

By making δ sufficiently small, if $\boldsymbol{G}$ is drawn from the planted distribution, the feasible solution Y that we constructed is PSD and satisfies the affine constraints regarding inner products with the $A_{\boldsymbol{G}}^{(s)}$ matrices with slack $\Omega(n)$. Every diagonal entry of Y is one, so by PSD-ness their off-diagonal entries have modulus at most one. Thus

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

Because of the $\Omega(n)$ slack, the Y we constructed from $\boldsymbol{G}$ will *still* satisfy the affine constraints to the SDP on input $\boldsymbol{H}$, for small enough ρ.

On the other hand, when $\boldsymbol{G}$ is drawn from the null model, again by making δ sufficiently small any putative solution Y with ones on the diagonal and zero inner product with $\mathbb{J}$ violates some linear combination of the above affine constraints by a margin of $\Omega(n)$. Thus, if we try these consraints with $\boldsymbol{H}$ instead of $\boldsymbol{G}$, this constraint will still be violated for ρ sufficiently small.

Remark 4. *The parameter ρ controlling the number of adversarial edge insertions and deletions made to random input $\boldsymbol{G}$ that the level-m LocalStatistic SDP can tolerate can be seen to decrease with m, which is indicative of a tradeoff between how close to the threshold an algorithm in this hierarchy works and how robust it is to perturbations.*

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 2.2.

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

$$\begin{aligned} G_{u,v}^2 &= G_{u,v} \\ x_{u,i}^2 &= x_{u,j} \\ \sum_i x_{u,i} &= 1 \end{aligned} \tag{5.11}$$

as in the introduction, we will write the set of equations in the last two lines as $\mathcal{B}_k$. It will at times be useful to write $\sigma_{\boldsymbol{x}} : [n] \to [k]$ for the labelling described by $\boldsymbol{x}$ in the zero locus of $\mathcal{B}_k$. 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) satisfies the polynomial equations $\mathcal{B}_k$, 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 4. (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 5.1. *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 (5.11), 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

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

where $|S| \leqslant 2$ and $|E(H)| \leqslant m$. The following theorem computes the right hand side of the above equation in any planted model, for this class of partially labelled subgraphs. We will discuss it briefly below and remit the proof to the full version.

DEFINITION 5. *Let (H, S) be a connected graph on $O(1)$ edges, with distinguished vertices S. Define $C_{H,S,d}$ to be the number of occurrences of (H, S) in an rooted, infinite d-regular tree in which some vertex in S is mapped to the root. If $S = \emptyset$, choose some distinguished vertex arbitrarily—the count will be the same no matter which one is chosen; we will at times use $C_{H,d}$ as shorthand in this case. Finally, if $(H, S) = (H_1, S_1) \sqcup \cdots \sqcup (H_\ell, S_\ell)$ has ℓ connected components, take $C_{H,S,d} = C_{H_1,S_1} \cdots C_{H_\ell,S_\ell}$. We note for later use that if H contains a cycle, $C_{H,S} = 0$, and if it is a path of length s with endpoints distinguished, $C_{H,S} = 2\|q_s\|^s_{\mathrm{KM}}$, the number of vertices at depth s in the tree. This factor of two arises because we have defined occurrences in terms of injective homomorphisms, so we are accumulate a factor accounting for the symmetries in the subgraph.*

THEOREM 5.1. (LOCAL STATISTICS) *If (H, S, τ) is a partially labelled graph with $O(1)$ edges, then in any planted model $\mathcal{P}_{d,k,M,\pi}$,*

1. *If H is unlabelled, i.e. $S = \emptyset$, then $n^{-\ell}\, \mathbb{E}\, p_{H,S,\tau}(\boldsymbol{G}, \boldsymbol{x}) \to C_{H,S,d}$*
2. *If H is labelled, with $S = \{\alpha, \beta\}$, $\tau(\alpha) = i$, and $\tau(\beta) = j$, then*

$$n^{-\ell}\, \mathbb{E}\, p_{H,S,\tau}(\boldsymbol{G}, \boldsymbol{x}) \to \pi(i) T_{i,j}^{\mathrm{dist}(\alpha,\beta)} C_{H,S,d},$$

and $p_{H,S,\tau}(\boldsymbol{G}, \boldsymbol{x})$ enjoys concentration up to an additive $\pm o(n^\ell)$. We say that $\mathrm{dist}(\alpha, \beta) = \infty$ if these two vertices lie in disjoint components of H, and we interpret $T^\infty_{i,j} = \pi(j)$.

