Title
Phase Transitions in Inference

Author
Mohanty, Sidhanth

Publication Date
2023

Peer reviewed|Thesis/dissertation

Phase Transitions in Inference

By

Sidhanth Mohanty

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Prasad Raghavendra, Chair
Professor Venkatesan Guruswami
Professor Nikhil Srivastava

Summer 2023

Phase Transitions in Inference

Abstract

Phase Transitions in Inference

By

Sidhanth Mohanty

Doctor of Philosophy in Computer Science

University of California, Berkeley

Professor Prasad Raghavendra, Chair

What makes an algorithmic problem easy or hard? Many general algorithmic techniques arising from decades of research, along with the theory of NP-completeness based on reductions between hard problems, offers a good answer for problems where the input is "worst-case".

However, this theory has very little to say when the input is random, and comprises of independent samples, as is frequently the case for problems in statistics. Statistical problems seemingly go through abrupt phase transitions in complexity, from hard to easy once the number of samples crosses a threshold. Understanding this boundary between "hard" and "easy" for statistical problems is still in nascent stages.

This thesis comprises recent progress in understanding these phase transitions from the lens of semidefinite programming.

To my parents, for their constant encouragement and sacrifices

Contents

Acknowledgments

First, I would like to thank the mentors I was blessed with in the past 5 years. I had a wonderful time during grad school being advised by Prasad Raghavendra. A particularly striking way I have been inspired by him is how his enthusiasm level for a research problem is the same as it was on day 1, even after months of no progress: the journey is at least as meaningful as the destination.

Ryan O'Donnell mentored me when I was an undergraduate at Carnegie Mellon University when I had just gotten started doing research, and very patiently listened to every idea I had and worked through them in the weeds with me. Some of my fondest research memories are working out small cases on the whiteboard in Ryan's office and trying to glean patterns, and it was always nice to go back and visit Ryan at CMU.

I am grateful for my many conversations with Nikhil Srivastava during my time at Berkeley. Interacting with him has shaped much of my aesthetic inclinations in mathematics — I have learned a new insight in spectral graph theory, random matrix theory, or geometry in almost all of our conversations.

I started working with Tselil Schramm at the start of my second year, and it is among the happiest things that happened during grad school. She has been a very supportive friend and mentor, who is seemingly always available to talk about math and life. I cherish all the outdoor pandemic meetings at her patio, ice cream trips, thinking about random geometric graphs and semidefinite programs together, and the Stanford visits.

I had the privilege of getting to know Sam Hopkins in my first year of grad school, and queried him about his papers on a daily basis: his contagious excitement for his research is a large reason I work on some of the things I do. It was also a pleasure to eventually collaborate on problems, and am grateful for the two amazing research visits at MIT.

It is always fun to talk to Pravesh Kothari about research, and hear his unique perspectives. I'm particularly grateful to Pravesh for generously including me in a vibrant online community he created for his students at CMU once the pandemic hit, and for the fun visits at CMU.

My favorite memories from grad school stemmed from the close friendships with fellow grad students. First, I would like to give a special mention to my close collaborators during graduate school — Siqi Liu, Tim Hsieh, Elizabeth Yang, and Jeff Xu. There is a lot to learn from each of them. A bulk of time spent doing

theory research is spent wading through a swamp until a picture emerges — gritty calculations, trying out small cases, searching for patterns with Python programs, reading abstract papers — and many of my swamp-wading memories are very fond and pleasant because of them.

Siqi has been a steadfast friend for the past five years, and has set a great example for me in how she always stays true to her principles. I appreciate how she has always been available to listen to random ideas and talk about life, and for frequently sharing her many insights.

Tim's positive attitude and excitement always makes my day bright when we talk about research or play Pokémon Showdown, and I appreciate all the fun activities during his visits to Berkeley and mine to Pittsburgh.

Elizabeth has been a great source of fun activities, conversation and advice. It was always fun to randomly take breaks to play word games in her office, and go climbing.

During my first two years of graduate school, Jeff was a source of high entropy, surprising and entertaining me in new ways everyday, while also teaching me new research-related things on almost a daily basis.

The many other friends who were an integral part of the graduate school journey that I would like to thank are: Nathan Ju, Omar Alrabiah, David Wu, Ishaq Aden-Ali and Ansh Nagda for being very joyful presences in Soda Hall, and for always inspiring me with their mathematical curiosity: I have become close friends with each of them and love their unique quirks and their propensity to act silly and joke around while also being very thoughtful and caring individuals; Nived Rajaraman for all our shared experiences through midnight conversations, listening to music, going on outdoor activities, playing games, cooking food, and for his infinite tolerance for my silly jokes; Yeshwanth Cherapanamjeri for always being down for spontaneous activities, and for setting an example for me to learn from through how he appreciates the present experiences in life over perpetually chasing goals; Abhishek Shetty for his endless stream of ideas, knowledge, and fun shared experiences; Seri Khoury, Morris Yau, and Antares Chen for their humor, mathematical conversations, and for their fascinating reflections about the world; Sam Gunn, Louis Golowich, Francisca Sousa Pereira Cruz de Vasconcelos, Christian Ikeokwu, Angelos Pelecanos and William He for their energy, fun vibes, and interesting conversations; Pedro Miguel Reis Bento Paredes, Theo McKenzie, and Jess Banks for all the fun collaborations on graphs.

I've also formed friendships with various people from the theory research

community beyond Berkeley: Ainesh Bakshi, Amit Rajaraman, Frederic Koehler, Sushrut Karmalkar, Shivam Nadimpalli, Xinyu Wu, and Goutham Rajendran.

I would like to thank various mentors from before graduate school who inspired me to pursue research further: Po-Shen Loh, Anupam Gupta, Anıl Ada, Bernhard Haeupler, and Ariel Procaccia.

Thanks to the Berkeley theory faculty for many interesting conversations: especially Venkat Guruswami, Jelani Nelson, and Avishay Tal.

Thanks to my undergraduate friends — Michael, Kumail, Anna, Phillip, Nancy, Maya, Adithya, Tom, Jacob, Rupal, Amal — and my high school friends — Preetham, Praful, Mehul, and Veda — who were always there to support me despite not always being in the same physical vicinity.

Finally, the greatest thanks to my parents Sunita and Sidhartha for all their selfless sacrifices while raising me, and for giving me the early stepping stones.

Chapter 1

What makes a problem easy or hard?

Checking if a graph is 2-colorable can be done in linear time, but the chances of even a subexponential time algorithm for checking if a graph is *3*-colorable look bleak. Solving an instance of 2SAT is easy, whereas our best algorithms for 3SAT run in exponential time. Indeed, friendly structure in the 2SAT and 2-COLORING problems can be utilized to design efficient algorithms. On the other hand, the lack of such structure can be used to reduce known NP-hard problems to 3SAT and 3-COLORING. The worst-case theory of algorithms and complexity, based on a large collection of algorithmic tools and NP-completeness has quite far-reaching predictions for whether a problem is easy or hard, *when the inputs are contrived by demons*.

However, when the input possesses more structure, for example, when the input is comprised of independent samples, like in problems from statistics, the predictions of this theory are moot. To give a concrete example, this theory doesn't have anything to say about the complexity of 3SAT when the input instance has n variables, and m independently drawn clauses. Some average-case algorithmic tasks of interest include:

- **Certification.** *Output a certificate that a random 3SAT formula has no satisfying assignment.*
- **Search.** *Find an assignment to a random 3SAT formula with high objective value.*
- **Recovery/Inference.** *Recover the hidden assignment x to a random 3SAT formula with x as a planted solution.*

- **Hypothesis testing/Distinguishing.** *Distinguish a random* 3*SAT formula from one with a planted solution.*
- **Counting.** *Output the number of satisfying assignments to a* 3*SAT formula.*
- **Sampling.** *Produce a uniformly random satisfying assignment to a random* 3*SAT formula.*

A predicted algorithmic terrain for random 3SAT has been charted out fairly precisely at this point, as denoted by the *phase diagram* in Figure 1.1. While we have similarly precise predictions for the algorithms and hardness landscape for numerous other *average-case* problems, a general encompassing theory where these predictions are on rigorous grounds is still in nascent stages.

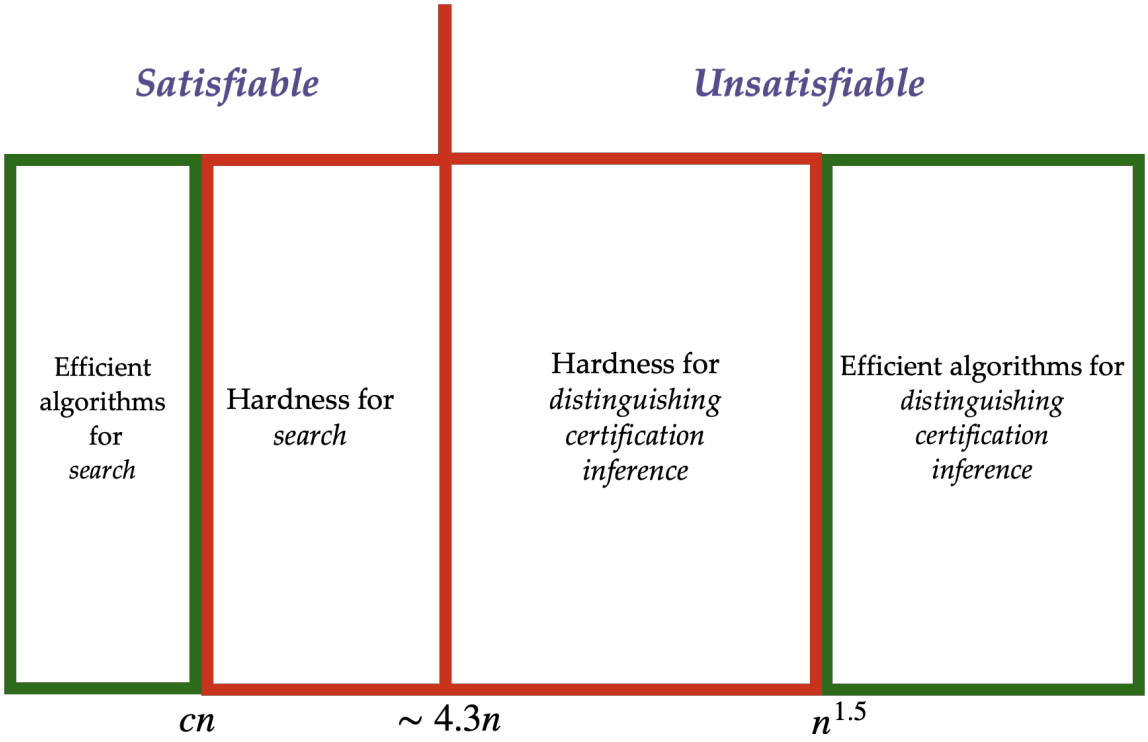


Figure 1.1: Predicted complexity of random 3SAT as function of number of clauses [MZ02, AOW15]

1.1 A general theory for complexity of average-case problems?

What would a general theory for the complexity of average-case problems look like? Ideally, we would like:

1. *Simple heuristics* we can use to predict if an average-case problem is tractable or intractable.
2. *Generic efficient algorithms* when these heuristics predict that a problem is tractable.
3. *Rigorous hardness evidence* when these heuristics predict that a problem is intractable. This evidence could be in the form of unconditional lower bounds in restricted computational models, or reductions from well-established hard problems.

Excellent examples of theories in the worst case that serve as inspiration for an average-case theory are theories for the complexity of *constraint satisfaction problems* (CSPs):

1. The *CSP dichotomy theorem* [Sch78, Zhu20] tells us that every CSP is either polynomial time solvable, or is NP-hard. Further, it is possible to tell if a CSP is tractable or intractable based on whether its solution space satisfies certain algebraic properties (the existence of "polymorphisms").
2. The *Unique Games Conjecture* lets us fully characterize the approximability of CSPs — there is a simple semidefinite programming based algorithm which attains the optimal approximation ratio for any CSP [Kho02, KKMO07, Rag08]. Thus approximation ratios for various CSPs can be understood by studying integrality gaps for this SDP.

One of the goals of this thesis is to contribute to the program for achieving a general theory of complexity of recovery and hypothesis testing, and we focus on random constraint satisfaction problems.

1.2 Technical overview

1.2.1 A generic algorithm for recovery?

Informally, in the class of problems we work with, $\boldsymbol{x}$ is a *hidden signal* drawn uniformly at random from some set Σ^n (for example, $\{\pm 1\}^n$), and we receive independent *observations* $\boldsymbol{E} = \boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$. The algorithmic task is then to estimate $\boldsymbol{x}|\boldsymbol{E}$, either by sampling from the distribution, or by producing (an approximation to) its low-degree moments.

As an example, consider the 2-*community stochastic block model* (SBM), where $\boldsymbol{x}$ is in $\{\pm 1\}^n$, and each $\boldsymbol{e}_i$ is a random edge $\{i, j\}$ for $i, j \in [n]$ chosen such that $\boldsymbol{x}_i = \boldsymbol{x}_j$ with probability $\frac{1+\varepsilon}{2}$ and $\boldsymbol{x}_i \neq \boldsymbol{x}_j$ with probability $\frac{1-\varepsilon}{2}$.

Given the success of semidefinite programming-based methods in giving optimal algorithms for combinatorial optimization in the worst-case and for robust statistics problems in the average-case, an SDP formulation is a natural candidate for a generic algorithm for recovery. However, the choice of SDP formulation is unclear. Indeed, even for the special case of the 2-community SBM, the traditional SDP which is a relaxation of the minimum bisection is suboptimal. That is, it is not known to recover communities better than random guessing even for parameter settings where other polynomial time algorithms exist.

The conceptual reason for this suboptimality is that this SDP focuses on optimizing an objective function, which in this case is the minimum bisection in the graph. However, for the task of estimating $\boldsymbol{x}|\boldsymbol{E}$, solving an optimization problem only makes sense if the distribution of $\boldsymbol{x}|\boldsymbol{E}$ is highly concentrated at the solution achieving the optimum. In the case of the SBM, the posterior is anti-concentrated and in fact the minimum bisection is not known to correlate with the low-degree moments of the posterior, which makes the choice of optimization SDP non-canonical.

Therefore, any general purpose SDP for inference must depart from the optimization paradigm and incorporate information about the prior. In Chapter 2, we give a semidefinite program hierarchy (the *Local Statistics hierarchy*) that bakes in the prior via concentration inequalities satisfied by $\boldsymbol{x}|\boldsymbol{E}$ as constraints, and illustrate its efficacy by showing that it succeeds at distinguishing stochastic block models from random graphs for all parameter regimes where the problem is believed to be tractable. The algorithm can be made more powerful by incorporating more information about the prior, at the expense of running time.

More concretely, the problem of distinguishing a stochastic block model from a random graph is believed to be efficiently tractable when the number of observations m exceeds the *Kesten–Stigum threshold*; for the special case of 2 communities, the problem is tractable when $\frac{m}{n} > \frac{1}{2\varepsilon^2}$.

Theorem 1.2.1 (Informal)**.** *The Local Statistics SDP can be used to distinguish the stochastic block model from an Erdős–Rényi graph when then number of edges exceeds the Kesten–Stigum threshold in polynomial time. Further, this algorithm is robust to any $o(n)$ arbitrary edge insertions and deletions.*

1.2.2 The cavity method

Precise predictions for the algorithmic threshold for recovery and hypothesis testing in stochastic block models came from the *cavity method* from statistical physics. It is a very simple and easy-to-implement heuristic to obtain predictions about the complexity of recovery and hypothesis testing for various problems.

We investigate whether we can rigorously say that the predictions produced by this method are accurate. To do so, we consider an expressive class of inference problems that capture stochastic block models and planted CSPs, which we call *Bayesian CSPs*, where the observations $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_m$ are "local", i.e., each observation depends only on a constant number of variables in the hidden signal. We defer a formal definition of this class of problems to Chapter 3, and instead provide some examples.

Stochastic Block Models. The hidden signal $\boldsymbol{x}$ is drawn from $[q]^n$, and each observation is an edge $\{i, j\}$ drawn between i, j with $\boldsymbol{x}_i = \boldsymbol{x}_j$ with probability $\frac{1+\varepsilon}{q}$ and with $\boldsymbol{x}_i \neq \boldsymbol{x}_j$ with probability $\frac{q-\varepsilon}{q}$.

Planted NotAllEquals3SAT. The hidden signal $\boldsymbol{x}$ is drawn from $\{\pm 1\}^n$, and each observation is a triple (a, b, c) such that $\boldsymbol{x}_a, \boldsymbol{x}_b, \boldsymbol{x}_c$ are not all equal.

The cavity method predictions are based on the assumption that the optimal algorithm for Bayesian CSPs is *belief propagation* (BP), an algorithm that aims to estimate marginals of $\boldsymbol{x}|\boldsymbol{E}$. The algorithm considers the bipartite graph $\mathcal{H}$ between variables and observations, and maintains a distribution $m^{a\to b}$, called a "message", over Σ on every directed edge (a, b). These messages are iteratively updated via an update rule that depends on the model from which the input is drawn, until the messages arrive at a fixed point. In particular, for every (a, b), there is a function $\Upsilon_{a\to b}$ such that: $m^{a\to b}$ is updated to $\Upsilon_{a\to b}\big(\{m^{c\to b} : c \in \partial a \setminus b\}\big)$.

The predictions of the cavity method are based on the behavior of BP around a certain set of fixed point messages $\overline{m}$, called the *uninformative fixed point*, since this set of messages does not depend on the graph $\mathcal{H}$, but only on the model that the instance is drawn from, and hence does not have any mutual information with the hidden assignment. When the uninformative fixed point is randomly perturbed, and BP is run on the perturbed messages, one of two things can happen: either running BP causes the perturbation to vanish and takes the messages back to the uninformative fixed point, in which case we say that the fixed point is *stable*; or running BP amplifies the magnitude of the perturbation and moves away from the uninformative fixed point (presumably towards the true solution). The hypothesis is that if the uninformative fixed point is *stable*, then recovery is hard, and if the uninformative fixed point is *unstable*, then recovery is easy.

In Chapter 3, we give a simple spectral algorithm, inspired by belief propagation and captured by the Local Statistics hierarchy, that succeeds at recovery whenever the cavity method predicts that recovery should be possible via an efficient algorithm.

Theorem 1.2.2 (Informal)**.** *Given an instance of a Bayesian CSP drawn from a model that the cavity method predicts is algorithmically tractable, there is a polynomial time algorithm to recover the hidden signal better than random guessing with high probability over the randomness of the instance.*

A striking feature of this method to obtain predictions is that the stability of the uninformative fixed point can be tested by checking if the top eigenvalue of a certain constant-sized matrix depending only on the model is at least 1. We show three examples below to illustrate its rich predictions.

Example 1.2.3. First, consider the problem of planted NAE3SAT wherein there is a uniformly random assignment in $\{0,1\}^n$ and Not-All-Equal clauses on 3 variables are sampled so that a ρ-fraction of them are satisfied. As one varies the average constraint-degree of a variable d and the approximation ρ, there is an explicit prediction of the region of parameters where the problem of recovering the planted assignment is computationally tractable (blue region in Figure 1.2).

Example 1.2.4. Next, we turn our attention to mixed planted CSPs. For concreteness, we consider one particular example: planted NAE-$(3,5)$-SAT. In this example, the variables are given a uniformly random assignment in $\{0,1\}^n$ and

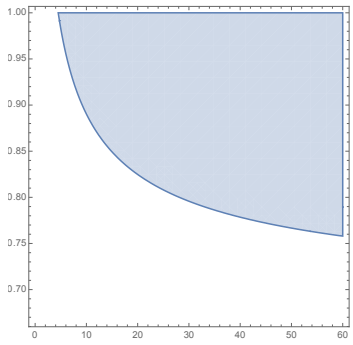


Figure 1.2: Easy region for planted NAE-3SAT shaded in blue. Average degree on x-axis, fraction of clauses satisfied on y-axis.

Not-All-Equals clauses are sampled to be on 3 variables with probability p and on 5 variables with probability $1-p$. As one varies the constraint-degree of a variable d and the proportion of NAE3SAT clauses p, we can plot a precise region of parameters where the recovery problem is computationally tractable (blue region in Figure 1.3).

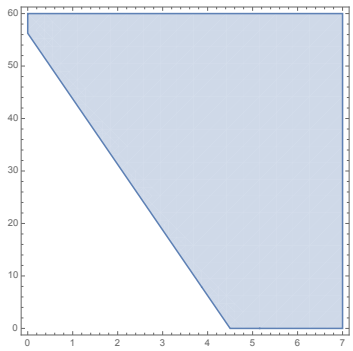
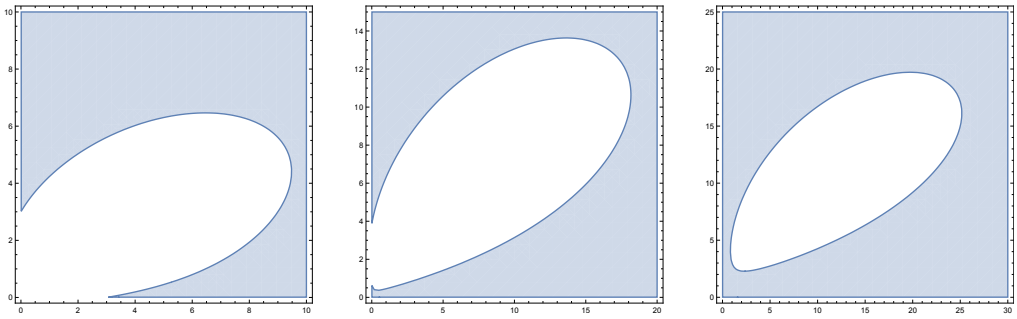


Figure 1.3: Easy region for planted NAE-$(3,5)$-SAT shaded in blue. Average NAE3-degree of vertex on x-axis, average NAE5-degree of vertex on y-axis.

Example 1.2.5. Consider the following version of 4-community stochastic block model with communities labeled $(0,0)$, $(0,1)$, $(1,0)$ and $(1,1)$ and 3 parameters d_0, d_1 and d_2. For a pair of vertices u and v from communities x and y we place an edge between u and v with probability $\frac{d_{\text{dist}(x,y)}}{n}$ where $\text{dist}(x,y)$ is the Hamming distance between x and y. For an additional twist, let us suppose that the first coordinate of

the community that every vertex belongs to is also revealed to the algorithm. What is the region of parameters d_0, d_1, d_2 for which an efficient algorithm can partially recover the second coordinate of the community labels? See Figure 1.4 for the hypothesized transition.



(a) $d_2 = 1$ (b) $d_2 = 5$ (c) $d_2 = 9$

Figure 1.4: Easy regions for (d_0, d_1) for variety of settings of d_2.

1.3 Open directions

Local algorithms. The spectral and semidefinite programming based algorithms for many of these inference problems find a solution positively correlated with the hidden solution, but are far from the optimal achievable overlap. Empirically, via heuristic calculations, and also in some special settings like the 2-community stochastic block model, it appears that running a local algorithm such as belief propagation or the Glauber dynamics Markov chain on top of the output of a "global" algorithm such as a spectral algorithm of semidefinite program boosts the overlap to optimality.

> *Can we give optimal algorithms for recovery with rigorously provable guarantees?*

Lower bounds. The broad goal here is to gather rigorous evidence for the validity of the hardness predictions. One such form of evidence is lower bounds against restricted families of algorithms. A concrete problem in this direction is:

Can we prove lower bounds against the subexponential time regime of the local statistics SDP hierarchy for hypothesis testing problems that the cavity method predicts are intractable?

Proving such a lower bound would require constructing a *pseudodistribution*[1] that fools the SDP into seeing a solution even when there is none. A candidate construction of such a pseudodistribution arises from the stable uninformative fixed point messages, and it is natural to investigate whether this pseudodistribution indeed fools the SDP.

The work of [HS17] proposes a heuristic called the *low-degree likelihood ratio* (LDLR) heuristic for predicting the computational complexity of general hypothesis testing problems. It was also shown in the same work that the heuristic predicts the same threshold as the cavity method for general stochastic block models, which suggests that the predictions of the LDLR heuristic captures those of the cavity method.

Does the LDLR heuristic predict the same threshold as the cavity method on Bayesian CSPs?

The LDLR heuristic is closely related to and was inspired from *pseudocalibration*, a technique to construct candidate pseudodistributions for SDPs for hypothesis testing problems pioneered in [BHK+19]. On many example problems, pseudocalibration can be used to prove that a problem is hard for SDPs when the LDLR heuristic predicts hardness. Strikingly, for Bayesian CSPs, pseudocalibration appears to produce the same pseudodistribution obtained from the BP fixed point. This raises the question as to whether pseudocalibration can be used to prove lower bounds against SDPs when the LDLR heuristic predicts that a problem is computationally hard.

Does the LDLR hardness for distinguishing problems imply hardness against semidefinite programs?

[1] A pseudodistribution is a collection of marginal probability distributions on a bounded number of variables satisfying some constraints but which do not need to be consistent with a global true probability distribution.

1.4 Connections

Much of the technical content in Chapter 2 and Chapter 3 involves understanding the spectra of sparse random matrices. The proof techniques are amenable to several other applications in theoretical computer science. In this thesis, we focus on two of these applications: constructing near-Ramanujan expander graphs, and in understanding girth-density tradeoffs in hypergraphs.

1.4.1 Expander constructions

Graph expansion has a variety of applications in theoretical computer science to error-correcting codes, pseudorandomness, metric embeddings, approximation algorithms, and more, for which we refer the interested reader to [HLW06].

Formally, a λ-*spectral expander* is a graph such that the second largest absolute eigenvalue of its normalized adjacency matrix is at most λ. A bound on the second largest eigenvalue is a powerful analytic handle on graphs as it controls various combinatorial quantities of interest, such as the mixing time of random walks, the sparsity of cuts, and the edge density within small subgraphs, which in turn lend such graphs to various applications. It is of interest to explicitly construct expanders with small value of λ, since the smaller the value of λ is, the more the expander resembles the complete graph with respect to these combinatorial properties. A second demand we have of expander constructions in the context of applications is *sparsity*: ideally, all the vertices should have constant degree.

In light of this, it is natural to try to understand the tradeoff between sparsity and spectral expansion.

> *For a fixed d, what is the smallest λ such that there is a family of d-regular expander graphs?*

The first result proved in this context was the Alon–Boppana bound [Nil91].

Theorem 1.4.1. *For any n-vertex d-regular graph G, $|\lambda|_2(G) \geqslant \frac{2\sqrt{d-1}-o_n(1)}{d}$.*

We say a graph is *Ramanujan* if it is an optimal spectral expander, i.e. if $|\lambda|_2(G) \leqslant \frac{2\sqrt{d-1}}{d}$. Understanding the existence of Ramanujan graphs has been a topic of extensive study. Here we list some key results.

1. Using number theoretic techniques, [LPS88, Mar88] and later [Mor94] gave constructions of d-regular Ramanujan Cayley graphs for d of the form $p^r + 1$ for prime p and integer r.

2. Friedman's theorem [Fri08] says that an n-vertex random d-regular graph is near-Ramanujan (i.e. has second eigenvalue $\frac{2\sqrt{d-1}+\varepsilon}{d}$ for arbitrarily small ε) with high probability for all $d \geqslant 3$. Eventually, a simpler proof was given by [Bor19].

3. [MSS15a, MSS15b] pioneered the *method of interlacing polynomials* and used it to nonconstructively show the existence of bipartite Ramanujan graphs of every degree (and therefore the minimum eigenvalue is -1). The latter was also made algorithmic by [Coh16].

While a random d-regular graph is the easiest to describe and doesn't require constraints on the degree or bipartiteness, the constructions based on Cayley graphs and interlacing polynomials are explicit, in the sense of being deterministic and constructible in polynomial time.

In Chapter 4, we give an explicit construction of near-Ramanujan graphs of every degree based on derandomizing Friedman's theorem, using the trace method for analyzing sparse random matrices.

Theorem 1.4.2. *For every $d \geqslant 3$ and $\varepsilon > 0$, there is an algorithm that takes in n as input, and in time $\mathrm{poly}(n)$, outputs a d-regular graph G on $\Theta(n)$ vertices such that $|\lambda|_2(G) \leqslant \frac{2\sqrt{d-1}+\varepsilon}{d}$.*

1.4.2 Girth-density tradeoffs

A common theme in extremal combinatorics is that of *Turán-type problems*, where one studies what subgraphs inevitably arise in dense enough graphs.

The following question is our starting point.

> *What is the maximum girth possible in a graph with average degree d?*

This question was answered by [AHL02], who showed that any graph with average degree d must have a cycle of length at most $2\log_{d-1} n + 2$.

Here, we study the hypergraph analog of this question. The generalization of a cycle that we consider here, relevant in the context of understanding *refutation*

of random constraint satisfaction problems, and in understanding rate-distance tradeoffs in low-density parity check codes is that of an *even cover*.

Definition 1.4.3. An *even cover* in a hypergraph $\mathcal{H}$ is a collection of hyperedges S such that each vertex is touched an even number of times by S.

The notion of girth of a hypergraph appropriately generalizes to being the size of the smallest even cover contained inside the hypergraph.

[Fei08] and [NV08] initiated the study of understanding the tradeoff between the density and girth in hypergraphs, and the full tradeoff was conjectured by Feige.

Conjecture 1.4.4. *A k-uniform hypergraph with $n \cdot \left(\frac{n}{r}\right)^{\frac{k}{2}-1}$ hyperedges must contain an even cover with $r \log n$ hyperedges.*

This conjecture was proved by [GKM21] up to a poly log n factor, with an exponent depending linearly in k, via a technique pioneered by them called "spectral double counting".

The main analytic handle on the hypergraph is via a generalized adjacency matrix encoding which hyperedges are present, known as its *Kikuchi matrix*. They showed that on one hand the density of edges in the hypergraph could be used to lower bound the maximum eigenvalue of the Kikuchi matrix, and the girth can be used to obtain an upper bound. The resulting constraints can be used to prove Feige's conjecture.

In Chapter 5, we give a significantly simpler proof of Feige's conjecture by carrying out the spectral double counting using techniques from analyzing sparse random matrices, which is tight up to a single $\log n$ factor.

Chapter 2

Local Statistics SDP

This chapter is adapted from [BMR21], a paper co-authored by the author of this thesis, Jess Banks, and Prasad Raghavendra.

We propose a hierarchy of semidefinite programming relaxations for inference and hypothesis testing problems and prove that it succeeds at robust community detection in stochastic 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.

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

In the case of the (irregular) Stochastic Block Model, it is known that efficient algorithms exist all the way down to this threshold, although none are robust to a linear number of adversarial perturbations of the graph when the average degree is small. More importantly, there is little complexity-theoretic evidence that detection is hard below the threshold.

2.1 Introduction

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

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

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

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

and a 'signal-to-noise ratio'

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

In this setup, it is believed that as we hold k and λ constant, then there is an *information-theoretic threshold* $d_{IT} \approx \frac{\log k}{k\lambda^2}$, in the sense that when $d < d_{IT}$ both detection and recovery are impossible for any algorithm. Moreover, Decelle et al. conjecture that efficient algorithms for both tasks exist only when the degree is larger than a point known as the *Kesten-Stigum threshold* $d_{KS} = \lambda^{-2}$. Much of this picture is now rigorous [MNS18, Mas14a, BLM15, ABH16]. Still, fundamental questions remain unanswered. What evidence can we furnish that detection and recovery are indeed intractible in the so-called 'hard regime' $d_{IT} < d < d_{KS}$? How robust are these thresholds to adversarial noise or small deviations from the model?

Zooming out, this discrepancy between information-theoretic and computational thresholds is conjectured to be quite universal among planted problems, where we are to reconstruct or detect a structured, high-dimensional signal observed through a noisy channel. The purpose behind our work is to begin developing a framework capable of providing evidence for average case computational intractability in such settings. To illustrate this broader motivation, consider a different average-case problem also conjectured to be computationally intractable: refutation of random 3-SAT. A random instance of 3-SAT with n literals and, say $m = 1000n$ clauses is unsatisfiable with high probability. However, it is widely conjectured that the problem of *certifying* that a given random 3-SAT instance is unsatisfiable is computationally intractable (all the way up to $n^{3/2}$ clauses) [Fei02a]. While proving intractability remains out of reach, the complexity theoretic literature now contains ample evidence in support of this conjecture. Most prominently, exponential lower bounds are known for the problem in restricted computational models such as linear and semidefinite programs [Gri01] and resolution based proofs [BSW01]. Within the context of combinatorial optimization, the Sum-of-Squares (SoS) SDPs yield a hierarchy of successively more powerful and complex algorithms which capture and unify many other known approaches. A lower bound against the SoS SDP hierarchy such as [Gri01] provides strong evidence that this refutation problem is computationally intractable. This paper is a step towards developing a similar framework to reason about the computational complexity of detection and recovery in stochastic block models specifically, and planted problems generally.

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

them. Their main result is a statistical lower bound indicating the phase transition for weak recovery changes in the presence of monotone errors. This still leaves open the question of whether there exist algorithms that weakly recover right at the threshold and are robust to $o(|E|)$ perturbations in the graph.

2.2 Main Results

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

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

Definition 2.2.1 (Detection and Robustness)**.** Let $\mathcal{P}_n$ and $\mathcal{N}_n$ denote two sequences of distributions on graphs. We say that an algorithm $\mathsf{A} : \text{Graphs} \to \{\mathrm{P}, \mathrm{N}\}$ *solves the detection problem*, or *can distinguish $\mathcal{P}_n$ and $\mathcal{N}_n$*. if

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

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

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

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

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

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 $T \triangleq M\mathrm{Diag}\pi$, noting that T is the transition matrix for a reversible Markov chain with stationary distribution π. For any vertex in group i, the label of a uniformly random neighbor is roughly distributed according to the ith row of T, and, more generally, the vertex labels encountered by a random non-backtracking random walk are approximately governed by the Markov process that T defines. As this process is stationary, the spectrum of T is real, and we will write its eigenvalues as $1 = \lambda_1 \geqslant |\lambda_2| \geqslant \cdots \geqslant |\lambda_k|$. The second eigenvalue λ_2 is a generalization of the signal-to-noise ratio λ from equation (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 theorem regarding the SBM is that, when $d > d_{\mathrm{KS}}$, the $\mathrm{LoSt}(2, D)$ SDP can robustly solve the detection problem for some $D = O(1)$ (albeit tending to infinity as $d \to d_{\mathrm{KS}}$).

Theorem 2.2.2. *Let $\mathcal{N}_n = \mathcal{G}(n, d/n)$, and $\mathcal{P}_n$ denote the n-vertex SBM with parameters (d, k, M, π). If $d > d_{\mathrm{KS}}$, then there exist $\delta > 0$, $D = O(1)$, and $\rho > 0$ (all dependent on d) for which the $\mathrm{LoSt}(2, D)$ SDP with error tolerance δ can ρ-robustly solve the detection problem.*

We additionally show that a simplified version of the $\mathrm{LoSt}(2, D)$ SDP (Definition 2.6.2) which is powerful enough to solve the detection problem above the KS

threshold, fails to do so below it at every constant level. This is the content of the forthcoming Theorem 2.6.3.

The Degree Regular Block Model We will parametrize the DRBM identically to the SBM, by a quadruple (d, k, M, π); this time we of course require that d is an integer. To sample a graph $\boldsymbol{G} = (V(\boldsymbol{G}), E(\boldsymbol{G}))$, first choose a uniformly random "π-balanced" partition $V(\boldsymbol{G}) = \bigsqcup_{i\in[k]} V_i(\boldsymbol{G})$, by which we mean that $|V_i(\boldsymbol{G})| = \pi(i)n$ for every i. Then, choose a uniformly random d-regular graph, conditioned on there being exactly $\pi(i)\pi(j)M(i,j)\cdot dn$ edges between each pair of distinct groups $i \neq j$, and $\pi(i)^2 M(i,j)\cdot dn/2$ edges internal to each group i. For simplicity, we will assume that the parameters are such that these group sizes and edge counts are integers. As with the SBM, we will call the model *symmetric* if the entries of M are constant on the diagonal and off-diagonal respectively. As a warm-up for the main technical arguments of the paper, we will study in Section 2.4 a simplified version of the Local Statistics SDP that can solve the detection problem on the symmetric DRBM.

Remark 2.2.3. The DRBM as we have defined it differs from the Regular Stochastic Block Model of [BDG+16], in which each vertex has a prescribed number of neighbors in every community. Although superficially similar, the behavior of this 'equitable' model (as it is known in the physics literature [NM14]) is quite different from ours. For instance, [BDG+16] show that whenever detection is possible in the two community case, one can *exactly* recover the planted labels. This is not true in our setting.

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

Our main result on the DRBM is analogous to Theorem 2.2.2 on the SBM.

Theorem 2.2.4. *Let $\mathcal{N}_n$ denote the uniform distribution on d-regular graphs with n-vertices, and $\mathcal{P}_n$ the DRBM with parameters (d, k, M, π). If $d > d_{\mathrm{KS}}$, then there exists a constant $m \in \mathbb{N}$, $\delta > 0$, and and $\rho > 0$ (all dependent on d) so that* $\mathrm{LoSt}(2, m)$ *with*

error tolerance δ can ρ-robustly solve the detection problem. Conversely, if $d < d_{\mathrm{KS}}$, then every constant level, no matter the error tolerance, fails to do so.

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

Corollary 2.2.5. *Let $\boldsymbol{G}$ be drawn from the DRBM with parameters (d, k, M, π) satisfying $d > d_{\mathrm{KS}} + \epsilon$. There exists some $\eta = \eta(\epsilon)$ such that, for each eigenvalue λ of M satisfying $|\lambda| > 1/\sqrt{d-1} + \epsilon$, the adjacency matrix $A_{\boldsymbol{G}}$ is guaranteed one eigenvalue μ satisfying $|\mu| > 2\sqrt{d-1} + \eta$.*

Future Work Regrettably, we do not solve the problem of recovery above Kesten-Stigum in either model. However, we will in Appendix 2.11 reduce recovery in the DRBM to the following conjecture regarding the spectrum of $A_{\boldsymbol{G}}$ for $\boldsymbol{G}$ drawn from the planted model.

Conjecture 2.2.6. *Let $\mathcal{P}_{(d,k,M,\pi)}$ be any DRBM with $|\lambda_1|, ..., |\lambda_\ell| > (d-1)^{-1/2}$. Then, for any η, with high probability, $A_{\boldsymbol{G}}$ has only ℓ eigenvalues with modulus larger than $2\sqrt{d-1} + \eta$.*

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

reader to [MMV16] for a more detailed survey. In the logarithmic-degree regime, robust algorithms for community detection are developed in [CL+15, KK10, AS12]. Far less is known in the case of regular graphs.

2.3 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, $\langle X, Y \rangle = \mathsf{Tr} X^T Y$ 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.

2.3.1 Optimization vs Inference

While it was suspected that a semidefinite programming relaxation could be used towards community detection in sparse stochastic block models, many earlier attempts at it [GV16, MS15] failed to detect communities right up to the KS threshold at a fixed degree. These works studied the Goemans-Williamson SDP relaxation for MaxCut applied to the problem of detecting two communities ($k = 2$). The idea being that if we consider a two community SBM with $p_{out} > p_{in}$, then the partition induced by the communities should have an unusually large number $(\frac{dn}{2} \cdot \frac{p_{out}}{p_{out}+p_{in}})$ of crossing edges. Hence an SDP relaxation of MaxCut could be harnessed towards detecting and possibly recovering the communities. Indeed, in this special case, the maximum bisection in the graph is a Maximum Likelihood Estimate (MLE) for the communities x given the graph G, i.e., $x = \arg\max_x p(x|G)$.

This approach of casting inference as optimization has its limitations. In particular, as one approaches the KS threshold, the number of crossing edges between

the two communities, namely $\frac{dn}{2} \cdot \frac{p_{out}}{p_{out}+p_{in}}$, is lower than the value of MaxCut in a random Erdos-Renyi graph! In other words, if we run an exponential-time algorithm that finds the maximum cut via a brute-force enumeration, then it will find a better MaxCut in a random Erdos-Renyi graph than the true communities in the planted model. It is therefore unclear whether an SDP relaxation of MaxCut can solve the problem.

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

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

2.3.2 Detection, Refutation, and Sum-of-Squares

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

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

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

$$\textit{Normalized} \qquad \mathbb{E}\, 1 = 1 \tag{2.2}$$

$$\textit{Satisfies of } \mathcal{S} \qquad \mathbb{E}\, f_i(x) \cdot p(x) = 0 \qquad \forall i \in [m], \forall p \in \mathbb{R}[x] \tag{2.3}$$

$$\textit{Positive} \qquad \mathbb{E}\, p(x)^2 \geqslant 0 \qquad \forall p \in \mathbb{R}[x] \tag{2.4}$$

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

Refuting the constraint satisfaction problem, e.g. proving that $\mathcal{S} = \emptyset$, is equivalent to showing that no operator obeying (2.2)-(2.4) can exist. The key insight of SoS is that often one can do this by focusing only on polynomials of some bounded degree. Writing $\mathbb{R}[x]_{\leqslant D}$ for the polynomials of degree at most D, we call an operator $\widetilde{\mathbb{E}} : \mathbb{R}[x]_{\leqslant D} \to \mathbb{R}$ a *degree-D pseudoexpectation* if it is normalized, positive, and satisfies $\mathcal{S}$ for every polynomial in its domain. It is well-known that one can search for a degree D pseudoexpectation with a semidefinite program of size $O(n^D)$, and if this smaller, relaxed problem is infeasible, we've shown that $\mathcal{S}$ is empty. This is the *degree-D Sum-of-Squares relaxation* of our CSP.

2.3.3 The Local Statistics Hierarchy

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

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

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

$$G_{u,v}^2 = G_{u,v} \qquad \forall u, v \in [n] \tag{2.5}$$

$$x_{u,i}^2 = x_{u,i} \qquad \forall u \in [n],\, i \in [k] \tag{2.6}$$

$$x_{u,1} + \cdots + x_{u,k} = 1 \qquad \forall u \in [n]. \tag{2.7}$$

For brevity, we will throughout the paper denote by $\mathcal{B}_k$ the set of polynomials constraints in the x variables appearing in (2.6) and (2.7). 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 2.3.1. Although we have stated it in the specific context of the DRBM, the local statistics framework extends readily to any planted problem involving a joint distribution μ on pairs (G, 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.

2.3.4 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:

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

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 $\overline{A}_G^{(m)} = (A - (d/n)\mathbb{J})^{(m)}$ 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 $\overline{A}_G^{(m)} = (A - (d/n)\mathbb{J})^{(m)}$ 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 [Fri03a] 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 $\overline{A}_G^{(m)}$, when $\boldsymbol{G}$ is a sparse Erdős-Rényi random graph, is the presence of of certain localized combinatorial structures which inflate the number of non-backtracking walks: high-degree vertices and small subgraphs with many cycles. 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 $|\mathsf{B}_t(v, G)| \geqslant (1+\varepsilon)^t d^t$. We will call a vertex v (t, r, ε)-*vexing* if either it participates in a cycle of length less than r or it is (t, ε)-heavy.

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

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

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

Theorem 2.3.2. *With probability* $1 - n^{-100}$,

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

2.3.5 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 2.3.3 (Linkages)**.** A closed walk W of length $k\ell$ is a $(k \times \ell)$-*linkage* if it can be split into k segments each of length-ℓ such that the walk W is nonbacktracking on each segment. Each ℓ-step non-backtracking segment is a "link".

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

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

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

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

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

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

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

The remainder of the paper will be laid out as follows. Before embarking on our investigation of the Local Statistics SDP in the DRBM and SBM in full generality, we will in Section 2.4 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 2.5 to the case of the general DRBM, proving Theorem 2.2.4 by way of a reduction to some key results from this simpler, symmetric case. Finally, in Section 2.6 we prove Theorem 2.2.2 regarding the SBM.

2.4 A Simplified SDP for the Symmetric DRBM

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

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

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

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

$$\boldsymbol{Y} \triangleq \frac{k}{k-1}\left(\boldsymbol{y}\boldsymbol{y}^* - \frac{1}{k}\mathbb{J}\right) \succeq 0. \tag{2.8}$$

This is a rank-$(k-1)$ positive semidefinite matrix that is n/k times the projector onto the subspace spanned by the indicator vectors for the k groups and orthogonal

to the all-ones vector. The inner product $\langle \boldsymbol{Y}, A_{\boldsymbol{G}}^{(s)} \rangle$ counts non-backtracking walks weighted according to the labels of their initial and terminal vertices.

Lemma 2.4.1. *Let $(\boldsymbol{G}, \boldsymbol{Y}) \sim \mathcal{P}$. Then for every $s \geqslant 1$,*

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

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

Definition 2.4.2. Fix a small number $\delta > 0$. The *level m symmetric path statistics SDP* with error tolerance $\delta > 0$, on input G_0, is the feasibility problem

$$\begin{aligned} \text{Find } Y \succeq 0 \text{ s.t.} \qquad Y_{u,u} &= 1 && \forall u \in [n] \\ \langle Y, \mathbb{J} \rangle &= 0 \\ \left| \langle Y, A_G^{(s)} \rangle - \lambda^s d(d-1)^{s-1} n \right| &\leqslant \delta n && \forall s \in [m] \end{aligned} \tag{2.9}$$

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

Theorem 2.4.3. *If $(d-1)\lambda^2 > 1$, then there exists constant $m \in \mathbb{N}$, $\delta > 0$, and $\rho > 0$ so that $SPS(m, \lambda)$ solves the detection problem ρ-robustly. Conversely if $(d-1)\lambda^2$ then no such m, δ, ρ exist.*

2.4.1 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,

$$\begin{aligned} A_G^{(0)} &= \mathbb{1} \\ A_G^{(1)} &= A_G \end{aligned}$$

$$A_G^{(2)} = A_G^2 - d\mathbb{1}$$
$$A_G^{(s)} = AA_G^{(s-1)} - (d-1)A_G^{(s-2)} \qquad s > 2,$$

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

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

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

with its associated inner product

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

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

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

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

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

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

Lemma 2.4.4. *For any $\varepsilon > 0$, there exists an $\eta > 0$ such that for $z \in [-2\sqrt{d-1} - \eta, 2\sqrt{d-1} + \eta]$,*

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

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

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

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

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

Lemma 2.4.5. *For any constant C and L, with high probability any graph $\boldsymbol{G} \sim \mathcal{P}$ has at most $O(\log n)$ bad vertices.*

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

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

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

2.4.2 Distinguishing

Let us now prove the first assertion in Theorem 2.4.3, namely that if $(d-1)\lambda^2 > 1$, then the $SPS(m, \lambda)$ SDP, for some $\delta > 0$ sufficiently large m, can distinguish the null and planted models. From Lemma 2.4.1, if $(\boldsymbol{G}, \boldsymbol{Y}) \sim \mathcal{P}$, then the matrix $\boldsymbol{Y}$ from equation (2.8) is with high probability a feasible solution to SDP (2.9). 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 2.4.6. *If there exists a degree-m polynomial $f \in \mathbb{R}[z]$ which is (i) strictly nonnegative on the interval $[-2\sqrt{d-1}, 2\sqrt{d-1}]$ and (ii) satisfies*

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

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

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

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

$$f = \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} q_m. \tag{2.10}$$

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

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

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

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

$$\leqslant \sum_{s=0}^{m} \frac{\langle f, q_s \rangle_{\mathrm{KM}}}{\|q_s\|_{\mathrm{KM}}^2} \cdot \lambda^s \|q_s\|_{\mathrm{KM}}^2 n + \delta \sum_{s=0}^{m} \frac{|\langle f, q_s \rangle_{\mathrm{KM}}|}{\|q_s\|_{\mathrm{KM}}^2}$$
$$\leqslant \langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle + \delta \sqrt{m} \|f\|_{\mathrm{KM}} < 0$$

□

The following proposition implies a proof of the first part of Theorem 2.4.3.

Proposition 2.4.7. *If $\lambda^2(d-1) > 1$, there exists a polynomial satisfying the hypotheses of Proposition 2.4.6.*

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

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

which by Lemma 2.4.4 has the desired positivity property. This choice of f satisfies

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

which is negative when

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

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

2.4.3 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 2.4.8. *If there exists a polynomial $g \in \mathbb{R}[z]$ that is (i) strictly positive on $(-2\sqrt{d-1}, 2\sqrt{d-1})$, and (ii) satisfies*

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

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

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

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

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

First, note that by expanding g in the non-backtracking basis and invoking Lemma 2.4.5, for any $0 \leqslant s \leqslant m$ we have

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

since $g(d)\|q_s\|_{\mathrm{KM}}^2$ is a constant. Moreover, as g is strictly positive between $-2\sqrt{d-1}$ and $2\sqrt{d-1}$ 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 2.4.5 we know that there are at most $O(\log n)$ of these *bad* vertices (keeping the terminology from the aforementioned Lemma). However, when we subtract $g(d)\mathbb{J}/n$, even the $\Omega(n - \log n)$ diagonal entries equal to one—those corresponding to *good* vertices—are shifted. Let us therefore define

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

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 $A_G^{(s)}$ are all $O(1)$. Thus, writing Γ for the set of good vertices, we know

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

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

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

Our final matrix Y will be the Gram matrix of these new β and remaining α vectors. We must finally check that the affine constraints against the $A_G^{(s)}$ matrices are still approximately satisfied. However, even starting from a bad vertex, there are at most a constant number of vertices within s steps of it, and at most a constant number of non-backtracking walks to any such vertex. Thus

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

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

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

for n sufficiently large. □

The second part of Theorem 2.4.3 ensues from the following proposition.

Proposition 2.4.9. *Whenever $\lambda^2(d-1) < 1$, there exists a polynomial satisfying the conditions of Proposition 2.4.8.*

Proof. Such a polynomial y is exactly of the form

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

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

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

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

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

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

One an quickly verify that

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

so we only need to check that $\lambda^2(d-1) < 1$ ensures $\lambda^{p+2}(d-1)q_p - \lambda^{p+1}q_{p+1} \to_p 0$. This follows immediately from Lemma 2.4.4, as $|q_p| \leqslant 2p\sqrt{d(d-1)^p}$. □

2.4.4 Robustness

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

Assume that $\boldsymbol{G}$ was drawn from either the planted or null distribution, and that $\tilde{\boldsymbol{H}} \approx_\rho \boldsymbol{G}$. When we defined the $SPS(m, \lambda)$ SDP, we stipulated that in the event

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

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

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

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

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

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

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

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

To prove (ii), assume $\boldsymbol{G} \sim \mathcal{N}$ and $(d-1)\lambda^2 > 1$. Infeasibility of the SDP on input $\boldsymbol{G}$ is witnessed by the polynomial f from Proposition 2.4.7. So, let $\boldsymbol{Y}$ be a putative solution to the SDP on input $\boldsymbol{H}$, at tolerance δ', seeking a contradiction: recycling some computations from the proof of Proposition 2.4.6

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

$$\leqslant \langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle_{\mathrm{KM}} + 2\rho\xi \sum_{s=0}^{m} |\langle f, q_s \rangle|_{\mathrm{KM}} + \delta' \sqrt{m} \|f\|_{\mathrm{KM}}$$
$$\leqslant \langle f, \sum_{s=0}^{m} \lambda^s q_s \rangle_{\mathrm{KM}} + (\delta' + 2\rho\xi)\sqrt{m}\|f\|_{\mathrm{KM}}.$$

Thus we have a contradiction if

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

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

2.5 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 2.4. 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 2.2.4.

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

$$\begin{aligned} G_{u,v}^2 - G_{u,v} &= 0 \\ x_{u,i}^2 - x_{u,i} &= 0 \\ \sum_i x_{u,i} - 1 &= 0 \end{aligned} \tag{2.11}$$

as in the introduction, we will write the ideal generated by the polynomials on the left of the second two equations as $\mathcal{B}_k$. Any point x in the vanishing locus

of $\mathcal{B}_k$ corresponds to a map $\sigma_x : [n] \to [k]$. Write $\mathbb{S}[G, x] \subset \mathbb{R}[G, x]$ for the vector subspace of multilinear polynomials, fixed under the action of the symmetric group $\mathfrak{S}_n$ on the index set $[n]$, and for which no monomial contains $x_{u,i}x_{u,j}$ for $i \neq j$. This contains some polynomials that vanish modulo the equations above, but is convenient to work with.

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

Definition 2.5.1 (Partially Labelled Subgraph)**.** A *partially labelled graph* (H, S, τ) consists of a graph H, distinguished subset of vertices $S \subset V(H)$, and a labelling $\tau : S \to [k]$. An *occurrence* of (H, S, τ) in a fully labelled graph (G, σ) is an injective homomorphism $\varphi : H \to G$ which respects the labelling. In other words, it is an injective map $\varphi : V(H) \to V(G)$ satisfying (i) $(\varphi(u), \varphi(v)) \in E(G)$ for every edge $(u, v) \in E$, and (ii) $\sigma(\varphi(v)) = \tau(v)$ for every $v \in S$.

Lemma 2.5.2 (Partially Labelled Subgraphs are a Basis)**.** *Let (H, S, τ) be a partially labelled subgraph. Then there is a symmetric polynomial $p_{H,S,\tau} \in \mathbb{R}[G, x]$ with degree $|S|$ in x and $|E(H)|$ in G that, for any $(\boldsymbol{G}, \boldsymbol{x})$ satisfying equations (2.11), counts occurrences of H in $(\boldsymbol{G}, \sigma_{\boldsymbol{x}})$. Furthermore, these polynomials form a basis for $\mathbb{S}[G, x]$.*

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

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

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

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

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

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

$$C_H(d) \triangleq \frac{\prod_{v \in V(H)} (d)_{\deg(v)}}{d^{|E(H)|}} \tag{2.12}$$

$$L_{(H,S,\tau)}(M, \pi) \triangleq \sum_{\hat{\tau}: \hat{\tau}|_S = \tau} \prod_{v \in v(H)} \pi(\hat{\tau}(v)) \prod_{(u,v) \in E(H)} M_{\hat{\tau}(u), \hat{\tau}(v)}. \tag{2.13}$$

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

Theorem 2.5.3 (Local Statistics)**.** *Let (H, S, τ) be a partially labelled graph with $O(1)$ edges. Then, with high probability over $(\boldsymbol{x}, \boldsymbol{G}) \sim \mathcal{P}$,*

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

The proof may be found in Appendix 2.8, but some comments are in order here. First, when H is a forest and $\chi(H) = c(H)$, we see that $p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G})$ concentrates, and that

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

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

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

π, and for each neighbor choose a label j with probability $T_{i,j} = M_{i,j}\pi(j)$. If one continues this until all of H is labelled, $L_{(H,S,\tau)}(M,\pi)$ gives the probability that every vertex $s \in S$ is given label $\tau(s)$.

We may finally define formally the Local Statistics algorithm.

Definition 2.5.4. The *degree* (D_x, D_G) *Local Statistics algorithm* with error tolerance $\delta > 0$, on input G_0, is the following SDP: find a pseudoexpectation $\widetilde{\mathbb{E}} : \mathbb{R}[x]_{\leqslant D_x} \to \mathbb{R}$ that is positive, normalized, satisfies $\mathcal{B}_k$, and for which

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

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

Lemma 2.5.5. *For any $\delta > 0$, the* $\mathrm{LoSt}(D_x, D_G)$ *algorithm is with high probability feasible on input $\boldsymbol{G} \sim \mathcal{P}$.*

Proof. Let $\boldsymbol{x}$ be the hidden signal; we will set $\widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G}) = p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G})$. This is clearly positive, satisfies $\mathcal{B}_k$, and from Theorem 2.5.3 it satisfies the affine constraints in Definition 2.5.4. □

2.5.2 Distinguishing

Let us prove the first part of Theorem 2.2.4: when $(d-1)\lambda_2^2 > 1$, there exist constant m, ρ, and $\delta > 0$ for which the $\mathrm{LoSt}(2,m)$ SDP at error tolerance δ solves the detection problem ρ-robustly. Since the SDP is with high probability feasible for any m and $\delta > 0$ when $\boldsymbol{G} \sim \mathcal{P}$, it remains only to show infeasibility for some m, ρ, δ when $\boldsymbol{G} \sim \mathcal{N}$.

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

Lemma 2.5.6. *For any i, j, and any $s = 0, ..., m$, recall that $A_G^{(s)}$ is the matrix counting non-backtracking walks of length s, and $\mathbb{J}$ is the all-ones matrix. For any $\delta' > \delta$,*

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

Proof. For the first assertion, let (H, S, τ) be the path of length s whose endpoints are labelled $i, j \in [k]$. In this case $C_H(d) = d(d-1)^{s-1} = \|q_s\|_{\mathrm{KM}}^2$, and one can quickly verify that $L_{(H,S,\tau)} = \pi(i) T_{i,j}^s$. Each *self-avoiding* walk of length s in G is an occurrence of H, so from Theorem 2.5.3

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

It is an easy consequence of Lemma 2.4.5 that for every constant s, $A_{\boldsymbol{G}}^{(s)}$ and $A_{\boldsymbol{G}}^{\langle s \rangle}$ differ only on $O(\log n)$ rows, and since each row has constant L_2 norm,

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

The matrix X has diagonal elements $X_{(u,i),(u,i)} = \widetilde{\mathbb{E}}\, x_{u,i}^2 = \widetilde{\mathbb{E}}\, x_{i,u}$ by the Boolean constraint, and $\widetilde{\mathbb{E}}\,(x_{u,1} + \cdots + x_{u,k}) = 1$ by the Single Color constraint. By PSDness 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 2.4.4

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

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

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

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

$$\langle X_{i,i}, \mathbb{J} \rangle = \widetilde{\mathbb{E}}\, p_{(H,S,\tau)}(x, \boldsymbol{G}) + \widetilde{\mathbb{E}}\, p_{(H',S',\tau')}(x, \boldsymbol{G})$$

$$= \pi(i)^2 n^2 + \pi(i) n \pm \delta n^2 = \pi(i)\pi(j) n^2 + \delta' n^2$$

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

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

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

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

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

Proof. Our block-wise change of basis commutes with taking inner products between the blocks $X_{i,j}$ and the non-backtracking walk matrices. In other words, invoking Lemma 2.5.6 with $\delta' > \delta$ and keeping track of how the additive errors compound as we take linear combinations,

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

$$= \left(F^T \mathrm{Diag}(\pi) F \Lambda^s\right)_{i,j} \cdot \|q_s\|_{\mathrm{KM}}^2 n \pm \|F\|^2 \sqrt{k}\delta' n$$
$$= \Lambda^s_{i,j} \cdot \|q_s\|_{\mathrm{KM}}^2 n \pm \|F\|^2 \sqrt{k}\delta' n.$$

A parallel calculation gives us

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

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 2.5.7 in hand, the remainder of the proof follows from Proposition 2.4.6 and Proposition 2.4.7 in the previous section. In particular, each block $\check{X}_{i,i}$ for $i = 2, \ldots k$ is a feasible solution to $SPS(m, \lambda_i)$ SDP with error tolerance $\delta'' > \|F\|^2 \sqrt{k}\delta$. We showed already that when $\lambda^2(d-1) > 1$, and for small enough error tolerance and large enough m, this SDP is w.h.p. infeasible on input $\boldsymbol{G} \sim \mathcal{N}$. Thus we need simply to make δ small enough so that δ'' is below the minimum tolerance in Proposition 2.4.6.

2.5.3 Spectral Distinguishing

Our argument in the previous section can be recast to prove Corollary 2.2.5, namely that above the Kesten-Stigum threshold the spectrum of the adjacency matrix can also be used to distinguish the null and planted distributions.

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

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

Think of $\check{\boldsymbol{X}}$ as a block matrix $(\boldsymbol{X}_{i,j})_{i,j \in [k]}$, as we did X in the previous section, and $\check{\boldsymbol{x}}$ as a block vector $(\check{\boldsymbol{x}}_i)_{i \in [k]}$. Applying Theorem 2.5.3 and repeating the calculations in Lemma 2.5.6 and Lemma 2.5.7 *mutatis mutandis* with $\boldsymbol{X}$ instead of X, we can show that w.h.p.

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

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

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

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

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

when $i \neq j$. In other words, the k vectors $\check{\boldsymbol{x}}_1, ..., \check{\boldsymbol{x}}_k$ 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 2.4.7 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 2.5.8. 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 2.5.5 that the Local Statistics SDP is robust to this kind of perturbation.

2.5.4 Lower Bounds

In this section, we prove the second half of Theorem 2.2.4, which gives a complementary lower bound: if every one of $\lambda_2, ..., \lambda_k$ has modulus at most $1/\sqrt{d-1}$ there exists some feasible solution to the Local Path Statistics SDP for every $m \geqslant 1$.

We can specify a pseudoexpectation completely by way of an $(nk+1) \times (nk+1)$ positive semidefinite matrix

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

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

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

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

Lemma 2.5.9. *The set of $\mathcal{B}_k$-satisfying pseudoexpectations is parameterized by pairs $(X, l) \in \mathbb{R}^{nk \times nk} \times \mathbb{R}^{nk}$ for which*

$$\begin{pmatrix} 1 & l^T \\ l & X \end{pmatrix} \succeq 0 \tag{2.14}$$

$$\operatorname{diag}(X) = l \tag{2.15}$$

$$\mathsf{Tr} X_{u,u} = e^T l = 1 \qquad \forall u \in [n] \tag{2.16}$$

$$X_{u,v} e = l_u \qquad \forall u, v \in [n] \tag{2.17}$$

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

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

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

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

Writing $X = \widetilde{\mathbb{E}}\, x^T x$ and $l = \widetilde{\mathbb{E}}\, x$ as above, the first constraint is equivalent to $l = \mathrm{diag}(X)$, 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 2.5.10. *Let $G \sim \mathcal{N}$, and let the pair $(X, l) \in \mathbb{R}^{nk\times nk} \times \mathbb{R}^{nk}$ satisfies (2.14)-(2.17) and*

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

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

We will defer the proof of Proposition 2.5.10 to Appendix 2.8. Its conclusion in hand, we can now set about constructing a pseudoexpectation. Since $(d-1)\lambda_2^2 < 1$, the $SPS(\lambda_i, m)$ SDP is feasible for every error tolerance $\delta' > 0$. Thus for each $i = 2, ..., k$ there exists a feasible solution in the form of a PSD matrix $Y(\lambda_i)$ satisfying

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

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

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

i.e. $\check{X}_{i,j} = 0$ when $i \neq j$, and the diagonal blocks are as above, and similarly let $\check{l} = (e, 0, ..., 0)^T$. We claim that the pair $X = (F^{-T} \otimes \mathbb{1})\check{X}(F^{-1} \otimes \mathbb{1})$ and $l = (F^{-1} \otimes \mathbb{1})\check{l}$ satisfies the conditions of Lemma 2.5.9 and Proposition 2.5.10.

First

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

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

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

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

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

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

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

It remains to verify the affine conditions in Proposition 2.5.10. As in the proof of Lemma 2.5.7, since each $Y(\lambda_i)$ is a feasible solution to the $SPS(\lambda_i, m)$ SDP with error tolerance δ',

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

$$= \pi(i)T^s_{i,j} \cdot \|q_s\|^2_{\mathrm{KM}} n \pm \|F^{-1}\|^2 \sqrt{k}\delta' n$$

and

$$\begin{aligned}
\langle X_{i,j}, \mathbb{J} \rangle &= \left(F^{-T} \begin{pmatrix} \langle \mathbb{J}, \mathbb{J} \rangle & & & \\ & \langle Y(\lambda_2), \mathbb{J} \rangle & & \\ & & \ddots & \\ & & & \langle Y(\lambda_k), \mathbb{J} \rangle \end{pmatrix} F^{-1} \right)_{i,j} \\
&= \left(F^{-T} e_1 e_1^T F^{-1} \right)_{i,j} \cdot n^2 \pm \|F^{-1}\|^2 \sqrt{k}\delta' n^2 \\
&= \left(\pi\pi^T \right)_{i,j} \cdot n^2 \pm \|F^{-1}\|^2 \sqrt{k}\delta' n^2 \\
&= \pi(i)\pi(j) \cdot n^2 \pm \|F^{-1}\|^2 \sqrt{k}\delta' n^2,
\end{aligned}$$

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

2.5.5 Robustness

The proof of robustness largely reduces to the discussion in Section 2.4.4. Recall that we need to produce a $\rho > 0$ for which (i) when $\boldsymbol{G} \sim \mathcal{P}$, or $\boldsymbol{G} \sim \mathcal{N}$ with $(d-1)\lambda_2^2 < 1$, the SDP with high probability remains feasible for any error tolerance δ, even after perturbing ρn edges, and (ii) that when $\boldsymbol{G} \sim \mathcal{N}$ and $(d-1) > \lambda_2^2$, if the SDP is infeasible at tolerance δ, it remains so at some tolerance $\delta' < \delta$ even after perturbing ρn edges.

For (i), assume that the SDP is feasible at error tolerance δ on input $\boldsymbol{G}$. We build the SDP as a linear combination of solutions $Y(\lambda_i)$ to the $SPS(m, \lambda_i)$ SDP, which we argued in Section 2.4.4 is robust in the desired sense. For (ii), when $\boldsymbol{G} \sim \mathcal{N}$ and $(d-1)\lambda_2^2 > 1$, we reduced infeasibility of the LoSt$(2, m)$ SDP to that of the $SPS(m, \lambda_2)$ SDP, which we showed already is infeasible. Moreover, from Section 2.4.4, the latter remains infeasible after a sufficiently small perturbation.

2.6 The Stochastic Block Model

We turn, finally, to the proof of Theorem 2.2.2 concerning the local statistics algorithm and Stochastic Block Model. For the sake of exposition, as we did for the

DRBM, we will first write down a simpler SDP that can robustly solve the detection problem above the KS threshold, and then show that feasibility of the full SDP implies feasibility of this simpler one.

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

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

$$\left(\overline{A}_G^{(s)}\right)_{u,v} = \sum_{\gamma : u \to v, |\gamma| = s} w_G(\gamma). \tag{2.18}$$

When $(\boldsymbol{G}, \boldsymbol{x}) \sim \mathcal{P}$, define as in Section 2.5.1 and Section 2.5.2 an $nk \times nk$ matrix $\boldsymbol{X} \triangleq \boldsymbol{x}\boldsymbol{x}^*$. As before, we will at times think of $\boldsymbol{X}$ as an $n \times n$ matrix of $k \times k$ blocks $\boldsymbol{X}_{u,v}$, and at others a $k \times k$ matrix of $n \times n$ blocks $\boldsymbol{X}_{i,j}$.

Claim 2.6.1. Let $\overline{T} = T - e^T \pi$. Then

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

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

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

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

There are $n(n-1)(n-2)\cdots(n-s)$ length-s self avoiding walks in the complete graph, so already the total expectation of $w_{\boldsymbol{G}}(\gamma)$ among these walks accounts for the

quantity in the claim. On the other hand, those γ's which contain a cycle contribute negligiblly: there are $O(n^v)$ γ's with v vertices, and for each one $\mathbb{E}\, w_G(\gamma) = O(n^{-e})$, so the total contribution of γ's with $e \geqslant v$ is at best $O(1)$. □

This motivates the following SDP:

Definition 2.6.2. For each $m \geqslant 1$, the *level m path statistics SDP* with error tolerance $\delta > 0$ is the feasibility problem

$$\text{Find } X = (X_{i,j}) \succeq 0 \text{ s.t. } (X_{i,i})_{u,u} \leqslant 1 \qquad \forall u \in [n], i \in [k] \tag{2.19}$$

$$\mathsf{Tr}(X_{u,u}) = 1 \qquad \forall u \in [n] \tag{2.20}$$

$$\langle X_{i,j}, \overline{A}_G^{(s)} \rangle = \pi(i)\overline{T}^s_{i,j} \cdot d^s n \pm \delta n \qquad \forall i, j \in [k], s = 0, ..., m. \tag{2.21}$$

Theorem 2.6.3. *When $\lambda_2^2 d > 1$, there exists $m = O(1)$ and $\delta > 0$ for which the level m path statistics SDP can solve the detection problem. Conversely, when $\lambda_2^2 d < 1$, no such m and δ exist.*

Recall that Λ is a $k \times k$ diagonal matrix containing the eigenvalues of T, sorted in descending order of modulus from the upper left corner. Since e and π^* are right and left eigenvectors, respectively, of T, $\overline{T}$ commutes with T and satisfies $\overline{T}F = F\overline{\Lambda}$, where $\overline{\Lambda}$ is obtained from Λ by deleting the upper left entry (which in our setup is equal to 1). We will accordingly take the same change-of-basis approach as in the DRBM. For any feasible solution X to this SDP, we can form an analogous matrix $\check{X} \triangleq (F^T \otimes \mathbb{1})X(F \otimes \mathbb{1})$, with blocks $\check{X}_{i,j}$. Following Lemma 2.5.7, observe that

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

$$= \overline{\Lambda}^s_{i,j} \cdot d^s n \pm \|F\|^2 \sqrt{k}\delta n.$$

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

Our observations about the matrices $\check{X}_{i,i}$ from Section 2.5 carry over here—namely each of these is PSD with ones on the diagonal. Thus we have shown that if the level-m Path Statistics SDP is feasible, then for some constant C,

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

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

Theorem 2.6.4. *Let $\boldsymbol{G} \sim \mathcal{N}$. Then for any $\epsilon, C > 0$ there exists $m \in \mathbb{N}$ so that with high probability*

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

This, and the preceding dicsussion, prove one half of Theorem 2.6.3, namely that whenever $\lambda_2^2 d > 1$, there are some $m = O(1)$ and $\delta > 0$ for which the level m Path Statistics SDP is with high probability infeasible on inout $\boldsymbol{G} \sim$, but feasible on input $\boldsymbol{G} \sim \mathcal{P}$. We will prove the other half in Section 2.7. Theorem 2.2.2, the analogous statement to Theorem 2.6.3 for the full local statistics algorithm, follows from a final observation:

Observation 2.6.5. With high probability over $\boldsymbol{G} \sim \mathcal{N}$, if the level-$m$ Path Statistics SDP is infeasible at error tolerance δ, then the LoSt$(2, m)$ SDP at some error tolerance $\delta'(\delta)$ is infeasible as well.

Proof. The quadratic block of the LoSt$(2, m)$ SDP concerns $nk \times nk$ matrices, and includes all hard constraints—bounds on diagonal entries, trace of diagonal blocks—present in the Path Statistics SDP. Moreover, it has access to affine constraints involving the counts of subgraphs with at most m edges. Since the entries of $\overline{A}^{(s)}_G$ for $s \leqslant m$ are simply linear combinations of such counts, LoSt$(2, m)$ has access to the affine constraints from the Local Path Statistics SDP as well. □

The promised robustness guarantee in Theorem 2.2.2 can be achieved by choosing B as per Theorem 2.10.4, deleting edges incident to all vertices of degree $> B$, and inputting the resulting graph into the LoSt$(2, m)$ SDP.

2.6.1 Local Statistics in the SBM

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

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

Theorem 2.6.6. *Let (H, S, τ) be a partially labelled graph with $O(1)$ edges and ℓ connected components. Then with high probability*

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

Proof. Fix (H, S, τ). There are

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

injective maps from $V(H) \hookrightarrow [n]$. The probability that each is an occurrence, once we condition on the labels σ of the relevant vertices, is given by

$$\prod_{(u,v)\in E(H)} M_{\sigma(u),\sigma(v)} \cdot d/n.$$

The probability of each labelling σ is $\prod_{u \in V(H)} \pi(\sigma(u))$, and we only consider labellings that agree with τ at the relevant vertices. Thus

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

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

To prove concentration, in the case when H has at least one cycle, $c(H) > \chi(H)$ and the assertion follows from Markov. Otherwise, let us consider

$$\mathbb{E}\, p_{H,S,\tau,U,W}(\boldsymbol{G}, \boldsymbol{x})^2.$$

This is a sum over pairs of maps $\phi, \psi : V(H) \to [n]$, and as H is acyclic, it is dominated by terms in which the images of these two maps are disjoint. Thus, to leading order, this variance is equal to the expected number of occurrences of two disjoint copies of (H, S, τ), which we just observed is $(\mathbb{E}\, p_{H,S,\tau,U,W}(\boldsymbol{G}, \boldsymbol{x}))^2$ to leading order. We finish by using Chebyshev and observing that $\chi(H) = c(H)$. $\square$

2.6.2 Proof of Theorem 2.6.4

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

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

We will also refer to $\boldsymbol{H}$ as the (t, r, ε)-*truncation of* G. In the sequel, we assume $r = \Theta\left(\frac{\log n}{(\log\log n)^2}\right)$. Then Theorem 2.6.4 is an immediate consequence of the following two results.

Theorem 2.6.7 (Truncated Spectral Norm Bound)**.** *For every $\epsilon > 0$, there exist t, m satisfying $m = t^3$ so that with high probability*

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

Proposition 2.6.8 (L_1 Bound)**.** *For every $\delta > 0$, and every $r = O(1)$, for any $t \geqslant \Omega(\frac{\log m - \log \delta}{\log(1+\epsilon)})$ so that with high probability*

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

With these two results in hand, Theorem 2.6.4 quickly follows: any matrix $Y \succeq 0$ with unit diagonal satisfies $|Y_{u,v}| \leqslant 1$, so

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

Theorem 2.6.7 is the heavier technical lift, so we will warm up with the proof of Proposition 2.6.8. The proof of Theorem 2.6.7 is deferred to Section 2.6.3.

2.6.2.1 Proof of Proposition 2.6.8

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

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

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

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

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

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

We will need the following simplified version of the forthcoming Lemma 2.9.8.

Lemma 2.6.9. *Let $\Gamma \subset K_n$, and write $\Gamma \subset \boldsymbol{G}$ to mean that every edge of Γ appears in $\boldsymbol{G}$. Then there exist universal C, c so that*

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

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

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

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

Thus we have shown that the event we care about is contained in the intersection of two independent ones: that $\Gamma \subset \boldsymbol{G}$, and that there exists a vertex in Γ that is (t,ϵ')-heavy in $\boldsymbol{G}^c$. We can bound this second probability by taking a union bound over all vertices in Γ, and noting that the probability of being heavy in $\boldsymbol{G}^c$ is at most the probability of being heavy in $\boldsymbol{G}$. From Lemma 2.9.8, the probability that a given vertex is (t,ϵ')-heavy in $\boldsymbol{G}$ is, for some universal C, c, at most $C\exp(-c(1+\epsilon')^t) \leqslant C\exp(-\frac{c}{|V(\Gamma)|+1}(1+\epsilon)^t)$. We then execute the union bound and assemble everything. □

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

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

Thus

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

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

Combining all of this,

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

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

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

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

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

Lemma 2.6.10. *Let $(Y, X_1, X_2, ..., X_T)$ be i.i.d. real random variables. Then for any function $f : \mathbb{R}^T \to \mathbb{R}$,*

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

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

Fact 2.6.11. *Let $\boldsymbol{G} \sim \mathcal{N}$. Then the probability that there is a vertex with degree larger than Δd is at most $n(e/\Delta)^{d\Delta}$. In particular, setting $\Delta = 2\log n$, this probability is $o(n^{-c})$ for every c.*

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

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

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

All told, then, we have that with high probability

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

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

2.6.3 Spectral norm bounds

In this section, we prove Theorem 2.6.7.

2.6.3.1 Setup

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

1. the hypothesis of Theorem 2.9.1 holds on parameters $d' := (1+\varepsilon)d, d$ and δ,
2. $((1+\varepsilon)d)^{1/t^2} t^{30/t^3} < 1+\varepsilon$,
3. $\delta^{-24/t} < 1+\varepsilon$,

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

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

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

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

Theorem 2.6.12. *With probability* $1 - n^{-100}$,

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

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

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

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

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

2.6.3.2 From High Trace to Counting

We borrow some more terminology from [MOP20]:

Definition 2.6.13 (Linkages)**.** We call a closed walk of length $k\ell$ on K_n a $(k \times \ell)$-*linkage* if it can be split into k segments each of length-ℓ such that W is nonbacktracking on each segment. We refer to each such length-ℓ nonbacktracking segment as a *link*. We use $V(W)$ to denote the vertices visited by W and $E(W)$ to denote the (undirected) edges visited by W.

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

$$\begin{aligned} \boldsymbol{F} &= \sum_{W \text{ is } (k\times\ell)\text{-linkage of } K_n} \prod_{ij \in W} \left(\boldsymbol{A}[i,j] - \frac{d}{n} \right) \cdot \mathbf{1}[\mathcal{E}(W)] \\ &= \sum_{W \text{ is } (k\times\ell)\text{-linkage of } K_n} \prod_{ij \in S(W)} \left(\boldsymbol{A}[i,j] - \frac{d}{n} \right) \prod_{ij \in D(W)} \left(\boldsymbol{A}[i,j] - \frac{d}{n} \right)^{a_{ij}(W)} \mathbf{1}[\mathcal{E}(W)] \end{aligned} \tag{2.22}$$

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

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

$$= \boldsymbol{A}[i,j]\left(\left(1-\frac{d}{n}\right)^{a_{ij}(W)} - \left(-\frac{d}{n}\right)^{a_{ij}(W)}\right) + \left(-\frac{d}{n}\right)^{a_{ij}(W)}.$$

Writing $\gamma_{ij} = \left(1-\frac{d}{n}\right)^{a_{ij}(W)} - \left(-\frac{d}{n}\right)^{a_{ij}(W)}$, we can rewrite (2.22) as

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

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

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

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

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

By Theorem 2.9.1:

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

$$
\begin{aligned}
& n^{.8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt} \\
\leqslant C(n) \sum_W & \left(\sum_{\substack{L \subseteq D(W) \\ L \text{ good}}} \prod_{ij \notin L} \left(\frac{d}{n}\right)^{a_{ij}(W)-1} \right) \cdot \\
& \left(\frac{d}{n}\right)^{|S(W)|+|D(W)|} \cdot n^{.8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt}
\end{aligned}
\tag{2.23}
$$

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

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

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

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

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

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

Plugging the above back into (2.23) and "absorbing" a factor of 2 into $C(n)$ tells us:

$$\text{(2.23)} \leqslant C(n) \sum_W \left(\frac{2d|D(W)|}{n}\right)^{\Delta(W)} \cdot \left(\frac{d}{n}\right)^{|S(W)|+|D(W)|} \cdot n^{\cdot 8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt}. \tag{2.24}$$

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

Definition 2.6.14. We say a step from u to v in a linkage W is *fresh* if v was never visited earlier in W. We will use $f(W)$ to denote the number of fresh steps in W.

Remark 2.6.15. For a linkage W, $|V(W)| = f(W) + 1$.

A consequence of Remark 2.6.15 along with the fact that $|S(W)| + |D(W)| = |E(W)|$, we get that $|E(W)| = f(W) + e(W)$. Thus, (2.24) is bounded by

$$C(n) \sum_W \left(\frac{2d|D(W)|}{n}\right)^{\Delta(W)} \cdot \left(\frac{d}{n}\right)^{e(W)+f(W)} \cdot n^{\cdot 8e(W)} \cdot 4^{|S(W)|} \cdot \delta^{|S(W)|-24kt}. \tag{2.25}$$

To complete the proof, we need the following, which is proved in Section 2.6.4:

Theorem 2.6.16. *The total number of (k,ℓ)-linkages with f fresh edges, e excess edges, s singleton edges and Δ profligate steps is at most:*

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

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

Now recall that we wished to obtain bounds on the following from (2.25):

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

From Theorem 2.6.16:

$$Q \leqslant C(n) \sum_{f,e,\Delta,s\geqslant 0} \left(\frac{2d|D(W)|}{n}\right)^{\Delta} \cdot \left(\frac{d}{n}\right)^{f} \cdot \left(\frac{d}{n^{\cdot 2}}\right)^{e} \cdot 4^{s} \cdot \delta^{s-24kt}.$$

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

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

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

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

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

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

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

We know that $\boldsymbol{F}$ is a nonnegative random variable since it is the trace of an even power of a Hermitian matrix, i.e., it is the trace of a positive semidefinite matrix.

Hence, by Markov's inequality, we know that except with probability n^{-100}, the random variable $\boldsymbol{F}$ defined in (2.22) is bounded by

$$n^{100} \cdot 8C'(n) \cdot ((1+\varepsilon)d)^{k\ell/2} \cdot ((\ell+1)^6((1+\varepsilon)d)^t)^k \cdot \delta^{-24kt} \cdot \mathrm{poly}(k,\ell)^{84\ln^6(k\ell)} \cdot (k\ell) \tag{2.26}$$

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

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

In particular, this means the k-th root of (2.26) is bounded by $\left((1+\varepsilon)^4\sqrt{d}\right)^\ell$ with probability $1 - n^{-100}$.

In summary, we have shown that whp:

$$\mathsf{Tr}\left(\left(\left(\boldsymbol{A} - \frac{d}{n}1_{[n]\setminus S}1_{[n]\setminus S}^\mathsf{T}\right)^{(\ell)}\right)^k\right)^{1/k} \leqslant \left((1+\varepsilon)^4\sqrt{d}\right)^\ell \tag{2.27}$$

thereby establishing:

Theorem 2.6.17 (Restatement of Theorem 2.6.12)**.** *With probability* $1 - n^{-100}$*:*

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

2.6.4 Counting Walks

This section is dedicated to proving Theorem 2.6.16. Let W be a $(k \times \ell)$-linkage with f fresh steps, e excess edges, s singleton edges, Δ profligate steps. In this section, we will give an efficient encoding of W which will help us upper bound the number of such linkages.

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

Remark 2.6.18. For any set of edges E', we will use $V(E')$ to denote the set of endpoints of edges in E'.

Remark 2.6.19. $|D(W)| = |D^*(W)| + \Delta$ and $|E(W)| = |D^*(W)| + \Delta + s$.

Remark 2.6.20. For any subset of edges $E' \subseteq E(W)$, the total number of times W walks on an edge in E' is given by

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

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

Our next step is to prove:

Claim 2.6.21. There is a spanning forest F of $D^*(W) \cup S(W)$ such that:

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

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

Fact 2.6.22. *Let P be the polytope in $\mathbb{R}^{D^*(W)}$ given by the convex hull of indicator vectors of spanning forests of $D^*(W)$. P is also the feasible region of the following linear program:*

$$\begin{aligned} & x \in \mathbb{R}^{D^*(W)} \\ & x \geqslant 0 \\ \sum_{ij \in R} x_{ij} & \leqslant |V(R)| - 1 \qquad \forall R \subseteq D^*(W). \end{aligned} \tag{2.28}$$

We additionally also state the following from [MOP19, Corollary 2.18] which is a consequence of the "irregular Moore bound" of [AHL02]:

Fact 2.6.23. *Let H be a graph with $v \geqslant 3$ vertices and girth $g \geqslant 20 \ln v$. Then $|E(H)| - v \leqslant \frac{2v \ln v}{g}$.*

Proof of Claim 2.6.21. Consider the following assignment to the variables of LP (2.28):

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

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

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

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

Case 2: $|V(R)| \geqslant r$**:** Since the girth of $(V(R), R)$ is at least r, by Fact 2.6.23,

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

If we augment the linear program (2.28) with the objective function

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

by Fact 2.6.22 we know that there is a spanning forest $\widetilde{F}$ such that maximum of the above objective is achieved at the indicator vector of $\widetilde{F}$. Since we showed $\widetilde{x}$ is feasible,

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

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

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

Now we extend $\widetilde{F}$ to a spanning forest F of $D^*(W) \cup S(W)$. Initially, we set $F = \widetilde{F}$ and process edges of $S(W)$ sequentially in an arbitrary order $i_1j_1, \ldots, i_sj_s$. We add i_tj_t to F if its addition does not create a cycle and "reject" it otherwise. The number of rejected edges is bounded by e and hence F must contain at least $s - e$ edges of S. Thus, $|S(W) \setminus F| \leqslant e$. Furthermore, since $D^*(W) \setminus F = D^*(W) \setminus \widetilde{F}$,

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

□

Claim 2.6.21 lends itself to a natural decomposition of the edges of $G(W)$ into *forest edges*, which we denote $F(W)$, and *crossing edges*, which we denote $C(W)$.

Remark 2.6.24. $C(W)$ can be written as the following natural disjoint union of sets:

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

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

Let's now express the linkage W in terms of the sequence of vertices walked on: in particular $W = w_0w_1w_2 \ldots w_{k\ell}$.

Definition 2.6.25. We call each consecutive pair w_iw_{i+1} a *step*. If the edge $\{w_i, w_{i+1}\}$ is a crossing edge, we call the step w_iw_{i+1} a *crossing step*, and a *forest step* otherwise. We call a maximal contiguous sequence of forest steps a *cruise*.

Remark 2.6.26. Any $(k \times \ell)$-linkage W can be expressed as

$$W = C_1s_1C_2s_2 \ldots C_{\gamma(W)}s_{\gamma(W)}C_{\gamma(W)+1}$$

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

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

Claim 2.6.27. $\gamma(W) \leqslant e + \frac{4k\ell \ln(k\ell)}{r} + \Delta$.

Proof. By Remark 2.6.24:

$$\begin{aligned}
\gamma(W) &= \sum_{ij \in S(W) \setminus F(W)} a_{ij} + \sum_{ij \in D^*(W) \setminus F(W)} a_{ij} + \sum_{ij \in D(W) \setminus D^*(W)} a_{ij} \\
&\leqslant \sum_{ij \in S(W) \setminus F(W)} 1 + \frac{4k\ell \ln(k\ell)}{r} + \Delta && \text{(by Claim 2.6.21)} \\
&\leqslant e + \frac{4k\ell \ln(k\ell)}{r} + \Delta. && \text{(by Claim 2.6.21)}
\end{aligned}$$

□

Definition 2.6.28. We refer to endpoints of edges in $C(W)$ as well as the start/end vertex of W[1] as *terminal vertices*. We use $T(W)$ to refer to the set of terminal vertices of $G(W)$.

Remark 2.6.29. $|T(W)| \leqslant 2|C(W)| + 1 \leqslant 2\gamma(W) + 1$. We will use $\lambda(W)$ to refer to $2\gamma(W) + 1$.

Remark 2.6.30. Each cruise starts and ends at terminal vertices.

Definition 2.6.31 (Skeleton forest)**.** We use $\mathsf{Skel}(F(W))$ to refer to the subforest of $F(W)$ given by the union of paths in $F(W)$ connecting terminal vertices. Formally,

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

Observation 2.6.32. Every leaf in $\mathsf{Skel}(F(W))$ is a terminal vertex and hence by Remark 2.6.29 the number of leaves in $\mathsf{Skel}(F(W))$ is at most $\lambda(W)$.

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

Lemma 2.6.33. *Let $\Gamma_{L,v}$ be the set of forests with at most L leaves on vertex set $\{1, \ldots, v\}$. There is a subset $Q \subseteq \Gamma_{L,v}$ of at most $(4Lv)^{2L+1}$ forests such that any forest in $\Gamma_{L,v}$ is isomorphic to a forest in Q.*

[1] which are the same since W is closed

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

Fact 2.6.34. *The number of labeled spanning trees on v vertices is v^{v-2}.*

We will also need the following fact about trees:

Fact 2.6.35. *Let T be a tree where v_i is the number of vertices of degree i. Then the following are true:*

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

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

$$\begin{aligned} 2\sum_{i \geqslant 1} v_i - 2 &= \sum_{i \geqslant 1} i v_i \\ v_1 + \sum_{i \geqslant 3} v_i - 2 &= \sum_{i \geqslant 3} i v_i \\ v_1 - 2 &= \sum_{i \geqslant 3} (i-2) v_i \end{aligned} \tag{2.29}$$

Lower bounding $i - 2$ by 1 in the RHS of (2.29), it follows that:

$$v_1 - 2 \geqslant \sum_{i \geqslant 3} v_i \tag{2.30}$$

Adding 2·(2.30) and (2.29) gives us:

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

□

Proof of Lemma 2.6.33. Let Ξ be any tree in $\Gamma_{L,v}$. If we split $V(\Xi)$ into leaves $V_1(\Xi)$, degree-2 vertices $V_2(\Xi)$, and degree-$\geqslant 3$ vertices $V_{\geqslant 3}(\Xi)$, we have the following from Fact 2.6.35:

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

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

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

Observe that $\widetilde{\Xi}$ has $\widetilde{v} \leqslant 2L$ vertices and the weight of an edge is an integer between 1 and $|V(\Xi)|$. By Cayley's formula (Fact 2.6.34) the number of labeled spanning trees on $\widetilde{v}$ vertices is at most $(\widetilde{v})^{\widetilde{v}-2}$. Consequently the number of spanning forests on $\widetilde{v}$ vertices is at most $(2\widetilde{v})^{\widetilde{v}}$ (since every spanning tree on $\widetilde{v}$ vertices has $2^{\widetilde{v}-1}$ subforests). Since $\widetilde{v} \leqslant 2L$ and there are at most $2L$ possibilities for $\widetilde{v}$, each labeled spanning forest on vertex set $[\widetilde{v}]$ that can be encoded by a number in $\left[(4L)^{2L+1}\right]$. In particular this gives us a way to encode the edge set of any $\widetilde{\Xi}$ by a number in $\left[(4L)^{2L+1}\right]$.

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

Lemma 2.6.36. $\mathsf{Skel}(F(W))$ *can be encoded by a number in*

$$\left[(4\lambda(W)k\ell)^{2\lambda(W)+1} \cdot n^{|V(\mathsf{Skel}(F(W)))|}\right].$$

Proof. At a high level, our proof uses Lemma 2.6.33 to encode an unlabeled version of $\mathsf{Skel}(F(W))$ in $\left[(4\lambda(W)k\ell)^{2\lambda(W)+1}\right]$ bits and encodes labels using a number in $\left[n^{|V(\mathsf{Skel}(F(W)))|}\right]$.

Encoding "unlabeled" version of $\mathsf{Skel}(F(W))$**.** Let π be an arbitrary function that maps $V(F(W))$ to $\{1, \ldots, |V(F(W))|\}$. Note that the graph $\pi(\mathsf{Skel}(F(W)))$ is isomorphic to $\mathsf{Skel}(F(W))$. By Observation 2.6.32, Lemma 2.6.33, and bounding $|V(\mathsf{Skel}(F(W)))|$ by $k\ell$, $\pi(\mathsf{Skel}(F(W)))$ can be encoded (up to isomorphism) by a number in $\left[(4\lambda(W)|V(\mathsf{Skel}(F(W)))|)^{2\lambda(W)+1}\right]$.

Encoding labels of $\mathsf{Skel}(F(W))$**.** From the encoding of $\pi(\mathsf{Skel}(F(W)))$, we can recover a graph on vertex set $\{1, \ldots, |V(\mathsf{Skel}(F))|\}$ isomorphic to $\mathsf{Skel}(F(W))$, which we call $\phi(\mathsf{Skel}(F(W)))$. We thus encode the map ϕ^{-1} as it is possible to reconstruct

$\mathsf{Skel}(F(W))$ from $\phi(\mathsf{Skel}(F(W)))$ and ϕ^{-1}; such a map can be encoded using a number in $\left[n^{|V(\mathsf{Skel}(F(W)))|}\right]$.