Let's take a moment and get a feel for Theorem 5.1. As a warm-up, consider the case when (H, S, τ) is a path of length $s \leqslant m$ with the endpoints labelled as $i, j \in [k]$, and we simply need to count the number of pairs of vertices in G with labels i and j respectively that are connected by a path of length s. As d-regular random graphs from models like $\mathcal{P}$ have very few short cycles, assume for simplicity that the girth is in fact much larger than m, so that the depth-s neighborhood about every vertex is a tree. If we start from a vertex i and follow a uniformly random edge, the transition matrix T from our model says that, on average at least, the probability of arriving at a vertex in group j is roughly $T_{i,j}$, and similarly if we take s (non-backtracking) steps, this probability is roughly $T^s_{i,j}$. There are $\pi(i)n$ starting vertices in group i, and $d(d-1)^{s-1}$ vertices at distance s from any such vertex.

If (H, S, τ) is a tree in which the two distinguished vertices are at distance s, then we can enumerate occurrences of (H, S, τ) in G by first choosing the image of the path connecting these two, and then counting the ways to place the remaining vertices. If we again assume that the girth is sufficiently large, it isn't too hard to see that the number of ways to do this second step is a constant independent of the number of ways to place the path, so we've reduced to the case above. The idea for the cases $|S| = 0, 1$ is similar. We'll prove Theorem 5.1 in the full version.

With this result in hand, we are now ready to state the full local statistics SDP.

DEFINITION 6. *Fix a small error tolerance $\delta > 0$, and write "$\simeq_\ell$" to mean "equal up to $\pm\delta n^\ell$." The level*

$(2,m)$ local statistics SDP, on input G_0, is the feasibility problem of finding a pseudoexpectation $\widetilde{\mathbb{E}} : \mathbb{R}[x]_{\leqslant 2} \to \mathbb{R}$ such that

1. *$\widetilde{\mathbb{E}}$ satisfies $\mathcal{B}_k$*
2. *$\widetilde{\mathbb{E}}\, p_{H,S,\tau}(G_0, x) \simeq_\ell \mathbb{E}_{\mathcal{P}}\, p_{H,S,\tau}(\boldsymbol{G}, \boldsymbol{x})$ for all (H, S, τ) with $|S| \leqslant 2$, $|E(H)| \leqslant m$, and ℓ connected components.*

There are two exceptions because of the rigidity of our model. When (H, S, τ) consists of a single vertex with label $i \in [k]$, we require $p_{H,S,\tau}(G_0, x) = \pi(i)n$. Similarly, when (H, S, τ) consists of two distinguished vertices labelled $i, j \in [k]$, we require $p_{H,S,\tau}(G_0, x) = \pi(i)\pi(j)n^2 - \pi(i)n\delta_{i,j}$. As in the symmetric case, we include as a contingency that, if G_0 is not d-regular, we greedily delete and add edges to make it so.

Note that, importantly, when $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathcal{P}$, the $\mathrm{LoSt}(2, m)$ SDP is feasible for every constant m: we simply define

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

Because of the concentration guarantees in Theorem 5.1, no matter our error tolerance δ, this will satisfy the constraints when n is sufficiently large.

5.2 Distinguishing Let us prove the first part of Theorem 2.2: when $(d-1)\lambda_2^2 > 1$, there exists a constant m for which the $\mathrm{LoSt}(2, m)$ SDP solves the detection problem. Since the SDP is with high probability feasible for $\boldsymbol{G} \sim \mathcal{P}$, it remains only to show infeasibility when $\boldsymbol{G} \sim \mathcal{N}$.