Combining the above two encodings proves the lemma. □

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

Lemma 2.6.37. *Given the encoding of* $\mathsf{Skel}(F(W))$ *from Lemma 2.6.36, the collection of start and end points of each cruise*

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

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

Proof. Let ϕ be the function from the proof of Lemma 2.6.36. The sequence

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

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

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

Remark 2.6.38. There is a unique path between t_{start} and t_{end} in F as follows:

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

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

Definition 2.6.39. Let C_i be a cruise. We say a contiguous subwalk of C_i is a *detour* if it starts and ends at the same vertex.

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

Claim 2.6.40. Cruise C_i can be constructed by taking the path from t_{start} to t_{end} as described in Remark 2.6.38 and inserting at most one detour after each vertex in the path. In particular, C_i can be written in the form

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

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

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

> If every vertex is visited once, the path from Remark 2.6.38 is the cruise. If there exists a vertex that occurs more than once, find the first such visited vertex v_{j_1}, and define Detour_{i,j_1} as the subwalk of C_i between the first and last occurrence of v_{j_1}; now repeat this procedure on the walk starting at the last occurrence of v_{j_1} and ending at the end of the cruise.

□

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

Definition 2.6.41. Given a detour Detour in W, we say the *timestamp* of Detour is the tuple (a, b) where a is the position of the start step of Detour in W and b is the position of the end step of Detour in W.

Lemma 2.6.42. *There is an encoding of the timestamps of all detours in W in $\left[(\ell+1)^{2k}\right]$.*

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

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

Claim 2.6.43. All the edges visited by any detour Detour are in $D^*(W) \cap F(W)$.

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

Corollary 2.6.44. *For any detour* Detour*, the graph* $G(\mathsf{Detour})$ *has no* (t, r, ε)*-vexing vertices.*

Observation 2.6.45. Any detour Detour can be decomposed into a sequence of links of length exactly ℓ, with the exception of the first and last link, which can both have any length between 1 and ℓ.

Definition 2.6.46. Any detour Detour starts and ends at some vertex v. We call v the *root* of Detour and denote it with $\mathsf{Root}(\mathsf{Detour})$.

Remark 2.6.47. One should think of a detour as a closed walk on a tree rooted at a distinguished vertex.

Definition 2.6.48. We call a step from u to v in Detour an *up-step* if v is closer to $\mathsf{Root}(\mathsf{Detour})$ than u. In similar spirit, we call that step a *down-step* if v is further from $\mathsf{Root}(\mathsf{Detour})$ than u.

Definition 2.6.49. We further classify down-steps in a detour Detour into three types:

1. We call a down-step from u to v a *fresh skeleton step* if the edge $\{u, v\}$ is part of $\mathsf{Skel}(F(W))$ and has not been traversed by any detour so far.

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

3. We call a down-step from u to v a *stale step* if it is not a fresh skeleton step or a fresh intrepid step.

Claim 2.6.50. Suppose there is a stale step from u to v at time T. Then there is an occurrence of a step from u to v as well as from v to u in a detour at an earlier time.

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

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

Definition 2.6.51. We call a (possibly empty) contiguous sequence of steps a *stretch*.

Observation 2.6.52. Due to nonbacktracking nature of links and the tree structure of detours, every link in a detour can be broken into 4 phases:

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

Lemma 2.6.53. *Given the encoding of* $\mathsf{Skel}(F(W))$ *from Lemma 2.6.36, the encoding of endpoints of cruises from Lemma 2.6.37, and the encoding of timestamps of detours from Lemma 2.6.42, it is possible to encode all detours in* W *using a number in*

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

Proof. Let $\mathsf{Detour}_1, \ldots, \mathsf{Detour}_b$ be the sequence of detours of W in order of time. We first specify how we encode detours, and then prove that the encoding is valid, i.e., recovery of all Detour_a from the given encoding is possible. As pointed out in Observation 2.6.45 each Detour_a can be broken into a sequence of links $L_1, \ldots, L_\tau$. We now describe how to encode each L_j.

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

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

Encoding a stale stretch. Given a stale stretch ζ, let E_ζ denote the set of edges visited before ζ starts. From Claim 2.6.43 ζ is completely contained in $D^*(W)$. Since ζ is a stale stretch, it must be contained in $E_\zeta \cap D^*(W)$. We first break ζ into $\left\lceil \frac{|\zeta|}{t} \right\rceil$ substretches $\zeta_1, \ldots, \zeta_{\left\lceil \frac{|\zeta|}{t} \right\rceil}$ each of length at most t and encode each substretch. Let v_i be the vertex at the start of ζ_i and v_i' be the end of ζ_i. Since $E_\zeta \cap D^*$ has no (t, r, ε)-vexing vertices, there are at most $((1+\varepsilon)d)^t$ vertices within distance t of v_i; in particular, there are at most $((1+\varepsilon)d)^t$ possible candidates for v_i'. We sort these candidates in increasing order of time first visited in a detour, and encode ζ_i with the index of v_i' in this list of candidates. Note that this index is a number in $\left[((1+\varepsilon)d)^t\right]$. To encode ζ, we specify $\left\lceil \frac{|\zeta|}{t} \right\rceil$ such numbers, one corresponding to each ζ_i.

Encoding a fresh skeleton stretch. For each step $u \to v$ of the fresh skeleton stretch, we don't specify any information if the degree of u within $\mathsf{Skel}(F(W))$ is $\leqslant 2$ and u is not a terminal. If the degree of u is at least 3 or if u is a terminal, we create a list of neighbors of u sorted in increasing order of their identities in K_n, and specify the index of v in this list. Note that this index is at most the degree of u within $\mathsf{Skel}(F(W))$, which from Fact 2.6.35 is at most $3 \times (\#$ leaves in $\mathsf{Skel}(F(W)))$, which in turn from Observation 2.6.32 is bounded by $3\lambda(W)$.

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

Recovery of detours. We now show how to recover the detours from the given encodings. First, it is possible to recover the root of every detour from the encodings given by Lemma 2.6.36, Lemma 2.6.37 and Lemma 2.6.42. We now show how to

recover the detours in order

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

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

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

Recovering the stale stretch in L_{j+1}. By Claim 2.6.50 every step in the stale stretch of L_{j+1} has been taken in a detour before. Since we know the the length of the stale stretch in L_{j+1} from the metadata encoding, and we have recovered all steps before the stale stretch in L_{j+1} that are part of a detour, we can infer a list of candidate endpoints of the stale stretch. Further, we also know the order in which these candidates were visited in detours, and hence we can recover the stale stretch in L_{j+1} from the encoding of stale stretches we described.

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

- Let x be a counter that is initially 0.
- Let v' be initially set to v (v' denotes the "current vertex" in our walk).
- While $x \leqslant \mathsf{LenSkel}$:
  - If the degree of v' within $\mathsf{Skel}(F(W))$ is $\leqslant 2$ and v' is not a terminal, then step along the unique unvisited edge incident to v' (called $v'w$) and update v' to w. *Note that if the first x steps of this walk and those in the fresh skeleton stretch coincide, then $v'w$ must be the $(x+1)$-th step in the fresh skeleton stretch.*

- If the degree of v' within $\mathsf{Skel}(F(W))$ is $\geqslant 3$ or v' is a terminal vertex: then assuming the first x steps of the current walk match those of the fresh skeleton stretch, we can recover the next step $v'w$ of the fresh skeleton stretch from the encoding of $\mathsf{Skel}(F(W))$ in Lemma 2.6.36 combined the encoding of fresh skeleton stretches described earlier in this proof. Thus, we update v' to w.
- Increment x by 1.

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

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

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

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

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

$$\prod_{L \in \mathrm{Links}(W)} \left(((1+\varepsilon)d)^t\right)^{\left\lceil \frac{|\zeta(L)|}{t} \right\rceil} \leqslant \left(((1+\varepsilon)d)^t\right)^{\sum_{L \in \mathrm{Links}(W)} \left(\frac{|\zeta(L)|}{t}+1\right)}. \tag{2.31}$$

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

$$\sum_{L\in\text{Links}(W)}\left(\frac{|\zeta(L)|}{t}+1\right)=k+\frac{1}{t}\sum_{L\in\text{Links}(W)}|\zeta(L)| \tag{2.32}$$

Note that $\sum_{L\in\text{Links}(W)}|\zeta(L)|$ is the total number of stale steps across all detours. From Claim 2.6.50 the (undirected) edge that a stale step is taken on is being traversed for *at least* the third time. Further, since the stale step is a down-step, there must be a corresponding up-step that is the reversal of the down-step in the detour. Thus, an edge $\{i,j\}$ is traversed by a stale step at most $\frac{a_{ij}-2}{2}$ times. Further, since there are multiple steps that traverse the same edge that a given stale step traverses, every stale step must traverse an edge in $D(W)$. Thus, we can bound (2.32) by:

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

Plugging in the above into (2.31) gives us a bound of:

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

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

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

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

$$\prod_{v\in P}(\deg_{\mathsf{Skel}(F(V))}(v)+1)^{\deg_{\mathsf{Skel}(F(V))}(v)}\leqslant(3\lambda(W)+1)^{\sum_{v\in P}\deg_{\mathsf{Skel}(F(V))}(v)} \tag{2.33}$$

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

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

From Fact 2.6.35 the first term is bounded by $3 \times$ # leaves in $\mathsf{Skel}(F(W))$, which from Observation 2.6.32 is bounded by $3\lambda(W)$. The second term is bounded by $2|T(W)|$, which from Remark 2.6.29 is at most $2\lambda(W)$. As an upshot we have:

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

Plugging this into (2.33) gives us a bound on the number of possible skeleton fresh stretch encodings of:

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

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

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

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

□

Since it is possible to recover a linkage W from $\mathsf{Skel}(W)$, the endpoints of its cruises and the order in which the cruises occur, the timestamps of the detours, and the detours, by a combination of Lemma 2.6.36, Lemma 2.6.37, Lemma 2.6.42 and Lemma 2.6.53 along with a bound on $\lambda(W)$ from Claim 2.6.27 we have the following bound:

Theorem 2.6.54 (Restatement of Theorem 2.6.16)**.** *The total number of (k,ℓ)-linkages with f fresh edges, e excess edges, s singleton edges and Δ profligate steps is at most:*

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

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

2.7 Lower Bounds in the Stochastic Block Model

In this section, we finish the proof of Theorem 2.6.3 by proving lower bounds for the level-M path statistics SDP (as described by Definition 2.6.2) for every constant M for detection in the stochastic block model under the Kesten-Stigum threshold.

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

2.7.1 Weighted Ihara-Bass and a Power Series Identity

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

Theorem 2.7.1 (Weighted Ihara-Bass)**.** *For any weights $c : E \to \mathbb{R}$, let $\hat{c} = c(1-c^2)^{-1}$. Then*

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

and

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

whenever this series converges.

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

Now consider the matrix

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

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

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

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

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

making

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

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

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

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

Considering the upper left block, we see

$$\begin{aligned} (1 - A_{\hat{c}} + D_{c\hat{c}})^{-1} &= 1 + S(1 - CB)^{-1}CT \\ &= 1 + \sum_{\ell \geqslant 0} S(CB)^{\ell}CT \\ &= \sum_{\ell \geqslant 0} A_c^{(\ell)} \end{aligned}$$

□

2.7.2 Construction of SDP solution

Let $\boldsymbol{G}$ be a $G(n, d/n)$ graph. Our goal is to construct a solution to the SDP given in Definition 2.6.2 when $\boldsymbol{G} \sim G(n, d/n)$, and d is under the KS threshold. We instead construct a solution to the following simpler SDP, and obtain a solution for the

SDP in Definition 2.6.2 via an identical procedure to the one described after the statement of Proposition 2.5.10. Given parameters λ, M, δ and graph G:

$$\begin{aligned} &\text{Find } n \times n \text{ matrix } Y \succeq 0 \text{ s.t.} \\ &\qquad Y_{i,i} = 1 \qquad\qquad \forall i \in [n] \\ &\left\langle Y, \left(A_G - \frac{d}{n} 1 1^\top\right)^{(\ell)} \right\rangle = d^\ell \lambda^\ell n \pm O(\delta n) \qquad \forall \ell \leqslant M. \end{aligned} \tag{2.34}$$

Our main technical result in this section is:

Theorem 2.7.2. *For $\boldsymbol{G} \sim G(n, d/n)$, for $|\lambda| < \frac{1}{\sqrt{d}}$, and for any $\delta, M > 0$, the SDP (2.34) is feasible with high probability.*

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

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

Define $A_c^{(m)}$ as $\mathbb{1}$ when $m = 0$ and akin to how $\overline{A}^{(m)}$ was defined in Section 2.6.2 when $m \geqslant 1$. And finally define B_c the way it is defined in Section 2.7.1. Our next ingredient is establishing an operator norm bound on $B_c^{\ell_0}$. Indeed:

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

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

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

which from (2.27) is bounded by:

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

which by our choice of ℓ_0 is at most

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

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

$$\|B_c^{\ell_0}\| \leqslant \left((1+\varepsilon)^5\sqrt{d}\right)^{\ell_0}. \tag{2.35}$$

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

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

for $\ell_i \in I$, we can conclude from a combination of submultiplicativity of operator norm and (2.35) that

$$\|B_c^{\ell}\| \leqslant \|B_c^{\ell_1}\| \cdots \|B_c^{\ell_s}\| \leqslant \left((1+\varepsilon)^5\sqrt{d}\right)^{\ell}. \tag{2.36}$$

Via the expression $A_c^{(\ell)} = SB_c^{\ell-1}CT$ in the proof of Theorem 2.7.1 and the fact that $\|S\| \leqslant n$ and $\|CT\| \leqslant n$, we know:

$$\|A_c^{(\ell)}\| \leqslant \left((1+\varepsilon)^6\sqrt{d}\right)^{\ell} \tag{2.37}$$

for all $\ell \geqslant \ell_0$. Another consequence of (2.36) is

$$\rho(B_c) \leqslant \|B_c^{\ell}\|^{1/\ell} \leqslant (1+\varepsilon)^5\sqrt{d}. \tag{2.38}$$

Now, let

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

Define $\hat{c}$ in terms of c identically to how it is defined in the statement of Theorem 2.7.1. From Theorem 2.7.1,

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

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

Proposition 2.7.3. *Suppose* $z \in \mathbb{R}$ *and* $|z| < \min\{1/\rho(B_c), 1\}$, *then* $M_\infty(z) \succeq 0$.

Proof. $M_\infty(0)$ is the identity matrix and hence is certainly positive definite, which means all its eigenvalues are positive. Additionally, by the fact that all edge weights $c(i,j)$ are bounded by 1 and the weighted Ihara–Bass formula (Theorem 2.7.1), we

can deduce that for all real z such that $|z| < \min\{1/\rho(B_c), 1\}$, $\det(M_\infty(z)) > 0$. Since the determinant (which is the product of eigenvalues) is strictly positive on a continuous interval, the eigenvalues of $M_\infty(z)$ are a continuous function of z on this interval, and the eigenvalues of M_∞ are strictly positive at one point in this interval, all eigenvalues of M_∞ must be positive for all real z where $|z| < \min\{1/\rho(B_c), 1\}$. Thus, the proposition follows. □

Our next goal will be to lower bound the minimum eigenvalue of $M_{r/2-1}(z)$, i.e. prove that the minimum eigenvalue is not too negative when z is in an appropriate range.

Proposition 2.7.4. *Suppose* $|z| < \frac{1}{(1+2\varepsilon)^6\sqrt{d}}$, *then* $\lambda_{\min}(M_{r/2-1}(z)) \geqslant -\delta(n)$ *where* $\delta(n) = o_n(1)$.

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

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

By a combination of Theorem 2.6.7 and (2.37), along with the assumption on $|z|$ we know the above is lower bounded by

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

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

Let $\delta(n)$ be the function in the statement of Proposition 2.7.4, and define

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

By Proposition 2.7.4, $X(z)$ is positive semidefinite when $|z| < \frac{1}{(1+2\varepsilon)^6\sqrt{d}}$. At this point, we state a fact that will be used later.

Fact 2.7.5. *With probability* $1 - o_n(1)$, *the maximum degree of a vertex in* $\boldsymbol{G}$ *is bounded by* $\log^2 n$.

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

Proposition 2.7.6. *Let $\ell < r$. The diagonal entries of $A_c^{(\ell)}$ are all bounded in magnitude by $\frac{\left(2\log^2 n\right)^\ell}{n}$ with probability $1 - o_n(1)$.*

Proposition 2.7.6 is a direct consequence of Fact 2.7.5 and the forthcoming Proposition 2.7.8 which we prove and use later.

Proposition 2.7.7. *The diagonal entries of $X(z)$ are all $1 \pm \delta'(n)$ with probability $1 - o_n(1)$ as long as $|z| < 1$ where $\delta'(n)$ is some function which is $o_n(1)$.*

Proof. As long as the maximum degree of G is bounded by $\log^2 n$, which happens with probability $1 - o_n(1)$, by Proposition 2.7.6 the diagonal entries of $X(z)$ are bounded by

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

which is $o_n(1)$. □

Finally for $|z| < \frac{1}{(1+2\varepsilon)^6\sqrt{d}}$ we define $Y(z) := (1 - \delta'(n))X(z) + \Gamma$ where Γ is a diagonal matrix chosen so that the diagonal of $Y(z)$ is all-ones. By Proposition 2.7.7, Γ is positive semidefinite and combined with the fact that $X(z)$ is positive semidefinite, we can conclude that $Y(z)$ is positive semidefinite as well.

2.7.3 Matching path statistics

In this section, we are interested in understanding the value of

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

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

To set up Proposition 2.7.8, we introduce some notation — let $\mathrm{NB}_{i,j,\ell}$ for $i, j \in [n]$ and $\ell \in \mathbb{N}$ denote the set of all nonbacktracking walks that start at i, end at j, and are of length-ℓ. Given a nonbacktracking walk W, we will interpret W as a set of tuples (i, ab) for $i \in [\ell]$ where ab is the edge walked on in the ith step of W. For any $S \subseteq W$, we will overload notation and use S to also denote the subset of edges with the timestep-indices removed.

Proposition 2.7.8. *Suppose G is a n-vertex graph with maximum degree bounded by $\Delta \geqslant d$, then for all $i, j \in [n]$:*

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

Proof.

$$\begin{aligned} \left| \sum_{\substack{(W,S): S \subseteq W, S \neq W \\ W \in \mathrm{NB}_{i,j,\ell} \\ S \subseteq E(G)}} \left(-\frac{d}{n}\right)^{\ell - |S|} \right| &\leqslant \sum_{k=1}^{\ell-1} \sum_{\substack{(W,S): S \subseteq W, |S| = k \\ W \in \mathrm{NB}_{i,j,\ell}}} \left(\frac{d}{n}\right)^{\ell - k} \\ &= \sum_{k=1}^{\ell-1} \left(\frac{d}{n}\right)^{\ell-k} \cdot \\ &\quad |\{(W,S) : S \subseteq W, |S| = k, W \in \mathrm{NB}_{i,j,\ell}, S \subseteq E(G)\}|. \end{aligned} \tag{2.39}$$

For each k, we bound the above summand via an encoding argument. Given a walk $W = u_0 u_1 u_2 \dots u_{\ell-1} u_\ell$ where $u_0 = i$ and $u_\ell = j$ along with a proper subset of steps S which are all contained in $E(G)$, we first find the last $u_{t-1}u_t$ on the segment which is not in S (which always exists since S is always a proper subset). Therefore the segment $u_t \dots u_\ell$ must be composed only of edges in G. In our encoding we specify:

- The set S via timestamps (for which there are $2^\ell - 1$ choices),
- The value of t (for which there are ℓ choices),

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

The total number of possible encodings is bounded by

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

Thus, the k-th summand is bounded by

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

Plugging in this bound into (2.39) implies the desired statement. □

As an upshot of Proposition 2.7.8 and Fact 2.7.5 we have:

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

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

$$\left\langle \left(A - \frac{d}{n}11^{\top}\right)^{(\ell)}, A_c^{(m)} \right\rangle = \langle A^{(\ell)}, A'^{(m)} \rangle + \langle A^{(\ell)}, R'_m \rangle + \langle A'^{(m)}, R_{\ell} \rangle + \langle R_{\ell}, R'_m \rangle. \tag{2.40}$$

The entrywise ℓ_1 norms of $A^{(\ell)}$ and $A'^{(m)}$ are the total number of nonbacktracking walks of length-ℓ and m in $\boldsymbol{G}$ and $\boldsymbol{G}_{t,r,\varepsilon}$ respectively, which by the degree-bound of $\log^2 n$ and the fact that $\ell, m \leqslant r$, are each at most $n(\log^2 n)^r$. Since the entries of R_{ℓ} and R'_m are bounded by $\frac{\ell \cdot (2\log^2 n)^r}{n}$, the second and third terms of (2.40) are bounded by $r \cdot (2\log^2 n)^{2r}$. It is easy to see that the fourth term is bounded by $r^2(2\log^2 n)^{2r}$. It remains to understand the first term.

By Lemma 2.6.9 and our choice of t, there is no (t, ε)-heavy vertex in the graph with high probability. Thus, with high probability the only vertices in $\boldsymbol{G}_{t,r,\varepsilon}$ which

were truncated are the ones that are part of a cycle of length at most r. To get a high probability bound on the number of such vertices, we observe that the number of cycles of length exactly k passing through a vertex v in the complete graph is at most n^{k-1}, whereas the probability of a fixed cycle occuring in $\boldsymbol{G}$ is $\left(\frac{d}{n}\right)^k$, which implies via a union bound that the probability that v is part of a k-cycle is at most $\frac{d^k}{n}$. Consequently, the probability that v is part of a length-$\leqslant r$ cycle is at most $\frac{rd^r}{n}$, which means the expected number of vertices that are part of a length-$\leqslant r$ cycle is at most rd^r. By Markov's inequality, with high probability the number of of such vertices is bounded by, say, $rd^r \cdot \log n$.

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

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

Thus,

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

By Fact 2.7.5 and the bound on $|S|$, the first term is bounded by $r\left(d\log^2 n\right)^{2r}$ with high probability. If $\ell \neq m$, then each of the second to fourth terms is equal to 0. If $\ell = m$, the sum of the second to fourth terms is sandwiched between the total number of length-ℓ self-avoiding walks in $\boldsymbol{G}$ that also avoid S and the total number of length-ℓ self-avoiding walks in $\boldsymbol{G}$. By Fact 2.7.5 and the bound on $|S|$ with high probability the two quantity differ by at most $r\left(d\log^2 n\right)^{2r}$, and by Theorem 2.6.6 the latter quantity is $(1 \pm o_n(1))d^\ell n$ with high probability. Since M from the statement of Theorem 2.7.2 is a constant, by a union bound, the latter quantity is $(1 \pm o_n(1))d^\ell n$ with high probability simultaneously for all $\ell \leqslant M$. Additionally observe that when $\ell \neq m$, the inner product quantity we wish to

bound remains bounded as long as the desired bounds on $|S|$ and the maximum degree in the graph hold. In conclusion, with high probability the following holds simultaneously for *all* $\ell \leqslant M$ and $m \leqslant r/2 - 1$:

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

As a consequence:

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

Since $Y(z)$ is PSD with high probability for all $|z| < \frac{1}{(1+\varepsilon)^6\sqrt{d}}$ and the choice of ε was arbitrary, Theorem 2.7.2 follows.

2.8 Local Statistics in the DRBM

In this section we will prove Theorem 2.5.3 and Proposition 2.5.10. Since the first posting of this paper, we have updated and streamlined the arguments using the framework developed by one of the authors in [BKW19]. Several lemmas below have analogues in that work with similar proofs, and we will point these out along the way.

We first prove Theorem 2.5.3 by computing the quantities $\mathbb{E}\, p_{(H,S,\tau)}(\boldsymbol{x}, \boldsymbol{G})$ in the planted model. Fix parameters d, k, M, π, recalling that M is symmetric, and $M\mathrm{Diag}(\pi)$ is stochastic. For any partially labelled graph (H, S, τ), let $\chi(H) = |V(H)| - |E(H)|$, $c(H)$ denote its number of connected components, and recall

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

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

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

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

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

Lemma 2.8.1. *Condition on a π-balanced labelling $\sigma : [n] \to [k]$, and let $\boldsymbol{P}$ be the random perfect matching on half-edges output by $\hat{\mathcal{P}}$. Let R be a simple partial matching involving constantly many half-edges, and for short write $(u, v) \in R$ if some pair (u_a, v_b) appears in the matching. Then*

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

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

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

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

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

For each edge $(u_a, v_b) \in R$, the numerator has a term $\pi(\sigma(u))M_{\sigma(u),\sigma(v)}$, and the denominator has two terms $\pi(\sigma(u))dn$ and $\pi(\sigma(v))dn$; the dropped subscripts are intentional here. Since R is simple, we can alternatively account for these terms by looking at pairs $(u, v) \in R$. Thus, again suppressing $o_n(1)$ fluctuations, we can

rewrite as

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

□

Proof of Theorem 2.5.3. Let's begin by computing the expectation of $p_{(H,S,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}})$ over $(\boldsymbol{x}, \hat{\boldsymbol{G}})$ sampled from the configuration model. This necessitates that we extend the quantity $p_{(HS,\tau)}(\boldsymbol{x}, \hat{\boldsymbol{G}})$ to the case when $\hat{\boldsymbol{G}}$ is a multigraph—we will simply take it to mean the evaluation of $p_{(H,S,\tau)}$ on the simple graph obtained by removing all self-loops and merging all multi-edges.

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

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

matchings that "collapse" to H. For each, Lemma 2.8.1 tells us the probability of inclusion in $\hat{\boldsymbol{G}} \sim \hat{\mathcal{P}}$. Finally, there are $\prod_{v\in(H)}(\pi(\tau(v))n)$ such injective maps ϕ. Putting this all together, and summing over all extensions $\hat{\tau}$,

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

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

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

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

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

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

It remains now to prove Proposition 2.5.10, the content of which is that in constructing a feasible pseudoexpectation in the planted model, it suffices to check only certain moment constraints. We first show that the moment constraints involving partially labelled subgraphs which contain a cycle are automatically satisfied. The argument below is essentially identical to [BBK+20, Lemma 5.19]

Lemma 2.8.2. *Let $\boldsymbol{G} \sim \mathcal{N}$, and assume that $\widetilde{\mathbb{E}}$ is a degree-D_x pseudoexpectation—perhaps dependent on $\boldsymbol{G}$—satisfying $\mathcal{B}_k$. For every $\delta > 0$, with high probability*

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

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

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

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

The latter is the number of occurrences of H in $\boldsymbol{G}$, with both regarded as unlabelled graphs; from the proof of Theorem 2.5.3 above, if H has a cycle, then this is $o(n^{c(H)}$. Thus with high probability

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

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

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

Definition 2.8.3. The *pruning* of a partially labelled tree (H, S, τ) is the unique maximal subtree with the property that all leaves are distinguished vertices; if (H, S, τ) is unlabelled, its pruning is empty. The pruning of a forest is obtained by taking the pruning of each tree.

Lemma 2.8.4. *Let (H, S, τ) be a partially labelled, tree, $(\tilde{H}, S, \tau)$ its pruning. Then*

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

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

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

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

Lemma 2.8.5. *Let $\boldsymbol{G} \sim \mathcal{N}$, and let (H, S, τ) and $(\tilde{H}, S, \tilde{\tau})$ be a partially labelled forest and its pruning, respectively. Then, with high probability,*

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

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

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

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

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

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

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

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

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

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

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

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

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

Lemma 2.8.6. *Let $\boldsymbol{G} \sim \mathcal{N}$, and assume that $\widetilde{\mathbb{E}}$ is is a degree-D_x pseudoexpectation, which may depend on $\boldsymbol{G}$. If $\widetilde{\mathbb{E}}$ satisfies the affine moment constraints for every pruned, partially labelled forest with at most D_G edges and D_x distinguished vertices, then with high probability*

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

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

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

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

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

To prove Proposition 2.5.10, we need to specialize this result to the case of pruned partially labelled forests with at most two distinguished vertices. These are exactly the paths with labelled endpoints, a pair of labelled vertices, and a single labelled vertex. Recalling the notation of $X \in \mathbb{R}^{nk \times nk}$ as the matrix $X_{(u,i),(v,j)} = \widetilde{\mathbb{E}}\, x_{u,i}x_{v,j}$ and $l \in \mathbb{R}^{nk}$ as the vector with $l_{(u,i)} = \widetilde{\mathbb{E}}\, x_{u,i}$, our argument in Lemma 2.5.6 may be rephrased to say that the moment constraints on $\widetilde{\mathbb{E}}$ for the first two cases at any error tolerance $\delta' > \delta$ are implied by

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

The third case, of a single labelled vertex, is implied by

$$\langle l_i, e \rangle = \pi(i) n \pm \delta n$$

Proposition 2.5.10 now follows from Lemma 2.8.2 and Lemma 2.8.6.

2.9 Bounding Singleton Expectation

Let

$$\zeta_W(A) \stackrel{\text{def}}{=} \prod_{i \in V(W)} \beta_i(A) \cdot \gamma_i(A)$$

where γ_i, β_i are boolean functions given by,

$$\beta_i(A) \stackrel{\text{def}}{=} 1[i \text{ is } (t, d') \text{ bounded in } A]$$
$$\gamma_i(A) \stackrel{\text{def}}{=} 1[i \text{ is NOT in a cycle of length} \leqslant r \text{ in } A]$$

This section is devoted to showing the following bound on the expectation of products involving singleton edges $S(W)$.

Theorem 2.9.1. *For every $d' > d > 1$ and $\delta \in (0,1)$, the following holds for all sufficiently large t. Suppose $S(W)$ is the singleton edges and $J \subseteq D(W)$ a set of duplicative edges in a (k,ℓ)-linkage W, and $g \leqslant \frac{\log n}{\log\log n}$ we have,*

$$\left| \mathbb{E} \left[\prod_{ij \in S(W)} \left(A_{ij} - \frac{d}{n} \right) \cdot \prod_{ij \in J} A_{ij} \cdot \zeta_W(A) \right] \right| \leqslant C \log^2 n \cdot \left(\frac{d}{n} \right)^{|S(W) \cup J|} \tag{2.41}$$

$$\cdot n^{0.8e(W)} \cdot 4^{|S(W)|}\delta^{|S(W)|-24kt} \tag{2.42}$$

for some absolute constant C. Here $\mathrm{exc}(W)$ *is the excess edges in the walk defined as* $e(W) = |E(W)| - |V(W)| + 1$.

We wish to emphasize that the key aspect of (2.41) is the term $\delta^{|S(W)|}$, showing that the expectation decays exponentially in $|S(W)|$.

Proof. Henceforth in this section, We will use S to denote $S(W)$. We begin the proof of the theorem by expanding out the expectation in (2.41).

$$\mathbb{E}\left[\prod_{ij\in S}\left(A_{ij}-\frac{d}{n}\right)\cdot\prod_{ij\in J}A_{ij}\cdot\zeta_W(A)\right]$$
$$=\sum_{\alpha\in\{0,1\}^S}\mathbf{Pr}[A_S=\alpha]\mathbb{E}\left[\left(1-\frac{d}{n}\right)^{|\alpha|}\left(-\frac{d}{n}\right)^{|S|-|\alpha|}\cdot\prod_{ij\in J}A_{ij}\cdot\zeta_W(A)\right].$$

Using $\mathbf{Pr}[A_S=\alpha]=\left(\frac{d}{n}\right)^{|\alpha|}\cdot\left(1-\frac{d}{n}\right)^{|S|-|\alpha|}$, we can simplify the above expression to,

$$=\left(1-\frac{d}{n}\right)^{|S|}\cdot\left(\frac{d}{n}\right)^{S\cup J}\cdot\sum_{\alpha\in\{0,1\}^S}(-1)^{\alpha}\mathop{\mathbb{E}}_{A^c}\left[\zeta_W\left(A^c,A_S=\alpha,A_J=1\right)\right],$$

where A^c denotes the random variables $\{A_{ij}|ij\in\overline{S\cup J}\}$, each of which is an independent Bernoulli random variable with expectation $\frac{d}{n}$.

We will now select a subset of edges $Q\subseteq S(W)$ such that the following two conditions hold:

1. Edges in Q are far from each other in the graph $G(W)$. Formally, for all $ij, i'j'\in Q$,
$$\mathrm{dist}_{G(W)}(i,i')\geqslant 4t\,.$$
2. Neighborhoods of each of the edges in Q have a small number of vertices. Specifically, for all $ij\in Q$, $|\mathsf{B}_{2t}(i,G_0)|\leqslant 2t+2$.

We will show in Lemma 2.9.2 that there exists such a set Q with $|Q|\geqslant |S(W)|/8t-3k-6e(W)$.

Let $R \stackrel{\text{def}}{=} S(W) \setminus Q$. Let $\alpha = \alpha_Q \cup \alpha_R$ where $\alpha_Q \in \{0,1\}^Q$ and $\alpha_R \in \{0,1\}^R$. We can upper bound the above term by,

$$\left(\frac{d}{n}\right)^{S\cup J} 2^{|R|} \cdot \max_{\alpha_R \in \{0,1\}^R} \left| \sum_{\alpha_Q \in \{0,1\}^Q} (-1)^{|\alpha_Q|} \mathop{\mathbb{E}}_{A^c} \left[\zeta_W \left(A^c, A_Q = \alpha_Q, A_R = \alpha_R, A_J = 1\right)\right] \right| \tag{2.43}$$

For any fixed choice of A^c, let $\zeta_{A^c,\alpha_R} : \{0,1\}^Q \to \{0,1\}$ denote the function,

$$\zeta_{A^c,\alpha_R}(z) \stackrel{\text{def}}{=} \zeta_W \left(A^c, A_Q = z, A_R = \alpha_R, A_J = 1\right) .$$

Rewriting the LHS of (2.43) in terms of ζ_{A^c,α_R},

$$\leqslant \left(\frac{d}{n}\right)^{S\cup J} 2^{|R|} \cdot \max_{\alpha_R \in \{0,1\}^R} \left| \mathop{\mathbb{E}}_{A^c} \left[\sum_{\alpha_Q \in \{0,1\}^Q} (-1)^{|\alpha_Q|} \cdot \zeta_{A^c,\alpha_R}(\alpha_Q) \right] \right|$$

Observe that for any function $\psi : \{0,1\}^Q \to \{0,1\}$, $\sum_{z\in\{0,1\}^Q} (-1)^{|z|}\psi(z) = 0$ if ψ is independent of any bit in z. Otherwise, the sum is upper bounded by $2^{|Q|}$. Therefore, we can rewrite the above bound as,

$$\leqslant \left(\frac{d}{n}\right)^{S\cup J} 2^{|Q\cup R|} \cdot \max_{\alpha_R \in \{0,1\}^R} \left(\mathop{\mathbf{Pr}}_{A^c} \left[\zeta_{A^c,\alpha_R} \text{ depends on all bits in } Q\right] \right) \tag{2.44}$$

Recall that $\zeta_W(A) = \beta_W(A) \cdot \gamma_W(A)$ where $\beta_W(A) = \prod_{i\in W} \beta_i(A)$ and $\gamma_W(A) = \prod_{i\in W} \gamma_i(A)$.

Analogous to the definition of ζ_{A^C,α_R}, define corresponding boolean functions β_{A^C,α_R} and γ_{A^C,α_R} over $\{0,1\}^Q$, i.e.,

$$\beta_{A^c,\alpha_R}(z) \stackrel{\text{def}}{=} \beta_W \left(A^c, A_Q = z, A_R = \alpha_R, A_J = 1\right) .$$

$$\gamma_{A^c,\alpha_R}(z) \stackrel{\text{def}}{=} \gamma_W \left(A^c, A_Q = z, A_R = \alpha_R, A_J = 1\right) .$$

By a simple union bound, we can write

$$\mathop{\mathbf{Pr}}_{A^c} \left[\zeta_{A^c,\alpha_R} \text{ depends on all bits in } Q\right]$$
$$\leqslant \sum_{Q'\subset Q} \mathop{\mathbf{Pr}}_{A^c} \left[\beta_{A^c,\alpha_R} \text{ depends on all bits in } Q' \bigwedge \gamma_{A^c,\alpha_R} \text{ depends on all bits in } Q \setminus Q'\right]$$

$$\leqslant \sum_{Q' \subset Q} \tag{2.45}$$

$$\min\left(\mathop{\mathbf{Pr}}_{A^c}\left[\beta_{A^c,\alpha_R} \text{ depends on all of } Q'\right], \mathop{\mathbf{Pr}}_{A^c}\left[\gamma_{A^c,\alpha_R} \text{ depends on all of } Q \setminus Q'\right]\right) \tag{2.46}$$

We will the probabilities in the above sum in Claim 2.9.4 and Claim 2.9.3 respectively. Substituting these bound on probabilities into (2.46)

$$\leqslant \sum_{Q' \subset Q} \min\left(\delta^{16t|Q'|}, C(\log^2 n) n^{-0.7(|Q\setminus Q'|/r - \#_c(Q\cup R\cup J)}\right) \tag{2.47}$$

$$\leqslant C(\log^2 n) n^{0.7\#_c(Q\cup R\cup J)} \sum_{Q' \subset Q} \cdot \min((\delta^{16t})^{|Q'|}, (n^{-0.7/g})^{|Q\setminus Q'|}) \tag{2.48}$$

$$\leqslant C(\log^2 n) n^{0.7e(W)} \cdot 2^{|Q|} \cdot (\delta^{16t})^{|Q|/2} \tag{2.49}$$

$$\leqslant C(\log^2 n) n^{0.7e(W)} \cdot 2^{|S(W)|} \cdot (\delta)^{|S(W)|-24kt-48e(W)} \tag{2.50}$$

$$\leqslant C \log^2 n \cdot n^{0.8e(W)} \cdot 2^{|S(W)|} \cdot \delta^{|S(W)|-24kt} \tag{2.51}$$

Substituting back in (2.44) we get the bound in the theorem.

□

Lemma 2.9.2. *For all $t < \ell$, in a $k \times \ell$-linkage there exists $Q \subset S(W)$ with $|Q| \geqslant \frac{|S(W)|}{8t} - 3\ell - 6e(W)$ such that,*

1. *For all $ij, i'j' \in Q$, $\mathrm{dist}_{G(W)}(i, i') \geqslant 4t$.*
2. *For all $ij \in Q$, $|\mathsf{B}_{2t}(i, G_0)| \leqslant 2t + 2$.*

Proof. All the steps of the walk are divided into consecutive segments of singleton edges (“singleton stretches") and duplicative edges (“duplicative stretch").

The walk can step from a singleton stretch into a duplicative stretch, either by a turn-around or at an edge that creates a cycle. The number of such transitions is therefore at most $\ell + 2e(W)$ where $2e(W)$ is the number of excess edges.

Hence there are $|S(W)|$ singletons split into $\ell + 2e(W)$ disjoint path segments. Given a path of length Δ, delete segments of length $8t$ from both end, and then pick edges at a regular intervals of length $8t$ from the each other in the remaining. This yields $\lfloor \frac{(\Delta - 16t)}{8t} \rfloor$ edges which are pairwise $8t$ away, and the $2t$ neighborhood

around each of them is a path and thus has only $2t+2$ edges. Perform this operation on each of the singleton segments to select a subset Q of singleton edges. By construction edges in Q satisfy the conditions of the Lemma above. It remains to lower bound the size of Q. If $\Delta_1, \ldots, \Delta_q$ are the lengths of the singleton stretches, we can write

$$|Q| \geqslant \sum_{i=1}^{q} \left\lfloor \frac{(\Delta_i - 16t)}{8t} \right\rfloor \geqslant \sum_{i=1}^{q} \frac{(\Delta_i - 16t)}{8t} - 1$$
$$\geqslant \frac{|S(W)|}{8t} - 3q \geqslant \frac{|S(W)|}{8t} - 3\ell - 6e(W)$$

edges. □

2.9.1 Away from short cycles

Claim 2.9.3. For any subset $Q^* \subset Q$,

$$\Pr_{A^c}\left[\gamma_{A^c,\alpha_R} \text{ depends on all bits in } Q^*\right] \leqslant C(\log^2 n) n^{-0.7(|Q^*|/r - \text{NumCycles}(Q \cup R \cup J)}$$

Proof. The function $\gamma_{A^c,\alpha_R}(z) = \gamma_W(A^c, A_Q = z, A_R = \alpha_R, A_J = 1)$ is a anti-monotone function of z.

For every pair $ij \in Q^*$, since γ_{A^c,α_R} depends on z_{ij} there is some setting of $z_{Q\setminus\{ij\}}$ such that $\gamma_{A^c,\alpha_R}(z_{ij} = 0, z_{Q\setminus\{ij\}}) = 1$ but $\gamma_{A^c,\alpha_R}(z_{ij} = 1, z_{Q\setminus\{ij\}}) = 0$. By definition of γ_W, this implies that addition of edge ij creates a cycle of length at most r.

Therefore, in the graph given by $A' = (A^c, A_R = \alpha_R, A_J = 1, A_Q = 1)$ every edge $ij \in Q^*$ is in a cycle of length at most r. There are at least $|Q^*|/r$ cycles in the graph A', and at least $|Q^*|/r - \text{NumCycles}(Q \cup R \cup J)$ involve edges of the random graph A^c.

Now we appeal to Lemma A.3 in [FM17] to conclude the claim. □

2.9.2 Heavy Vertices

The goal of this section is to prove the following claim, a component in the proof of Theorem 2.9.1.

Claim 2.9.4. Given $d' > d > 1$ and $\delta > 0$, for all sufficiently large value of t the following holds for every subset $Q^* \subset Q$,

$$\Pr_{A^c}\left[\beta_{A^c,\alpha_R} \text{ depends on all bits in } Q^*\right] \leqslant (\delta)^{t|Q^*|}$$

First, let us setup some notation. For a graph G and a set of vertices S, we make the following definitions.

$$\mathsf{B}_r(S, G) \stackrel{\text{def}}{=} \{i | \operatorname{dist}_G(i, S) \leqslant r\}$$

$$\mathsf{N}_r(S, G) \stackrel{\text{def}}{=} \{i | \operatorname{dist}_G(i, S) = r\}$$

Here dist_G refers to the shortest path distance on the graph G. We borrow the following tail bound on the sizes of neighborhoods in $\mathcal{G}(n, \frac{d}{n})$ from [FM17].

Lemma 2.9.5. *Fix $d > 1$ and consider the Erdős–Rényi graph $G \sim \mathcal{G}(n, \frac{d}{n})$. Then there exists $C, c > 0$ such that for any $s \geqslant 0$, $t \geqslant 1$ and $v \in [n]$*

$$\mathbf{Pr}\left[|\mathsf{B}_t(v; G)| \geqslant sd^t\right] \leqslant Ce^{-cs}$$

Critical to the proof of Claim 2.9.4 is the notion of being heavy vertex, and close-to heavy vertices. A heavy vertex is any vertex with $\mathsf{B}_t(v, G) \geqslant (d')^t$. A vertex is marked as close-to-heavy if it is within distance t of a heavy vertex. Formally, we have the following definition

Definition 2.9.6. A vertex v in a graph $G = (V, E)$ is (t, d')*-close to heavy* if there exists v' such that $\operatorname{dist}_G(v, v') \leqslant t$ such that $|\mathsf{B}_t(v'; G)| > (d')^t$.

First, we will bound the probability that a vertex in an Erdős-Rényi graph is (t, d')-close to heavy.

Lemma 2.9.7. *There exists absolute constant C, c such that for all t and $d' > d$, for a graph $G \sim \mathcal{G}(n, \frac{d}{n})$ and a vertex v,*

$$\mathop{\mathbf{Pr}}_{G}\left[v \text{ is } (t, d')\text{-close to heavy}\right] \leqslant Cd^t e^{-c(d'/d)^t}$$

Proof. Let X be the random variable denoting the number of vertices that are (t, d')-close to heavy in a graph $G \sim \mathcal{G}(n, \frac{d}{n})$. Clearly the above probability is given by $\frac{1}{n}\mathbb{E}[X]$. Suppose a vertex v has $|\mathsf{B}_t(v; G)| = \gamma(d')^t$ for some $\gamma > 1$. Then every vertex $u \in \mathsf{B}_t(v; G)$ is (t, d')-close to heavy. Therefore, we can upper bound the expected number of vertices that are (t, d')-close to heavy in a graph $G \sim \mathcal{G}(n, \frac{d}{n})$ by,

$$\frac{1}{n}\mathbf{E}[X]$$

$$\leqslant \int_{\gamma=1}^{\infty} \mathbf{Pr}[|\mathsf{B}_t(v;G)| = \gamma(d')^t] \cdot (\gamma(d')^t) d\gamma$$
$$\leqslant \int_{s=(d'/d)^t}^{\infty} \mathbf{Pr}[|\mathsf{B}_t(v;G)| = s(d)^t] \cdot (sd^t) \cdot \left(\frac{d^t}{(d')^t} ds\right)$$
$$\leqslant \frac{d^{2t}}{(d')^t} \int_{s=(d'/d)^t}^{\infty} Ce^{-cs} s ds \leqslant \frac{C}{c^2} \cdot \frac{d^{2t}}{(d')^t} \cdot \left[-e^{-z}z - e^{-z}\right]_{c(d'/d)^t}^{\infty} < C' d^t e^{-c(d'/d)^t}$$

where the last inequality holds whenever $(d'/d) > 1$ and $C' > \frac{C(1+c)}{c^2}$. □

The following Lemma upper bounds the probability of a vertex v being close to heavy in a more complicated setup. Here a subgraph $G' = (V', E')$ is chosen to be included in the graph, and a set of vertices $\mathcal{F}$ are forbidden in the neighborhood of v.

Lemma 2.9.8. *There exists absolute constants C, c such that the following holds for all t and $d' > d > 1$.*

Suppose $G' = (V', E')$ be a subgraph of the complete graph and let $v \in V'$ be a vertex in G'.

Let $\mathcal{F} \subseteq [n]$ be a set of vertices disjoint from V', i.e., $\mathcal{F} \cap V' = \emptyset$. Suppose we draw $G^c \sim \mathcal{G}(n, \frac{d}{n})$ and set $G = G' \cup G^c$ then,

$$\mathop{\mathbf{Pr}}_{G}\left[v \text{ is } (t,d')\text{-close to heavy in } G | \mathsf{B}_{2t}(v,G) \cap \mathcal{F} = \emptyset\right] \leqslant C|V'|d^t e^{-c\frac{1}{|V'|+1}\cdot\left(\frac{d'}{d}\right)^t}$$

Proof. Notice that the indicator of the event

$$\mathcal{E}_1 = 1[v \text{ is } (t,d')\text{-close to heavy in } G]$$

is a monotone function of the edges G^c. On the other hand, the event $\mathcal{E}_2 = 1[\mathsf{B}_{4t}(v,G) \cap \mathcal{F} = \emptyset]$ is an anti-monotone function.

By FKG inequality, the two events are negatively correlated and therefore conditioning on $\mathcal{E}_2$ reduces the chance of $\mathcal{E}_1$, i.e.,

$$\mathop{\mathbf{Pr}}_{G}\left[v \text{ is } (t,d')\text{-close to heavy in } G | \mathsf{B}_{4t}(v,G) \cap \mathcal{F} = \emptyset\right]$$
$$\leqslant \mathop{\mathbf{Pr}}_{G}\left[v \text{ is } (t,d')\text{-close to heavy in } G\right].$$

Now we make the following claim which will prove subsequently.

Claim 2.9.9. Let $\rho \stackrel{\text{def}}{=} \left(\frac{1}{|V'|+1}\right)^{1/t}$. If no vertex $w \in V'$ is $(t, \rho d')$-close to heavy in G^c, then v is not (t, d')-close to heavy in G.

Assuming the above claim, we can use the union bound to argue

$$\begin{aligned}
&\mathop{\mathbf{Pr}}_{G}\left[v \text{ is } (t,d')\text{-close to heavy in } G\right] \\
&\leqslant \mathop{\mathbf{Pr}}_{G^c}\left[\exists u \in V' \text{ which is } (t,\rho d')\text{-close to heavy in } G^c\right] \\
&\leqslant \sum_{u\in V'} \mathop{\mathbf{Pr}}_{G^c}\left[u \text{ is } (t,\rho d')\text{-close to heavy in } G^c\right] \\
&\leqslant C|V'|d^t e^{-c\frac{1}{|V'|+1}\cdot\left(\frac{d'}{d}\right)^t}
\end{aligned}$$

where the last inequality follows from Lemma 2.9.7 □

Now we return to proving Claim 2.9.9.

Proof. (Proof of Claim 2.9.9) Suppose $v \in V'$ is (t, d')-close to heavy in G, and let $u \in [n]$ be the heavy vertex with $\text{dist}_G(u, v) \leqslant t$.

Now we will lower bound $|\mathsf{B}_t(u, G^c)|$. To this end, consider any $u' \in [n]$ with $\text{dist}_G(u, u') \leqslant t$. The path from $u \to u'$ is either completely contained in G^c in which case $\text{dist}_{G^c}(u, u') \leqslant t$ or the path from $u \to u'$ uses edges in G' which implies that $u' \in \mathsf{B}_t(V', G^c)$. Therefore, we can write

$$|\mathsf{B}_t(u, G)| \leqslant |\mathsf{B}_t(u, G^c)| + |\mathsf{B}_t(V', G^c)|.$$

Since u is (t, d')-heavy, $|\mathsf{B}_t(u, G)| \geqslant (d')^t$. If no vertex $w \in V'$ is $(t, \rho d')$-close to heavy in G^c, then

$$|\mathsf{B}_t(V', G^c)| \leqslant \sum_{w\in V'} |\mathsf{B}_t(w, G^c)| \leqslant |V'| \cdot \left(\rho d'\right)^t$$

One can thus conclude that,

$$|\mathsf{B}_t(u, G^c)| \geqslant (d')^t \left(1 - \rho^t |V'|\right) \geqslant (\rho d')^t .$$

Finally since $\text{dist}_G(v, u) \leqslant t$, there exists some vertex $w \in V'$ such that

$$\text{dist}_{G^c}(w, u) \leqslant t.$$

Thus w is $(t, \rho d')$-close to heavy in G^c. □

Lemma 2.9.10. *For every $d' > d$ and $\delta > 0$, there exists t such that the following holds. Fix a subset $V_0 \subset [n]$ of vertices and a graph $G_0 = (V_0, E_0)$ with at most $|E_0| < \log^2 n$ edges. Suppose $V^* \subset V_0$ be such that,*

1. *For every vertex $i \in V^*$, $|\mathsf{B}_{2t}(i; G_0)| < t^2$.*
2. *$\mathrm{dist}_{G_0}(i, j) \geqslant 4t$ for all $i, j \in V^*$.*

Then if we sample a graph G by including each of the remaining edges $\binom{n}{2} - E_0$ independently with probability $\frac{d}{n}$,

$$\mathbf{Pr}\left[\forall v \in V^*, v \text{ is } (t, d')\text{-close to heavy in } G\right] \leqslant \delta^{t|V^*|}$$

Proof. Let $G^c = ([n], E^c)$ denote the graph consisting of edges in $\binom{n}{2} - E_0$ each of which is included independently with probability $\frac{d}{n}$.

Consider the neighborhood $\mathsf{B}_{2t}(v; G)$ around a vertex $v \in V^*$. Clearly, the neighborhood contains the sub-graph $\mathsf{B}_{2t}(v, G_0)$ since $G_0 \subset G$. All the additional vertices (and edges) in $\mathsf{B}_{2t}(v; G)$ are those reachable by taking the newly sampled edges in G^c.

Intuitively, up to constant distances, the graph G^c will be "tree-like". More specifically, for a typical sample, one would expect that the neighborhood can be decomposed as,

$$\mathsf{B}_{2t}(v, G) = \mathsf{B}_{2t}(v, G_0) \cup \bigcup_{w \in \mathsf{B}_{2t}(v, G_0)} \mathsf{T}_w$$

where T_w is a tree with vertex w as root, and no other vertices in V_0. Call a vertex $v \in V^*$ to be *typical* if the above assumptions hold.

We will first show that there is a significant fraction of vertices in $|V^*|$ are *typical* with all but negligible probability.

To this end, consider the graph $\mathcal{H}$ formed by the edges in

$$E[\mathsf{B}_{2t}(V^*; G)] - E[\mathsf{B}_{2t}(V^*; G_0)]\,,$$

where $E[\mathcal{S}]$ denotes the set of edges contained in a set of vertices $\mathcal{S}$.

Consider a vertex $v \in V^*$. For every vertex $w \in \mathsf{B}_{2t}(v; G_0)$ and $\mathrm{dist}(w, v) = d$, the graph $\mathcal{H}$ contains the subgraph $\mathsf{B}_{2t-d}(w, G) - \mathsf{B}_{2t-d}(w, G_0)$. In fact, in a typical vertex $v \in V^*$, this would be a tree of depth $2t - d$ with vertex w as root.

Claim 2.9.11. The number of typical vertices is at least $|V^*| - 2s$ where $s \stackrel{\text{def}}{=} \#_c(\mathsf{B}_{2t}(V_0; G)) - \#_c(G_0)$.

Proof. Consider the execution of a depth-first-traversal on the graph $\mathcal{H}$. More precisely, consider the execution of the following algorithm:

- ExploreGraph()
  - Set $visited[w] = false$ for all $w \in \mathcal{H} \cup V_0$
  - For each vertex $w \in \mathsf{B}_{2t}(V^*; G_0)$
    - If $visited[w] = false$ then Mark w as *isolated* and $Explore(w)$
- Explore(v)
  - for each edge $(v, w) \in \mathcal{H}$ do
    - If $w \in V_0$, mark (v, w) as *stale edge* and set $visited[w] = true$.
    - If $w \notin V_0$ and $visited[w] = true$, mark the edge (v, w) as *back edge*
    - If $w \notin V_0$ and $visited[w] = false$ set visited[w] = true and call Explore(w)

Execution of ExploreGraph will consist of a sequence of DFS traversals each producing a connected component of $\mathcal{H}$. Each traversal starts at some node $w \in \mathsf{B}_{2t}(V^*, G_0)$ that has not been visited yet. The traversal goes through edges in $\mathcal{H}$, visiting new nodes, marking some edges as back and stale.

Observe that every stale edge or a back-edge increases the cycle number of $\mathsf{B}_2t(V_0; G)$ by adding an edge, but no new vertex. Therefore, the total number of stale/back edges is at most $\#_c(\mathsf{B}_{2t}(V_0; G)) - \#_c(G_0)$. For brevity, let us denote $s \stackrel{\text{def}}{=} \#_c(\mathsf{B}_{2t}(V_0; G)) - \#_c(G_0)$.

A vertex $v \in V^*$ is typical if the following hold:

1. Every vertex $w \in \mathsf{B}_{2t}(v; G_0)$ is marked *isolated* (never visited via a stale edge).
2. For every vertex $w \in \mathsf{B}_{2t}(v; G_0)$, the corresponding call $Explore(w)$ did not produce a *stale* or *back* edge in one of its descendants.

   Since there

As there are at most s-stale edges, at most s vertices $v \in V^*$ have some vertex $w \in \mathsf{B}_{2t}(v, G_0)$ visited by a stale edge. Furthermore, at most s vertices $v \in V^*$ have a vertex $w \in \mathsf{B}_{2t}(v, G_0)$ that produced a *stale* or *back* edge. Hence at least $|V^*| - 2s$ vertices are typical. □

Returning to the proof of Lemma 2.9.10, let $V^*_{\mathsf{typ}} = \{i_1, \ldots, i_R\} \subseteq V^*$ denote the set of *typical* vertices in V^*. Now we will describe how to sample a graph $G = G^c \cup G_0$ from the conditional distribution: $\left(G \mid V^*_{\mathsf{typ}} \text{are typical}\right)$.

- For $j = 1$ to R
  - Sample $\mathsf{B}_{2t}(i_j; G)$ conditioned on i_j being *typical* or equivalently,

    $$\mathsf{B}_{2t}(i_j; G) \cap G_{j-1} = \mathsf{B}_{2t}(i_j, G_0)$$

    .

    For sake of concreteness, we will outline how to sample $\mathsf{B}_{2t}(i_j; G)$. For each vertex $w \in \mathsf{B}_{2t}(i_j, G_0)$, with distance $\mathrm{dist}(w, i_j) = D$, sample the local neighborhood tree T_w of depth D in a breadth first manner, while avoiding vertices in G_{j-1}.
  - $G_j = G_{j-1} \bigcup \mathsf{B}_{2t}(i_j; G)$.
- Sample all the remaining unrevealed edges by including them independently with probability $\frac{d}{n}$, conditioned on V^*_{typ} being typical.

By virtue of the above order of sampling, we can write

$$\mathbf{Pr}\left[\forall v \in V^*_{\mathsf{typ}}, v \text{ is } (t, d')\text{-close to heavy in } G \middle| V^*_{\mathsf{typ}} \text{ is typical}\right] \tag{2.52}$$

$$= \prod_{j=1}^{R} \mathbf{Pr}\left[i_j \text{ is } (t, d')\text{-close to heavy in } G_j \middle| G_{j-1}\right] \tag{2.53}$$

where recall that G_j is sampled conditioned on i_j being typical.

By virtue of being typical, the neighborhood of i_j, $\mathsf{B}_{2t}(i_j, G_{j-1})$ is same as its original neighborhood $\mathsf{B}_{2t}(i_j, G_0)$ in G_0. Let $\mathcal{F}_j$ be the remaining vertices in G_{j-1} namely,

$$\mathcal{F}_j = G_{j-1} - \mathsf{B}_{2t}(i_j, G_0)$$

Since i_j conditioned on being typical, vertices in $\mathcal{F}_j$ are forbidden to be chosen in the neighborhood $\mathsf{B}_{2t}(i_j; G_j)$. Therefore deleting all edges among $\mathcal{F}_j$ has no effect on whether i_j is (t, d')-close to heavy. That is,

$$i_j \text{ is } (t,d')\text{-close to heavy in } G_j \iff i_j \text{ is } (t,d')\text{-close to heavy in } \mathsf{B}_{2t}(v, G_0) \cup G^c$$

We will bound the probability of the latter event using Lemma 2.9.8. Specifically, apply Lemma 2.9.8 with $G' = \mathsf{B}_t(i_j, G_0)$ and $\mathcal{F}_j$ we get that,

$$\begin{aligned}\mathbf{Pr}\left[i_j \text{ is } (t,d')\text{-close to heavy in } |\mathsf{B}_{2t}(v,G) \cap \mathcal{F} = \emptyset\right] &\leqslant Ct^2 \cdot d^t e^{-c\left(\frac{d'}{d}\right)^t \cdot \frac{1}{1+t^2}} \\ &\stackrel{\text{def}}{=} \Delta(d,d',t)\end{aligned}$$

Substituting back in (2.52),

$$\mathbf{Pr}\left[\forall v \in V^*_{\mathsf{typ}}, v \text{ is } (t,d')\text{-close to heavy in } G | V^*_{\mathsf{typ}} \text{ is typical}\right] \leqslant \Delta(d,d',t)^{|V^*_{\mathsf{typ}}|} \tag{2.54}$$

$$\leqslant \Delta(d,d',t)^{|V^*|-2s} \tag{2.55}$$

where recall that $s \stackrel{\text{def}}{=} \#_c(\mathsf{B}_{2t}(V_0, G)) - \#_c(G_0)$. By Lemma A.3 in [FM17], we know that or all $k < \log^2 n$,

$$\mathbf{Pr}[\#_c(\mathsf{B}_{2t}(V_0,G)) - \#_c(G_0) \geqslant k] \leqslant C(\log^2 n) n^{-0.7k}$$

Along with (2.54), this implies that

$$\begin{aligned}&\mathbf{Pr}\left[\forall v \in V^*, v \text{ is } (t,d')\text{-close to heavy in } G\right] \\ &\leqslant \mathbf{Pr}\left[s > \frac{|V^*|}{4}\right] + \Delta(d,d',t)^{|V^*|-2\left(\frac{|V^*|}{4}\right)} \leqslant 2\Delta(d,d',t)^{|V^*|/2}\end{aligned}$$

for large enough n. Finally, the lemma follows by observing that for all fixed d, d', δ, we can make $\Delta(d,d',t) \leqslant \delta^{4t}$ for sufficiently large t.

□

We now have all the pieces need to prove Claim 2.9.4.

Proof. (Proof of Claim 2.9.4) The function $\beta_{A^c,\alpha_R}(z) = \beta_W(A^c, A_Q = z, A_R = \alpha_R, A_J = 1)$ is a anti-monotone function of z.

For every pair $ij \in Q^*$, since β_{A^c,α_R} depends on z_{ij}, there is some setting of $z_{Q\setminus\{ij\}}$ such that $\beta_{A^c,\alpha_R}(z_{ij} = 0, z_{Q\setminus\{ij\}}) = 1$ but $\beta_{A^c,\alpha_R}(z_{ij} = 1, z_{Q\setminus\{ij\}}) = 0$. This implies that addition of edge ij creates a vertex v that is not (t, d')-bounded. The vertex v is within distance t of both endpoints i and j, since otherwise the addition of the edge ij has no effect on the (t, d')-boundedness of vertex v.

Therefore we can upper bound,

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{A^c}\left[\beta_{A^c,\alpha_R} \text{ depends on all bits in } Q^*\right]\\ &\leqslant \mathop{\mathbf{Pr}}_{A^c}\left[\forall ij \in Q^*, i \text{ is } (t, d')\text{-close to heavy in graph } A^c \cup J \cup Q \cup R\right]\end{aligned}$$

By construction of the set of edges Q, the set $V^* = \{i | ij \in Q^*\}$ satisfy the conditions in the hypothesis of Lemma 2.9.10 in the graph formed by $Q \cup R \cup J$. Hence the claim follows by appealing to Lemma 2.9.10. □

2.10 Robustness in the Stochastic Block Model

In this section, we will show that the local statistic SDP relaxation yields a robust algorithm. Throughout this section, let $\boldsymbol{G}$ be drawn from either the Erdős-Rényi or Stochastic Block Model on n vertices, with average degree d. We will prove that an adversarial modification of ϵn edges, for sufficiently small ϵ, cannot meaningfully later subgraph occurrences, except by creating vertices of high degree. Therefore, if we run the Local Statistics SDP after deletion of sufficiently high-degree vertices, the resulting algorithm is robust to adversarial edge meddling.

Let us make this intuition precise. In a similar vein to the partially labelled graph formalism from the main body of the paper, let us define now a *pinned graph* to be a pair (H, R) where $R \subset V(H)$ contains exactly one vertex from each connected component of H. Write $\ell(H) = |R|$ for the number of such components. Given a graph G and a subset T of $\ell(H)$ vertices, an *occurrence* of (H, R) in (G, T) is an (injective?) homomorphism that maps the pinned vertices R to the target set T. Let's write $\Gamma_{H,R}(G, T)$ for the set of such occurrences.

Claim 2.10.1. For every pinned graph (H, R) and any then for any $T \subset [n]$,

$$\lim_{n\to\infty} \mathop{\mathbf{E}}_{G}[|\Gamma_{H,W(H)}(G, T)|^2] = c_H$$

for a constant c_H dependent only on H.

Proof. First let us consider the following expectation:

$$\mathbf{E}[|\Gamma_{H,R}(\boldsymbol{G}, S)|] = \sum_{\phi: V(H)\to[n]} \mathbb{P}\left[\phi \text{ is an occurrence}\right]$$

The number of nonzero terms in the summation is $\binom{n}{|V(H)|-\ell(H)}(|V(H)| - \ell(H))! = n^{|V(H)|-\ell(H)} + O(n^{|V(H)|-\ell(H)})$. For each term, the probability that ϕ is an occurrence is $O(n^{-|E(H)|})$ Since $|E(H)| \geqslant |V(H)| - \ell(H)$ in a graph with at most ℓ connected components, the above expectation is a constant depending on graph H.

Now we turn our attention to $\mathbf{E}[|\Gamma_{H,R}(\boldsymbol{G}, T)|^2]$, which we can expand as

$$\mathbf{E}[|\Gamma_{H,R}(\boldsymbol{G}, T)|^2] = \sum_{\phi,\psi: V(H)\to[n]} \mathbb{P}[\phi, \psi \text{ are occurrences}]$$

Each nonzero term gives rise to a graph H^* obtained by taking the union of the images of $\phi(H)$ and $\psi(H)$; this union is a graph with ℓ connected components, each of which contains one of the target vertices T. There are only finitely many graphs on at most $2|V(H)|$ vertices that have this form, so we can write the expectation of concern to us as a sum of expected occurrence counts of these types, and apply our initial observation. □

Lemma 2.10.2. *Fix $d > 0, \epsilon \in (0,1)$, and a finite pinned graph (H, R). There exists $\Delta(H, R, d, \epsilon) > 0$ such that the following holds: with probability $1 - \epsilon$ for all $Q \subset [n]$ with $|Q| \leqslant \Delta(H, R, \epsilon)n$,*

$$\left|\bigcup_{T\cap Q\neq\emptyset} \Gamma_{H,R}(\boldsymbol{G}, T)\right| \leqslant \epsilon n^{\ell(H)}$$

Proof. We can expand the size of this union as a sum over subsets $T \in \binom{[n]}{\ell(H)}$:

$$\begin{aligned}\left|\bigcup_{T\cap Q\neq\emptyset} \Gamma_{H,R}(\boldsymbol{G}, T)\right| &= \sum_{T\in\binom{[n]}{\ell(H)}} |\Gamma_{H,R}(\boldsymbol{G}, T)| \cdot \mathbf{1}\left[T \cap Q \neq \emptyset\right] \\ &\leqslant \left(\sum_{T\in\binom{[n]}{\ell}} |\Gamma_{H,R}(\boldsymbol{G}, T)|^2\right)^{1/2} \cdot \left(\sum_{T\in\binom{[n]}{\ell}} \mathbf{1}\left[T \cap Q \neq \emptyset\right]^2\right)^{1/2}\end{aligned} \tag{2.56}$$

With probability at least $1-\epsilon$, we have

$$\sum_{T\in\binom{[]n]}{\ell(H)}} |\Gamma_{H,R}(\boldsymbol{G},T)|^2 \leqslant \frac{1}{\epsilon}\mathbb{E}\left[\sum_{T\in\binom{[n]}{\ell(H)}} |\Gamma_{H,R}(\boldsymbol{G},T)|^2\right] \leqslant \frac{c_H}{\epsilon}n^{\ell(H)}$$

where c_H is the constant depending on H from Claim 2.10.1. Set $\Delta(H,d,\epsilon)=\frac{\epsilon^3}{\ell c_T}$. Notice that for a set Q smaller than $\Delta(H,d,\epsilon)$, the number of $T\subset\binom{[n]}{\ell(H)}$ is at most $\ell\cdot\frac{\epsilon^3}{\ell(H)c_H}n^{\ell(H)}=\frac{\epsilon^3}{c_H}n^{\ell(H)}$. Conditioned on this event of probability $1-\epsilon$, we can use (2.56) to conclude that,

$$\left|\bigcup_{T\cap Q\neq\emptyset}\Gamma_{H,R}(\boldsymbol{G},T)\right| \leqslant \epsilon n^{\ell(H)}$$

whenever $|Q|\leqslant\Delta(H,d,\epsilon)n$. □

By taking a union bound over all trees of size k and all choices of designated vertices, we have the following corollary.

Corollary 2.10.3. *For every $d,k>0$ and $\epsilon\in(0,1)$, there exists η such that following holds. Denoting by $\mathcal{H}$ the set of all graphs with at most m edges, then with probability $1-\epsilon$, for all $Q\subset[n], |Q|\leqslant\eta n$ and $H\in\mathcal{H}$ we have*

$$|\Gamma_H(\boldsymbol{G},Q)| \leqslant \epsilon n^{\ell(H)}.$$

Now we are ready to prove the main theorem of this section, namely robustness of local statistics SDP relaxation.

Theorem 2.10.4. *(Robustness of Local Statistics SDP) For every d,ϵ,k, there exist B and γ such that, with probability at least $1-\epsilon$ over $\boldsymbol{G}=(\boldsymbol{G},\boldsymbol{E})$, the following holds:*

Let $\tilde{G}=([n],\tilde{E})$ be an arbitrary graph such that $|\boldsymbol{E}\triangle\tilde{E}|\leqslant\gamma n$; write $G^*=([n],E^*)$ for the graph obtained by deleting edges incident to all vertices of degree $>B$ in $\tilde{G}$. Then for every graph H with at most m edges,

$$|\Gamma_H(G)\triangle\Gamma_H(G^*)| \leqslant \epsilon n^{\ell(H)}$$

Consequently, if $\widetilde{\mathbf{E}}:\mathbb{R}[x]_{\leqslant 2}\to\mathbb{R}$ is a pseudoexpectation that is a feasible solution to the level $(2,m)$ local statistics SDP on G^* (or G) with tolerance δ, then $\widetilde{\mathbf{E}}$ is a feasible solution

on level $(2,m)$ local statistics SDP with tolerance $\delta+\epsilon$ on G (or G^*). Further, if $\widetilde{\mathbf{E}}$ is infeasible for the level $(2,m)$ local statistics SDP on G^* (or G) by a margin of δ, then $\widetilde{\mathbf{E}}$ remains infeasible on the level $(2,m)$ local statistics SDP by margin of $\delta-\epsilon$ on G (or G^*).

Proof. Let $\eta>0$ be the choice for which Corollary 2.10.3 holds given $d,k,\epsilon/4$. Set $B \stackrel{\text{def}}{=} \lceil \frac{2d}{\eta} \rceil$ and $\gamma = \frac{\epsilon}{4m2^m B^{m^3}}$. We will express $\Gamma_H(\boldsymbol{G}) \triangle \Gamma_H(G^*) = \Gamma_{del} \cup \Gamma_{trunc} \cup \Gamma_{add}$ and bound the size of each of the three sets.

- $\Gamma_{del} = \Gamma_H(\boldsymbol{G}) - \Gamma_H(\tilde{G})$ are the occurrences of H in $\boldsymbol{G}$ that were deleted by the adversarial corruption of edges.

  Since the corruption deletes at most γn edges, which are incident on at most $2\gamma n < \eta n$ vertices, we can use Corollary 2.10.3 to conclude that this set is at most $\epsilon n^{\ell(H)}/4$

- $\Gamma_{trunc} = (\Gamma_H(\boldsymbol{G}) \cap \Gamma_H(\tilde{G})) \setminus \Gamma_H(G^*)$ are the occurrences of H that were deleted due to the removal of edges incident to high-degree vertices while constructing G^*.

  The average degree of the graph $\boldsymbol{G}$ is $d+o(1)$ with $1-o_n(1)$. Therefore, the average degree of $\tilde{G}$ is at most $d+2\gamma < 2d$. Hence, the number of vertices of degree $> B$ is at most $(2d/B)\cdot n < \eta n$. Again by Corollary 2.10.3, $|\Gamma_{trunc}| \leqslant \epsilon n/4$.

- $\Gamma_{add} = \Gamma_H(G^*) \setminus \Gamma_H(\boldsymbol{G})$ are the occurrences of H in $\boldsymbol{G}$ that were added by the adversarial corruption, and survived the truncation of high-degree vertices.

  Every occurrence in Γ_{add} includes one of the γn edges in $\tilde{E}-E$.

  Since the degree of each vertex of G^* is at most B, there are at most B^m vertices in their neighborhood of radius m around every vertex v. Hence, for any given connected component $\mathcal{C} \subseteq H$, the number of occurrences of $\mathcal{C}$ that contain a vertex $i \in [n]$ is at most $|\mathcal{C}| \cdot (B^m)^{|\mathcal{C}|}$.

  For every edge $e=(u,v) \in \tilde{E}-E$, there are at most $2B^m$ vertices in their neighborhood of radius m. The number of occurrences of any connected component $\mathcal{C}$ in this neighborhood is thus at most $(2B^m)^{|\mathcal{C}|}$.

  Hence the number of occurrences that use at least one edge in $|\tilde{E}-E|$ is at most

$$\sum_{\mathcal{C}\subset H} \left(n^{\ell(H)-1} \cdot (B^m)^{|V(H)|-|\mathcal{C}|} \right) \cdot \left(|(\tilde{E}-E) \cap E^*| \cdot (2B^m)^{|\mathcal{C}|} \right.$$

$$\leqslant \ell(H) \cdot 2^m B^{m^2 \ell(H)} \gamma n^{\ell(H)}$$

By the choice of γ, the desired bound follows.

Conditioned on the event that assertion in Corollary 2.10.3 holds, for every choice of corruptions, we have that

$$\Gamma_H(\boldsymbol{G}) \triangle \Gamma_H(G^*) \leqslant \epsilon n/4 + \epsilon n/4 + \epsilon n/4 < \epsilon n$$

The claim about the solution to the level $(2, m)$-local statistics SDP is immediate by observing that for any partially labelled subgraph (H, S, τ),

$$\widetilde{\mathbf{E}}[|p_{H,S,\tau}(G^*, x) - p_{H,S,\tau}(G, x)|] \leqslant |\Gamma_H(G) \triangle \Gamma_H(G^*)|$$

for any $\widetilde{\mathbf{E}}$ that satisfies $\mathcal{B}_k$. □

2.11 Conjectural recovery in the DRBM

As discussed in the introduction, this paper will not settle fully the question of recovering the planted communities. However, we can at least reduce some key aspects of this problem to Conjecture 2.2.6 regarding the spectrum of $A_{\boldsymbol{G}}$ when $\boldsymbol{G} \sim \mathcal{P}_{(d,k,M,\pi)}$.

There are numerous ways to pose the recovery task, and as many metrics of success, but let us set ourselves the modest goal of, given $\boldsymbol{G}$ drawn from a planted model with $\lambda_1^2, ..., \lambda_\ell^2 > (d-1)^{-1}$ and knowledge of the parameters (d, k, M, π), recovering a vector in $\mathbb{R}^n$ with constant correlation to each of the vectors $\check{\boldsymbol{x}}_1, ..., \check{\boldsymbol{x}}_\ell$ from the Section 2.5. If $\ell = k$, we can use this and our knowledge of M to apply the change-of-basis F^{-1} and recover vectors correlated to the indicators $\boldsymbol{x}_1, ..., \boldsymbol{x}_k$ for each of the k communities.

Our first claim is that, assuming Conjecture 2.2.6, the eigenvectors of $A_{\boldsymbol{G}}$ can be used to approximate the $\check{\boldsymbol{x}}_i$'s. In Section 2.4 we showed that there exists a polynomial f strictly positive on $(-2\sqrt{d-1}, 2\sqrt{d-1}) \cup \{d\}$ with the property that

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

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

$$-\delta n > \sum_{u \in [n]} f(\mu_u) \check{\boldsymbol{x}}_i^T \Pi_u \check{\boldsymbol{x}}_i$$

$$
\begin{aligned}
&= \sum_{|\mu_u|<2\sqrt{d-1}} f(\mu_u)\check{\boldsymbol{x}}_i^T \Pi_u \check{\boldsymbol{x}}_i + \sum_{|\mu_u|\geqslant 2\sqrt{d-1}} f(\mu_u)\check{\boldsymbol{x}}_i^T \Pi_u \check{x}_i \\
&\geqslant \sum_{|\mu_u|\geqslant 2\sqrt{d-1}} f(\mu_u)\check{\boldsymbol{x}}_i^T \Pi_u \check{\boldsymbol{x}}_i \\
&\geqslant \inf_{|x|\leqslant d} f(x) \cdot \check{\boldsymbol{x}}_i^T \left(\sum_{|\mu_u|\geqslant 2\sqrt{d-1}} \Pi_u \right) \check{\boldsymbol{x}}_i.
\end{aligned}
$$

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

In order to recover *robustly* we will lean on the results of Section 2.5.5. If we begin with $\boldsymbol{G}$ from the planted model, perform ϵn adversarial edge insertion or deletions, and then run the SDP again, we showed that the old SDP solution will *still* be feasible. Thus, if we take $\check{X}$ from the SDP run on the corrupted graph, we will still have

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

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

Chapter 3

Efficient algorithms from unstable belief propagation fixed points

This chapter is adapted from [LMR22], a paper co-authored by the author of this thesis, Siqi Liu, and Prasad Raghavendra.

Many statistical inference problems correspond to recovering the values of a set of hidden variables from sparse observations on them. For instance, in a planted constraint satisfaction problem such as planted 3-SAT, the clauses are *sparse observations* from which the hidden assignment is to be recovered. In the problem of community detection in a stochastic block model, the community labels are hidden variables that are to be recovered from the edges of the graph.

Inspired by ideas from statistical physics, the presence of a stable fixed point for belief propogation has been widely conjectured to characterize the computational tractability of these problems. For community detection in stochastic block models, many of these predictions have been rigorously confirmed.

In this chapter, we consider a general model of statistical inference problems that includes both community detection in stochastic block models, and all planted constraint satisfaction problems as special cases. We carry out the cavity method calculations from statistical physics to compute the regime of parameters where detection and recovery should be algorithmically tractable. At precisely the predicted tractable regime, we give:

(i) a general polynomial-time algorithm for the problem of *detection*: distinguishing an input with a planted signal from one without;

(ii) a general polynomial-time algorithm for the problem of *recovery*: outputting a vector that correlates with the hidden assignment significantly better than a random guess would.

Analogous to the spectral algorithm for community detection [KMM+13, BLM15], the detection and recovery algorithms are based on the spectra of a matrix that arises as the derivatives of the belief propagation update rule. To devise a spectral algorithm in our general model, we obtain bounds on the spectral norms of certain families of random matrices with correlated and matrix valued entries. We then demonstrate how eigenvectors of various powers of the matrix can be used to partially recover the hidden variables.

3.1 Introduction

In the PLANTED-q-COLORING problem, a hidden coloring $\boldsymbol{c} : [n] \to \{1, \ldots, q\}$ is sampled from the uniform distribution over $[q]^n$. A random graph $G = ([n], \boldsymbol{E})$ is drawn from the Erdős-Rényi distribution conditioned on $\boldsymbol{c}$ being a legitimate coloring. So every edge (i, j) is included in the graph with probability $\frac{d}{n} \cdot 1[\boldsymbol{c}(i) \neq \boldsymbol{c}(j)]$ independently at random. Given the edges $\boldsymbol{E}$ as input, the goal of an inference algorithm is to recover (even partially) the hidden coloring $\boldsymbol{c}$.

PLANTED-q-COLORING is the archetypal example of a broad class of statistical inference problems where the goal is to recover a set of hidden variables from sparse observations on it (see [Mon08]). A large number of inference problems ranging from decoding LDPC codes to community detection in random graphs fall into this broad framework. Broadly speaking, the setup in these inference problems is as follows. A set of *hidden variables* $\{\boldsymbol{c}(1), \ldots \boldsymbol{c}(n)\}$ are drawn from a known prior product distribution $\mathbb{P}_{\boldsymbol{c}}$. A sequence of *observations* (a.k.a. hyperedges) $\boldsymbol{E}$ on these hidden variables are revealed to the algorithm. Each hyperedge $(i_1, \ldots, i_k)$ is included with probability $\frac{1}{n^{k-1}} \cdot \Phi(\boldsymbol{c}(i_1), \ldots, \boldsymbol{c}(i_k))$ for some constant $\Phi(\boldsymbol{c}(i_1), \ldots, \boldsymbol{c}(i_k))$ that depends on the values of hidden variables $\boldsymbol{c}(i_1), \ldots, \boldsymbol{c}(i_k)$. Thus the inference algorithm receives $\Theta(n)$ observations with high probability and its goal is to partially recover the values of the hidden coloring.

The key computational task is to recover the values of the hidden variables. In a sparse setup where the number of observations is linear, it is typically impossible to recover the hidden variables exactly. Therefore, one settles for the relaxed goal

of *weak* recovery where the algorithm is required to produce an assignment which correlates better than random with hidden variables.

It is often useful to also define a related decision problem of "detection" . Here, the algorithm is required to distinguish between a set of observations consistent with a single fixed assignment to hidden variables (planted distribution) or a set of observations each sampled independently by drawing a new assignment to the hidden variables (null distribution).

In this work, we will be considering a more general model that will permit constantly many *types* of variables and observations. The prior distribution of each variable depends on its type, and the probability of sampling an observation depends on the types and values of variables involved. We defer the formal description of our general model to Section 3.2.1, but instead present a few examples of these problems.

Example 3.1.1. (Stochastic Block Models) A natural generalization of the PLANTED-q-COLORING problem is the stochastic block model (SBM). The stochastic block model is defined by a parameter q (the number of labels), a distribution $\mathbb{P}_{\boldsymbol{c}}$ over $[q]$ (the expected fraction of vertices with a specific label), and a matrix $P \in \mathbb{R}^{[q]\times[q]}$ such that $P[c,d]$ gives the probability of an edge between two vertices with labels c and d. In the community detection problem, a hidden labelling $\boldsymbol{c} : [n] \to \{1,\dots,q\}$ is sampled from the product distribution $\mathbb{P}_{\boldsymbol{c}}^n$. Given $\boldsymbol{c}$, a random graph $G = ([n], \boldsymbol{E})$ is drawn by including each edge (u,v) independently with probability $P[\boldsymbol{c}(u), \boldsymbol{c}(v)]$ depending on the labels of the endpoints. The goal of the problem is to recover the labelling $\boldsymbol{c}$ from the graph G.

Example 3.1.2. (Planted CSPs) In a planted CSP over a domain $[q]$, an assignment $x \in [q]^n$ is chosen at random and clauses are sampled conditioned on being satisfied by the planted assignment x. Depending on the predicate used, one obtains different planted CSPs such as Planted NAE-k-SAT and Planted k-SAT.

Many more examples of problems that fit our framework will be presented in the rest of the paper. Alternatively, this class of problems can be viewed as *"Bayesian CSPs"*. Traditionally, a constraint satisfaction problem involves variables taking values over finite domain and a set of local constraints on them. The goal is to find an assignment that satisfies either all the constraints (exact CSPs) or the largest fraction of constraints (approximate CSP). The key difference in this setup

is that there is a prior distribution associated with assignment on the variables and the constraints.

Constraint satisfaction problems (CSP) lie at the bedrock of worst-case complexity theory tracing back all the way to SAT and NP-completeness and by now there is a rich and comprehensive theory that correctly predicts the computational complexity of the traditional CSPs, with (i) the CSP dichotomy conjecture [Sch78, Zhu20] for exact CSPs, which cleanly classifies a constraint satisfaction problem as polynomial-time solvable or NP-hard depending on whether a pair of solutions could be combined to form a third solution via a function called a polymorphism, and (ii) the Unique Games Conjecture for approximate CSPs, which characterizes the best approximation ratio possible in polynomial time with an integrality gap of a semidefinite program [Kho02, KKMO07, Rag08]. There is also a well understood picture of the complexity of refutation of random CSPs from the lens of the Sum-of-Squares semidefinite programming hierarchy [AOW15, RRS17, KMOW17]. On the other hand, our understanding of the complexity of Bayesian CSPs is still in its nascent stages. Bayesian CSPs are a rich and natural class of average case problems, and understanding their complexity would be a good test-bed for average case complexity theory. Indeed, Goldreich's pseudorandom generator [Gol11] is precisely based on harnessing the computational intractability of certain Bayesian CSPs.

A naive exponential-time algorithm for the problem would be to use the Bayes rule to compute/sample from the conditional distribution $\boldsymbol{c}|\boldsymbol{E}$. The fundamental question here is to understand the limits of efficient algorithms for this class of statistical inference problems. Furthermore, both exact and approximate versions of traditional CSPs exhibit abrupt transitions wherein the computational complexity of the problem changes from polynomial to exponential. It is a compelling question whether Bayesian CSPs also exhibit similar abrupt transitions in computational complexity, and whether there exist broadly applicable optimal algorithms for them.

3.1.1 Belief Propogation and Cavity Method

A natural candidate for an optimal algorithm for Bayesian CSPs (especially in the sparse case) is belief propogation (BP). BP is often hypothesized to be theoretically optimal, and is also very efficient in practice. There is a vast body of literature on belief propogation (BP) drawing ideas from statistical physics (see [MM09a,

Chapter 14] and [ZK16] for a comprehensive treatment). It is often very difficult to analyze BP as a standalone algorithm and we are quite far from demonstrating its optimality among polynomial-time algorithms. However, there has been a growing body of work in the past decade which suggest a very general and precise theory to predict the computational complexity of Bayesian CSPs.

To the best of our knowledge, it appears that the work of Krzakala and Zdeborova [KZ09] is the first to hypothesize a precise computational phase transition for planted problems based on ideas from statistical physics. Specifically, Krzakala and Zdeborova [KZ09] hypothesized that for a broad class of planted distributions, the problem of distinguishing the planted vs null distributions becomes computationally intractable at a well-defined threshold. In the case of community detection, this threshold coincides with the so-called Kesten-Stigum threshold. More broadly, in this work, we will often refer to this threshold of intractability for Bayesian CSPs as *the stable fixed point barrier* for reasons that will be soon clear.

Building on the ideas from [KZ09], [DKMZ11b, DKMZ11a] made a fascinating set of conjectures on community detection. For example, they conjectured that the k-coloring problem is easy exactly when the average degree of a vertex in the model satisfies $d > k^2$. Their conjectures fuelled a flurry of work, leading to algorithms that match the conjectured computational thresholds [MNS18, Mas14b, BLM15, AS15].

The *stable fixed point barrier* suggested by [KZ09, DKMZ11a] is applicable beyond the setting of community detection. For instance, Krzakala and Zdeborova point out that this stable fixed point barrier is shared by problems such as hypergraph bicoloring and locked CSPs. Here locked CSPs are those wherein every pair of assignments to a predicate have Hamming distance at least 2 (analogous to pairwise-independence leading to approximation resistance [AM09]). More broadly, there is a heuristic cavity method calculation to pinpoint the location of the stable fixed point barrier in general (see Section 3.2.4 to Section 3.2.7).

Unfortunately, we are still far from establishing the veracity of these heuristic predictions. For most of these problems, BP has not been proven to succeed in the blue region of parameters, nor is any other polynomial time algorithm known. There is no roadmap to establishing intractability of these problems when the parameters are chosen in the white region.

Our main result takes a step towards establishing these predictions by giving a spectral algorithm to partially recover the hidden variables whenever the parameters are in the blue region. Specifically, we devise a spectral algorithm that uses a

linearization of BP, an approach that has been succesfully carried out for the case of community detection in [KMM+13, BLM15].

3.1.2 Stable Fixed Point Barrier

Belief propogation (BP) aims to estimate the marginals of the hidden variables, in our case $\boldsymbol{c}(v)$ for $v \in [n]$. To visualize BP, it will be useful to consider the bipartite graph $\mathcal{H}$ with variables $[n]$ on one side and the factors (a.k.a. observations) $\boldsymbol{E}$ on the other. There is an edge between a variable v and an observation $e \in \boldsymbol{E}$ if $v \in e$. The execution of BP is divided into rounds where in each round, the variable nodes send messages to factor nodes or vice versa.

Let $m^{v\to e}$ denote the message sent by a variable v to a factor $e \in \boldsymbol{E}$ and let $m^{e\to v}$ denote the message from a factor $e \in \boldsymbol{E}$ to a variable v. All messages exchanged are distributions over the domain $[q]$, i.e., $m^{v\to e} = (m_1^{v\to e}, \ldots, m_q^{v\to e})$ and similarly $m^{e\to v} = (m_1^{e\to v}, \ldots, m_q^{e\to v})$. Intuitively speaking, $m_c^{v\to e}$ is an estimate of the marginal probability that v is assigned the color c when the factor e is absent, and $m_c^{e\to u}$ is an estimate of the marginal probability that u has color c when all other factors involving u are absent.

The general schema of a BP algorithm is to start BP with some intialization of the messages

$$\{m^{v\to e}[0], m^{e\to v}[0]\}_{v\in[n], e\in\boldsymbol{E}}$$

and iteratively update the messages as specified by the functions Y, until the messages stabilize into a fixed point, i.e., a set of messages $\{\hat{m}^{v\to e}, \hat{m}^{e\to v}\}$ so that,

$$\hat{m}^{v\to e} = \mathrm{Y}_{v\to e}\left(\{\hat{m}^{f\to v} \mid f \in \partial v \backslash e\}\right)$$
$$\hat{m}^{e\to v} = \mathrm{Y}_{e\to v}\left(\{\hat{m}^{u\to e} \mid u \in \partial e \backslash v\}\right)$$

There is a canonical starting point $\overline{m}$ for the BP iterations where the messages $m^{e\to v}$ correspond to uniform distribution over the possible values $[q]$. Conjecturally, this canonical initialization $\overline{m}$ plays a critical role in characterizing the computational complexity of inferring the hidden variables in model M. There appear to be three possible cases with regards to this canonical initialization.

Case 1: $\overline{m}$ is not a fixed point Suppose $\overline{m}$ is not a fixed point for the BP iteration over the model M, then BP iteration can be expected to make progress, thereby yielding a weak recovery of hidden variables.

In fact, we will present a self-contained algorithm that weakly-recovers the hidden coloring in this case. Formally, we will show the following in Appendix 3.10:

Lemma 3.1.3. *If $\overline{m}$ is not a fixed point for the BP iteration on model* M*, then there is a polynomial time algorithm $\mathcal{A}$ and an $\epsilon > 0$ such that*

1. *if $(\boldsymbol{E}, \boldsymbol{\tau}) \sim$ M: $\mathcal{A}$ outputs a coloring that beats the correlation random guessing achieves with the hidden coloring by ε,*
2. *$\mathcal{A}$ solves the* M *vs.* $\mathrm{M}^{\times}$ *(the null distribution) distinguishing problem with high probability.*

In light of the above lemma, it is natural to restrict our attention to the case where $\overline{m}$ is a fixed point for the BP iteration.

Case 2: $\overline{m}$ is an unstable fixed point $\overline{m}$ is an *unstable fixed point* if arbitrarily small perturbations of $\overline{m}$ will lead to the BP iteration moving away from the fixed point $\overline{m}$. This case was marked by the blue region in Figure 1.2 and Figure 1.4. In this case, our main algorithmic result is a spectral algorithm to recover a coloring c' that beats the correlation random guessing achieves with the hidden coloring. Alternatively, the spectral algorithm can be used to distinguish between the planted and the null distributions.

Case 3: $\overline{m}$ is a stable fixed point $\overline{m}$ is a *stable fixed point* if there exists a neighborhood U around $\overline{m}$ such that for any initialization $\hat{m} \in U$, BP iteration converges to the canonical fixed point $\overline{m}$. In this case, the canonical fixed point $\overline{m}$ clearly highlights a potential failure of BP algorithm. The hypothesis of Krzakala and Zdeborova [KZ09] asserts that existence of this stable fixed point marks the onset of computational intractability in general.

3.1.3 Related Work

Ideas from statistical physics have long been brought to bear on inference problems. We refer the reader to [Nis01, MM09b, ZK16, RTSZ19] for an introduction to the phase transitions that mark changes in statistical and computational properties of these problems.

Planted models Special cases of the planted model we consider have appeared extensively in literature. The conditional probability of the hidden vector given the noisy observations takes the form of a graphical model, i.e. factorizes according to an hypergraph whose nodes correspond to variables and hyperedges correspond to noisy observations. Such graphical models have been studied by many authors in machine learning [LMP01] under the name of *conditional random fields*. We highlight a few among the extensive body of literature on information-theoretic and structural properties of these planted models. Montanari [Mon08] characterized the posterior marginals in terms of fixed points of the associated density evolution operator. Subsequently, Abbe and Montanari [AM13] show concentration for the conditional entropy per hidden variable given the observations. More recently, Coja-Oghlan et. al. [COHKL+20] study the information theoretic limits to recovery and confirm a conjectured formula for the mutual information between the observations and the planted assignment.