Let $\boldsymbol{G} \sim \mathcal{P}$, and assume we have a viable pseudoexpectation $\widetilde{\mathbb{E}}$ for the $\mathrm{LoSt}(2, m)$ SDP. 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 5.2. *For any i, j, and any $s = 0, ..., m$, recalling that $A_G^{(s)}$ is the matrix counting non-backtracking walks of length s, and $\mathbb{J}$ is the all-ones matrix,*

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

Proof. For the first assertion, let (H, S, τ) be the path of length s whose endpoints are labelled $i, j \in [k]$. Each *self-avoiding* walk of length s in G is an occurrence of H, so from Theorem 5.1

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

It is an easy consequence of Lemma 4.3 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 4.2

$$\begin{aligned}\langle X_{i,j}, A_{\boldsymbol{G}}^{(s)}\rangle &= \langle X_{i,j}, A_{\boldsymbol{G}}^{\langle s\rangle}\rangle + \langle X_{i,j} A_{\boldsymbol{G}}^{\langle s\rangle} - A_{\boldsymbol{G}}^{(s)}\rangle \\ &= \langle X_{i,j}, A_{\boldsymbol{G}}^{\langle s\rangle}\rangle \pm O(\log n) \simeq \pi(i) T_{i,j}^s \|q_s\|_{\mathrm{KM}}^2\end{aligned}$$

for $s = 0, ..., m$. 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. From Lemma 5.1 we have

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

When $i = j$, take (H, S, τ) as above and (H', S', τ') to be a single vertex labelled i. ∎

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 $F^{-1}\,\mathrm{Diag}(\pi)F = \mathbb{1}$.

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 5.3. *For any $s = 0, ..., m$, if $i \neq j$ $\langle \check{X}_{i,j} A_{\boldsymbol{G}}^{(s)} \rangle \simeq 0$, and*

$$\langle \check{X}_{i,i}, A_{\boldsymbol{G}}^{(s)} \rangle \simeq \lambda_i^s \|q_s\|_k^2 n.$$

Furthermore,

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

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,

$$\begin{pmatrix} \langle \check{X}_{1,1}, A_{\boldsymbol{G}}^{(s)} \rangle & \cdots & \langle \check{X}_{1,k}, A_{\boldsymbol{G}}^{(s)} \rangle \\ \vdots & \ddots & \vdots \\ \langle \check{X}_{k,1}, A_{\boldsymbol{G}}^{(s)} \rangle & \cdots & \langle \check{X}_{k,k}, A_{\boldsymbol{G}}^{(s)} \rangle \end{pmatrix}$$
$$= F^T \begin{pmatrix} \langle X_{1,1}, A_{\boldsymbol{G}}^{(s)} \rangle & \cdots & \langle X_{1,k}, A_{\boldsymbol{G}}^{(s)} \rangle \\ \vdots & \ddots & \vdots \\ \langle X_{k,1}, A_{\boldsymbol{G}}^{(s)} \rangle & \cdots & \langle X_{k,k}, A_{\boldsymbol{G}}^{(s)} \rangle \end{pmatrix} F$$
$$\simeq F^T \operatorname{Diag}(\pi) T^s F \cdot \|q_s\|_{\text{KM}}^s n$$
$$= F^T \operatorname{Diag}(\pi) F \Lambda^s \cdot \|q_s\|_{\text{KM}}^s n$$
$$= \Lambda^s \cdot \|q_s\|_{\text{KM}}^s n$$

A parallel calculation gives us

$$\begin{pmatrix} \langle \check{X}_{1,1}, \mathbb{J} \rangle & \cdots & \langle \check{X}_{1,k}, \mathbb{J} \rangle \\ \vdots & \ddots & \vdots \\ \langle \check{X}_{k,1}, \mathbb{J} \rangle & \cdots & \langle \check{X}_{k,k}, \mathbb{J} \rangle \end{pmatrix}$$
$$= F^T \begin{pmatrix} \langle X_{1,1}, \mathbb{J} \rangle & \cdots & \langle X_{1,k}, \mathbb{J} \rangle \\ \vdots & \ddots & \vdots \\ \langle X_{k,1}, \mathbb{J} \rangle & \cdots & \langle X_{k,k}, \mathbb{J} \rangle \end{pmatrix} F$$
$$= F^T \pi \pi^T F \cdot n^2$$
$$= e_1 e_1^T \cdot n^2,$$

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 5.3 in hand, the remainder of the proof follows from Proposition 4.1 and Proposition 4.2 in the previous section. In particular, each block $\check{X}_{i,i}$ for $i = 2, ...k$ is a feasible solution to the Symmetric Path Statistics SDP, and we showed already that when $\lambda^2(d-1) > 1$, this SDP is infeasible when $\boldsymbol{G} \sim \mathcal{N}$ for m sufficiently large.