Spectral algorithms via non-backtracking operator The idea of using the spectra of non-backtracking matrix for recovery in planted problems can be traced back to the seminal work of Krzakala et al. [KMM+13] in the context of community detection. While this work provided heuristic arguments supporting the correctness of the algorithm, it was rigorously established in the work of Bordenave et. al. [BLM15]. Subsequently, [SLKZ15] devised spectral algorithms for solving the recovery problem in the censored block model, a variant of community detection wherein the edges are weighted and the weights carry the information about the community labels, but the edges don't. Building on the result of [BLM15], this work shows that the eigenvectors of non-backtracking matrix can be used to partially recover the communities, right up to the threshold. Finally, Angelini et al. [ACKZ15] consider a model of sparse hypergraphs that includes planted CSPs as a special case. The paper proposes a spectral algorithm based on a generalization of a non-backtracking matrix to hypergraphs, and gives a heuristic argument that the algorithm solves detection whenever belief propogation succeeds. Unlike our work, the algorithm proposed in [ACKZ15] uses an unweighted non-backtracking matrix that is independent of the prior probabilities. While it is a desirable feature that the algorithm is *non-parametric*, i.e., does not rely on the knowledge of prior distributions generating the instance, it is unclear if such a non-parametric algorithm can achieve detection up to the threshold in general.

Apart from recovery in planted models, the non-backtracking operator and the closely related Bethe-Hessian matrix have also been applied towards computing upper bounds for the log-partition function in ferro-magnetic Ising models on general graphs [SKZ17].

Quiet Planting Planted distributions that are indistinguishable from their random counterparts are often referred to as "quiet planting", though the terminology is not often consistent on whether the distributions are computationally or statistically indistinguishable.

A quiet planting that is statistically indistinguishable from random was used as a technical tool to study random instances in [AC08]. Krzakala and Zdeborova [KZ09] studied the existence of quiet plantings for graph coloring problem and were the first to hypothesize that under the Kesten-Stigum threshold, the planted ensembles are a computationally indistinguishable from random. Subsequently, the authors [ZK11] considered planted distributions for *locked CSPs*, wherein every pair of assignments to a predicate have Hamming distance at least 2 and showed that problem is easy above a threshold that coincides with the Kesten-Stigum threshold and hypothesize that non-trivial recovery is computationally hard under it. Finally, a statistically quiet planting for the random k-SAT problem has been proposed in [KMZ12].

Community Detection Extensive work on community detection for stochastic block models has led to the confirmation [MNS18, Mas14b, BLM15, AS15] of conjectures of Decelle et. al. [DKMZ11b, DKMZ11a]. As predicted, existing algorithms [Mas14b, BLM15, AS15] can partially recover community labels up to the Kesten-Stigum threshold, but no lower. For $q = 2$ communities, the Kesten-Stigum threshold also matches the information theoretic threshold beyond which recovery is impossible. However, for $q \geqslant 3$ communities, the problem is believed to exhibit a statistical-vs-computational gap, in that there is a range of parameters where partial recovery is possible but is computationally intractable. The presence of a gap between the Kesten-Stigum threshold and the information theoretic threshold for all $q > 5$ was established in [Sly09]. More recently, Ricci-Tersenghi et al. [RTSZ19] predicted the existence of such a gap for $q = 4$ communities for some degree distributions, and also identifies a threshold beyond which there is a hard phase in asymmetric SBM. Furthermore, this work predicts the existence of

hybrid-hard phases where it is computationally easy to reach a non-trivial inference accuracy, but computationally hard to match the information theoretically optimal one. Specifically, there are stable fixed points for BP that are not the trivial fixed point, but also don't correspond to optimal recovery.

Spectral norm bounds Technically, our work draws on ideas from Bordenave, Lelarge and Massoulie [BLM15] who established spectral norm bounds for non-backtracking matrices associated with Erdős-Renyi random graphs. Closer to our own setup, Stephan et al. [SM20] show eigenvalue bounds for the non-backtracking matrices of random graphs that have independent and bounded edge weights, and bounded model complexity (measured by the rank of the expected adjacency matrix). However, in our model the edges have correlated matrix weights instead of independent scalar weights, so their eigenvalue bounds do not generally apply to our model. Another work we draw several ideas from is that of Bordenave and Collins [BC19], who prove that the spectra of a wide family of random graphs, namely those arising from matrix-weighted noncommutative polynomials of random permutation matrices (see [OW20] for a comprehensive characterization and examples in this family), are roughly contained within the spectrum of an appropriately defined infinite graph. The key techniques useful in our work are the ones they employ to bound the spectral norms of the non-backtracking matrices of random regular graphs whose the edges are endowed with varying matrix weights.

3.1.4 Technical Overview

We define a general model for sparse observations on a hidden vector, and carry out the cavity method calculations in full generality following [DKMZ11a]. We obtain a criterion for computational tractability of the recovery and detection problems on this model, and provide spectral algorithms for recovery (Theorem 3.2.10) and detection (Theorem 3.2.11) in the tractable regime. The key technical ingredient in our work is tight eigenvalue bounds for nonbacktracking matrices of sparse random hypergraphs with (possibly varying) matrix-valued edge weights (Theorem 3.3.8).

In this section, we will attempt a brief technical outline of our result specialized to the case of distinguishing a random NAE3SAT instance from one with a hidden satisfying assignment. Concretely, consider the problem distinguishing $\mathcal{D}_{\text{null}}$ from

$\mathcal{D}_{\text{planted}}$ where:[1]

- An instance $\boldsymbol{\mathcal{I}} \sim \mathcal{D}_{\text{null}}$ is obtained by sampling each triple of distinct vertices (u, v, w) in $[n]^3$ independently with probability $\frac{d}{3!n^2}$ and then placing uniformly random negations $(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v, \boldsymbol{\sigma}_w)$ on each variable.
- An instance $\boldsymbol{\mathcal{I}} \sim \mathcal{D}_{\text{planted}}$ is sampled in a two-stage process: (1) sample a hidden assignment $\boldsymbol{x} \sim \{\pm 1\}^n$, (2) sample each triple of distinct vertices (u, v, w) in $[n]^3$ independently with probability $\frac{d}{3!n^2}$ and place uniformly random negations $(\boldsymbol{\sigma}_u, \boldsymbol{\sigma}_v, \boldsymbol{\sigma}_w)$ *conditioned* on $\text{NotAllEquals}(\boldsymbol{\sigma}_u \boldsymbol{x}_u, \boldsymbol{\sigma}_v \boldsymbol{x}_v, \boldsymbol{\sigma}_w \boldsymbol{x}_w) = 1$.

First, let us map out the statistical physics prediction of the smallest value of d at which the problem becomes computationally tractable. In particular, we need to work out the value of d for which the trivial fixed point for belief propogation is unstable. To this end, one emulates the cavity method heuristic calculations analogous to the one carried out in [DKMZ11a] for stochastic block models. Oversimplifying for the sake of presentation, the cavity method heuristic amounts to carrying out the calculation by treating the neighborhood of each variable to be an infinite tree (see Section 3.4 for more details).

Concretely, the setup in the cavity method calculation is as follows. The neighborhood of a variable v in the NAE3SAT instance is modelled as an infinite tree with alternating layers of variable and NAE3SAT constraint nodes. The tree is generated by a Galton-Watson process where each variable v picks a degree $d_w \sim \text{Poisson}(d)$ from the Poisson distribution, and has d_w NAE3SAT constraint nodes as children, and each constraint node has exactly 2 children. For each path of length 2, $u \to C \to w$ from a variable u to its constraint node C followed by another variable w in the constraint, there is an associated constant sized matrix M_{uCw} depending on the prior distribution. For the case of NAE3SAT, all of the matrices M_{uCw} are given by $M_{uCw} = \sigma_u \sigma_w M$ where

$$M := \begin{bmatrix} -1/6 & 1/6 \\ 1/6 & -1/6 \end{bmatrix}.$$

[1] Strictly speaking, this model and the distribution over NAE3SAT instances our generic model yields differ slightly. Nevertheless they are contiguous and so the phenomena in one carry to the other.

For any depth t, consider the following quantity ρ_t where the expectation is over the choice of the infinite tree $\mathcal{T}$,

$$\rho_t(d) = \mathop{\mathbf{E}}_{\text{tree } \mathcal{T}} \left[\sum_{\text{paths} u_0 = v \to C_0 \to u_1 \to C_1 \to u_2 \to \cdots \to u_t} \text{Tr} \left(\left(\prod_{i=0}^{t-1} M_{u_i C_i u_{i+1}} \right) \left(\prod_{i=0}^{t-1} M_{u_i C_i u_{i+1}} \right)^* \right) \right]$$

The threshold d^* predicted by the cavity method is precisely the smallest value of d for which $\lim_{t \to \infty} \rho_t(d) > 1$.

This characterization of d^* is a little unwieldy in that it is not immediate that the value of the threshold d^* is decidable. Fortunately, through ideas from the work of Bordenave and Collins [BC19], the above characterization can be equivalently written in terms of the spectral radius of an associated finite matrix. Specifically, for NAE3SAT, d^* is the smallest d for which the spectral radius of L exceeds 1 where:

$$L = d \cdot \begin{bmatrix} 1/18 & -1/18 & -1/18 & 1/18 \\ -1/18 & 1/18 & 1/18 & -1/18 \\ -1/18 & 1/18 & 1/18 & -1/18 \\ 1/18 & -1/18 & -1/18 & 1/18 \end{bmatrix}.$$

(see Section 3.2.9 for an overview of how to construct L in general, and Section 3.4 for details). Hence, for NAE3SAT, the problem is hypothesized to become algorithmically tractable once $d > 4.5$. Our main results are algorithms for distinguishing the null and planted distributions, and for partially recovering a hidden assignment in the general model we consider whenever the spectral radius $\rho(L)$ of the matrix L corresponding to the model exceeds 1. In the case of NAE3SAT, we prove:

Theorem 3.1.4. *When $d > 4.5$, given $\boldsymbol{\mathcal{I}} \sim \mathcal{D}_{null}$ or $\mathcal{D}_{planted}$:*

1. *There is an efficient algorithm to distinguish $\mathcal{D}_{null}$ from $\mathcal{D}_{planted}$ with probability $1 - o(1)$.*
2. *There is an efficient algorithm to produce $\Theta(1)$ unit vectors V where $\langle v, x \rangle \geqslant \Omega(\sqrt{n})$ for some $v \in V$.*

We now describe the distinguishing algorithm, which is spectral in nature, and briefly survey the techniques to analyze the matrix involved. The matrix we employ is a power of the so-called non-backtracking matrix obtained by linearizing

belief propogation. For each clause C and pair of variables u, v in the clause signed by σ_u, σ_v we define matrix $M_{uCv} := \sigma_u \sigma_v M$. The s-th nonbacktracking power matrix is a $n \times n$ block matrix where each block is 2×2:

$$A^{(s)}[a,b] := \sum_{uC_1u_1C_2u_2\ldots u_{s-1}C_sv} M_{uC_1u_1} M_{u_1C_2u_2} \cdots M_{u_{s-1}C_sv}.$$

The algorithm is then fairly simple:

- Let $s = \lfloor \sqrt{\log n} \rfloor$, and let κ be strictly between $\sqrt{\rho(L)}$ and $\rho(L)$.
- If $\|A^{(s)}\| < \kappa^s$ output $\mathcal{D}_{\text{null}}$, otherwise output $\mathcal{D}_{\text{planted}}$.

In order to prove that the algorithm is correct, there are two key technical steps: (1) to prove that in the planted model, the operator norm is large, (2) to prove that in the null model, the operator norm is bounded.

The key insight in proving (1) is that the large operator norm of $A^{(s)}$ arises from the hidden assignment to the planted instance $\mathcal{I}$ itself. In particular, denoting $\boldsymbol{y} := \boldsymbol{x} \otimes \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ we prove:

Lemma 3.1.5. *With probability* $1 - o(1)$*:*

$$\frac{\langle \boldsymbol{y}, A^{(s)} \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2} \geqslant \Omega(\rho(L)^s).$$

This is proved in full generality in Section 3.6.

The main technical difficulty is in proving (2) in the general model, as is done in Section 3.7. We prove:

Lemma 3.1.6. *With probability* $1 - o(1)$*:*

$$\|A^{(s)}\| \leqslant \left((1 + o(1))\sqrt{\rho(L)}\right)^s.$$

Our proof is largely inspired by the works of [BLM15, BC19, SM20]. On one hand, [BLM15] and [SM20] show tight eigenvalue bounds for the nonbacktracking matrices of sparse (possibly inhomogenous) Erdős-Rényi graphs with scalar edge weights. The proof exploits the commutativity of scalar products, i.e. the product of edge weights along a walk is invariant under reordering. However, the graphs

we consider have matrix-valued weights, which in general don't commute under multiplication. Therefore the product of edge weights changes depending on the order of multiplication. On the other hand, random regular graphs with matrix-valued edge weights is handled in the work of [BC19]. However, the proof in [BC19] heavily exploits the regularity of the model — each vertex has exactly d adjacent edges and these edges have the exact same set of matrix weights. This leads to every vertex having isomorphic neighborhoods, and simplifies the analysis, which does not occur in our setting due to the lack of regularity. Our situation is further complicated by the fact that due to hyperedges of size greater than 2 even the random matrix weights in different blocks are not independent, which introduces mild correlations. The proof follows the general framework of the trace method and gives a more fine-grained analysis for nonbacktracking walks based on their shapes.

While the spectral radius of non-backtracking powers $A^{(s)}$ serve as a distinguisher, recovering the hidden assignment from the eigenvectors is little more subtle. In particular, this requires proving a converse of Lemma 3.1.5 that the every vector v for which $\langle v, A^{(s)}v\rangle$ is large, is actually correlated with the planted assignment. Instead, we bypass this issue by collating information from eigenvectors of $A^{(s)}$ for a range of values of s (see Section 3.8 for details).

3.1.5 Discussion and Future Work

In this work, we have shown that for a very general class of planted problems, the problem is computationally tractable whenever the trivial fixed point is unstable. This establishes the algorithmic side of the predictions of Krzakala and Zdeborova [KZ09] for all these problems. Several compelling open questions remain, we list a few here.

Reductions. From the standpoint of average case complexity, the main open question is to establish or refute the stable fixed point barrier. Given that all Bayesian CSPs have a uniform onset of intractability as specified by the stable fixed point barrier, perhaps these problems are reducible to one another. Traditional CSPs are very amenable to reductions, it is compelling to see if there are reductions between Bayesian CSPs, and stable fixed point barrier can be obtained as a consequence of the intractability of a single Bayesian CSP. The main challenge here is in coming up with reductions between problems that are *distribution-preserving* and we spec-

ulate that the ideas in [BBH18, BB20], which are examples of recent successes in reductions between average case problems, might be useful.

Hardness evidence in restricted computational models. Evidence on the stable fixed point barrier would also be very interesting. [HS17] showed that an algorithm based on low-degree polynomials solves the distinguishing problem in community detection up to Kesten-Stigum threshold, and also proves matching hardness in that low-degree polynomials fail to solve the problem under the Kesten-Stigum threshold. Recent work introduced the local statistics SDP hierarchy [BMR21] and showed the same algorithmic result for this class of algorithms and proved a negative result for the degree-2 SOS version of this algorithm. It will be useful to show that low-degree method and local statistics SDP hierarchy fail to solve the detection problem in the general model we consider in the presence of a stable fixed point. It will also be interesting to see if conditional hardness results for the problem can be obtained in other models such as statistical query algorithms [FGR$^+$17].

Another direction in the spirit of the recent work of [BBH$^+$20] which establishes an equivalence between the predictions of statistical query algorithms and the low-degree polynomials method would be to formally establish the equivalence of the predictions for the stable fixed point based on the cavity method with the other restricted models of computation such as the ones mentioned above.

Goldreich's PRG for 1-wise independent, balanced, local predicates Goldreich proposed a construction of pseudorandom generators from random CSPs with balanced local predicates [Gol00]. The generator mapping n variables to $\{0,1\}^m$ is constructed as follows: let $E_1, \ldots, E_m$ be a randomly chosen set of constraints on n input variables, then the i-th bit of the output string indicates whether E_i is satisfied by the input or not.

The constraints in Goldreich's generator can be sampled from the null distribution of the model that we study. Then on any input $\boldsymbol{c}$ (analogous to the hidden variables in the model), the output of the generator together with the constraints can be viewed as observations from the model's planted distribution. Roughly speaking, we say that this generator produces pseudorandom strings if and only if the detection problem for this model is intractable.

For any random CSP with 1-wise independent predicates, the cavity method yields a concrete predicate density threshold above which the detection problem should be tractable (indeed the threshold is always of order $O(n)$). Our distinguishing algorithm confirms tractability in this regime, and therefore provides a concrete linear upper bound on the stretch of the Goldreich's PRG constructed from the random CSP. Indeed, the upper bounds would be tight if the stable fixed point barrier hypothesis holds.

NP problem If the stable fixed point barrier hypothesis holds, then these Bayesian CSPs are excellent examples of average-case hard problems that are easy to sample. Their intractability can be harnessed to build cryptographic and pseudorandom primitives whose security depends on the existance of average-case hard problems. To this end, it is important that the underlying intractable problem is in NP, i.e., given the true hidden assignment, an efficient algorithm must be able to recognize it. Formally, this motivates the following NP-version of the problem:

Problem 3.1.7. (NP version) Devise an efficient verification algorithm $\mathcal{A}$ that, given observations $\boldsymbol{E}$ from the planted model M and a candidate assignment $\boldsymbol{c} : [n] \to [q]$, has the following property:

- If $(\boldsymbol{c}, \boldsymbol{E})$ are generated from the model M, the algorithm $\mathcal{A}$ accepts $(\boldsymbol{c}, \boldsymbol{E})$ with high probability.
- If the observations $\boldsymbol{E}$ are generated from the null model M, then for every assignment $\boldsymbol{c} : [n] \to [q]$, the algorithm rejects $(\boldsymbol{c}, \boldsymbol{E})$ with high probability.

Dense models The focus of this paper has been the sparse settings, where in the underlying variable-observation graph is constant degree on average. Stability of trivial fixed point is also hypothesized to indicate computational intractability in dense problems such as spiked Wigner matrix (see [MV17] for some rigorous results). In this setting, it is the stability of fixed points of the *approximate message passing* (AMP) algorithm. A natural open question is whether the spectral algorithm based on linearizing AMP can be shown to generically hold in the dense setting.

Optimal Recovery Finally, in the region where weak recovery is possible, BP is conjectured to achieve the optimal recovery rate, i.e., achieve the maximum possible correlation with the hidden communities. While spectral algorithms provably

achieve weak recovery, there has only been partial progress on the problem of achieving the optimal recovery rate [MNS14] — in particular, optimal recovery even in the 2-community block model close to the Kesten-Stigum threshold is open. In analogy with traditional CSPs, stable fixed point barrier marks the onset of *"approximation resistance"* for some problems, while the recovery rate corresponds to the approximation ratio.

3.2 Preliminaries

3.2.1 Observation Model

We will now formally define the *observation model* that is used throughout this work. The basic setup consists of a set of hidden variables $\boldsymbol{c}(1), \ldots, \boldsymbol{c}(n)$ taking values over a finite domain $[q] = \{1, \ldots, q\}$. Borrowing terminology from the PLANTED COLORING problem, we will refer to $[n] = \{1, \ldots, n\}$ as the set of variables, $[q]$ as the set of colors and $\boldsymbol{c} : [n] \to [q]$ as the *hidden coloring*.

The hidden coloring $\boldsymbol{c} : [n] \to [q]$ is drawn from a prior distribution $\mathbb{P}_{\boldsymbol{c}}$. A sequence of hyperedges $\boldsymbol{E}$ on the vertex set $[n]$ are drawn, and we will refer to these hyperedges as *observations*. More precisely, an observation is effectively a hyperedge $e \in \boldsymbol{E}$ with a type $\boldsymbol{\tau}(e)$.

Definition 3.2.1. An observation model $\mathrm{M} = ([q], T, \mathcal{T}, \{\mathbb{P}_\tau\}_{\tau \in T}, \Phi)$ describes a distribution on n-vertex hypergraphs M_n for every $n \geqslant 1$ and is specified by,

- **(Variable Types T).** A set of types T for the hidden variables and a distribution $\mathcal{T}$ over them.

  Each variable is assigned a random type sampled from $\mathcal{T}$ and is described by $\boldsymbol{\tau} : [n] \to T$; in aggregate there are $\approx \mathcal{T}(\tau) \cdot n$ variables of type τ.

- **(Prior Distributions $\{\mathbb{P}_\tau\}_{\tau \in T}$).** For each variable of type $\tau \in T$, a prior distribution $\mathbb{P}_\tau$.

  The prior distribution of hidden coloring $\boldsymbol{c} : [n] \to [q]$ is the product distribution,

$$\mathbb{P}_{\boldsymbol{c}} = \mathbb{P}_{\boldsymbol{\tau}(1)} \times \mathbb{P}_{\boldsymbol{\tau}(2)} \cdots \times \mathbb{P}_{\boldsymbol{\tau}(n)}$$

- **(Observation Types Φ).** Set of observation types $\Phi = \{\phi_1, \ldots, \phi_F\}$. The arity of a type i observation is denoted by $\mathrm{a}(i)$.

  Each observation on the variables is a hyperedge with a type from Φ. Specifically, the set of all observations is a set of hyperedges $\boldsymbol{E}$ partitioned as $\boldsymbol{E} = \cup_{i \in [F]} \boldsymbol{E}_i$ where $\boldsymbol{E}_i$ is a set of $\mathrm{a}(i)$-tuples of distinct elements in $[n]$.

- **(Observation Distributions).** For each observation type $\phi_i \in \Phi$, we have a bounded function $\phi_i : T^{\mathrm{a}(i)} \times [q]^{\mathrm{a}(i)} \to \mathbb{R}^+$.

  For every $\mathrm{a}(i)$-tuple $(v_1, \ldots, v_{\mathrm{a}(i)})$ of distinct elements in $[n]$, the observation $\phi_i(v_1, \ldots, v_{\mathrm{a}(i)})$ is included independently with probability

$$\mathbf{Pr}\left[\phi_i(v_1, \ldots, v_{\mathrm{a}(i)}) \in \boldsymbol{E}_i\right] \stackrel{\text{def}}{=} \frac{\phi_i\left((\boldsymbol{\tau}(v_1), \boldsymbol{c}(v_1)), \ldots, (\boldsymbol{\tau}(v_{\mathrm{a}(i)}), \boldsymbol{c}(v_{\mathrm{a}(i)}))\right)}{n^{\mathrm{a}(i)-1}}$$

  Notice that the probability of drawing an observation $\phi_i(v_1, \ldots, v_{\mathrm{a}(i)})$ depends both on the types of the variables and their colors.

We refer the reader to the work of Montanari [Mon08], where this model has been previously used for a wealth of concrete examples captured in this framework. Here we will exhibit a few examples.

Example 3.2.2. (Stochastic Block Model in semi-supervised setting)

In this variant of community detection, a graph $G = (V, E)$ is drawn from a $[q]$-community SBM and in addition an α-fraction of the vertex labels are revealed. [ZMZ14] study the KS threshold in this model using the cavity method.

To encode this problem into our framework, we will have vertex types $T = [q] \cup \{\perp\}$ wherein the type of a vertex v is $\boldsymbol{\tau}(v) \in [q]$ if the label of v is revealed, and $\boldsymbol{\tau}(v) = \perp$ if it is unrevealed.

We have a single observation type namely the edges of the SBM, and the probability of an edge (u, v) is clearly $\phi\left((\boldsymbol{\tau}(u), \boldsymbol{c}(u)), (\boldsymbol{\tau}(v), \boldsymbol{c}(v))\right)/n$ for a function ϕ depending on the types and colors of two vertices.

More generally, the model can encode variants of SBM wherein there is additional attributes revealed about the vertices or edges or both. For example, SBM with labelled edges [HLM12] are subsumed by different types of observations, while SBM with vertex features [DSMM18] are captured by vertex types.

Further, the model can also be used to express geometric SBM [GMPS18] in restricted cases. In a geometric SBM, the vertices are distributed on a compact metric space like the sphere, and the probability of including an edge between vertices u, v is a function of the distance between the two. If the metric space is compact, say a sphere in a constant dimensional space, then one can use an ϵ-net of the compact set as a finite set of vertex types to model the SBM in our framework.

3.2.1.1 Miscellaneous simplifying notation

Class function Cl**:** For notational convenience, we will make a modification to our Definition 3.2.1 that does not affect the generality of our results. We will enforce that each observation type ϕ_i have a fixed tuple of variable types on which it applies. Formally, each observation type ϕ_i has an associated *class* type $\mathrm{Cl}(i) \in T^{\mathrm{a}(i)}$ such that all occurrences of the observation ϕ_i have input variable types given by $\mathrm{Cl}(i)$. It is clear that this restriction is a special case of Definition 3.2.1 with the additional restriction that,

$$\phi_i\left((\tau_1, c_1), \ldots, (\tau_{\mathrm{a}(i)}, c_{\mathrm{a}(i)})\right) = 0 \text{ if } (\tau_1, \ldots, \tau_{\mathrm{a}(i)}) \neq \mathrm{Cl}(i)$$

Conversely, given a general model M as per Definition 3.2.1, for each observation type ϕ_i and each tuple $\boldsymbol{\tau} = (\tau_1, \ldots \tau_{\mathrm{a}(i)})$, introduce an observation type $\phi'_{i,\boldsymbol{\tau}}$ that is identical to ϕ_i, but restricted to variable types $\boldsymbol{\tau}$, i.e., set $\mathrm{Cl}(i) = \boldsymbol{\tau}$. It is easy to see that this transformation creates a model M′ that is equivalent to M. Without loss of generality we will henceforth use $\phi_i(c_1, \ldots, c_{\mathrm{a}(i)})$ to denote $\phi_i((\tau_1, c_1), \ldots, (\tau_{\mathrm{a}(i)}, c_i))$ where $\boldsymbol{\tau} = \mathrm{Cl}(i)$.

Average factor density: We will use $\overline{\phi_i}$ to denote the average density of a factor:

$$\overline{\phi_i} := \sum_{(c_1,\ldots,c_{\mathrm{a}(i)}) \in [q]^{\mathrm{a}(i)}} \left(\prod_{k=1}^{\mathrm{a}(i)} \mathbb{P}_{\mathrm{Cl}(i)_k}\right) \cdot \phi_i(c_1, \ldots, c_{\mathrm{a}(i)}).$$

Bipartite view: Given a collection of sampled observations $\boldsymbol{E} = \cup_{i=1}^{F} \boldsymbol{E}_i$, we associate a bipartite graph $\boldsymbol{G}$ where the left vertex set is given by the variables $[n]$ and the right vertex set is given by the collection of all (i, γ) for γ in $\boldsymbol{E}_i$.

Index function: For $e = (i, (v_1, \ldots, v_{\mathrm{a}(i)}))$ we define $\mathrm{i}_e(v_s)$ as s and $e[s]$ as v_s. When e is clear from context we will drop the e and just use $\mathrm{i}(v_s)$.

Definition 3.2.3. For an observation model $\mathrm{M} = ([q], T, \mathcal{T}, \{\mathbb{P}_\tau\}_{\tau \in T}, \Phi)$ the corresponding *null model* $\mathrm{M}^\times$ is the observation distribution where for every $\mathrm{a}(i)$-tuple $(v_1, \ldots, v_{\mathrm{a}(i)})$ a hidden coloring $\boldsymbol{c}$ is sampled independently, and the observation $\phi_i(v_1, \ldots, v_{\mathrm{a}(i)})$ is included with probability

$$\mathbf{Pr}\left[\phi_i(v_1, \ldots, v_{\mathrm{a}(i)}) \in \boldsymbol{E}_i\right] \stackrel{\text{def}}{=} \frac{\phi_i\left((\boldsymbol{\tau}(v_1), \boldsymbol{c}(v_1)), \ldots, (\boldsymbol{\tau}(v_{\mathrm{a}}(i)), \boldsymbol{c}(v_{\mathrm{a}(i)}))\right)}{n^{\mathrm{a}(i)-1}}.$$

Equivalently, in the null model for every $(v_1, \ldots, v_{\mathrm{a}(i)})$ the observation $\phi_i(v_1, \ldots, v_{\mathrm{a}(i)})$ is included independently with probability:

$$\mathbf{Pr}\left[\phi_i(v_1, \ldots, v_{\mathrm{a}(i)}) \in \boldsymbol{E}_i\right] \stackrel{\text{def}}{=} \frac{\overline{\phi_i}}{n^{\mathrm{a}(i)-1}}.$$

Remark 3.2.4. For a model M we will refer to it as the *planted model* and we will refer to $\mathcal{N}$ as the *null model*. Two computational problems we are interested in are distinguishing whether a sample is drawn from M or $\mathcal{N}$, and inferring the hidden coloring for a sample drawn from M.

3.2.2 Bayesian Inference

Given the variable types $\boldsymbol{\tau} : [n] \to T$ and the observations $\boldsymbol{E}$, the canonical algorithm to infer the hidden coloring $\boldsymbol{c}$ is to use the Bayes rule to compute the conditional distribution $\mathbb{P}_{\boldsymbol{c}|\boldsymbol{E}}$. Formally, the probability that a model $\mathrm{M} = ([q], T, \mathcal{T}, \{\mathbb{P}_\tau\}_{\tau \in T}, \Phi)$ generates a hidden coloring $\boldsymbol{c}$ and observations $\boldsymbol{E}$ is

$$\begin{aligned}
&\mathbf{Pr}[\boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{c} \mid \boldsymbol{\tau}] = \mathbf{Pr}[\boldsymbol{c} \mid \boldsymbol{\tau}] \cdot \mathbf{Pr}[\boldsymbol{E}_1, \ldots, \boldsymbol{E}_F \mid \boldsymbol{c}, \boldsymbol{\tau}] \\
&= \left(\prod_{v \in [n]} \mathbb{P}_{\boldsymbol{\tau}(v)}(\boldsymbol{c}(v))\right) \cdot \\
&\prod_{i \in [F]} \\
&\left(\prod_{(v_j)_j \in [n]^{\mathrm{a}(i)}} \left[\frac{\phi_i(\boldsymbol{c}(v_1), \ldots, \boldsymbol{c}(v_{\mathrm{a}(i)}))}{n^{\mathrm{a}(i)-1}}\right]^{\mathbf{1}_{(v_j)_j \in \boldsymbol{E}_i}} \left[1 - \frac{\phi_i(\boldsymbol{c}(v_1), \ldots, \boldsymbol{c}(v_{\mathrm{a}(i)}))}{n^{\mathrm{a}(i)-1}}\right]^{\mathbf{1}_{(v_j)_j \notin \boldsymbol{E}_i}}\right) .
\end{aligned}$$

By applying Bayes rule,

$$\mathbf{Pr}[\boldsymbol{c} \mid \boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{\tau}] = \frac{\mathbf{Pr}[\boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{c} \mid \boldsymbol{\tau}]}{\sum_{\boldsymbol{c}^*} \mathbf{Pr}[(\boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{c}^* \mid \boldsymbol{\tau}]} \ .$$

Ignoring the normalizing constant, we can write

$$\mathbf{Pr}[\boldsymbol{c} \mid \boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{\tau}] \propto e^{-H(c|\boldsymbol{E}_1,\ldots,\boldsymbol{E}_F,\boldsymbol{\tau})} \ ,$$

where

$$\begin{aligned}
&\mathbf{H}(\boldsymbol{c} \mid \boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{\tau}) = \\
&- \sum_{i \in [F]} \sum_{(v_j)_j \in [n]^{\mathrm{a}(i)}} \mathbf{1}_{(v_j)_j \in \boldsymbol{E}_i} \log \left(\frac{\phi_i(\boldsymbol{c}(v_1), \ldots, \boldsymbol{c}(v_{\mathrm{a}(i)}))}{n^{\mathrm{a}(i)-1}} \right) + \\
&\mathbf{1}_{(v_j)_j \notin \boldsymbol{E}_i} \log \left(1 - \frac{\phi_i(\boldsymbol{c}(v_1), \ldots, \boldsymbol{c}(v_{\mathrm{a}(i)}))}{n^{\mathrm{a}(i)-1}} \right) \\
&- \sum_{v \in [n]} \log \mathbb{P}_{\boldsymbol{\tau}(v)}(\boldsymbol{c}(v)).
\end{aligned}$$

The function $\mathbf{H}(\boldsymbol{c}|\boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{\tau})$ is referred to as the Hamiltonian, and the distribution is the Boltzmann distribution with Hamiltonian $\mathbf{H}$ and inverse temperature $\beta = 1$.

Since in our setting, the hypergraph is sparse, i.e., $\overline{\phi_i} = O(1)$, the terms

$$\log \left(1 - \frac{\phi_i(\boldsymbol{c}(v_1), \ldots, \boldsymbol{c}(v_{\mathrm{a}(i)}))}{n^{\mathrm{a}(i)-1}} \right) \approx 0 \tag{3.1}$$

for all $(v_j)_j \notin \boldsymbol{E}_i$. So the these terms can be dropped to simplify the Hamiltonian to

$$\mathbf{H}(\boldsymbol{c} \mid \boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{\tau}) = - \sum_{i \in [F]} \sum_{(v_j)_j \in \boldsymbol{E}_i} \log \Big(\phi_i(\boldsymbol{c}(v_1), \ldots, \boldsymbol{c}(v_{\mathrm{a}(i)})) \Big) - \tag{3.2}$$

$$\sum_{v \in [n]} \log \mathbb{P}_{\boldsymbol{\tau}(v)}(\boldsymbol{c}(v)). \tag{3.3}$$

The Hamiltonian $\mathbf{H}$ is a sum of *local* terms each depending on a constant number of variables. The observations $\boldsymbol{E}$ and the variables $\boldsymbol{c}$ together form what is termed

as *factor graphs* (see [MM09a]), where each observation is a *factor* of the Boltzmann distribution. Recall that the Boltzmann distribution is given by

$$\mathbf{Pr}[\boldsymbol{c} \mid (\boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{\tau})] = \frac{e^{-\mathbf{H}(\boldsymbol{c}|\boldsymbol{E}_1,\ldots,\boldsymbol{E}_F,\boldsymbol{\tau})}}{\sum_{\boldsymbol{c}^*} e^{-\mathbf{H}(\boldsymbol{c}^*|(\boldsymbol{E}_1,\ldots,\boldsymbol{E}_F,\boldsymbol{\tau})}}$$

The normalization term in the denominator is called the partition function of the distribution and is denoted $Z(\mathrm{M})$. Notice that a naive algorithm to infer the hidden coloring via the Bayes rule as described above would take exponential time.

3.2.3 Belief Propogation

The algorithm of choice to infer the hidden variables in a sparse factor model would be belief propogation. We refer the reader to [MM09a] for a detailed exposition of belief propogation, and restrict ourselves to a broad outline.

Belief propogation (BP) aims to estimate the marginals of the hidden variables, in our case $\boldsymbol{c}(v)$ for $v \in [n]$. BP draws its inspiration from a dynamic programming algorithm to compute the marginals when the underlying factor graph is a tree, and is broadly applicable to sparse settings where the local neighborhood of a vertex is tree-like. In particular, while BP computes the marginals exactly on a tree, it is very succesful in practice over sparse factor models that are locally tree-like.

To visualize BP, it will be useful to consider the bipartite graph $\mathcal{H}$ with variables $[n]$ on one side and the factors (a.k.a. observations) $\boldsymbol{E}$ on the other. There is an edge between a variable v and an observation $e \in \boldsymbol{E}$ if $v \in e$. The execution of BP is divided into rounds where in each round, the variable nodes send messages to factor nodes or vice versa.

Let $m^{v\to e}$ denote the message sent by a variable v to a factor $e \in \boldsymbol{E}$ and let $m^{e\to v}$ denote the message from a factor $e \in \boldsymbol{E}$ to a variable v. All messages exchanged are marginal distributions over the domain $[q]$, i.e., $m^{v\to e} = (m_1^{v\to e}, \ldots, m_q^{v\to e})$ and similarly $m^{e\to v} = (m_1^{e\to v}, \ldots, m_q^{e\to v})$. Intuitively speaking, $m_c^{v\to e}$ is an estimate of the marginal probability that v is assigned the color c when the factor e is absent, and $m_c^{e\to u}$ is an estimate of the marginal probability that u has color c when all other factors involving u are absent.

BP specifies an update rule for every variable/factor node to update its outgoing messages each round, depending on its incoming messages. Let ∂e denotes the set of variables incident a factor e and let ∂v denote the set of factors incident on

a variable v. BP specifies functions $\Upsilon_{v\to e}$, $\Upsilon_{e\to v}$ so that if $\{m^{v\to e}[t], m^{e\to v}[t]$ denote the messages in round t, then the updated messages are given by

$$m^{v\to e}[t+1] = \Upsilon_{v\to e}\left(\{m^{f\to v}[t] \mid f \in \partial v \backslash e\}\right) \tag{3.4}$$

$$m^{e\to v}[t+1] = \Upsilon_{e\to v}\left(\{m^{u\to e}[t] \mid u \in \partial e \backslash v\}\right) \tag{3.5}$$

We will describe the specific form of the functions Υ in Appendix 3.9, but there are two salient details that we would like to highlight at this time. First, the functions Υ are smooth rational functions that map marginals over $[q]$ to a marginal distribution over $[q]$. Second, the updated outgoing message $m^{v\to e}[t+1]$ depends on all messages incoming to variable v **except** the message $m^{e\to v}[t]$. Similarly, the updated outgoing message $m^{e\to v}[t+1]$ is independent of the incoming message $m^{v\to e}[t]$.

The general schema of a BP algorithm is to start BP with some intialization of the messages

$$\{m^{v\to e}[0], m^{e\to v}[0]\}_{v\in[n], e\in \boldsymbol{E}}$$

and iteratively update the messages as specified by the functions Υ, until the messages stabilize into a fixed point, i.e., a set of messages $\{\hat{m}^{v\to e}, \hat{m}^{e\to v}\}$ so that,

$$\hat{m}^{v\to e} = \Upsilon_{v\to e}\left(\{\hat{m}^{f\to v} \mid f \in \partial v \backslash e\}\right)$$

$$\hat{m}^{e\to v} = \Upsilon_{e\to v}\left(\{\hat{m}^{u\to e} \mid u \in \partial e \backslash v\}\right)$$

While it can often be difficult at times to show convergence to a fixed point, BP is very succesful in practice over locally tree-like factor models.

3.2.4 Stable Fixed Point Barrier

A natural starting point for BP iteration for a model M is given by the following:

$$\overline{m}^{v\to e} \stackrel{\text{def}}{=} \text{prior distribution } \mathbb{P}_{\boldsymbol{\tau}(v)} \tag{3.6}$$

$$\overline{m}^{e\to v} \stackrel{\text{def}}{=} \text{uniform distribution over support of } \mathbb{P}_{\boldsymbol{\tau}(v)} \tag{3.7}$$

Conjecturally, this canonical initialization $\overline{m}$ plays a critical role in characterizing the computational complexity of inferring the hidden variables in model M.

There appear to be three possible cases with regards to this canonical initialization.

Case 1: $\overline{m}$ is not a fixed point Suppose $\overline{m}$ is not a fixed point for the BP iteration over the model M, then BP iteration can be expected to make progress, thereby yielding a weak recovery of hidden variables.

In fact, we will present a self-contained algorithm that weakly-recovers the hidden coloring in this case. Formally, we will show the following in Appendix 3.10:

Lemma 3.2.5. *If $\overline{m}$ is not a fixed point for the BP iteration on model* M*, then there is a polynomial time algorithm $\mathcal{A}$ and an $\epsilon > 0$ such that*

1. *if $(\boldsymbol{E}, \boldsymbol{\tau}) \sim$ M: $\mathcal{A}$ outputs a coloring that beats the correlation random guessing achieves with the hidden coloring by ε,*
2. *$\mathcal{A}$ solves the* M *vs.* $\mathrm{M}^{\times}$ *distinguishing problem with high probability.*

In light of the above lemma, it is natural to restrict our attention to the case where $\overline{m}$ is a fixed point for the BP iteration. $\overline{m}$ being a fixed point of BP is equivalent to a "detailed balance" condition holding (in the sense of (3.47)).

Case 2: $\overline{m}$ is an unstable fixed point $\overline{m}$ is an *unstable fixed point* if arbitrary small perturbations of $\overline{m}$ will lead to the BP iteration moving away from the fixed point $\overline{m}$. BP is conjectured to succeed in weak-recovery of hidden coloring and distinguishing between M vs. $\mathrm{M}^{\times}$ in this case, and this has been extensively demonstrated experimentally [DKMZ11a, ZMZ14].

Case 3: $\overline{m}$ is a stable fixed point $\overline{m}$ is a *stable fixed point* if there exists a neighborhood U around $\overline{m}$ such that for any initialization $\hat{m} \in U$, BP iteration converges to the canonical fixed point $\overline{m}$. In this case, the canonical fixed point $\overline{m}$ clearly highlights a potential failure of BP algorithm. A priori, it is conceivable that by using BP with an alternative starting point or an entirely different algorithm, one could still efficiently infer the hidden coloring in this case.

Surprisingly, it is conjectured that the existence of this canonical fixed point that is stable marks the onset of computational intractability! Inspired by ideas from statistical physics, Krzakala and Zdeborova [KZ09] were the first to hypothesize that the existence of a trivial fixed point that is stable marks computational intractability. Building on these intuitions, Decelle et. al. [DKMZ11a] outlined a fascinating set of conjectures on community detection problem which fuelled a flurry of activity, resulting in algorithms matching the conjectured computational thresholds [MNS18, Mas14b, BLM15, AS15].

3.2.5 Analyzing Stability

The stability of the canonical fixed point $\overline{m}$ under BP iteration can be analyzed using derivatives of the BP update rule. Suppose Γ denote the map associated with running two rounds of BP iteration to produce the messages, i.e.,

$$\{m^{v\to e}[t+2]\}_{v\in[n],e\ni v} = \Gamma\left(\{m^{v\to e}[t]\}_{v\in[n],e\ni v}\right)$$

In other words, Γ is given by the composition of the functions in (3.4) and (3.5). If $\overline{m}$ is a fixed point of BP, then we will have,

$$\Gamma(\{\overline{m}^{v\to e}\}) = \{\overline{m}^{v\to e}\}$$

To analyze the stability of the fixed point $\overline{m}$, one uses the linear approximation of Γ in a neighborhood of $\overline{m}$, by setting

$$\Gamma(\overline{m}+\epsilon) = \overline{m} + B\epsilon$$

where B is the matrix of partial derivatives, i.e.,

$$B[m^{u\to e}, m^{u'\to e'}] = \frac{\partial\Gamma(m)^{u'\to e'}}{\partial m^{u\to e}}\,|_{\overline{m}}$$

With this linear approximation $\Gamma^{\ell}(\overline{m}+\epsilon) \approx \overline{m} + B^{\ell}\epsilon$. Therefore, the stability of the fixed point is characterized by the spectral radius of the operator B.

Specifically, $\overline{m}$ is a stable fixed point if and only if $\rho(B) \leqslant 1$ where $\rho(B) \stackrel{\text{def}}{=} \max_i |\lambda_i(B)|$ is the largest magnitude of an eigenvalue of B.

Notice that B is an asymmetric random matrix depending on the set of observations $\boldsymbol{E}$. The *cavity method* is a heuristic to guess the spectral radius of a typical derivative matrix B in terms of the spectral radius of some constant sized linear operator L. In the rest of the section we first use the cavity method to obtain a precise condition on M for $\rho(B) \leqslant 1$, then state our main theorem that the distinguishing problem and the weak recovery problem are efficiently solvable when $\rho(B) > 1$, and finally define the operator L whose spectral bound λ_L satisifies that $\rho(B) = \lambda_L^{1/2}$.

3.2.6 The local distributions of M

Before diving into the calculation, we define a few local distributions of M that would be used later.

The color assigment distribution μ_i For each factor $\phi_i \in \Phi$, define a local distribution μ_i over $[q]^{\mathrm{a}(i)}$ as,

$$\mu_i(c_1, \ldots, c_{\mathrm{a}(i)}) \propto \left(\prod_{j \in [\mathrm{a}(i)]} \mathbb{P}_{\mathrm{Cl}(i)_j}(c_j) \right) \cdot \phi_i(\boldsymbol{c}) \tag{3.8}$$

For each $\phi_i \in \Phi$ and $a, b \in [\mathrm{a}(i)]$ define a matrix $\boldsymbol{\Psi}_{i,a|b} \in \mathbb{R}^{[q] \times [q]}$ by fixing,

$$\boldsymbol{\Psi}_{i,a|b}(\alpha, \beta) \stackrel{\text{def}}{=} \underset{(c_1,\ldots,c_{\mathrm{a}(i)}) \sim \mu_i}{\mathbf{Pr}} [c_a = \alpha | c_b = \beta] \tag{3.9}$$

This says that conditioned on that ϕ_i is in the observations $\boldsymbol{E}$, the matrix $\boldsymbol{\Psi}_{i,a|b}$ encodes the color distribution of a conditioned on the color of b. Finally for $\phi_i \in \Phi$ and $a, b \in [\mathrm{a}(i)]$ we define a matrix that is useful later,

$$\overline{\mathbf{M}}_{i,a|b} = (\mathbf{I} - \mathbb{P}_{\mathrm{Cl}(i)_a} \mathbf{1}^T) \Psi_{i,a|b}. \tag{3.10}$$

The neighbor factor distribution of a variable We now take a closer look at a type τ variable's neighbor factor distribution. Here a variable's neighbor factors refer to all factors that are connected to the variable in the factor graph.

To study this neighborhood distribution, we first define random variables $\deg_{i,j}(\tau)$ for a type τ variable v.

Definition 3.2.6. For $\tau \in T, \phi_i \in \Phi$, $\deg_{i,j}(\tau)$ is the random variable denoting the number of type ϕ_i factors in the neighborhood of the type τ variable v such that the index of v in all these factor is j.

From the definition, we see that each $\deg_{i,j}(\tau)$ is the sum of many binomial variables each of which indicates whether a specific type ϕ_i factor exisits in the factor graph. We formally define these binomial variables.

Definition 3.2.7. For $\tau \in T, (v_1, \ldots, v_{\mathrm{a}(i)}) \in [n]^{\mathrm{a}(i)}$, $b_\tau^{v_1,\ldots,v_{\mathrm{a}(i)}}$ is the indicator variable of whether the type ϕ_i factor e whose j-th variable is v_j for all $j \in [\mathrm{a}(i)]$ is in the observations $\boldsymbol{E}$.

We can compute the probability of $b_\tau^{v_1,\ldots,v_{\mathrm{a}(i)}} = 1$ in M_n.

$$\underset{\mathrm{M}_n}{\mathbf{Pr}} \left[b_\tau^{v_1,\ldots,v_{\mathrm{a}(i)}} = 1 \right] = \frac{1}{\mathcal{T}(\tau)} \cdot \prod_{j=1}^{\mathrm{a}(i)} \mathcal{T}(\mathrm{Cl}(i)_j) \cdot$$

$$\sum_{(c_1,\ldots,c_{\mathrm{a}(i)})\in[q]^{\mathrm{a}(i)}} \prod_{j=1}^{\mathrm{a}(i)} \mathbb{P}_{\mathrm{Cl}(i)_j}(c_j) \cdot \frac{\phi_i(c_1,\ldots,c_{\mathrm{a}(i)})}{n^{\mathrm{a}(i)-1}}$$

$$= \frac{1}{\mathcal{T}(\tau)} \cdot \frac{\overline{\phi_i}}{n^{\mathrm{a}(i)-1}} \cdot \prod_{j=1}^{\mathrm{a}(i)} \mathcal{T}(\mathrm{Cl}(i)_j)$$

Thus $b_\tau^{v_1,\ldots,v_{\mathrm{a}(i)}}$ has distribution Binomial $\left(\frac{\overline{\phi_i}}{\mathcal{T}(\tau)n^{\mathrm{a}(i)-1}} \cdot \prod_{j=1}^{\mathrm{a}(i)} \mathcal{T}(\mathrm{Cl}(i)_j)\right)$.

Now we can express $\deg_{i,j}(\tau)$ as the sum of $n^{\mathrm{a}(i)-1}$ binomial random variables.

$$\deg_{i,j}(\tau) = \sum_{(v_j)_j\in[n]^{\mathrm{a}(i)}|v_j=v} b_\tau^{v_1,\ldots,v_{\mathrm{a}(i)}}.$$

We also note that most of the $b_\tau^{v_1,\ldots,v_{\mathrm{a}(i)}}$s are independent. Two random variables $b_\tau^{v_1,\ldots,v_{\mathrm{a}(i)}}$ and $b_\tau^{v'_1,\ldots,v'_{\mathrm{a}(i)}}$ are not independent only if there exist $j, j' \in [\mathrm{a}(i)]$ such that $v_j = v'_{j'}$ but $\mathrm{Cl}(i)_j \neq \mathrm{Cl}(i)_{j'}$. That is the two factors share some variable but require the variable to have different types. However, only $O\left(n^{-(\mathrm{a}(i)-1)}\right)$ fraction of the pairs are correlated. Thus, when n is large we can treat the $n^{\mathrm{a}(i)-1}$ random variables as being independent. Then each $\deg_{i,j}(\tau)$ has a Poisson distribution.

Claim 3.2.8. $\deg_{i,j}(\tau) \sim \mathrm{Poisson}\left(\frac{\overline{\phi_i}}{\mathcal{T}(\tau)} \cdot \prod_{j=1}^{\mathrm{a}(i)} \mathcal{T}(\mathrm{Cl}(i)_j)\right)$.

For similar reason as above, when n is large we can treat the random variables $\{\deg_{i,j}(\tau)\}_{i\in[F],j\in[\mathrm{a}(i)]}$ of a variable v as independent. Therefore for large n, the neighbor factor distribution of a type τ variable v is very close to the product distribution of the random variables $\{\deg_{i,j}(\tau)\}_{i\in F,j\in[\mathrm{a}(i)]}$.

3.2.7 The stability condition

Now we continue to explore the condition on M that makes $\overline{m}$ stable. We focus on sparse models whose average factor degrees $\overline{\phi_i} = O(1)$ for all $\phi_i \in \Phi$. In such models, a variable node is contained in constant number of factors, and its $o(\log n)$-neighborhood is locally tree-like with high probability. Set the tree depth ℓ be a function such that $\ell(n) \in o(\log n)$, and consider the distance $(2\ell+1)$ neighborhood of a variable v_0. Assume each level-$(2\ell+1)$ factor node e_ℓ's outgoing message to some level-2ℓ variable v_ℓ is perturbed to $m^{e_\ell\to v_\ell} = \overline{m}^{e_\ell\to v_\ell} + \epsilon_{e_\ell}$. Recall that

$\overline{m}^{e_\ell \to v_\ell}$ is the trivial fixed point message, and ϵ_{e_ℓ} is the random perturbation that is independent across different edges $e_\ell \to v_\ell$. We want to compute the expected influence of the perturbations on the messages to the root v_0.

We first consider the distance $2\ell + 1$ neighborhood of a variable v_0. The treelike neighborhood can be constructed by the following process.

1. Sample the type of the root variable $\boldsymbol{\tau}(v_0)$ from $\mathcal{T}$.
2. Sample the level-1 factors: for each type of factor $\phi_i \in \Phi$ sample the number of type ϕ_i neighbor factors of v_0 by sampling $\{\deg_{i,j}(\boldsymbol{\tau}(v_0))\}_{i \in F, j \in [\mathrm{a}(i)]}$ independently. Add v_0's neighbor factors to level 1. Add the other variables in these factors to the next level of the tree, assuming that there is no shared variables other then v_0. Note that these variables already have types.
3. Repeat step 2 for the new variables until we get a depth-$(2\ell + 1)$ tree $\mathbb{T}_\ell$.

We next use the tree $\mathbb{T}_\ell$ to give a precise condition on M for the fixed point $\overline{m}$ to be stable.

In a $\mathbb{T}_\ell$, a leaf node e_ℓ is connected to the root node v_0 via a path $e_\ell, v_\ell, \ldots, e_0, v_0$. A perturbation on the leaf message $m^{e_\ell \to v_\ell}$ influence the next level message $m^{e_{\ell-1} \to v_{\ell-1}}$ via the partial deriviative matrix $\frac{\partial \Gamma(m)^{e_{\ell-1} \to v_{\ell-1}}}{\partial m^{e_\ell \to v_\ell}}$. We can express the partial deriviative matrix evaluated at the fixed point using the matrix defined in (3.10).

Claim 3.2.9 (Claim 3.4.1)**.**

$$\frac{\partial \Gamma(m)^{e_{\ell-1} \to v_{\ell-1}}}{\partial m^{e_\ell \to v_\ell}} \Big|_{\overline{m}} = \overline{\mathbf{M}}_{\theta(e_{\ell-1}), \mathrm{i}_{e_{\ell-1}}(v_{\ell-1}) | \mathrm{i}_{e_{\ell-1}}(v_\ell)}.$$

This claim is proved in Appendix 3.11. When writing the matrix $\overline{\mathbf{M}}_{\theta(e), \mathrm{i}_e(v) | \mathrm{i}_e(v')}$, it's clear that the index function is associated with the factor e, so we drop the subscript e in i_e for simplicity.

Using the chain rule, we can compose the partial deriviative matrices along a path, and conclude that each path influences the root message by

$$\left(\prod_{j=1}^{\ell} \overline{\mathbf{M}}_{\theta(e_{j-1}), \mathrm{i}(v_{j-1}) | \mathrm{i}(v_j)} \right) \epsilon_{e_\ell}.$$

Thus the influence of all paths in $\mathbb{T}_\ell$ is

$$\sum_{(e_\ell, v_\ell, \ldots, e_0, v_0) \in \mathbb{T}_\ell} \left(\prod_{j=1}^{\ell} \overline{\mathbf{M}}_{\theta(e_{j-1}), \mathrm{i}(v_{j-1}) | \mathrm{i}(v_j)} \right) \epsilon_{e_\ell}$$

To decide if the fixed point is stable, we compute the variance α_ℓ of this influence.

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\mathbb{T}_\ell,\epsilon}\left[\left\|\sum_{(e_\ell,v_\ell,\ldots,e_0,v_0)\in\mathbb{T}_\ell}\left(\prod_{j=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)\epsilon_{e_\ell}\right\|^2\right]\\
&=\mathop{\mathbf{E}}_{\mathbb{T}_\ell,\epsilon}\left[\sum_{(e_\ell,v_\ell,\ldots,e_0,v_0)\in\mathbb{T}_\ell}\epsilon_{e_\ell}^\top\left(\prod_{j=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)^*\left(\prod_{j=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)\epsilon_{e_\ell}\right]\\
&=\mathop{\mathbf{E}}_{\mathbb{T}_\ell}\left[\sum_{(e_\ell,v_\ell,\ldots,e_0,v_0)\in\mathbb{T}_\ell}\mathsf{Tr}\left(\left(\prod_{j=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)\left(\prod_{j=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)^*\right)\right]\cdot\\
&\quad\frac{\mathbf{E}_{\epsilon_{e_\ell}}[\|\epsilon_{e_\ell}\|^2]}{q}
\end{aligned}$$

So the squared norm of the perturbations ϵ_{e_ℓ} is amplified by

$$\alpha_\ell \stackrel{\text{def}}{=} \mathop{\mathbf{E}}_{\mathbb{T}_\ell}\sum_{(e_\ell,v_\ell,\ldots,e_0,v_0)\in\mathbb{T}_\ell}\mathsf{Tr}\left(\left(\prod_{j=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)\left(\prod_{j=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)^*\right).$$

We note that for a model M, by definition of α_ℓ and the operator B in Section 3.2.5, $\rho(B) = \lim_{n\to\infty}\alpha_\ell^{1/2\ell}$ (recall that ℓ is a function of n). Thus when $\lim_{n\to\infty}\alpha_\ell^{1/2\ell}\leqslant 1$, $\overline{m}$ is a stable fixed point.

3.2.8 Efficient recovery and detection when the fixed point is unstable

For a model M whose fixed point $\overline{m}$ is unstable, it is conjectured that the BP algorithm can successfully weak-recover the hidden coloring. We provide a BP-inspired spectral algorithm that solves the weak recovery problem in this regime. We state the result somewhat informally below; the full formal statement can be found in Theorem 3.8.1. However, before we state the result we go on a small digression on how to set the benchmark for weak recovery. A first attempt might be:

> For a fixed type $\tau \in T$ and color $c \in [q]$, produce a vector $\overline{u} \in \mathbb{R}^n$ such that w correlates with the following "centered indicator vector" of (τ, c): $\underline{\chi}^{\tau,c}$ where $\underline{\chi}^{\tau,\alpha}$ is an n-dimensional vector with i-th coordinate $\mathbf{1}[\tau(i) = \tau] \cdot (\mathbf{1}[\boldsymbol{c}(i) = c] - \mathbb{P}_\tau(c))$.

However, this benchmark is unattainable since for a problem such as PLANTED-q-COLORING there is no way to statistically discriminate between a given coloring and a different coloring obtained by permuting the names of the colors. Thus, to account for this complication we consider the following modification of the above benchmark, which we first state in words.

> Produce a vector $\overline{u} \in \mathbb{R}^n$ such that after some permutation is applied to the names of the colors, for some type $\tau \in T$ and color $c \in [q]$, $\overline{u}$ correlates with the centered indicator vector of (τ, c).

More formally:

Theorem 3.2.10. *If a model* M *has* $\lim_{\ell\to\infty} \alpha_\ell^{1/2\ell} > 1$*, there is a spectral algorithm* A *that solves the weak-recovery problem. A bit more concretely, for* $\boldsymbol{G} \sim \mathrm{M}_n$*, the algorithm* $\mathsf{A}(\boldsymbol{G})$ *produces* $O_{\mathrm{M}}(1)$ *vectors* $\{\overline{u}_1, \ldots, \overline{u}_r\}$ *such that one of these vectors* $\overline{u}_j$ *has constant correlation with the planted coloring in the following sense:*

> *There is a type* $\tau \in T$*, and a color* $\alpha \in [q]$ *such that if we construct* $\underline{\chi}^{\tau,\alpha} \in \mathbb{R}^n$ *as*
>
> $$\underline{\chi}^{\tau,\alpha}[i] = \mathbf{1}[\tau(i) = \tau](\mathbf{1}[\boldsymbol{c}(i) = \alpha] - \mathbb{P}_\tau(\alpha))$$
>
> *then*
>
> $$\langle \overline{u}_j, \underline{\chi}^{\tau,\alpha} \rangle \geqslant \Omega_{\mathrm{M}}(1) \cdot \sqrt{n}.$$

For a model M with $\overline{m}$ as an unstable fixed point, it is conjectured that the BP algorithm can successfully distinguish it from the null model $\mathcal{N}$. We provide a BP-inspired spectral algorithm that solves the detection problem in this regime. We state the result below but leave the proof sketch to the technical overview section.

Theorem 3.2.11. *If a planted model* M *has* $\lim_{\ell\to\infty} \alpha_\ell^{1/2\ell} > 1$*, there is a spectral algorithm* A : *factor graphs* $\to \{\mathrm{P}, \mathrm{N}\}$ *that solves the detection problem in the following sense*

$$\underset{\mathrm{M}_n}{\mathbf{Pr}}[\mathsf{A}(\textit{factor graph}) = \mathrm{P}] = 1 - o_n(1) \quad \textit{and} \quad \underset{\mathcal{N}_n}{\mathbf{Pr}}[\mathsf{A}(\textit{factor graph}) = \mathrm{N}] = 1 - o_n(1).$$

3.2.9 The stability condition via a finite linear operator L

In this part we define a finite linear operator L whose spectral radius gives a criterion for when $\lim_{\ell\to\infty} \alpha_\ell^{1/2\ell}$ is greater than 1 or less than 1.

Naively, computing $\lim_{\ell\to\infty} \alpha_\ell^{1/2\ell}$ requires us to consider trees whose size grows with ℓ. However we can simplify the expression for $\alpha_\ell^{1/2\ell}$ via an insight of [BC19] by observing that the tree $\mathbb{T}_\ell$ is constructed recursively. For any even level variable node v_k in the tree, the distribution of its children factor nodes depends only on v_k's type. Furthermore, the factor node distribution $\{\deg_{i,j}(\boldsymbol{\tau}(v_k))\}_{i\in F, j\in[\mathrm{a}(i)]}$ also fully describes the distance 2 neighborhood of v_k. For a type τ variable v, define the random variable $\mathrm{num}_\tau(i,j,j')$ to be the number of variables u that are connected to v via some type ϕ_i factor, and additionally u have index j' in the factor and v has index j. These random variables give a way to concisely express the total influence of the distance 2 type τ' variables to the type τ variable. This influence is

$$\sum_{i,j,j'|\mathrm{Cl}(i)_{j'}=\tau'} \mathrm{num}_\tau(i,j,j')\cdot \overline{\mathbf{M}}_{i,j|j'} \in \mathbb{R}^{[q]\times[q]}.$$

Then we can build a 2 step quadratic influence operator $L : \mathbb{R}^{T\cdot[q]\times T\cdot[q]} \to \mathbb{R}^{T\cdot[q]\times T\cdot[q]}$ such that the (τ,τ) block of $L(M)$ is

$$L(M)^{\tau,\tau} = \sum_{\tau'} \sum_{i\in F, j,j'\in[\mathrm{a}(i)]} \sum_{i,j,j'|\mathrm{Cl}(i)_{j'}=\tau'} \mathrm{num}_\tau(i,j,j')\cdot \overline{\mathbf{M}}_{i,j|j'}\, M^{\tau',\tau'}\, \overline{\mathbf{M}}^*_{i,j|j'},$$

and the off-diagonal blocks are $L(M)^{\tau,\tau'} = \mathbf{0}$.

Suppose the input M is such that each $M^{\tau',\tau'}$ captures the quadratic influence of some path $v_1, e_1, \ldots, v_\ell, e_\ell$ such that the endpoint v_1 has type τ',

$$\left(\prod_{j=2}^{\ell} \overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)\left(\prod_{j=2}^{\ell} \overline{\mathbf{M}}_{\theta(e_{j-1}),\mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)^*.$$

And all other blocks in M are 0. Then after applying the operator, every diagonal block $L(M)^{\tau,\tau}$ captures the expected quadratic influence of all paths $v_0, e_0, \ldots, v_\ell, e_\ell$ which are 2-step extensions of the path $v_1, e_1, \ldots, v_\ell, e_\ell$ and whose endpoint v_0 has type τ.

Using this operator L we can rewrite α_ℓ as

$$\begin{aligned}\alpha_\ell &= \mathop{\mathbf{E}}_{\mathbb{T}_\ell} \sum_{(e_\ell, v_\ell, \ldots, e_0, v_0) \in \mathbb{T}_\ell} \mathsf{Tr}\left(\left(\prod_{j=1}^{\ell} \overline{\mathbf{M}}_{\theta(e_{j-1}), \mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)\left(\prod_{j=1}^{\ell} \overline{\mathbf{M}}_{\theta(e_{j-1}), \mathrm{i}(v_{j-1})|\mathrm{i}(v_j)}\right)^*\right) \\ &= \mathsf{Tr}\left(L^\ell(\mathrm{Diag}(\mathbb{P}))\right),\end{aligned}$$

where $\mathrm{Diag}(\mathbb{P}) \in \mathbb{R}^{T\cdot[q]\times T\cdot[q]}$ is a diagonal matrix whose $((\tau, c), (\tau, c))$ entry has value $\mathbb{P}_\tau(c)$.

Use λ_L to denote the maximum eigenvalue of L. Then by standard linear algebra fact,

$$\lim_{n\to\infty} \mathsf{Tr}\left(L^\ell(\mathrm{Diag}(\mathbb{P}))\right) \to \lambda_L^\ell.$$

Therefore we obtain the following equivalence relation between α_ℓ and λ_L.

Lemma 3.2.12. *For a model* M, $\lim_{\ell\to\infty} \alpha_\ell^{1/2\ell} = \lambda_L^{1/2}$. *Thus the fixed point* $\overline{m}$ *of* M *is stable if and only if* $\lambda_L \leqslant 1$.

3.3 Technical Overview

3.3.1 Algorithm for distinguishing

We now describe our algorithm for distinguishing if an instance $\boldsymbol{G}$ was sampled from the null distribution $\mathcal{N}$ from the planted distribution $\mathcal{P}$. Our algorithm constructs a matrix $M_{\boldsymbol{G}}$ obtained from linearizing ℓ rounds of the belief propagation algorithm at the uninformative fixed point on input $\boldsymbol{G}$ and tests if its largest eigenvalue exceeds a chosen threshold κ. If it does then the algorithm declares that $\boldsymbol{G}$ came from the planted distribution, and otherwise claims $\boldsymbol{G}$ was sampled from the null distribution. A bulk of the technical work is in proving that this particular matrix $M_{\boldsymbol{G}}$ has all its eigenvalues bounded by the chosen threshold κ when $\boldsymbol{G}$ is sampled from the null model, and in illustrating that $M_{\boldsymbol{G}}$ has an "outlier" eigenvalue exceeding κ otherwise. In this section, we delve more into the description of $M_{\boldsymbol{G}}$ and then give a brief description of how we prove the statements about the eigenvalues of $M_{\boldsymbol{G}}$ in the null and planted models.

More concretely, given a random instance $\boldsymbol{G}$ sampled either from $\mathcal{N}$ or $\mathcal{P}$, we set $M_{\boldsymbol{G}}$ as the following matrix $\underline{A}_{\boldsymbol{G}}^{(\ell)}$ called the *length-ℓ centered nonbacktracking walk power* of $\boldsymbol{G}$, for which we provide a slightly informal description below.

Definition 3.3.1 (Centered nonbacktracking power (slightly informal)). $\underline{A}_{\boldsymbol{G}}^{(\ell)}$ is a $nq \times nq$ matrix which we treat as a $n \times n$ grid of $q \times q$ blocks. The block rows and columns are indexed by $[n]$. In the (i,j)-th block, we place the following $q \times q$ matrix:

$$\sum_{\substack{ie_1v_1e_2v_2\ldots e_\ell j \in \\ \text{all nonbacktracking walks} \\ \text{from } i \text{ to } j \text{ in complete} \\ \text{factor graph}}} \overline{\mathbf{M}}_{e_1,i|v_1} \cdot \overline{\mathbf{M}}_{e_2,v_1|v_2} \cdots \overline{\mathbf{M}}_{e_\ell,v_{\ell-1}|j} \cdot$$
$$(\mathbf{1}[e_1 \in \boldsymbol{G}] - \Pr_{\mathcal{N}}[e_1 \in \boldsymbol{G}]) \cdots (\mathbf{1}[e_\ell \in \boldsymbol{G}] - \Pr_{\mathcal{N}}[e_\ell \in \boldsymbol{G}]).$$

Recall the matrix L from Section 3.2.8 which the stability prediction of belief propagation was based on, and let λ_L denote its largest eigenvalue. The two main technical theorems we prove about $\underline{A}_{\boldsymbol{G}}^{(\ell)}$ in service of proving that our algorithm is correct with high probability are:

Theorem 3.3.2 (Local statistics in planted model). *When $\boldsymbol{G} \sim \mathcal{P}$, with probability $1 - o_n(1)$: $\lambda_{\max}(\underline{A}_{\boldsymbol{G}}^{(\ell)}) \geqslant \frac{\lambda_L^\ell}{q}$.*

The proof of this is carried out in Section 3.6 and uses two ingredients: the first is recognizing that $\underline{A}_{\boldsymbol{G}}^{(\ell)}$ is self-adjoint under a certain inner product $\langle \cdot, \cdot \rangle_{\mathbf{H}}$, due to which for any vector x:

$$\langle x, \underline{A}_{\boldsymbol{G}}^{(\ell)} x \rangle_{\mathbf{H}} \leqslant \lambda_{\max}\left(\underline{A}_{\boldsymbol{G}}^{(\ell)}\right) \cdot \langle x, x \rangle_{\mathbf{H}}.$$

The second ingredient is in identifying a vector x depending on the planted solution the instance $\boldsymbol{G}$ was sampled with which makes the above quadratic $\langle x, \underline{A}_{\boldsymbol{G}}^{(\ell)} x \rangle_{\mathbf{H}}$ larger than the desired lower bound of $\frac{\lambda_L^\ell}{q}$ with high probability.

The second main technical theorem, which is proved in Theorem 3.3.3 is:

Theorem 3.3.3 (Eigenvalue bound in null model). *When $\boldsymbol{G} \sim \mathcal{N}$ and $(\log\log n)^2 \leqslant \ell \leqslant \frac{\log n}{(\log\log n)^2}$, for every constant $\varepsilon > 0$ with probability $1 - o_n(1)$, all eigenvalues of $\underline{A}_{\boldsymbol{G}}^{(\ell)}$ are bounded in magnitude by $((1+\varepsilon)\sqrt{\lambda_L})^\ell$.*

When $\lambda_L > 1$, we choose $\delta > 0$ so that $1 + \delta < \sqrt{\lambda_L}$, ℓ as $(\log\log n)^2$, and κ as $((1+\delta)\sqrt{\lambda_L})^\ell$. Then as an immediate consequence of Theorem 3.3.2 and

Theorem 3.3.3 we know that the algorithm correctly distinguishes between $\mathcal{N}$ and $\mathcal{P}$ with high probability.

We now elaborate on Theorem 3.3.3 and elucidate the exact random matrix concentration statement.

3.3.2 Matrix concentration vignette

Consider an Erdős-Rényi graph $\boldsymbol{H}$ sampled from $G\left(n, \frac{d}{n}\right)$. Pick a random vertex v in $\boldsymbol{H}$ and observe a "large" radius neighborhood around v. Typically, this neighborhood around v will be a tree, and additionally, and in the large-n limit the distribution of this tree is a *Galton-Watson process* – a random (possibly infinite) tree $\mathbb{T}$ generated by starting at a root vertex r, attaching Poisson(d) children to r, and then attaching Poisson(d) children to each child of r and so on.[2] So this tells us that there is some sense in which $\mathbb{T}$ "approximates" the finite random graph. This intuition is spectrally articulated by a theorem which is (implicitly) due to [BLM15] (see also [FM17] and [BMR21]). Before we state the theorem, we bring up a natural quantity to associate to the random tree: the *growth rate* which is defined as

$$\lim_{\ell \to \infty} \mathbf{E}[\#\text{ of vertices at depth-}\ell]^{1/\ell},$$

which is equal to d for the aforementioned Galton-Watson process.

Theorem 3.3.4 ([BLM15, FM17, BMR21]). *Let $\underline{A}_{\boldsymbol{H}} := A_{\boldsymbol{H}} - \mathbf{E}\, A_{\boldsymbol{H}}$ be the centered adjacency matrix of $\boldsymbol{H}$. Suppose $d > 1$, then:*

$$|\lambda|_{\max}(\underline{A}_{\boldsymbol{H}}^{(\ell)}) \leqslant ((1+o(1))\sqrt{d})^{\ell} = ((1+o(1))\sqrt{\text{growth rate of } \mathbb{T}})^{\ell}$$

for $\ell \in \left[(\log\log n)^2, \frac{\log n}{\log\log n}\right]$.

Now, let's add a small twist: sample two Erdős-Rényi graphs $\boldsymbol{H}_1 \sim G\left(n, \frac{d_1}{n}\right)$ and $\boldsymbol{H}_2 \sim G\left(n, \frac{d_2}{n}\right)$ and consider the weighted graph $\boldsymbol{H} = 0.9\boldsymbol{H}_1 - \boldsymbol{H}_2$. The random tree that $\boldsymbol{H}$ locally resembles is the following *different* Galton-Watson process $\mathbb{T}'$: start at a root vertex r, connect Poisson(d_1) children with edges of

[2] The reader is advised to not pay too much attention to the fact that the number of children are distributed according to a *Poisson* random variable. The important property is that a vertex has d children on average.

weight 0.9 and Poisson(d_2) children with edges of weight -1 to r, then repeat the same for each child vertex, and keep going. The following quantity is the correct generalization of growth rate to weighted graphs, which we call the *weighted growth rate* of the tree:

$$\mathrm{wgr}(\mathbb{T}') := \lim_{\ell \to \infty} \mathbf{E} \left[\sum_{\substack{P \in \text{length-}\ell \text{ paths} \\ \text{starting at root}}} \prod_{e \in P} w_e^2 \right].$$

For illustrative purposes, one subcase of our matrix concentration result is:

Theorem 3.3.5. *Let $\underline{A}_{\boldsymbol{H}} := A_{\boldsymbol{H}} - \mathbf{E}\, A_{\boldsymbol{H}}$ be the centered adjacency matrix of $\boldsymbol{H}$. Suppose* $\mathrm{wgr}(\mathbb{T}') > 1$*, then:*

$$|\lambda|_{\max}(\underline{A}_{\boldsymbol{H}}^{(\ell)}) \leqslant ((1 + o(1)) \sqrt{\mathrm{wgr}(\mathbb{T}')})^{\ell}$$

for $\ell \in \left[(\log \log n)^2, \frac{\log n}{(\log \log n)^2} \right]$

We now discuss our full matrix concentration theorem which captures both of the above mentioned theorems. Before doing so, it is worth noting that the picture for random graphs being spectrally approximated by infinite graphs is far more well understood in the setting of models of random *regular* graphs through works of [Fri03b, Bor19, BC19, MOP20, OW20] but we defer the readers to [OW20] for an extensive discussion of what is known in that setting.

3.3.3 Matrix concentration statement

Let $\langle \cdot, \cdot \rangle_\nu$ be an inner product on $\mathbb{R}^{nq}$ and let M^* denote the adjoint of a matrix M under this inner product. We consider $nq \times nq$ random matrices sampled according to the following model (whose notation the reader should treat independently from the preceding notation related to distinguishing instances from the null and planted distributions).

Definition 3.3.6 (Random matrix model (slightly informal))**.** The model has an underlying *left vertex set* which is equal to $[n]$. First, every vertex v is assigned a *type* $\boldsymbol{\tau}(v)$ in $[T]$ sampled according to a distribution π. There are F types of right vertices, given by set $[F]$. Each right vertex type i comes with an *arity* k_i, which is a positive integer, a *profile* χ_i which is a tuple in $[T]^{k_i}$, a collection of $k_i(k_i - 1)$

matrices $\{M_{i,(a,b)}\}_{(a,b)\in[k_i]^2:a\neq b}$, and a *density* ϕ_i. A random instance $\boldsymbol{H}$ is sampled in the following way: for every $(i,(v_1,\ldots,v_{k_i}))$ for $i\in[F]$ and tuple $(v_1,\ldots,v_{k_i})$ in $[n]^{k_i}$ of distinct elements such that $(\boldsymbol{\tau}(v_1),\ldots,\boldsymbol{\tau}(v_{k_i}))=\chi_i$ we add $(i,(v_1,\ldots,v_{k_i}))$ as a right vertex with probability $\frac{\phi_i}{n^{k_i-1}}$, connect edges to $v_1,\ldots,v_{k_i}$ and mark the edge to v_t with number t. We use $\mathcal{K}_n$ to refer to the bipartite graph with left vertex set $[n]$, and the right vertex set containing every potential right vertex. Now let $\gamma=(i,(v_1,\ldots,v_{k_i}))$; for a two-step $v_a\to\gamma\to v_b$ in the complete graph for $a\neq b$ we use $M_{v_a\gamma v_b}$ to denote the matrix $M_{i,(a,b)}$. The random matrix we are interested in, which we denote $\underline{A}_{\boldsymbol{H}}^{(\ell)}$, is the matrix where the uv entry contains:

$$\underline{A}_{\boldsymbol{H}}^{(\ell)}[i,j] \coloneqq \sum_{\substack{i\gamma_1v_1\ldots\gamma_\ell j\\ \text{nonbacktracking walks in } \mathcal{K}}} M_{i\gamma_1v_1}\cdots M_{v_{\ell-1}\gamma_\ell j}\cdot(\mathbf{1}[\gamma_1\in\boldsymbol{H}]-\mathbf{Pr}[\gamma_1\in\boldsymbol{H}])\cdots(\mathbf{1}[\gamma_\ell\in\boldsymbol{H}]-\mathbf{Pr}[\gamma_\ell\in\boldsymbol{H}]).$$

Definition 3.3.7 (Galton-Watson tree approximating random matrix (informal))**.** For a given setting of parameters for the random model from Definition 3.3.6, the bipartite Galton-Watson tree $\mathbb{T}$ which "locally resembles" an instance $\boldsymbol{H}$ sampled from the model is as follows:

1. Start with a left root vertex r and assign it type $\boldsymbol{\tau}(r)\sim\pi$.
2. For each $i\in[F]$ and each $j\in[k_i]$ such that $(\chi_i)_j=\boldsymbol{\tau}(r)$ sample
$$n_{i,j}\sim\text{Poisson}\left(\frac{\phi_i}{\pi_{\boldsymbol{\tau}(v)}}\prod_{t=1}^{k_i}\pi_{(\chi_i)_t}\right),$$
and attach $n_{i,j}$ right vertices of type i to r and mark the corresponding edge with j. Then to each such right vertex, attach k_i-1 (left vertex) children and mark the edges with numbers from $[k_i]\setminus\{j\}$. To each added child vertex v with edge marked with t, assign it type $(\chi_i)_t$.
3. Repeat step 2 for each added left vertex child.

We define the *matrix weighted growth rate* of $\mathbb{T}$ to be the following:

$$\text{mwgr}(\mathbb{T})\coloneqq\lim_{\ell\to\infty}$$

$$\mathbf{E}\left[\mathsf{Tr}\left(\sum_{\substack{a\gamma_1 v_1 \ldots \gamma_\ell b \\ \text{nonbacktracking walks in } \mathcal{K}}} M_{a\gamma_1 v_1} \ldots M_{v_{\ell-1}\gamma_\ell b} M_{b\gamma_\ell v_{\ell-1}} \ldots M_{v_1\gamma_1 a}\right)\right]^{1/\ell}.$$

We prove:

Theorem 3.3.8 (Main matrix concentration theorem (slightly informal))**.** *Let $\boldsymbol{H}$ be a random instance of a setting of parameters for the model from Definition 3.3.6 and let $\mathbb{T}$ be the tree which locally approximates $\boldsymbol{H}$ in the sense of Definition 3.3.7. Suppose:*

1. *For every right vertex $\gamma = (i, (v_1, \ldots, v_{k_i}))$ in $\mathcal{K}_n$ and every $1 \leqslant a, b \leqslant k_i$ for distinct a, b the $nq \times nq$ matrix obtained by placing $M_{v_a \gamma v_b}$ in the (v_a, v_b) block and zeros everywhere else is adjoint to the $nq \times nq$ matrix obtained by placing $M_{v_b \gamma v_a}$ in the (v_b, v_a) block and zeros everywhere else under $\langle \cdot, \cdot \rangle_\nu$.*

2. *There is a constant C such that for any nonbacktracking walk $v_0 \gamma_1 v_1 \ldots \gamma_s v_s$ in $\mathcal{K}_n$:*

$$\|M_{v_0 \gamma_1 v_1} \ldots M_{v_{s-1} \gamma_s v_s}\| \leqslant C.$$

3. $\mathrm{mwgr}(\mathbb{T}) \geqslant 1$.

Then if $(\log \log n)^2 \leqslant \ell \leqslant \frac{\log n}{(\log \log n)^2}$, with probability $1 - o_n(1)$:

$$|\lambda|_{\max}(\underline{A}_{\boldsymbol{H}})^{(\ell)} \leqslant ((1 + o(1))\sqrt{\mathrm{mwgr}(\mathbb{T})})^\ell.$$

The full formal set-up for Theorem 3.3.8 along with its proof is in Section 3.7.

3.4 A conjectured detection/recovery threshold

In Section 3.2.9 we see the connection between λ_L and α_ℓ of a model M. In this section we prove this connection more rigorously, and conclude with a conjectured weak-recovery threshold in terms of λ_L. We start by quickly going through the definitions of the partial derivative matrices $\overline{\mathbf{M}}_{\theta(e_j), \mathbf{i}_{e_j}(v_j) | \mathbf{i}_{e_j}(v_{j+1})}$, the color distribution matrix $\mathbf{D}_\tau$, the influence variance α_ℓ, and their connections.

Claim 3.4.1. The following matrices satisfy:

1. $\overline{\mathbf{M}}_{\theta(e_j),\mathbf{i}_{e_j}(v_j)|\mathbf{i}_{e_j}(v_{j+1})} = \left(\mathbf{I} - \mathbb{P}_{\boldsymbol{\tau}(v_j)}\mathbf{1}^\top\right)\boldsymbol{\Psi}_{\theta(e_j),\mathbf{i}_{e_j}(v_j)|\mathbf{i}_{e_j}(v_{j+1})}$.
2. Let $\mathbf{D}_\tau := \mathrm{Diag}(\mathbb{P}_\tau)$. Then $\mathbf{D}^\dagger_{\boldsymbol{\tau}(v_j)}\overline{\mathbf{M}}_{\theta(e_j),\mathbf{i}(v_j)|\mathbf{i}(v_{j+1})} = \overline{\mathbf{M}}^\top_{\theta(e_j),\mathbf{i}(v_{j+1})|\mathbf{i}(v_j)}\mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})}$ where the † in the superscript denotes the pseudoinverse of the matrix.

This claim is proved in Appendix 3.11. From here on we drop the subscript e from the index function $\mathbf{i}_e$ whenever it is clear from the context which factor is being considered.

Using these notations we can express the influence variance, or the *amplification factor*, α_ℓ as follows:

$$\alpha_\ell := \mathop{\mathbf{E}}_{\mathbb{T}}\left[\mathsf{Tr}\left(\sum_{(e_\ell,v_\ell,\ldots,e_0,v_0)\in\mathbb{T}}\left(\prod_{i=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{i-1}),\mathbf{i}(v_{i-1})|\mathbf{i}(v_i)}\right)\left(\prod_{i=1}^{\ell}\overline{\mathbf{M}}_{\theta(e_{i-1}),\mathbf{i}(v_{i-1})|\mathbf{i}(v_i)}\right)^*\right)\right].$$

The key property of the above amplification factor we are interested in is its limiting behavior as k goes to infinity. We say that the uninformative fixed point of the belief propagation update rule is *stable* if $\lim_{n\to\infty}\alpha_\ell^{1/2\ell} \leqslant 1$, and *unstable* otherwise. Furthermore, if the fixed point is stable, the problem of weak-recovering the hidden coloring of M is conjectured to be hard, and if the fixed point is unstable, this problem is conjectured to be easy.

In the remainder of this section, we focus on obtaining a simpler criterion for stability by simplifying the expression for the amplification factor. In particular, we give a constant dimensional linear transformation whose top eigenvalue is greater than 1 if the fixed point is stable and is less than 1 if the fixed point is unstable. For a leaf-to-root path $e_\ell v_\ell e_{\ell-1}v_{\ell-1}\ldots e_0v_0$ we say its type is

$$\theta_\ell \to \mathrm{out}_\ell \to \tau_\ell \to \cdots \to \mathrm{in}_0 \to \theta_0 \to \mathrm{out}_0 \to \tau_0$$

where θ_t is the factor type of e_t, τ_t is the variable type of v_t, out_t is the index of v_t in e_t, and in_t is the index of v_{t+1} in e_t. The amplification factor can then be written as:

$$\alpha_\ell = \mathsf{Tr}\left(\sum_{\theta_\ell\to\cdots\to\tau_0} M(\theta_\ell \to \cdots \to \tau_0)\right)$$

where the sum is enumerated over all leaf-to-root path types and $M(\theta_\ell \to \cdots \to \tau_0)$ is defined as follows:

$$M(\theta_\ell \to \cdots \to \tau_0) = \mathcal{T}(\tau_0).$$

$$\left(\prod_{t=0}^{\ell-1}\frac{\phi_{\theta_t}}{\mathcal{T}(\tau_t)}\cdot\prod_{s=1}^{\mathrm{a}(\theta_t)}\mathcal{T}(\mathrm{Cl}(\theta_t)_s)\cdot\mathbf{1}[\mathrm{Cl}(\theta_t)_{\mathrm{out}_t}=\tau_t]\cdot\mathbf{1}[\mathrm{Cl}(\theta_t)_{\mathrm{in}_t}=\tau_{t+1}]\right)\cdot$$
$$\left(\prod_{t=0}^{\ell-1}\overline{\mathbf{M}}_{\theta_t,\mathrm{out}_t|\mathrm{in}_t}\right)\left(\prod_{t=0}^{\ell-1}\overline{\mathbf{M}}_{\theta_t,\mathrm{out}_t|\mathrm{in}_t}\right)^{*}\cdot\frac{\phi_{\theta_\ell}}{\mathcal{T}(\tau_\ell)}\prod_{s=1}^{\mathrm{a}(i_\ell)}\mathcal{T}(\mathrm{Cl}(\theta_\ell)_s)\cdot\mathbf{1}[\mathrm{Cl}(\theta_\ell)_{\mathrm{out}_\ell}=\tau_\ell].$$

Now, let's define $V_{\ell,\tau}$ as:

$$V_{\ell,\tau}\coloneqq\sum_{\substack{\theta_\ell\to\cdots\to\tau_0\\ \tau_0=\tau}}M(\theta_\ell\to\cdots\to\tau_0)$$

and V_ℓ as the following $|T|\cdot q\times|T|\cdot q$ block diagonal matrix comprised of $q\times q$-dimensional blocks with block rows and columns indexed by T:

$$V_\ell[\tau,\tau]=V_{\ell,\tau}.$$

Finally, we define a linear transformation L on the space of $|T|\cdot q\times|T|\cdot q$ matrices. To define $L(M)$ we treat M as a block matrix comprised of $q\times q$-dimensional blocks with blocks rows and columns indexed by T.

$$L(M)[\tau,\tau]\coloneqq$$
$$\mathcal{T}(\tau)\sum_{\tau'\to\mathrm{in}\to\theta\to\mathrm{out}\to\tau}\frac{\phi_\theta}{\mathcal{T}(\tau)}\prod_{s=1}^{\mathrm{a}(\theta)}\mathcal{T}(\mathrm{Cl}(\theta)_s)\cdot\mathbf{1}[\mathrm{Cl}(\theta)_{\mathrm{out}}=\tau]\cdot\mathbf{1}[\mathrm{Cl}(\theta)_{\mathrm{in}}=\tau']\cdot$$
$$\overline{\mathbf{M}}_{\theta,\mathrm{out}|\mathrm{in}}\cdot M[\tau',\tau']\cdot\overline{\mathbf{M}}^{*}_{\theta,\mathrm{out}|\mathrm{in}}.$$
$$L(M)[\tau_1,\tau_2]\coloneqq 0\quad\text{for }\tau_1\neq\tau_2.$$

Now observe that $V_{\ell+1}=L(V_\ell)$; consequently $V_\ell=L^\ell(V_0)$ and $\alpha_\ell=\mathsf{Tr}\left(L^\ell(V_0)\right)$.

We now connect the limiting behavior of the amplification factor α_ℓ to the eigenvalues of L. We start by making a few observations:

Observation 3.4.2. If M is a positive semidefinite matrix, then $L(M)$ is also a positive semidefinite matrix.

Observation 3.4.3. V_0 is a diagonal matrix with strictly positive entries on its diagonal and hence is positive definite.

Observation 3.4.4. Since $L^\ell(V_0)$ is positive semidefinite:

$$\|L^\ell(V_0)\|_F\leqslant\mathsf{Tr}\left(L^\ell(V_0)\right)\leqslant\sqrt{q\cdot|T|}\|L^\ell(V_0)\|_F.$$

Our final ingredient is a lemma that appears in [BC19, Theorem 16, part (ii)].

Lemma 3.4.5. *If L is a linear operator on $r \times r$ matrices with spectral radius λ_L such that for any positive semidefinite matrix M, $L(M)$ is also positive semidefinite, then for any positive definite M',*

$$\lambda_L = \lim_{\ell \to \infty} \|L^\ell(M')\|_F^{1/\ell}.$$

By Observation 3.4.2, Observation 3.4.3 and Lemma 3.4.5:

$$\lambda_L = \lim_{\ell \to \infty} \|L^\ell(V_0)\|_F^{1/\ell}.$$

Observation 3.4.4 lets us conclude:

$$\lambda_L = \lim_{\ell \to \infty} \mathsf{Tr}\left(L^\ell(V_0)\right)^{1/\ell} = \lim_{\ell \to \infty} \alpha_\ell^{1/\ell}.$$

Thus we prove Lemma 3.2.12, and make the following conjecture:

Conjecture 3.4.6. *If $\lambda_L > 1$, then $\lim_{\ell \to \infty} \alpha_\ell$ goes to ∞ we conjecture that recovery is easy and if $\lambda_L \leqslant 1$, then $\lim_{\ell \to \infty} \alpha_\ell = 0$ and we conjecture that it is hard.*

3.5 A spectral distinguishing algorithm

We now describe the spectral distinguisher we use, which is based on linearizing the belief propagation algorithm outlined in Section 3.4. Recall that given $\boldsymbol{G}$ sampled from either $\mathcal{N}$ or $\mathcal{P}$ our goal is to output "null" if $\boldsymbol{G} \sim \mathcal{N}$ and "planted" if $\boldsymbol{G} \sim \mathcal{P}$ with probability $1 - o(1)$. Further, the messages given by (3.6) and (3.7) are a fixed point for the BP update rule for $\mathcal{P}$ (which is equivalent to the detailed balanced condition (3.47) holding). The sample $\boldsymbol{G}$ is given by the tuple $([n], \boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{\tau})$. Our algorithm constructs a matrix called the *null-centered nonbacktracking power matrix* and thresholds on its largest eigenvalue against a particular value t which is a function of the null and planted models and outputs "planted" if the largest eigenvalue exceeds t and "null" otherwise.

Recall the definition of the matrix $\underline{A}_{\boldsymbol{G}}^{(\ell)}$.

$$\sum_{\substack{ie_1v_1e_2v_2\ldots e_\ell j \in \\ \text{all nonbacktracking walks} \\ \text{from } i \text{ to } j \text{ in complete} \\ \text{factor graph}}} \overline{\mathbf{M}}_{e_1,i|v_1} \cdot \overline{\mathbf{M}}_{e_2,v_1|v_2} \cdots \overline{\mathbf{M}}_{e_\ell,v_{\ell-1}|j}.$$

$$(\mathbf{1}[e_1 \in \boldsymbol{G}] - \mathop{\mathbf{Pr}}_{\mathcal{N}|\boldsymbol{\tau}}[e_1 \in \boldsymbol{G}]) \cdots (\mathbf{1}[e_\ell \in \boldsymbol{G}] - \mathop{\mathbf{Pr}}_{\mathcal{N}|\boldsymbol{\tau}}[e_\ell \in \boldsymbol{G}])$$

where the complete factor graph is defined as follows.

Definition 3.5.1. We define the *complete factor graph* $\mathcal{K}_n = ([n], E_1, \ldots, E_F)$ where E_i denotes the collection of all potential type-i factors that could appear in $\boldsymbol{G}$.