5.3 Spectral Distinguishing Our argument in the previous section can be recast to prove Corollary **??**, 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 5.1 and repeating the calculations in Lemma 5.2 and Lemma 5.3 *mutatis mutandis* with $\boldsymbol{X}$ instead of X, we can show that w.h.p.

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

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

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

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 $\check{\boldsymbol{x}}_1, ..., \check{\boldsymbol{x}}_k$ are orthogonal.

We can show that $A_{\boldsymbol{G}}$ has an eigenvalue with a separation $\eta > 0$ from the bulk spectrum by proving

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

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 4.2 works here. As the $\check{\boldsymbol{x}}_i$ are orthogonal, we get one distinct eigenvalue outside the bulk for each eigenvalue of T satisfying this property.

REMARK 5. *To distinguish the null model from the planted one using the spectrum of $A_{\boldsymbol{G}}$, simply return* PLANTED *if $A_{\boldsymbol{G}}$ 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 $\boldsymbol{G} \sim \mathcal{N}$, 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 2.2, 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 $\mathrm{LoSt}(2,m)$ SDP, it suffices for it to satisfy

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

when (H,S,τ) is a path of length $0, ..., m$ with labelled endpoints. 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 5.4. *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*

$$\begin{pmatrix} 1 & l^T \\ l & X \end{pmatrix} \succeq 0 \tag{5.12}$$

$$\operatorname{diag}(X) = l \tag{5.13}$$

$$\operatorname{tr} X_{u,u} = e^T l = 1 \qquad \forall u \in [n] \tag{5.14}$$

$$X_{u,v} e = l_u \qquad \forall u, v \in [n] \tag{5.15}$$

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

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

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

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

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 5.1. *It suffices to check*

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

in the cases (i) (H,S,τ) is a path of length $s = 0, ..., m$ with labelled endpoints, and (ii) when (H,S,τ) is a graph with no edges on one or two labelled vertices.

We will defer the proof of Proposition 5.1 to the full version. Its conclusion in hand, we can now set about constructing a pseudoexpectation. We'll construct $l \in \mathbb{R}^{nk}$ and $X \in \mathbb{R}^{nk\times nk}$ so that (i) the $\mathcal{B}_k$ constraints in Lemma 5.4 hold, and (ii)

$$\langle e, l_i\rangle = \pi(i) n$$
$$\langle X_{i,j}, A_{\boldsymbol{G}}^{(s)}\rangle \simeq \pi(i) T_{i,j}^s n$$
$$\langle X_{i,j}, \mathbb{J}\rangle = \pi(i)\pi(j) n^2.$$

It will simplify things immensely to use the same change of basis as we did in Section Section 5.2. Namely, letting F be the matrix of right eigenvectors, we will produce a pair $\check{l} \in \mathbb{R}^{nk}$ and $\check{X} \in \mathbb{R}^{nk\times nk}$ so that $l = (F^{-T}\otimes\mathbb{1})\check{l}$ and $X = (F^{-T}\otimes\mathbb{1})\check{X}(F^{-1}\otimes\mathbb{1})$ satisfy the above conditions. Recycling the relevant calculations from Section 5.2, the above moment conditions translate to

$$\langle e, \check{l}_i\rangle = \begin{cases} n & i = 1 \\ 0 & \text{else} \end{cases}$$
$$\langle \check{X}_{i,j}, A_{\boldsymbol{G}}^{(s)}\rangle \simeq \lambda_i^s \|q_s\|_{\mathrm{KM}}^2 n \mathbb{1}[i = j]$$

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

As in the proof of distinguishing, we have already executed the two key steps of the proof in our lower bound argument for the Symmetric Path Statistics SDP. From Proposition 4.3 and Proposition 4.4, we know that if $(d-1)\lambda^2 < 1$, there exists a PSD matrix $Y(\lambda)$ such that

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

Since $(d-1)\lambda_2^2 < 1$, we can 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 $\check{l} = (e, 0, ..., 0)^T$. This way, certainly

$$\begin{pmatrix} 1 & \check{l}^T \\ \check{l} & \check{X} \end{pmatrix} \succeq 0 \tag{5.16}$$

(by taking a Schur complement), and the three inner product conditions above are satisfied on every block. We now need to check carefully that

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

is a pseudoexpectation satisfying $\mathcal{B}_k$. The above construction guarantees PSD-ness, since we have multiplied a matrix and its transpose on the right and left respectively of a PSD matrix. Since π is the first row of F^{-1}, we know $l_i = \pi(i)e$. On the other hand, 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

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

as desired. Similarly, because $\check{X}$ is diagonal, $\check{X}_{u,u} = \mathbb{1}$, 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

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

This completes the construction of our pseudoexpectation.