We now describe our algorithm.[3]

- Compute a matrix representation of the linear operator L from the statement of Conjecture 3.4.6.
- Let λ_L be the spectral radius of L.
- Choose κ strictly in between $\sqrt{\lambda_L}$ and λ_L.
- Let $s = \lceil \sqrt{\log n} \rceil$. Compute $\underline{A}_{\boldsymbol{G}}^{(s)}$ and compute its largest eigenvalue ρ.
- If $\rho > \kappa^s$, output "planted", otherwise output "null".

To prove that the above algorithm works it suffices to prove that when $\boldsymbol{G}$ is sampled from the null distribution, all its eigenvalues are all less κ^s and when $\boldsymbol{G}$ is sampled from the planted distribution there is an eigenvalue greater than κ^s. Henceforth we assume $\lambda_L > 1$. To prove both of these facts under the hypothesis that $\lambda_L > 1$, one ingredient we need is that the matrix $\underline{A}_{\boldsymbol{G}}^{(s)}$ is self-adjoint under an appropriate inner product.

Given a vector in $\mathbb{R}^{nq}$ we treat it as a block vector comprising of n blocks of dimension q each where each block corresponds to a vertex in $[n]$. Now, we define a $nq \times nq$-dimensional positive diagonal matrix $\mathbf{H}_{\boldsymbol{\tau}}$ where the (v, v) block is equal to:

$$\mathbf{H}_{\boldsymbol{\tau},(v,v)}[c, c] := \begin{cases} \mathbb{P}_{\boldsymbol{\tau}(v)}(c) & \text{if } \mathbb{P}_{\boldsymbol{\tau}(v)}(c) > 0 \\ 1 & \text{otherwise.} \end{cases}$$

We will use the following inner product on $\mathbb{R}^{nq}$:

$$\langle x, y \rangle_{\mathbf{H}} := x^\top \mathbf{H}_{\boldsymbol{\tau}}^{-1} y.$$

[3]The details for why each step can be carried out efficiently are briefly discussed at the end of this section.

Remark 3.5.2. Ideally, we would like to simply place $\mathbb{P}_{\boldsymbol{\tau}(v)}(c)$ in every diagonal entry $[(v,c),(v,c)]$ and use the pseudoinverse of $\mathbf{H}$ instead. But doing so leads to some complications of the defined bilinear form not necessarily satisfying strict positive definiteness required of an inner product. The choice of 1 is arbitrary and does not influence any of the statements since all the vectors we work with have zeros in the (v,c) coordinates where $\mathbb{P}_{\boldsymbol{\tau}(v)}(c)$ is 0, and also that coordinate subspace is in the kernel of every matrix we work with.

Claim 3.5.3. The matrix $\underline{A}_{\boldsymbol{G}}^{(s)}$ is self-adjoint under $\langle\cdot,\cdot\rangle_{\mathbf{H}}$.

Proof. For any vectors x, y:

$$\langle x, \underline{A}_{\boldsymbol{G}}^{(s)} y\rangle_{\mathbf{H}} = \sum_{u,v\in[n]} \sum_{\substack{ue_1v_1e_2v_2\\ \ldots e_\ell v\in\mathcal{K}_n}} x[u]^\top \mathbf{H}^{-1}\overline{\mathbf{M}}_{e_1,u|v_1}\cdot\overline{\mathbf{M}}_{e_2,v_1|v_2}\cdots\overline{\mathbf{M}}_{e_\ell,v_{\ell-1}|v} y[v]\cdot$$
$$(\mathbf{1}[e_1\in\boldsymbol{G}] - \Pr_{\mathcal{N}|\boldsymbol{\tau}}[e_1\in\boldsymbol{G}])\cdots(\mathbf{1}[e_\ell\in\boldsymbol{G}] - \Pr_{\mathcal{N}|\boldsymbol{\tau}}[e_\ell\in\boldsymbol{G}])$$

From Part 2 of Claim 3.4.1,

$$x[u]^\top \mathbf{H}^{-1}[u,u]\overline{\mathbf{M}}_{e_1,u|v_1}\cdots\overline{\mathbf{M}}_{e_\ell,v_{\ell-1}|v} y[v] =$$
$$x[u]^\top (\overline{\mathbf{M}}_{e_\ell,v|v_{\ell-1}}\cdots\overline{\mathbf{M}}_{e_1,v_1|u})^\top \mathbf{H}^{-1}[v,v] y[v].$$

Plugging this back into the above gives:

$$= \sum_{u,v\in[n]} \sum_{\substack{ue_1v_1e_2v_2\\ \ldots e_\ell v\in\mathcal{K}_n}} x[u]^\top (\overline{\mathbf{M}}_{e_\ell,v|v_{\ell-1}}\cdots\overline{\mathbf{M}}_{e_1,v_1|u})^\top \mathbf{H}^{-1}[v,v] y[v]\cdot$$
$$(\mathbf{1}[e_1\in\boldsymbol{G}] - \Pr_{\mathcal{N}|\boldsymbol{\tau}}[e_1\in\boldsymbol{G}])\cdots(\mathbf{1}[e_\ell\in\boldsymbol{G}] - \Pr_{\mathcal{N}|\boldsymbol{\tau}}[e_\ell\in\boldsymbol{G}])$$
$$= \sum_{u,v\in[n]} \sum_{\substack{ue_1v_1e_2v_2\\ \ldots e_\ell v\in\mathcal{K}_n}} (\overline{\mathbf{M}}_{e_\ell,v|v_{\ell-1}}\cdots\overline{\mathbf{M}}_{e_1,v_1|u} x[u])^\top \mathbf{H}^{-1}[v,v] y[v]\cdot$$
$$(\mathbf{1}[e_1\in\boldsymbol{G}] - \Pr_{\mathcal{N}|\boldsymbol{\tau}}[e_1\in\boldsymbol{G}])\cdots(\mathbf{1}[e_\ell\in\boldsymbol{G}] - \Pr_{\mathcal{N}|\boldsymbol{\tau}}[e_\ell\in\boldsymbol{G}])$$
$$= \langle \underline{A}_{\boldsymbol{G}}^{(s)} x, y\rangle_{\mathbf{H}}$$

which proves the claim. □

We first focus on obtaining spectral norm bounds on $\underline{A}_{\boldsymbol{G}}^{(s)}$ in the null model. We obtain these bounds from Theorem 3.7.4 so we verify that the matrix $\underline{A}_{\boldsymbol{G}}^{(s)}$ indeed

meets the hypothesis of the theorem statement. 1 is satisfied due to Claim 3.5.3. As a consequence of the first part of Claim 3.4.1, all the matrices $M(p)$ are Markov transition matrices with an eigenspace projected away, and hence have all their entries bounded by 1. Since these matrices have dimension $q \times q$, their operator norm is bounded by some constant C depending only on q, and hence 2 is also satisfied. Next, $\rho(\mathrm{Bl}, \mathrm{M}^{\times})$ is exactly equal to $\sqrt{\lambda_L}$, which by our assumption is greater than 1. Finally, we chose s in the range handled by the theorem statement and thus Theorem 3.7.4 implies:

Theorem 3.5.4. *Suppose $\boldsymbol{G} \sim \mathcal{N}$. For every constant $\varepsilon > 0$, with probability $1 - o(1)$:*

$$|\lambda|_{\max}\left(\underline{A}_{\boldsymbol{G}}^{(s)}\right) \leqslant ((1+\varepsilon)\sqrt{\lambda_L})^s.$$

We can choose ε small enough so that $|\lambda|_{\max}\left(\underline{A}_{\boldsymbol{G}}^{(s)}\right) \leqslant \kappa^s$ for $\boldsymbol{G} \sim \mathcal{N}$ whp.

Finally, to prove that there is an eigenvalue greater than κ^s when $\boldsymbol{G} \sim \mathcal{P}$, by Claim 3.5.3 it suffices to illustrate a vector $x \in \mathbb{R}^{nq}$ such that $\frac{\langle x, \underline{A}_G x\rangle_{\mathbf{H}}}{\langle x,x\rangle_{\mathbf{H}}} \geqslant \kappa^s$. Then as a direct consequence of Theorem 3.6.1:

Theorem 3.5.5. *Suppose $\boldsymbol{G} \sim \mathcal{P}$. There is an absolute constant C such that with probability $1 - o(1)$:*

$$|\lambda|_{\max}\left(\underline{A}_{\boldsymbol{G}}^{(s)}\right) \geqslant C\lambda_L^s.$$

Since s is super-constant and κ is strictly less than λ_L, it is indeed true that $|\lambda|_{\max}\left(\underline{A}_{\boldsymbol{G}}^{(s)}\right) \geqslant \kappa^s$ for $\boldsymbol{G} \sim \mathcal{P}$ whp. Consequently, we can summarize our main theorem on distinguishing the null distribution from the planted distribution:

Theorem 3.5.6. *When $\lambda_L > 1$, the task of distinguishing $\mathcal{N}$ from $\mathcal{P}$ with high probability can be done in polynomial time.*

3.5.1 Implementation details

Our first goal is to explain how to efficiently choose κ which strictly between $\sqrt{\lambda_L}$ and λ_L when $\lambda_L > 1$. First note that there is a small enough ε such that if $\widetilde{\lambda}_L$ is an additive ε-approximation of λ_L, then $\frac{\widetilde{\lambda}_L + \sqrt{\widetilde{\lambda}_L}}{2}$ lies strictly in between $\sqrt{\lambda_L}$ and λ_L. By Lemma 3.4.5 there is large enough constant C such that $\|L^C(\mathbf{I})\|_F^{1/C}$ is ε-close to

λ_L. Thus, if our estimator $\widetilde{\lambda}_L$ is $\|L^{\log n}(\mathbf{I})\|_F^{1/\log n}$, then for large enough n, choosing κ as $\frac{\widetilde{\lambda_L}+\sqrt{\widetilde{\lambda_L}}}{2}$ would give us a number strictly between $\sqrt{\lambda_L}$ and λ_L.

Our second goal is to explain how to efficiently compute the matrix $\underline{A}_{\boldsymbol{G}}^{(s)}$. Towards doing so, define the *nonbacktracking walk generator matrix* as the matrix with rows and columns indexed by vev' for variable vertices v and v' and constraint vertex e':

$$B_{\boldsymbol{G}}[(v_1e_1v_2),(v_3e_2v_4)] = \begin{cases} \overline{\mathbf{M}}_{e_2,v_3|v_4} \cdot \left(\mathbf{1}[e_2 \in \boldsymbol{G}] - \mathbf{Pr}_{\boldsymbol{G}\sim\mathcal{P}|\boldsymbol{\tau}}[e_2 \in \boldsymbol{G}]\right) & \text{if } e_1 \neq e_2, v_2 = v_3 \\ 0 & \text{otherwise.} \end{cases}$$

Let $S_{\boldsymbol{G}}$ be the matrix with rows indexed by variables in $[n]$ and columns indexed by all vev' where the (v, vev') entry contains $\overline{\mathbf{M}}_{e,v|v'} \cdot (\mathbf{1}[e \in \boldsymbol{G}] - \mathbf{Pr}_{\boldsymbol{G}\sim\mathcal{P}|\boldsymbol{\tau}}[e \in \boldsymbol{G}])$ and the remaining entries contain 0, and let $T_{\boldsymbol{G}}$ be the matrix with rows indexed by all vev' and columns indexed by variables in $[n]$ where the (vev', v') entry is $\mathbf{I}$ and the remaining entries contain 0. Then $S_{\boldsymbol{G}}B_{\boldsymbol{G}}^{s-1}T_{\boldsymbol{G}} = \underline{A}_{\boldsymbol{G}}^{(s)}$, and it is apparent that the LHS can be computed efficiently.

Finally, since $\underline{A}_{\boldsymbol{G}}^{(s)}$ is self-adjoint under $\langle\cdot,\cdot\rangle_{\mathbf{H}}$ by Claim 3.5.3 its largest eigenvalue can be efficiently computed via standard methods such as the power iteration method to a precision necessary for the distinguishing algorithm.

3.6 Statistics for the planted model

Consider the planted model $\mathrm{M}_n = (n, T, \mathcal{T}, C, \mathbb{P}, \phi)$. Use $\boldsymbol{G} = ([n], \boldsymbol{E}_1, \ldots, \boldsymbol{E}_F, \boldsymbol{\tau}, \boldsymbol{c})$ to denote a sample from $\mathrm{M}_{\mathrm{P}}^{[n]}$. Recall that in this section our goal is to prove a lower bound on the spectral radius of $\underline{A}_{\boldsymbol{G}}^{(s)}$ for $s \leqslant \frac{\log n}{(\log\log n)^2}$ by illustrating a "witness" vector with large quadratic form.

We define the *local statistics vector* associated to $\boldsymbol{G}$ denoted $g \in \mathbb{R}^{q\cdot n}$ to be the concatenation of vectors $g^v = u_{\boldsymbol{c}_v}$ for all $v \in [n]$, where $u_c \in \mathbb{R}^q$ is the indicator of vector of color c. We will shorten $\langle g, g\rangle_{\mathbf{H}}$ to $\|g\|^2$ in this section.

In this section we prove:

Theorem 3.6.1. *Let $s \leqslant \frac{\log n}{(\log\log n)^2}$. There exists a constant γ such that with probability $1 - o_n(1)$:*

$$\frac{\left\langle g, \underline{A}_{\boldsymbol{G}}^{(s)} g \right\rangle_{\mathbf{H}}}{\|g\|^2} = \gamma \lambda_L^s.$$

To prove Theorem 3.6.1 we will introduce and recall some notation to streamline the proofs. First recall that we defined $\mathcal{K}_n$ in Definition 3.5.1 as the instance on variable set $[n]$ with all potential factors that could appear in a graph sampled from M_n. Let p be some length-$2s$ walk $(v_0 \to e_0 \to v_1 \dots v_{s-1} \to e_{s-1} \to v_s)$ in $\mathcal{K}_n$ starting and ending at a variable node. Use p_e to denote the factor nodes in p and p_v to denote the variable nodes in ∂p_e.[4] We use $p_v[i]$ to denote v_i and $p_e[i]$ to denote e_i. Recall that we use $\theta(e)$ to denote the type of a factor node e. We use $\chi(p)$ to denote the number of excess edges in p. Concretely:

Definition 3.6.2. $\chi(p) = |p_v| - \sum_{j \in [s]} (\mathrm{a}(\theta(p_e[j])))$.

Definition 3.6.3. Let $\mathrm{NB}(s)$ denote the set of all length-s nonbacktracking walks in $\mathcal{K}_n$.

Definition 3.6.4. $\overline{\mathbf{M}}_p := \prod_{j=0}^{s-1} \overline{\mathbf{M}}_{\theta(p_e[j]),\mathrm{i}(p_v[j])|\mathrm{i}(p_v[j+1])}$.

Definition 3.6.5. $\mathbf{\Psi}_p := \prod_{j=0}^{s-1} \mathbf{\Psi}_{\theta(p_e[j]),\mathrm{i}(p_v[j])|\mathrm{i}(p_v[j+1])}$.

For $\boldsymbol{G} \sim \mathrm{M}_n$, we define the following notation:

Definition 3.6.6. $\mathbf{w}(\boldsymbol{G}, p) := \overline{\mathbf{M}}_p \prod_{j=0}^{s-1} \left(\mathbf{1}[p_e[j] \in \boldsymbol{G}] - \mathbf{Pr}_{\boldsymbol{G} \sim \mathcal{P}|\boldsymbol{\tau}}[p_e[j] \in \boldsymbol{G}] \right)$.

Definition 3.6.7. We will use $\mathrm{wt}(p)$ to denote

$$\prod_{j=0}^{s-1} \left(\mathbf{1}[p_e[j] \in \boldsymbol{G}] - \mathop{\mathbf{Pr}}_{\boldsymbol{G} \sim \mathcal{P}|\boldsymbol{\tau}}[p_e[j] \in \boldsymbol{G}] \right).$$

Remark 3.6.8. In the above language, the centered nonbacktracking power matrix of $\boldsymbol{G}$ is:

$$\underline{A}_{\boldsymbol{G}}^{(s)}[u, v] = \sum_{p \in \mathrm{NB}(s) | p_v[0]=u, p_v[s]=v} \mathbf{w}(\boldsymbol{G}, p).$$

[4] We would like to stress that p_v also includes vertices that are *not* walked on, but are incident to factor nodes which are walked on.

Recall that a block (u, v) in the centered nonbacktracking power matrix $\underline{A}_{\boldsymbol{G}}^{(s)}$ captures the matrix weight of all length-$2s$ nonbacktracking walks from u to v. Although the paths are nonbacktracking, there could be variables that are visited multiple times and there could also be off-path variables in p_v connected to multiple factors. It is hard to pinpoint the statistics precisely for this kind of walks, but luckily their contribution is negligible. Thus, we first remove these "bad" walks from $\underline{A}_{G}^{(s)}$ and analyze the resulting matrix $\overline{A}_{\boldsymbol{G}}^{(s)}$, and then bound the contribution of these bad walks.

We give a formal definition of the "nice" walks that are kept in $\overline{A}_{\boldsymbol{G}}^{(s)}$.

Definition 3.6.9. A path p is *self-avoiding* if $\chi(p) = |p_v| - \sum_{j\in[s]} \left(\mathrm{a}(\mathrm{Cl}_p(j)) - 1\right) = 1$.

In any self-avoiding path, every variable node in the interior of the path has degree 2 (i.e. is contained in 2 factors in p) and each of the other variable nodes has degree 1 (i.e. is contained in 1 factor in p).

Definition 3.6.10. Let $\mathrm{SA}(s)$ denote the set of all length-$2s$ self-avoiding walks in $\mathcal{K}_n$.

Definition 3.6.11. The centered self-avoiding-walk matrix of $\boldsymbol{G}$ is $\overline{A}_{\boldsymbol{G}}^{(s)} \in \mathbb{R}^{qn\times qn}$ where the (u, v)-th block of the matrix is

$$\overline{A}_{\boldsymbol{G}}^{(s)}[u, v] = \sum_{p\in\mathrm{SA}(s)\,|\,p_v[0]=u, p_v[s]=v} \mathbf{w}(\boldsymbol{G}, p).$$

3.6.1 Statistics for the centered self-avoiding-walk matrix

Claim 3.6.12. For any self-avoiding path p in $\mathcal{K}_n$,

$$\mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g^{p_v[0]}, \mathbf{w}(\boldsymbol{G}, p) g^{p_v[s]}\right\rangle_{\mathbf{H}}\right] = \Pr_{\boldsymbol{G}\sim\mathrm{M}|\boldsymbol{\tau}}[p \in \boldsymbol{G}] \cdot \mathsf{Tr}\left(\overline{\mathbf{M}}_p \overline{\mathbf{M}}_p^*\right).$$

Proof. The proof is via a chain of equalities. To lighten notation we use e_j to denote $p_e[j]$ and v_j to denote $p_v[j]$. We use $\mathrm{Int}(p_v)$ to denote the interior vertices of p.

$$\mathop{\mathbf{E}}_{\boldsymbol{G}|\boldsymbol{\tau}}\left[\langle g^{v_0}, \mathbf{w}(\boldsymbol{G}, p) g^{v_s}\rangle_{\mathbf{H}}\right] =$$

$$\sum_{c: p_v\to[q]} \prod_{w\in p_v} \mathbb{P}_{\boldsymbol{\tau}(w)}(c(w)) \cdot \mathop{\mathbf{E}}_{\boldsymbol{G}|\boldsymbol{\tau},\boldsymbol{c}}\Big[\mathbf{H}^{-1}[(v_0, c(v_0)), (v_0, c(v_0))]\cdot$$

$$\prod_{j=0}^{s-1}(\mathbf{1}[e_j \in \boldsymbol{G}] - \mathop{\mathbf{Pr}}_{\boldsymbol{G}\sim \mathrm{M}|\tau}[e_j \in \boldsymbol{G}])\overline{\mathbf{M}}_p[c(v_0), c(v_s)]\Big]$$

$$= \sum_{c:p_v\to[q]} \prod_{w\in p_v} \mathbb{P}_{\boldsymbol{\tau}(w)}(c(w)) \cdot \prod_{j=0}^{s-1}(\mathop{\mathbf{Pr}}_{\boldsymbol{G}\sim \mathrm{M}|\boldsymbol{\tau},\boldsymbol{c}}[e_j \in \boldsymbol{G}] - \mathop{\mathbf{Pr}}_{\boldsymbol{G}\sim \mathrm{M}|\boldsymbol{\tau}}[e_j \in \boldsymbol{G}])\cdot$$

$$\frac{1}{\mathbb{P}_{\boldsymbol{\tau}(v_s)}(c(v_s))} \cdot \overline{\mathbf{M}}_p^*[c(v_s), c(v_0)]$$

$$= \sum_{c:p_v\to[q]} \prod_{w\in \mathrm{Int}(p_v)} \frac{1}{\mathbb{P}_{\boldsymbol{\tau}(w)}(c(w))} \cdot \prod_{j=0}^{s-1} \mathop{\mathbf{Pr}}_{\boldsymbol{G}\sim \mathrm{M}|\boldsymbol{\tau}}[e_j \in \boldsymbol{G}]\cdot$$

$$\left(\mu_{e_j}(c(\partial e_j)) - \prod_{w\in\partial e_j} \mathbb{P}_{\boldsymbol{\tau}(w)}(c(w))\right) \cdot$$

$$\frac{1}{\mathbb{P}_{\boldsymbol{\tau}(v_s)}(c(v_s))} \cdot \overline{\mathbf{M}}_p^*[c(v_s), c(v_0)]$$

$$= \mathop{\mathbf{Pr}}_{\boldsymbol{G}\sim \mathrm{M}|\boldsymbol{\tau}}[p \in \boldsymbol{G}] \sum_{c:\{v_0,\ldots,v_s\}\to[q]} \prod_{j=0}^{s-1}(\boldsymbol{\Psi}_{e_j,v_j|v_{j+1}}[c(v_j), c(v_{j+1})] - \mathbb{P}_{\boldsymbol{\tau}(v_j)})\cdot$$

$$\overline{\mathbf{M}}_p^*[c(v_s), c(v_0)]$$

$$= \mathop{\mathbf{Pr}}_{\boldsymbol{G}\sim \mathrm{M}|\boldsymbol{\tau}}[p \in \boldsymbol{G}] \sum_{c(v_0),c(v_s)} \overline{\mathbf{M}}_p[c(v_0), c(v_s)] \cdot \overline{\mathbf{M}}_p^*[c(v_s), c(v_0)]$$

$$= \mathop{\mathbf{Pr}}_{\boldsymbol{G}\sim \mathrm{M}|\boldsymbol{\tau}}[p \in \boldsymbol{G}] \cdot \mathsf{Tr}\left(\overline{\mathbf{M}}_p\overline{\mathbf{M}}_p^*\right)$$

□

Now we give precise estimates for the statistics of the centered self-avoiding walk matrix.

Lemma 3.6.13.

$$\mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g, \overline{A}_{\boldsymbol{G}}^{(s)} g\right\rangle_{\mathbf{H}}\right] = (1 - o_n(1))\lambda_L^s n.$$

Proof. Expanding out $\overline{A}_{\boldsymbol{G}}^{(s)}$ as a sum by its definition in Definition 3.6.11 gives:

$$\mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g, \overline{A}_{\boldsymbol{G}}^{(s)} g\right\rangle_{\mathbf{H}}\right] = \sum_{p\in \mathrm{SA}(s)} \mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g^{p_v[0]}, \mathbf{w}(\boldsymbol{G}, p) g^{p_v[s]}\right\rangle_{\mathbf{H}}\right]$$

$$= \sum_{p \in \mathrm{SA}(s)} \Pr_{\boldsymbol{G} \sim \mathrm{M}|\boldsymbol{\tau}}[p \in \boldsymbol{G}] \cdot \mathsf{Tr}\left(\overline{\mathbf{M}}_p \overline{\mathbf{M}}_p^*\right) \quad \text{(via Claim 3.6.12)}$$

$$= n \cdot (1 - o_n(1))\lambda_L^s \quad \text{(by the definition of } \lambda_L\text{)}.$$

□

We next bound the variance of the local statistics.

Lemma 3.6.14. *There is an absolute constant C such that:*

$$\mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g, \overline{A}_{\boldsymbol{G}}^{(s)} g\right\rangle_{\mathbf{H}}^2\right] - \mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g, \overline{A}_{\boldsymbol{G}}^{(s)} g\right\rangle_{\mathbf{H}}\right]^2 \leqslant n(Cs)^{Cs}.$$

Proof. If two walks p and $\tilde{p}$ do not share any vertices, then their contribution to $\mathbf{E}_{\boldsymbol{G}}\left[\left\langle g, \overline{A}_{\boldsymbol{G}}^{(s)} g\right\rangle_{\mathbf{H}}^2\right]$ satisifies

$$\begin{aligned}\text{contribution of } p, \tilde{p} &= \mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g^{p_v[0]}, \mathbf{w}(\boldsymbol{G}, p) g^{p_v[s]}\right\rangle_{\mathbf{H}} \left\langle g^{\tilde{p}_v[0]}, \mathbf{w}(\boldsymbol{G}, \tilde{p}) g^{\tilde{p}_v[s]}\right\rangle_{\mathbf{H}}\right] \\ &= \mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g^{p_v[0]}, \mathbf{w}(\boldsymbol{G}, p) g^{p_v[s]}\right\rangle_{\mathbf{H}}\right] \mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g^{\tilde{p}_v[0]}, \mathbf{w}(\boldsymbol{G}, \tilde{p}) g^{\tilde{p}_v[s]}\right\rangle_{\mathbf{H}}\right].\end{aligned}$$

So this contribution cancels out with the identical term in $\mathbf{E}_{\boldsymbol{G}}\left[\left\langle g, \overline{A}_{\boldsymbol{G}}^{(s)} g\right\rangle_{\mathbf{H}}\right]^2$. Thus it suffices for us to consider self-avoiding walks p and $\tilde{p}$ that share some vertices.

We write $p \parallel \tilde{p}$ if they share some variable or factor nodes and the shared nodes have consistent types and use $p^{\cup}$ to denote the union of the two walks. Now, note that:

1. Conditioned on $\boldsymbol{\tau}$ and $\boldsymbol{c}$ every factor node with arity k is chosen independently with probability at most $\frac{\alpha}{n^{k-1}}$ for some constant α. Thus, for any subgraph of $\mathcal{K}_n$ on e edges and r factor nodes, the probability of it occurring in $\boldsymbol{G}$ is at most $\alpha^r n^{e-r}$.

2. The matrix weight $\overline{\mathbf{M}}_p$ of any self-avoiding path has entries bounded in magnitude by 1 since it is a product of projected stochastic matrices.

Using the above facts, a straightforward calculation tells us:

$$\left|\mathop{\mathbf{E}}_{\boldsymbol{G}}\left[\left\langle g^{p_v[0]}, \mathbf{w}(\boldsymbol{G}, p) g^{p_v[s]}\right\rangle_{\mathbf{H}} \left\langle g^{\tilde{p}_v[0]}, \mathbf{w}(\boldsymbol{G}, \tilde{p}) g^{\tilde{p}_v[s]}\right\rangle_{\mathbf{H}}\right]\right| \leqslant (\alpha')^s \, n^{-(\sum_{e \in p_e^{\cup}} \mathrm{a}(e) - 1)}. \tag{3.11}$$

We say $p_1^{\cup} \sim p_2^{\cup}$ if the subgraphs induced by them are isomorphic. $\sim$ partitions the space of all $p^{\cup}$ into equivalence classes. We use $[p^{\cup}]$ to denote the equivalence class of $p^{\cup}$. The number of equivalence classes can be bounded by $(C's)^{C's}$ for some constant $C' > 1$ (since the graph representing the equivalence class of $p \cup \tilde{p}$ is on $O(s)$ vertices can be specified by a list of $O(s)$ edges). Due to the shared vertices, $p^{\cup}$ is connected and $\chi(p^{\cup}) \leqslant 1$. Thus, we now bound the variance as follows:

$$\begin{aligned}
&\mathop{\mathbf{E}}_{G}\left[\left\langle g, \overline{A}_{G}^{(s)} g\right\rangle_{\mathbf{H}}^{2}\right] - \mathop{\mathbf{E}}_{G}\left[\left\langle g, \overline{A}_{G}^{(s)} g\right\rangle_{\mathbf{H}}\right]^{2} \\
=&\sum_{p\|\tilde{p}} \mathop{\mathbf{E}}_{G}\left[\left\langle g^{p_v[0]}, \mathbf{w}(G,p) g^{p_v[s]}\right\rangle_{\mathbf{H}} \left\langle g^{\tilde{p}_v[0]}, \mathbf{w}(G,\tilde{p}) g^{\tilde{p}_v[s]}\right\rangle_{\mathbf{H}}\right] \\
\leqslant&\sum_{p^{\cup}} \sum_{p\|\tilde{p}: p\cup\tilde{p}=p^{\cup}} (\alpha')^{s} n^{-(\sum_{e\in p_e^{\cup}} \mathrm{a}(e)-1)} \\
\leqslant&\sum_{[p^{\cup}]} \sum_{p_i^{\cup}\in[p^{\cup}]} (9\alpha' s)^{s} n^{-(\sum_{e\in p_e^{\cup}} \mathrm{a}(e)-1)} \\
=&\sum_{[p^{\cup}]} (9\alpha' s)^{s} n^{|p_v^{\cup}|-(\sum_{e\in p_e^{\cup}} \mathrm{a}(e)-1)} \\
=&\sum_{[p^{\cup}]} (9\alpha' s)^{s} n^{\chi(p^{\cup})} \\
\leqslant&(9C'\alpha' s^2)^{C's} n.
\end{aligned}$$

Thus, the claim follows. □

3.6.2 Bounding contribution of non-self-avoiding walks

In comparison to the self-avoiding walks, the non-self-avoiding walks have negligible contributions to the expectation and the variance of the statistics. We prove this statement using the following claim.

Claim 3.6.15. We can bound the statistics of non-self-avoiding walks as follows:

$$\left|\mathop{\mathbf{E}}_{G}\left[\sum_{p \text{ non-self-avoiding}} \left\langle g^{p_v[0]}, \mathbf{w}(G,p) g^{p_v[s]}\right\rangle_{\mathbf{H}}\right]\right| \leqslant (Cs)^{Cs}.$$

$$\mathop{\mathbf{E}}_{G}\left[\sum_{\substack{p\|\tilde{p} \\ p \text{ or } \tilde{p} \text{ non-self-avoiding}}} \left\langle g^{p_v[0]}, \mathbf{w}(G,p) g^{p_v[s]}\right\rangle_{\mathbf{H}} \cdot \left\langle g^{\tilde{p}_v[0]}, \mathbf{w}(G,\tilde{p}) g^{\tilde{p}_v[s]}\right\rangle_{\mathbf{H}}\right] \leqslant (Cs)^{Cs}$$

for some absolute constant C.

Proof. The first part is derived by applying Item 1 in the proof of Lemma 3.6.14.

$$\begin{aligned}
&\left|\mathop{\mathbf{E}}_{G}\left[\sum_{p \text{ non-self-avoiding}} \left\langle g^{p_v[0]}, \mathbf{w}(G,p) g^{p_v[s]} \right\rangle_{\mathbf{H}}\right]\right| \\
\leqslant &\sum_{p \text{ non-self-avoiding}} \left|\mathop{\mathbf{E}}_{G}\left[\left\langle g^{p_v[0]}, \mathbf{w}(G,p) g^{p_v[s]} \right\rangle_{\mathbf{H}}\right]\right| \\
\leqslant &\sum_{p \text{ non-self-avoiding}} \mathop{\mathbf{Pr}}_{G}[p \in G] \cdot O(1) \\
\leqslant &\sum_{p \text{ non-self-avoiding}} (\alpha')^s n^{\chi(p)-|p_v|} \\
\leqslant &\sum_{p \text{ non-self-avoiding}} (\alpha')^s n^{-|p_v|} \\
\leqslant &(Cs)^{Cs}.
\end{aligned}$$

where the equality from the second to third line is a consequence of Item 1 in the proof of Lemma 3.6.17, and the last inequality is due to $\chi(p) \leqslant 0$ for a non-self-avoiding walk.

The second expression is equal to:

$$\sum_{\substack{p \| \tilde{p} \\ p \text{ or } \tilde{p} \text{ non-self-avoiding}}} \mathop{\mathbf{E}}_{G}\left[\left\langle g^{p_v[0]}, \mathbf{w}(G,p) g^{p_v[s]} \right\rangle_{\mathbf{H}} \cdot \left\langle g^{\tilde{p}_v[0]}, \mathbf{w}(G,\tilde{p}) g^{\tilde{p}_v[s]} \right\rangle_{\mathbf{H}}\right].$$

By (3.11) the above can be bounded by:

$$\sum_{\substack{p \| \tilde{p} \\ p \text{ or } \tilde{p} \text{ non-self-avoiding}}} (\alpha')^s \, n^{\chi(p^\cup)-|p^\cup_v|}$$

where $p^\cup$, recall, is the union of p and $\tilde{p}$. Since p and $\tilde{p}$ share vertices, $\chi(p^\cup) \leqslant \min\{\chi(p), \chi(\tilde{p})\}$, and since at least one of the two walks is non-self-avoiding, $\chi(p^\cup) \leqslant 0$. This lets us bound the above by:

$$\sum_{\substack{p \| \tilde{p} \\ p \text{ or } \tilde{p} \text{ non-self-avoiding}}} (\alpha')^s \, n^{-|p^\cup_v|} \leqslant \sum_{p^\cup} \sum_{p \| \tilde{p} : p \cup \tilde{p} = p^\cup} (\alpha')^s \, n^{-|p^\cup_v|}$$

$$
\begin{aligned}
&= \sum_{[p^{\cup}]} \sum_{p_i^{\cup} \in [p^{\cup}]} (9\alpha' s)^s n^{-|p_v^{\cup}|} \\
&= \sum_{[p^{\cup}]} (9\alpha' s)^s \\
&\leqslant (9C'\alpha' s^2)^{C's}
\end{aligned}
$$

which gives us the desired bound in the second part of the statement. □

3.6.3 Wrapping up estimates of statistics

The following claim about the statistics of the centered non-backtracking-walk matrix $\underline{A}_G^{(s)}$ is an immediate consequence of combining Lemma 3.6.13 and the first part of Claim 3.6.15.

Lemma 3.6.16. *For $s \leqslant \frac{\log n}{(\log\log n)^2}$:*

$$
\mathop{\mathbf{E}}_{G}\left[\left\langle g, \underline{A}_G^{(s)} g\right\rangle_{\mathbf{H}}\right] = (1 \pm o_n(1))\lambda_L^s n.
$$

Lemma 3.6.17. *For some absolute constant C:*

$$
\mathop{\mathbf{E}}_{G}\left[\left\langle g, \underline{A}_G^{(s)} g\right\rangle_{\mathbf{H}}^2\right] - \mathop{\mathbf{E}}_{G}\left[\left\langle g, \underline{A}_G^{(s)} g\right\rangle_{\mathbf{H}}\right]^2 \leqslant n(Cs)^{Cs}.
$$

Proof. Using the observation that only pairs of walks that share some vertices contribute to the variance, we have:

$$
\begin{aligned}
&\mathop{\mathbf{E}}_{G}\left[\left\langle g, \underline{A}_G^{(s)} g\right\rangle_{\mathbf{H}}^2\right] - \mathop{\mathbf{E}}_{G}\left[\left\langle g, \underline{A}_G^{(s)} g\right\rangle_{\mathbf{H}}\right]^2 \\
&= \mathop{\mathbf{E}}_{G}\left[\left\langle g, \overline{A}_G^{(s)} g\right\rangle_{\mathbf{H}}^2\right] - \mathop{\mathbf{E}}_{G}\left[\left\langle g, \overline{A}_G^{(s)} g\right\rangle_{\mathbf{H}}\right]^2 \\
&\quad + \sum_{\substack{p \| \tilde{p} \\ p \text{ or } \tilde{p} \text{ non-self-avoiding}}} \\
&\quad \mathop{\mathbf{E}}_{G}\left[\left\langle g^{\eta(p_v^0)}, \mathbf{w}(G, p, \eta) g^{\eta(p_v^s)}\right\rangle_{\mathbf{H}} \cdot \left\langle g^{\eta(\tilde{p}_v^0)}, \mathbf{w}(G, \tilde{p}, \tilde{\eta}) g^{\tilde{\eta}(\tilde{p}_v^s)}\right\rangle_{\mathbf{H}}\right]
\end{aligned}
$$

We can conclude the desired bound immediately from Lemma 3.6.14 and the second part of Claim 3.6.15. □

Finally, to establish Theorem 3.6.1 we first use Chebyshev's inequality to conclude that when $s \leqslant \frac{\log n}{(\log\log n)^2}$,

$$\left\langle g, \underline{A}_{\boldsymbol{G}}^{(s)} g \right\rangle_{\mathbf{H}} = (1 \pm o_n(1))\lambda_L^s n$$

with probability $1 - o_n(1)$. Now, since the $\|g\|^2$ is $(1 \pm o_n(1))\frac{n}{\gamma}$ with probability $1 - o_n(1)$ for some constant γ, we can conclude that with probability $1 - o_n(1)$,

$$\frac{\left\langle g, \underline{A}_{\boldsymbol{G}}^{(s)} g \right\rangle_{\mathbf{H}}}{\|g\|^2} = (1 \pm o_n(1))\gamma\lambda_L^s$$

thereby finishing the proof.

3.7 Eigenvalue bounds

In this section we show an eigenvalue upper bound for the centered nonbacktracking-walk matrix in the null model. We first describe the matrix distribution in detail.

Recall the definition of a null model $\mathrm{M}^{\times}$ (Definition 3.2.3). Let $\boldsymbol{H} := \bigcup_{i=1}^{F} \boldsymbol{E}_i$ be an observation sampled from $\mathrm{M}^{\times}$. As discussed in Section 3.2.1.1, the observation $\boldsymbol{H}$ has an associated bipartite graph which we denote $\mathrm{Bip}(\boldsymbol{H})$. The left vertex set is given by the variables $[n]$ and the right vertex set is given by the factors $\gamma \in \boldsymbol{H}$. We will use $L(\boldsymbol{H})$ and $R(\boldsymbol{H})$ to denote the left and right vertex sets of $\mathrm{Bip}(\boldsymbol{H})$.

Next associate with each triple $v\gamma u$, where $v, u \in \gamma$ and $v \neq u$, a $q \times q$ matrix $M_{v\gamma u}$. Like before, the value of the matrix only depends on the factor type $\theta(\gamma)$ and the two variables' indices $\mathrm{i}(v), \mathrm{i}(u)$ in γ. We use Bl to denote the collection of these $q \times q$ matrices $\{M_{a\phi_i b}\}_{i\in[F], a\neq b\in[\mathrm{a}(i)]}$. Now we are ready to define the matrix distribution.

Definition 3.7.1. The matrix distribution is defined as follows. First sample an observation H from some null model $\mathrm{M}^{\times}$. We define the *length-ℓ $\mathrm{M}^{\times}$-centered nonbacktracking power* $\underline{A}_{\boldsymbol{H}}^{(\ell)}$ is the $n \times n$ block matrix where the (i,j)-block as the following $q \times q$ matrix:

$$\underline{A}_{\boldsymbol{H}}^{(\ell)}[i,j] := \sum_{\substack{(v_0\gamma_1 v_1 \ldots v_{\ell-1}\gamma_\ell v_\ell) \\ \in \mathrm{NB}(\mathcal{K}_n, \ell, i, j)}} \prod_{t=1}^{\ell} M_{v_{2t-2}\gamma_t v_{2t-1}} \left(\mathbf{1}[\gamma_t \in \boldsymbol{H}] - \underset{\mathrm{M}^{\times}}{\mathbf{Pr}}[\gamma_t \in \boldsymbol{H}] \right),$$

where $\mathrm{NB}(\mathcal{K}_n, \ell, i, j)$ denote the set of all length-2ℓ nonbacktracking walks in the complete bipartite factor graph $\mathrm{Bip}(\mathcal{K}_n)$ starting at variable i and ending at variable j.

We are interested in obtaining a high probability upper bound on $\|\underline{A}_{\boldsymbol{H}}^{(\ell)}\|$ in terms of a particular quantity depending on Bl and $\mathrm{M}^{\times}$, which we denote by $\rho(\mathrm{Bl}, \mathrm{M}^{\times})$. Before giving its definition, we simplify the notation a bit:

Definition 3.7.2. Given a length-2ℓ nonbacktracking walk $W = v_0\gamma_1 \dots v_{\ell-1}\gamma_\ell v_\ell$ in $\mathrm{Bip}(\boldsymbol{H})$, we define the weight of the walk with M_W to be

$$M_W \coloneqq M_{v_0\gamma_1 v_1} M_{v_1\gamma_2 v_2} \dots M_{v_{\ell-1}\gamma_\ell v_\ell}.$$

We use $\gamma_i(W)$ to denote the i-th factor visited by W.

We now define $\rho(\mathrm{Bl}, \mathrm{M}^{\times})$.

Definition 3.7.3. For a positive integer m, construct a length-m path P with random matrix weights in the following way. Start with a vertex v_0 and assign to it a random label $\boldsymbol{\tau}(v_0)$ sampled from $\mathcal{T}$. We iteratively construct a path with edges weighed by matrices until its length is equal to m. Suppose we already have a path $v_0 v_1 \dots v_t$ where each v_i has a label $\boldsymbol{\tau}(v_i)$, and for an edge $\{v_i, v_{i+1}\}$ its two directed edges (v_i, v_{i+1}) and (v_{i+1}, v_i) have matrix weights $\boldsymbol{W}_{i,i+1}$ and $\boldsymbol{W}_{i+1,i}$ respectively. To grow this path, we sample (s, a) where $s \in [F]$ and $a \in [\mathrm{a}(s)]$ with probability proportional to $\overline{\phi_s} \cdot \mathbf{1}[\mathrm{Cl}(s)_a = \boldsymbol{\tau}(v_t)] \cdot \prod_{j=1, j\neq a}^{t} \mathcal{T}(\mathrm{Cl}(s)_j)$, followed by a uniformly random b in $[\mathrm{a}(s)] \setminus \{a\}$. Add vertex v_{t+1} and set $\boldsymbol{\tau}(v_{t+1}) = \mathrm{Cl}(s)_b$. Then add edge $\{v_t, v_{t+1}\}$ and let the matrix weight of the directed edge (v_t, v_{t+1}) be $\boldsymbol{W}_{t,t+1} = M_{a\phi_s b} \in \mathrm{Bl}$, and matrix weight of directed edge (v_{t+1}, v_t) be $\boldsymbol{W}_{t+1,t} = M_{b\phi_s a} \in \mathrm{Bl}$.[5] We then define w_m as:

$$w_m \coloneqq \mathbf{E}\,\mathrm{Tr}(\boldsymbol{W}_{0,1}\boldsymbol{W}_{1,2} \dots \boldsymbol{W}_{m-1,m}\boldsymbol{W}_{m,m-1} \dots \boldsymbol{W}_{2,1}\boldsymbol{W}_{1,0}).$$

Now define $r(\mathrm{Bl}, \mathrm{M}^{\times})$ as

$$r(\mathrm{Bl}, \mathrm{M}^{\times}) \coloneqq \limsup_{m\to\infty} w_m^{\frac{1}{2m}}.$$

[5]For the sake of intuition the reader should think of the distribution of v_{t+1} as first sampling $\boldsymbol{H} \sim \mathrm{M}^{\times}$, then choosing a random vertex v with label $\boldsymbol{\tau}(v_t)$ and finally choosing as a random neighbor w of v within $\boldsymbol{H}$. $\boldsymbol{\tau}(v_{t+1})$ is then set to the label of w and the matrix weight on edge (v_t, v_{t+1}) is chosen as the matrix weight on (v, w).

Next, define $d(\mathrm{M}^{\times})$ (which, intuitively, is the average degree of a vertex in a sample from $\mathrm{M}^{\times}$) as

$$d(\mathrm{M}^{\times}) := \sum_{t=1}^{T} \mathcal{T}(t) \sum_{i=1}^{F} \sum_{j=1}^{k_i} \overline{\phi_i} \mathbf{1}[\mathrm{Cl}(i)_j = t].$$

Finally, we define $\rho(\mathrm{Bl}, \mathrm{M}^{\times})$ as

$$\rho(\mathrm{Bl}, \mathrm{M}^{\times}) := r(\mathrm{Bl}, \mathrm{M}^{\times}) \sqrt{d(\mathrm{M}^{\times})}.$$

We remark that if the weight collection Bl is defined such that $M_{b\phi_i a} = \overline{\mathbf{M}}_{i,a|b}$ (defined in (3.10)) for all $i \in [F], a \neq b \in [\mathrm{a}(i)]$, then $\rho(\mathrm{Bl}, \mathrm{M}^{\times}) = \sqrt{\lambda_L}$.

The main result of this section is that $\|\underline{A}_{\boldsymbol{H}}^{(\ell)}\| \leqslant ((1 + o_n(1))\rho(\mathrm{Bl}, \mathrm{M}^{\times}))^{\ell}$ for a wide range of ℓ when $\boldsymbol{H} \sim \mathrm{M}^{\times}$.

Theorem 3.7.4. *Suppose:*

1. *There is an inner product $\langle \cdot, \cdot \rangle_v$ on $\mathbb{R}^{nq}$ such that for every right vertex $\gamma = (v_1, \ldots, v_{\mathrm{a}(i)})$ in $\mathcal{K}_{\mathrm{Bl},n}$, for all $1 \leqslant a, b \leqslant \mathrm{a}(i)$, the $nq \times nq$ matrix obtained by placing $M_{v_a \gamma v_b}$ in the (v_a, v_b) block and zeros everywhere else is the adjoint of the $nq \times nq$ matrix obtained by placing $M_{v_b \gamma v_a}$ in the (v_b, v_a) block and zeroes everywhere else under $\langle \cdot, \cdot \rangle_v$.*

2. *There is a constant $C \geqslant 1$ such that the weight M_W every nonbacktracking walk W in $\mathcal{K}_n$ satisfies:*

$$\|M_W\| \leqslant C$$

   where $\| \cdot \|$ is the operator norm induced by $\langle \cdot, \cdot \rangle$.