5.5 Robustness The proof of robustness follows almost immediately from the discussion in Section 4.4. In particular, if $\boldsymbol{G}$ is drawn from the planted distribution, by making δ sufficiently small, the matrices $Y(\lambda_i)$ are PSD and satisfy the affine constraints regarding inner products with the $A^{(s)}_{\boldsymbol{G}}$ matrices with slack $\Omega(n)$. Every diagonal entry of $Y(\lambda_i)$ is one, so by PSDness their off-diagonal entries have modulus at most one. Thus

$$\begin{aligned} \left| \langle A_s^{(\boldsymbol{G})} Y(\lambda_i) \rangle - \langle A_s^{(\boldsymbol{H})} \rangle \right| &= \left| \langle A_s^{(\boldsymbol{G})} - A_s^{(\boldsymbol{H})} \rangle \right| \\ &\leqslant \left\| A_s^{(\boldsymbol{G})} - A_s^{(\boldsymbol{H})} \right\|_1 \\ &\leqslant 2\xi\rho d(d-1)^{s-1}. \end{aligned}$$

Because of the $\Omega(n)$ slack, if we construct $Y(\lambda_i)$ from $\boldsymbol{H}$ instead of $\boldsymbol{G}$, the constraints will *still* be satisfied for small enough ρ. We can then use these $Y(\lambda_i)$ to build the full feasible solution as before.

On the other hand, when $\boldsymbol{G}$ is drawn from the null model, again by making δ sufficiently small, any pseudoexpectation satisfying the Boolean and Single Color constraints violates some linear combination of the above affine constraints by a margin of $\Omega(n)$, and this constraint will still be violated for ϵ sufficiently small.

References

[ABARS20] Emmanuel Abbe, Enric Boix-Adserà, Peter Ralli, and Colin Sandon, *Graph powering and spectral robustness*, SIAM Journal on Mathematics of Data Science **2** (2020), no. 1, 132–157. 2

[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. 2, 4

[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. 9

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

[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). 6

[AR20] Emmanuel Abbe and Peter Ralli, *An alon-boppana theorem for powered graphs and generalized ramanujan graphs*, arXiv preprint arXiv:2006.11248 (2020). 2

[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. 4

[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. 4

[AS18] ———, *Proof of the achievability conjectures for the general stochastic block model*, Communications on Pure and Applied Mathematics **71** (2018), no. 7, 1334–1406. 2

[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. 3

[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. 6

[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. 6

[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. 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. 4

[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*.

[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. 4

[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. 4

[CO04] Amin Coja-Oghlan, *Coloring semirandom graphs optimally*, International Colloquium on Automata, Languages, and Programming, Springer, 2004, pp. 383–395. 4

[CO07] ———, *Solving np-hard semirandom graph problems in polynomial expected time*, Journal of Algorithms **62** (2007), no. 1, 19–46. 4

[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. 4

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

[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. 6

[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. 6

[Fri08] ———, *A proof of Alon's second eigenvalue conjecture and related problems*. 9

[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, 4

[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. 6

[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, 6

[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. 4

[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. 6

[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. 4

[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. 4

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

[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. 2

[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. 4

[MMV16] ———, *Learning communities in the presence of errors*, Conference on Learning Theory, 2016, pp. 1258–1291. 2, 4

[MNS18] Elchanan Mossel, Joe Neeman, and Allan Sly, *A proof of the block model threshold conjecture*, Combinatorica **38** (2018), no. 3, 665–708. 2

[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).

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

[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, 4

[NM14] MEJ Newman and Travis Martin, *Equitable random graphs*, Physical Review E **90** (2014), no. 5, 052824. 3

[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). 6

[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. 6

[SM19] Ludovic Stephan and Laurent Massoulié, *Robustness of spectral methods for community detection*, Conference on Learning Theory, 2019, pp. 2831–2860. 2

[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

[ZK16] Lenka Zdeborová and Florent Krzakala, *Statistical physics of inference: Thresholds and algorithms*, Advances in Physics **65** (2016), no. 5, 453–552. 6