3. *$\rho(\mathrm{Bl}, \mathrm{M}^{\times}) \geqslant 1$.*

Then for every $\varepsilon > 0$ and $(\log \log n)^2 \leqslant \ell \leqslant \frac{\log n}{(\log \log n)^2}$, with probability $1 - o_n(1)$:

$$|\lambda|_{\max}\left(\underline{A}_{\boldsymbol{H}}^{(\ell)}\right) \leqslant \left((1 + \varepsilon)\rho(\mathrm{Bl}, \mathrm{M}^{\times})\right)^{\ell}.$$

3.7.1 Proof of Theorem 3.7.4

The proof of Theorem 3.7.4 is via the *trace method*. One preliminary observation is that 1 implies that $\underline{A}_{\boldsymbol{H}}^{(\ell)}$ is self-adjoint and hence all its eigenvalues are real.

Consequently, for any positive even integer k:

$$\|\underline{A}_{\boldsymbol{H}}^{(\ell)}\|^k \leqslant \mathsf{Tr}\left(\left(\underline{A}_{\boldsymbol{H}}^{(\ell)}\right)^k\right). \tag{3.12}$$

Our goal is now to obtain a handle on $\boldsymbol{S} := \mathsf{Tr}\left(\left(\underline{A}_{\boldsymbol{H}}^{(\ell)}\right)^k\right)$ and obtain a high probability bound on it. We borrow some terminology from [MOP20]:

Definition 3.7.5 (Linkages)**.** A $(k \times 2\ell)$-nonbacktracking Bl-linkage is a length-$2k\ell$ closed walk in $\mathsf{Bip}(\mathcal{K}_n)$ that starts and ends in the left vertex set and can be expressed as a concatenation of k nonbacktracking walks of length-2ℓ each. Each length-2ℓ nonbacktracking segment is called a *link*. We use $\mathsf{Lkgs}(\mathsf{Bl}, n, k, \ell)$ to denote the collection of all $(k \times 2\ell)$-nonbacktracking Bl-linkages.

Definition 3.7.6. Given a $(k \times 2\ell)$-nonbacktracking Bl-linkage W, we use $L(W)$ to denote the set of left vertices visited by W, $R(W)$ to denote the set of right vertices visited by W, $V(W)$ to denote $L(W) \cup R(W)$, $E(W)$ to denote the set of edges visited by W, and $G(W)$ to denote the graph $(V(W), E(W))$ induced by W.

With the above terminology and notation in hand, we can write $\boldsymbol{S}$ as:

$$\boldsymbol{S} = \sum_{W \in \mathsf{Lkgs}(\mathsf{Bl},n,k,\ell)} \mathsf{Tr}(M_W) \prod_{t=1}^{k\ell} \left(\mathbf{1}[\gamma_t(W) \in \boldsymbol{H}] - \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} \mathbf{1}[\gamma_t(W) \in \boldsymbol{H}]\right).$$

A natural strategy to obtaining a high probability bound on $\boldsymbol{S}$ is to bound $\mathbf{E}[\boldsymbol{S}]$ by some Z and use Markov's inequality to conclude that $\boldsymbol{S}$ is bounded by, say, nZ with high probability. However, $\mathbf{E}[\boldsymbol{S}]$ is not as small as we would hope due to blowing up in magnitude owing to the occurrences of certain rare and problematic subgraphs in $\mathsf{Bip}(\boldsymbol{H})$. So this suggests a natural tweak of conditioning away these rare subgraphs and trying to carry out the same strategy. This tweak is an idea that occurs in many previous papers in the line of work on getting eigenvalue bounds on sparse random matrices [Fri03b, BLM15, Bor19, BC19, MOP20]. These problematic subgraphs all share one common trait – having multiple cycles in a small neighborhood.

Definition 3.7.7. We say a graph Γ is *r-bicycle free* if for every vertex v, the radius-r ball around v contains at most one cycle. We say Γ is an r-bicycle if it has at most r edges and has at least two cycles, and we say Γ is an r-bicycle frame if it is an r-bicycle such that no subgraph of it is an r-bicycle.

Lemma 3.7.8. *With probability* $1 - o_n(1)$, $\mathrm{Bip}(\boldsymbol{H})$ *is* r*-bicycle free for* $r = \frac{\log n}{\log \log n}$.

We refer the reader to Corollary 3.12.9 for a proof of this fact.

Henceforth, we use $\boldsymbol{\mathcal{E}}$ to denote the event that $\boldsymbol{H}$ is r-bicycle free for $r = \frac{\log n}{\log \log n}$. Now define $\boldsymbol{U} \coloneqq \boldsymbol{S} \cdot \mathbf{1}[\boldsymbol{\mathcal{E}}]$. By Lemma 3.7.8 with probability $1 - o_n(1)$, $\boldsymbol{S} = \boldsymbol{U}$ so if we can prove that $\boldsymbol{U} \leqslant Z$ with probability $1 - o_n(1)$, we can also show that $\boldsymbol{S} \leqslant Z$ with probability $1 - o_n(1)$. Thus, we turn our attention to bounding $\mathbf{E}[\boldsymbol{U}]$.

$$\boldsymbol{U} = \sum_{W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)} \mathsf{Tr}(M_W) \prod_{t=1}^{k\ell} \left(\mathbf{1}[\gamma_t(W) \in \boldsymbol{H}] - \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} \mathbf{1}[\gamma_t(W) \in \boldsymbol{H}] \right) \mathbf{1}[\boldsymbol{\mathcal{E}}]. \quad (3.13)$$

We now study the quantity $\prod_{t=1}^{k\ell} \left(\mathbf{1}[\gamma_t(W) \in \boldsymbol{H}] - \mathbf{E}_{\boldsymbol{H}|\boldsymbol{\tau}} \mathbf{1}[\gamma_t(W) \in \boldsymbol{H}] \right)$.

Definition 3.7.9. For a given right vertex γ of $\mathcal{K}_n$ the *multiplicity* $m_W(\gamma)$ of γ in W is the number of times γ is visited by W. $S(W)$ denotes the set of all right vertices that are visited exactly once and are called *singleton right vertices*. $D(W)$ denotes the set of all right vertices that are visited more than once, and are called *duplicative right vertices*.

Henceforth, we shorten $\mathbf{1}[\gamma \in \boldsymbol{H}]$ to $\mathbf{1}_\gamma$ and $\mathbf{E}_{\boldsymbol{H}|\boldsymbol{\tau}} \mathbf{1}[\gamma \in \boldsymbol{H}]$ to μ_γ. Thus:

$$\begin{aligned}
&\prod_{t=1}^{k\ell} \left(\mathbf{1}_{\gamma_t(W)} - \mu_{\gamma_t(W)} \right) \\
&= \prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in D(W)} (\mathbf{1}_\gamma - \mu_\gamma)^{m_W(\gamma)} \\
&= \prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in D(W)} \left(\mathbf{1}_\gamma \cdot \sum_{i=1}^{m_W(\gamma)} (-\mu_\gamma)^{m_W(\gamma)-i} \binom{m_W(\gamma)}{i} + (-\mu_\gamma)^{m_W(\gamma)} \right)
\end{aligned}$$

To lighten notation, we use α_γ to denote $\sum_{i=1}^{m_W(\gamma)} (-\mu_\gamma)^{m_W(\gamma)-i} \binom{m_W(\gamma)}{i}$.

$$\begin{aligned}
&= \prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in D(W)} (\mathbf{1}_\gamma \alpha_\gamma + (-\mu_\gamma)^{m_W(\gamma)}) \\
&= \prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \sum_{L \subseteq D(W)} \prod_{\gamma \in L} \mathbf{1}_\gamma \alpha_\gamma \prod_{\gamma \in D(W) \setminus L} (-\mu_\gamma)^{m_W(\gamma)} \\
&= \sum_{L \subseteq D(W)} \prod_{\gamma \in L} \alpha_\gamma \prod_{\gamma \in D(W) \setminus L} (-\mu_\gamma)^{m_W(\gamma)} \prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in L} \mathbf{1}_\gamma. \quad (3.14)
\end{aligned}$$

Plugging in (3.14) into (3.13) gives:

$$\boldsymbol{U} = \sum_{W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)} \mathsf{Tr}(M_W) \sum_{L \subseteq D(W)} \prod_{\gamma \in L} \alpha_\gamma \prod_{\gamma \in D(W) \setminus L} (-\mu_\gamma)^{m_W(\gamma)} \prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in L} \mathbf{1}_\gamma \mathbf{1}[\boldsymbol{\mathcal{E}}]$$

We are interested in understanding $\mathbf{E}[\boldsymbol{U}]$. Note that this is equal to $\mathbf{E}_{\boldsymbol{\tau}} \mathbf{E}_{\boldsymbol{H}|\boldsymbol{\tau}}[\boldsymbol{U}]$. We will first focus our attention on understanding $\mathbf{E}_{\boldsymbol{H}|\boldsymbol{\tau}}[\boldsymbol{U}]$. We have:

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}}[\boldsymbol{U}] = \\
&\sum_{W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)} \mathsf{Tr}(M_W) \sum_{L \subseteq D(W)} \prod_{\gamma \in L} \alpha_\gamma \prod_{\gamma \in D(W) \setminus L} (-\mu_\gamma)^{m_W(\gamma)} \\
&\mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} \left[\prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in L} \mathbf{1}_\gamma \mathbf{1}[\boldsymbol{\mathcal{E}}] \right] \\
&\leqslant \sum_{W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)} \mathsf{Tr}(M_W) \sum_{L \subseteq D(W)} \prod_{\gamma \in L} |\alpha_\gamma| \prod_{\gamma \in D(W) \setminus L} \mu_\gamma^{m_W(\gamma)} \\
&\left| \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} \left[\prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in L} \mathbf{1}_\gamma \mathbf{1}[\boldsymbol{\mathcal{E}}] \right] \right| .
\end{aligned}$$

Notice that $\alpha_\gamma = (1-\mu_\gamma)^{m_W(\gamma)} - (-\mu_\gamma)^{m_W(\gamma)}$ and hence $|\alpha_\gamma| \leqslant (1-\mu_\gamma)^{m_W(\gamma)} + \mu_\gamma^{m_W(\gamma)} \leqslant (1-\mu_\gamma) + \mu_\gamma = 1$ where the second inequality is a consequence of $\mu_\gamma \in [0,1]$. The result is:

$$\begin{aligned}
\mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}}[\boldsymbol{U}] \leqslant &\sum_{W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)} \mathsf{Tr}(M_W) \sum_{L \subseteq D(W)} \prod_{\gamma \in D(W) \setminus L} \mu_\gamma^{m_W(\gamma)} \cdot \\
&\left| \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} \left[\prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in L} \mathbf{1}_\gamma \mathbf{1}[\boldsymbol{\mathcal{E}}] \right] \right| . \qquad (3.15)
\end{aligned}$$

Next, we would like to obtain a bound on $\left| \mathbf{E}_{\boldsymbol{H}|\boldsymbol{\tau}} \left[\prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in L} \mathbf{1}_\gamma \mathbf{1}[\boldsymbol{\mathcal{E}}] \right] \right|$.

Towards doing so, we first set up a couple of definitions and an observation.

Definition 3.7.10 (Closure of subgraph)**.** Given a subgraph Γ of $\mathrm{Bip}(\mathcal{K}_n)$ with right vertex set $R(\Gamma)$, we define its *closure* $\mathrm{Clos}(\Gamma)$ as the induced subgraph on the vertex set $\left(\bigcup_{\gamma \in R(\Gamma)} N(\gamma)\right) \bigcup R(\Gamma)$. We say Γ is *closed* if $\mathrm{Clos}(\Gamma) = \Gamma$.

Definition 3.7.11 (Excess)**.** Given a graph Γ on e edges, v vertices and c connected components, we define the *excess* of Γ, denoted $\mathrm{Exc}(\Gamma)$, to be $e - v + c$.

The following is immediate from the observation that the excess of a graph cannot decrease on adding a new vertex or a new edge.

Lemma 3.7.12. *If $\Gamma = (V, E)$ and $\Gamma' = (V', E')$ are two graphs such that Γ is a subgraph of Γ', i.e. $V \subseteq V'$ and $E \subseteq E'$, then $\mathrm{Exc}(\Gamma') \geqslant \mathrm{Exc}(\Gamma)$.*

If $\mathrm{Clos}(L)$ is not r-bicycle free, then $\prod_{\gamma \in L} \mathbf{1}_\gamma \mathbf{1}[\boldsymbol{\mathcal{E}}]$ is equal to 0. Otherwise Lemma 3.12.11 then shows that:

$$\left| \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} \left[\prod_{\gamma \in S(W)} (\mathbf{1}_\gamma - \mu_\gamma) \prod_{\gamma \in L} \mathbf{1}_\gamma \mathbf{1}[\boldsymbol{\mathcal{E}}] \right] \right| \leqslant \prod_{\gamma \in S \cup L} \mu_\gamma \cdot 2^{|S(W)|} \left(\frac{1}{n^{.5}} \right)^{\frac{|S(W)|}{r} - \mathrm{Exc}(\mathrm{Clos}(S(W) \cup L))}.$$

Plugging the above into (3.15) tells us:

$$\mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}}[\boldsymbol{U}] \leqslant \sum_{W \in \mathrm{Lkgs}(\mathrm{Bl}, n, k, \ell)} \mathsf{Tr}(M_W) \sum_{\substack{L \subseteq D(W) \\ \mathrm{Clos}(L)\ r\text{-bicycle free}}} \prod_{\gamma \in R(W)} \mu_\gamma \prod_{\gamma \in D(W) \setminus L} \mu_\gamma^{m_W(\gamma) - 1} 2^{|S(W)|} \left(\frac{1}{n^{.5}} \right)^{\frac{|S(W)|}{r} - \mathrm{Exc}(\mathrm{Clos}(S(W) \cup L))}.$$

Henceforth, we will shorten $\mathrm{Exc}(\mathrm{Clos}(S(W) \cup D(W)))$ to Exc_W for simplicity of notation. Since $\mathrm{Clos}(S(W) \cup L)$ is a subgraph of $\mathrm{Clos}(S(W) \cup D(W))$, by Lemma 3.7.12 we have $\mathrm{Exc}_W \geqslant \mathrm{Exc}(\mathrm{Clos}(S(W) \cup L))$, which means:

$$\mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}}[\boldsymbol{U}] \leqslant \sum_{W \in \mathrm{Lkgs}(\mathrm{Bl}, n, k, \ell)} \mathsf{Tr}(M_W)$$

$$\sum_{\substack{L\subseteq D(W)\\ \mathrm{Clos}(L)\ r\text{-bicycle free}}} \prod_{\gamma\in R(W)} \mu_\gamma \prod_{\gamma\in D(W)\setminus L} \mu_\gamma^{m_W(\gamma)-1} 2^{|S(W)|} \left(\frac{1}{n^{.5}}\right)^{\frac{|S(W)|}{r}-\mathrm{Exc}_W}$$

$$= \sum_{W\in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)} \mathsf{Tr}(M_W)\cdot 2^{|S(W)|}\left(\frac{1}{n^{.5}}\right)^{\frac{|S(W)|}{r}-\mathrm{Exc}_W}$$

$$\prod_{\gamma\in R(W)} \mu_\gamma \sum_{\substack{L\subseteq D(W)\\ \mathrm{Clos}(L)\ r\text{-bicycle free}}} \prod_{\gamma\in D(W)\setminus L} \mu_\gamma^{m_W(\gamma)-1} \tag{3.16}$$

Next, we focus on bounding $\sum_{\substack{L\subseteq D(W)\\ L\ r\text{-bicycle free}}} \prod_{\gamma\in D(W)\setminus L} \mu_\gamma^{m_W(\gamma)-1}$. For starters, observe that:

$$\sum_{\substack{L\subseteq D(W)\\ \mathrm{Clos}(L)\ r\text{-bicycle free}}} \prod_{\gamma\in D(W)\setminus L} \mu_\gamma^{m_W(\gamma)-1} \leqslant \sum_{\substack{L\subseteq D(W)\\ \mathrm{Clos}(L)\ r\text{-bicycle free}}} \prod_{\gamma\in D(W)\setminus L} \left(\frac{\overline{\phi}_{\max}}{n}\right)^{m_W(\gamma)-1}$$

We proceed to bound this in a manner identical to [BMR21]. We define a weight function w on subsets of $D(W)$ as follows: $w(K) = \sum_{\gamma\in K} m_W(\gamma) - 1$. Choose $D^*(W)$ as a maximum weight subset (according to w) of W such that $\mathrm{Clos}(D^*(W))$ is r-bicycle free, and let $\Delta(W) := w(D(W)) - w(D^*(W))$. Note that for any $L \subseteq D(W)$ such that $\mathrm{Clos}(L)$ is r-bicycle free, $\Delta(W) \leqslant w(D(W)\setminus L)$.

$$\sum_{\substack{L\subseteq D(W)\\ \mathrm{Clos}(L)\ r\text{-bicycle free}}} \prod_{\gamma\in D(W)\setminus L} \left(\frac{\overline{\phi}_{\max}}{n}\right)^{m_W(\gamma)-1}$$

$$\leqslant \sum_{\substack{L\subseteq D(W)\\ \mathrm{Clos}(L)\ r\text{-bicycle free}}} \left(\frac{\overline{\phi}_{\max}}{n}\right)^{w(D(W)\setminus L)}$$

Since $L \subseteq D(W)$ every $m_W(\gamma) - 1 \geqslant 1$. Using this along with $\Delta(W) \leqslant w(D(W)\setminus L)$ we can bound the above by:

$$\leqslant \sum_{L\subseteq D(W)} \left(\frac{\overline{\phi}_{\max}}{n}\right)^{\max\{|D(W)\setminus L|,\Delta(W)\}}$$

$$
\begin{aligned}
&= \sum_{i \leqslant \Delta(W)} \left(\frac{\overline{\phi}_{\max}}{n}\right)^{\Delta(W)} \cdot \binom{|D(W)|}{i} + \\
&\sum_{i > \Delta(W)} \left(\frac{\overline{\phi}_{\max}}{n}\right)^{i} \binom{|D(W)|}{i} \\
&\leqslant (\Delta(W)+1)\left(\frac{\overline{\phi}_{\max}}{n}\right)^{\Delta(W)} + \sum_{i > \Delta(W)} \left(\frac{\overline{\phi}_{\max}|D(W)|}{n}\right)^{i} \\
&\leqslant (\Delta(W)+2)\left(\frac{\overline{\phi}_{\max}}{n}\right)^{\Delta(W)} \\
&\leqslant 2\left(\frac{2\overline{\phi}_{\max}}{n}\right)^{\Delta(W)}
\end{aligned}
$$

Plugging this back into (3.16) gives us:

$$
\leqslant 2 \sum_{W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)} \mathrm{Tr}(M_W) \cdot 2^{|S(W)|} \cdot \left(\frac{1}{n^{.5}}\right)^{\frac{|S(W)|}{r} - \mathrm{Exc}_W} \cdot \left(\frac{2\overline{\phi}_{\max}}{n}\right)^{\Delta(W)} \cdot \prod_{\gamma \in R(W)} \mu_\gamma \tag{3.16}
$$

As a first step towards simplifying the above quantity we make the following definition:

Definition 3.7.13 (Shape of a linkage)**.** Given a $(k \times 2\ell)$-nonbacktracking linkage $W = v_0 v_1 \dots v_{2k\ell}$ that visits v distinct vertices, we say the *shape* of W denoted $\mathrm{Sh}(W)$ is the $(k \times 2\ell)$-nonbacktracking linkage on graph on vertex set $[v]$ obtained by first constructing map $\xi : V(W) \to [2k\ell]$ where $\xi(v) = i$ where v is the i-th distinct vertex visited by W and defining the t-th step of the walk $\mathrm{Sh}(W)$ to be $\xi(v_{t-1})\xi(v_t)$. We say the left vertex set of $\mathrm{Sh}(W)$ is $\xi(L(W))$ and the right vertex set is $\xi(R(W))$. For $a \in V(\mathrm{Sh}(W))$ we will use the notation $W[a]$ to denote $\xi^{-1}(a)$. We use $\mathrm{Shps}(k,\ell)$ to denote the set of all distinct shapes of linkages in $\mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)$, and $\mathrm{Shps}(k,\ell,v,e)$ to denote the set of all shapes in $\mathrm{Shps}(k,\ell)$ on v vertices and e edges.

We can rewrite the bound on (3.16) as:

$$
(3.16) \leqslant 2 \sum_{\mathrm{Sh} \in \mathrm{Shps}(k,\ell)} 2^{|S(\mathrm{Sh})|} \cdot \left(\frac{1}{n^{.5}}\right)^{\frac{|S(\mathrm{Sh})|}{r} - \mathrm{Exc}_{\mathrm{Sh}}} \cdot
$$

$$\left(\frac{2\overline{\phi}_{\max}}{n}\right)^{\Delta(\mathrm{Sh})} \sum_{\substack{W:W\in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell)\\ \mathrm{Sh}(W)=\mathrm{Sh}}} \mathsf{Tr}(M_W)\cdot \prod_{\gamma\in R(W)} \mu_\gamma. \tag{3.17}$$

For a given linkage W, we begin by deriving an upper bound on $\mathsf{Tr}(M_W)$. A preliminary observation is:

Observation 3.7.14. $\mathsf{Tr}(M_W) \leqslant q\|M_W\|$.

Our next step is to decompose W into simpler "subwalks". This segment of the argument follows [BC19, OW20]

Definition 3.7.15. We call a vertex v in $L(W)$ a *landmark* of W if it satisfies at least one of the following conditions: (i) is an endpoint of a link, (ii) $\deg_{G(W)}(v) \geqslant 3$, (iii) $\deg_{G(W)}(w) \geqslant 3$ for some $w \in R(W)$ which is incident to v within $G(W)$. We refer to the set of all landmark vertices in W as $\mathrm{Lm}(W)$ We call any path between two landmark vertices v_1 and v_2 with no intermediate landmark vertices a *trail*. We call a trail a *forked trail* if it has an intermediate vertex w in $R(W)$ such that $\deg_{G(W)}(v) \geqslant 3$, and an *unforked trail* otherwise. We use $\mathrm{Trs}(W)$ to denote the collection of all trails in W, $\mathrm{UTrs}(W)$ to denote the collection of all unforked trails in W and $\mathrm{FTrs}(W)$ to denote the collection of all forked trails in W.

Observation 3.7.16. Any forked trail must be a single two-step of the form $u\gamma v$ where u and v are left vertices and γ is a right vertex.

Any $W \in \mathrm{Lkgs}(\mathrm{Bl}, n, k, \ell)$ is a sequence of nonbacktracking walks on trails. W can be written as the sequence of vertices visited $v_0\gamma_1 v_1 \dots \gamma_{k\ell} v_{k\ell}$. Let T be the set of all times t such that v_t is a landmark. Using T, we construct a set of *pause* times P in the following way:

> For each $t \in T$, if the trail starting or ending at time t is visited for the first or second time, we add t to P.

Recall from Definition 3.7.2 that M_W is the product of $k\ell$ matrices $M_{v_0\gamma_1 v_1} \dots M_{v_{k\ell-1}\gamma_{k\ell}v_{k\ell}}$. Let $p_1 < \dots < p_s$ be the sequence of all pause times. By submultiplicativity of the operator norm,

$$\|M_W\| \leqslant \|M_{v_0\gamma_1 v_1}\cdots M_{v_{p_1-1}\gamma_{p_1}v_{p_1}}\| \cdot \|M_{v_{p_1}\gamma_{p_1+1}v_{p_1+1}}\cdots M_{v_{p_2-1}\gamma_{p_2}v_{p_2}}\| \cdots \|M_{v_{p_s}\gamma_{p_s+1}v_{p_s+1}} \cdots M_{v_{k\ell-1}\gamma_{k\ell}v_{k\ell}}\|.$$

Each segment between consecutive pauses p_i and p_{i+1} falls into one of the following categories:

- $\mathrm{Seg}_{\leqslant 2}(W)$: the segment is composed of exactly one trail and is the first or second visit to the trail.
- $\mathrm{Seg}_{>2}(W)$ the segment is a union of trails and each of these trails has already been visited at least twice before.

Rewriting the above upper bound, we now have:

$$\|M_W\| \leqslant \prod_{\omega \in \mathrm{Seg}_{\leqslant 2}(W)} \|M_\omega\| \cdot \prod_{\omega \in \mathrm{Seg}_{>2}(W)} \|M_\omega\|.$$

Given a trail T with endpoints u and v, there are two nonbacktracking walks ω_1 and ω_2 that cover T, one from u to v and another from v to u. By 1, $M_{\omega_1}^* = M_{\omega_2}$ where the $*$ in the superscript refers to the adjoint induced by the inner product $\langle \cdot, \cdot \rangle_v$ and so $\|M_{\omega_1}\| = \|M_{\omega_2}\|$. Henceforth we use $\|M_T\|$ to denote $\|M_{\omega_1}\| = \|M_{\omega_2}\|$. Using the notation $\mathrm{UTrs}_{\geqslant 2}(W)$ for unforked trails that are visited more than once, we can write the above as:

$$\|M_W\| \leqslant \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot \prod_{\omega \in \mathrm{Seg}_{>2}(W) \cup \mathrm{Seg}_1(W) \setminus \mathrm{UTrs}_{\geqslant 2}(W)} \|M_\omega\|.$$

2 further lets us get the following bound:

$$\|M_W\| \leqslant C^{|\mathrm{Seg}_1(W)| + |\mathrm{Seg}_{\geqslant 2}(W)|} \cdot \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \leqslant C^{|P|+1} \cdot \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \tag{3.18}$$

We now turn our attention to bounding $|P|$. Given a landmark vertex v and an edge e incident to it, there are at most $\mathrm{a}_{\max}$ trails that start at v and tread on e on their first step, and hence the number of distinct trails starting at v is at most $\mathrm{a}_{\max} \cdot \deg_{G(W)}(v)$. By the construction of P, the number of pauses at vertex v is at most twice the number of distinct trails starting at v, and hence is at most $2\mathrm{a}_{\max} \cdot \deg_{G(W)}(v)$. Thus:

$$|P| \leqslant 2\mathrm{a}_{\max} \sum_{v \in \mathrm{Lm}(W)} \deg_{G(W)}(v) \leqslant 2\mathrm{a}_{\max} \left(2|\mathrm{Lm}_{\leqslant 2}(W)| + \sum_{v \in \mathrm{Lm}_{\geqslant 3}(W)} \deg_{G(W)}(v) \right)$$

where $\mathrm{Lm}_{\leqslant 2}(W)$ and $\mathrm{Lm}_{\geqslant 3}(W)$ denote the sets of landmark vertices of degree-$\leqslant 2$ and degree-$\geqslant 3$ respectively. Each vertex in $\mathrm{Lm}_{\leqslant 2}(v)$ is either an endpoint of a link or a neighbor of a degree-$\geqslant 3$ right vertex. There are exactly $k+1$ endpoints of links, and at most $\sum_{v\in G(W):\deg_{G(W)}(v)\geqslant 3}\deg_{G(W)}(v)$ landmark vertices induced as neighbors of degree-$\geqslant 3$ right vertices, and hence:

$$\begin{aligned}&|P|\\&\leqslant 2\mathrm{a}_{\max}\cdot\\&\left(2(k+1)+2\sum_{v\in G(W):\deg_{G(W)}(v)\geqslant 3}\deg_{G(W)}(v)+\sum_{v\in G(W):\deg_{G(W)}(v)\geqslant 3}\deg_{G(W)}(v)\right)\end{aligned}\tag{3.19}$$

It remains to bound $\sum_{v\in G(W):\deg_{G(W)}(v)\geqslant 3}\deg_{G(W)}(v)$. Let X be a set of edges of size Exc_W such that $T(W,X) := (V(W), E(W)\setminus X)$ is a tree. Since $G(W)$ has at most k leaves, $T(W,X)$ has at most $k+2\mathrm{Exc}_W$ leaves. We now state the following well known fact about trees and refer the reader to [BMR21, Fact 6.35] for a proof.

Fact 3.7.17. *Let T be a tree with l leaves. Then $3l \geqslant \sum_{v\in V(T):\deg_T(v)\geqslant 3}\deg_T(v)$.*

As a consequence of Fact 3.7.17:

$$\sum_{v\in V(T(W,X)):\deg_{T(V,X)}(v)\geqslant 3}\deg_{T(V,X)}(v)\leqslant 3(k+2\mathrm{Exc}_W).$$

Now observe that for any graph Γ and graph Γ' obtained by adding a single edge to Γ:

$$\sum_{v\in V(\Gamma'):\deg_{\Gamma'}(v)\geqslant 3}\deg_{\Gamma'}(v)\leqslant\left(\sum_{v\in V(\Gamma):\deg_{\Gamma}(v)\geqslant 3}\deg_{\Gamma}(v)\right)+6,$$

and thus

$$\sum_{v\in G(W):\deg_{G(W)}(v)\geqslant 3}\deg_{G(W)}(v)\leqslant 3(k+2\mathrm{Exc}_W)+6\mathrm{Exc}_W=3k+12\mathrm{Exc}_W.\tag{3.20}$$

Plugging this back into (3.19) and using $k\geqslant 2$ gives:

$$|P|\leqslant 2\mathrm{a}_{\max}\left(12k+36\mathrm{Exc}_W\right)=24\mathrm{a}_{\max}\left(k+3\mathrm{Exc}_W\right).$$

Remark 3.7.18. Observe that the above also proved the following, which will be of utility later in the proof:

$$\sum_{v \in \mathrm{Lm}(W)} \deg_{G(W)}(v) \leqslant 12k + 36\mathrm{Exc}_W.$$

Next, plugging the bound on $|P|$ back into (3.18) along with Observation 3.7.14 gives:

$$\mathsf{Tr}(M_W) \leqslant qC^{24\mathrm{a}_{\max}(k+3\mathrm{Exc}_W)+1}.$$
$$\prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \leqslant qC^{25\mathrm{a}_{\max}(k+3\mathrm{Exc}_W)} \cdot \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2.$$

And via (3.17) and introducing the expected value over the randomness of $\boldsymbol{\tau}$:

$$\mathop{\mathbf{E}}_{\boldsymbol{\tau}} \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} [\boldsymbol{U}] \leqslant 2q \sum_{\mathrm{Sh} \in \mathrm{Shps}(k,\ell)} \left(\frac{2^r}{n^{.5}}\right)^{\frac{|S(\mathrm{Sh})|}{r}} \cdot \left(n^{.5} C^{75\mathrm{a}_{\max}}\right)^{\mathrm{Exc}_{\mathrm{Sh}}} C^{25\mathrm{a}_{\max}k} \cdot \left(\frac{2\overline{\phi}_{\max}}{n}\right)^{\Delta(\mathrm{Sh})} \sum_{\substack{W : W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell) \\ \mathrm{Sh}(W) = \mathrm{Sh}}} \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot \mathop{\mathbf{E}}_{\boldsymbol{\tau}} \prod_{\gamma \in R(W)} \mu_\gamma \tag{3.21}$$

Thus, for a fixed shape Sh we now restrict our attention to bounding:

$$\sum_{\substack{W : W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell) \\ \mathrm{Sh}(W) = \mathrm{Sh}}} \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot \mathop{\mathbf{E}}_{\boldsymbol{\tau}} \prod_{\gamma \in R(W)} \mu_\gamma. \tag{3.22}$$

We now define the notion of Cl-consistent.

Definition 3.7.19. Let Γ be a subgraph of the complete bipartite factor graph $\mathrm{Bip}(\mathcal{K}_n)$. We say Γ is Cl-consistent if there exists a τ such that every

$$\gamma = (v_1, \ldots, v_{\mathrm{a}(i)}) \in R(\Gamma)$$

satisfies $(\tau(v_1), \ldots, \tau(v_{\mathrm{a}(i)})) = \mathrm{Cl}(\theta(\gamma))$. If Γ is Cl-consistent we use τ_Γ to refer to the unique τ such that Γ is (τ, Cl)-consistent.

Observation 3.7.20. For $W \in \mathrm{Lkgs}(\mathrm{Bl}, n, k, \ell)$, $\prod_{\gamma \in R(W)} \mu_\gamma$ is equal to 0 if W is not Cl-consistent.

Thus:

$$(3.22) = \sum_{\substack{W: W \in \mathrm{Lkgs}(\mathrm{Bl},n,k,\ell) \\ \mathrm{Sh}(W)=\mathrm{Sh} \\ W \text{ Cl-consistent}}} \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot$$

$$\prod_{\gamma \in R(W)} \frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}} \cdot \prod_{v \in L(\mathrm{Clos}(W))} \mathcal{T}(\tau_W(v)).$$

Definition 3.7.21. We call two walks W_1 and W_2 *equivalent* denoted $W_1 \sim W_2$ if

- $\mathrm{Sh}(W_1) = \mathrm{Sh}(W_2) =: \mathrm{Sh}$,
- for any $a \in R(\mathrm{Sh})$, $\theta(W_1[a]) = \theta(W_2[a])$,
- for any edge $\{v, \gamma\}$ in Sh for $v \in L(\mathrm{Sh})$ and $\gamma \in R(\mathrm{Sh})$, $\mathrm{i}(W_1[\gamma], W_1[v]) = \mathrm{i}(W_2[\gamma], W_2[v])$.

We say W_1 and W_2 are *closure equivalent* denoted $W_1 \sim_{\mathrm{Clos}} W_2$ if $W_1 \sim W_2$ and the graphs induced by $\mathrm{Clos}(W_1)$ and $\mathrm{Clos}(W_2)$ are isomorphic.

The relationship $\sim$ partitions the space of all Cl-consistent W in $\mathrm{Lkgs}(\mathrm{Bl}, n, k, \ell)$ with shape Sh into a collection of equivalence classes $\mathcal{C}$. We use $[W]$ to denote the equivalence class it is contained in. $\sim_{\mathrm{Clos}}$ further partitions each equivalence class $[W] \in \mathcal{C}$ into a collection of sub-equivalence classes $\mathcal{C}_{[W]}$, and we denote the sub-equivalence class of W with $[[W]]$. We use $J(W)$ to denote $|E(\mathrm{Clos}(W))| - |E(W)|$. With these definitions and notation in hand, we can write:

$$(3.22) =$$

$$\sum_{[W] \in \mathcal{C}} \sum_{[[W]] \in \mathcal{C}[W]} \sum_{W' \in [[W]]} \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot \prod_{\gamma \in R(W)} \frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}} \cdot$$
$$\prod_{v \in L(\mathrm{Clos}(W))} \mathcal{T}(\tau_W(v))$$

$$\leqslant \sum_{[W] \in \mathcal{C}} \sum_{[[W]] \in \mathcal{C}[W]} \prod_{T \in \mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot \prod_{\gamma \in R(W)} \frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}} \cdot$$
$$\prod_{v \in L(\mathrm{Clos}(W))} \mathcal{T}(\tau_W(v)) \cdot n^{|L(\mathrm{Clos}(W))|}$$

$$= \sum_{[W]\in\mathcal{C}} \sum_{t=0}^{J(W)} \sum_{\substack{[[W]]\in\mathcal{C}[W] \\ |L(\mathrm{Clos}(W))|=|L(W)|+t}} \prod_{T\in\mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot \prod_{\gamma\in R(W)} \frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}} \cdot$$
$$\prod_{v\in L(\mathrm{Clos}(W))} \mathcal{T}(\tau_W(v)) \cdot \frac{n^{|L(W)|+|J(W)|}}{n^{|J(W)|-t}}$$

To enumerate the innermost sum, first observe that any W can be changed to $\mathrm{Clos}(W)$ where $|L(\mathrm{Clos}(W))| - |L(W)| = t$ via the following procedure. First add $J(W)$ new vertices and to each $\gamma \in R(W)$ attach an edge from γ to $\mathrm{a}(\theta(\gamma)) - \deg_W(\gamma)$ of the new vertices. There exists a sequence of $J(W) - t$ "merge" operations on the left vertices, and a labeling of the newly added left vertices in $[n]$ that would result in $\mathrm{Clos}(W)$. The number of possible sequences of merge operations is at most $(\mathrm{a}_{\max}k\ell)^{2(|J(W)|-t)}$. Thus, the above sum can be bounded by:

$$\leqslant \sum_{[W]\in\mathcal{C}} \sum_{t=0}^{J(W)} \prod_{T\in\mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot \prod_{\gamma\in R(W)} \frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}} \cdot$$
$$\prod_{v\in L(\mathrm{Clos}(W))} \mathcal{T}(\tau_W(v)) \cdot n^{|L(W)|+|J(W)|} \cdot$$
$$\left(\frac{(\mathrm{a}_{\max}k\ell)^2}{n}\right)^{|J(W)|-t}$$
$$= \sum_{[W]\in\mathcal{C}} \sum_{t=0}^{J(W)} \prod_{T\in\mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot$$
$$\prod_{\gamma\in R(W)} \left(\frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}} \cdot \prod_{v\in L(\mathrm{Clos}(W))\setminus L(W)} \mathcal{T}(\tau_W(v))\right) \cdot$$
$$\prod_{v\in L(W)} \mathcal{T}(\tau_W(v)) \cdot$$
$$\prod_{v\in L(\mathrm{Clos}(W))\setminus L(w)} \frac{1}{\mathcal{T}(\tau_W(v))^{\deg_W(v)-1}} \cdot n^{|L(W)|+|J(W)|} \cdot \left(\frac{(\mathrm{a}_{\max}k\ell)^2}{n}\right)^{|J(W)|-t}$$
$$\leqslant \sum_{[W]\in\mathcal{C}} \sum_{t=0}^{J(W)} \prod_{T\in\mathrm{UTrs}_{\geqslant 2}(W)} \|M_T\|^2 \cdot$$

$$\prod_{\gamma\in R(W)}\left(\frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}}\cdot\prod_{v\in L(\mathrm{Clos}(W))\setminus L(W)}\mathcal{T}(\tau_W(v))\right)\cdot$$
$$\prod_{v\in L(W)}\mathcal{T}(\tau_W(v))\cdot n^{|L(W)|+|J(W)|}\cdot\left(\frac{(\mathrm{a}_{\max}k\ell)^2}{\mathcal{T}_{\min}n}\right)^{|J(W)|-t}$$

We can use the bound $\sum_{t=0}^{J(W)}\left(\frac{(\mathrm{a}_{\max}k\ell)^2}{\mathcal{T}_{\min}n}\right)^{|J(W)|-t}\leqslant 2$ to deduce:

$$\leqslant 2\sum_{[W]\in\mathcal{C}}\prod_{T\in\mathrm{UTrs}_{\geqslant 2}(W)}\|M_T\|^2\cdot$$
$$\prod_{\gamma\in R(W)}\left(\frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}}\cdot\prod_{v\in L(\mathrm{Clos}(W))\setminus L(W)}\mathcal{T}(\tau_W(v))\right)\cdot$$
$$\prod_{v\in L(W)}\mathcal{T}(\tau_W(v))\cdot n^{|L(W)|+|J(W)|}$$
$$=2\sum_{[W]\in\mathcal{C}}\prod_{T\in\mathrm{UTrs}_{\geqslant 2}(W)}\|M_T\|^2\cdot\prod_{\gamma\in R(W)}\left(\overline{\phi_{\theta(\gamma)}}\cdot\prod_{v\in L(\mathrm{Clos}(W))\setminus L(W)}\mathcal{T}(\tau_W(v))\right)\cdot$$
$$\prod_{v\in L(W)}\mathcal{T}(\tau_W(v))\cdot n^{1-\mathrm{Exc}_W}. \tag{3.23}$$

We decompose $\prod_{\gamma\in R(W)}\left(\overline{\phi_{\eta(\gamma)}}\cdot\prod_{v\in L(\mathrm{Clos}(W))\setminus L(W)}\mathcal{T}(\tau_W(v))\right)$ into three parts: the contribution of singleton right vertices

$$\prod_{\gamma\in S(W)}\left(\overline{\phi_{\eta(\gamma)}}\cdot\prod_{v\in L(\mathrm{Clos}(W))\setminus L(W)}\mathcal{T}(\tau_W(v))\right)$$

which can be bounded by $\overline{\phi}_{\max}^{|S(\mathrm{Sh})|}$, the contribution of duplicative degree-$\geqslant 3$ right vertices $\prod_{\substack{\gamma\in D(W)\\ \deg_W(\gamma)\geqslant 3}}\left(\overline{\phi_{\eta(\gamma)}}\cdot\prod_{v\in L(\mathrm{Clos}(W))\setminus L(W)}\mathcal{T}(\tau_W(v))\right)$ which via (3.20) can be bounded by $\overline{\phi}_{\max}^{3k+12\mathrm{Exc}_{\mathrm{Sh}}}$, and finally the contribution of duplicative degree-2 right vertices $\prod_{\substack{\gamma\in D(W)\\ \deg_W(\gamma)=2}}\left(\overline{\phi_{\eta(\gamma)}}\cdot\prod_{v\in L(\mathrm{Clos}(W))\setminus L(W)}\mathcal{T}(\tau_W(v))\right)$. For each γ considered in the final case we can identify a unique $T\in\mathrm{UTrs}_{\geqslant 2}(W)$ such that γ is in T, and likewise for every $T\in\mathrm{UTrs}_{\geqslant 2}(W)$, every right vertex γ in T is duplicative and has

degree exactly 2 and hence appears in the product. Thus, the third product can be written as

$$\prod_{T\in \mathrm{UTrs}_{\geqslant 2}(W)} \prod_{\gamma\in R(T)} \left(\overline{\phi_{\theta(\gamma)}} \cdot \prod_{v\in L(\mathrm{Clos}(T))\setminus L(T)} \mathcal{T}(\tau_W(v)) \right).$$

Next, observe that using the facts that each $\pi_i \in [0,1]$ and every interior left vertex of a trail occurs in no other trail, $\prod_{v\in L(W)} \mathcal{T}(\tau_W(v))$ can be upper bounded by

$$\prod_{T\in \mathrm{UTrs}_{\geqslant 2}(W)} \prod_{v\in L(T)} \mathcal{T}(\tau_W(v)) \cdot \prod_{v\in \mathrm{Lm}(W)} \frac{1}{\mathcal{T}(\tau_W(v))^{\deg_{G(W)}(v)}},$$

which by Remark 3.7.18 is at most

$$\prod_{T\in \mathrm{UTrs}_{\geqslant 2}(W)} \prod_{v\in L(T)} \mathcal{T}(\tau_W(v)) \cdot \left(\frac{1}{\mathcal{T}_{\min}}\right)^{12k+36\mathrm{Exc}_{\mathrm{Sh}}}.$$

$$(3.23) \leqslant 2\overline{\phi}_{\max}^{|S(\mathrm{Sh})|+3k+12\mathrm{Exc}_{\mathrm{Sh}}} \left(\frac{1}{\mathcal{T}_{\min}}\right)^{12k+36\mathrm{Exc}_{\mathrm{Sh}}} n^{-\mathrm{Exc}_{\mathrm{Sh}}+1}.$$

$$\sum_{[W]} \prod_{T\in \mathrm{UTrs}_{\geqslant 2}(W)} \left(\|M_T\|^2 \cdot \prod_{\gamma\in R(T)} \overline{\phi_{\theta(\gamma)}} \cdot \prod_{v\in L(\mathrm{Clos}(T))\setminus L(T)} \mathcal{T}(\tau_W(v)) \cdot \prod_{v\in L(T)} \mathcal{T}(\tau_W(v)) \right)$$

$$\leqslant 2\overline{\phi}_{\max}^{|S(\mathrm{Sh})|+3k+12\mathrm{Exc}_{\mathrm{Sh}}} \left(\frac{1}{\mathcal{T}_{\min}}\right)^{12k+36\mathrm{Exc}_{\mathrm{Sh}}} n^{-\mathrm{Exc}_{\mathrm{Sh}}+1}.$$

$$\prod_{T\in \mathrm{UTrs}_{\geqslant 2}(\mathrm{Sh})} \sum_{[U]:\mathrm{Sh}(U)=T} \left(\|M_U\|^2 \cdot \prod_{\gamma\in R(U)} \overline{\phi_{\theta(\gamma)}} \cdot \prod_{v\in L(\mathrm{Clos}(U))\setminus L(U)} \mathcal{T}(\tau_W(v)) \cdot \prod_{v\in L(U)} \mathcal{T}(\tau_W(v)) \right)$$

For any constant ε, there exists a constant C_ε such that the above is at most:

$$\leqslant 2\overline{\phi}_{\max}^{|S(\mathrm{Sh})|+3k+12\mathrm{Exc}_{\mathrm{Sh}}}.$$

$$\left(\frac{1}{\mathcal{T}_{\min}}\right)^{12k+36\mathrm{Exc}_{\mathrm{Sh}}} n^{-\mathrm{Exc}_{\mathrm{Sh}}+1} \prod_{T\in \mathrm{UTrs}_{\geqslant 2}(\mathrm{Sh})} C_\varepsilon \cdot \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\right)^{2|T|}$$

Since $|\mathrm{UTrs}_{\geqslant 2}(\mathrm{Sh})|$ is at most the sum of degrees of landmark vertices on which we have an upper bound by Remark 3.7.18, and since $\sum_{T\in \mathrm{UTrs}_{\geqslant 2}(\mathrm{Sh})} |T| \leqslant \frac{k\ell}{2}$:

$$\leqslant 2\overline{\phi}_{\max}^{|S(\mathrm{Sh})|+3k+12\mathrm{Exc}_{\mathrm{Sh}}} \cdot \left(\frac{1}{\mathcal{T}_{\min}}\right)^{12k+36\mathrm{Exc}_{\mathrm{Sh}}} n^{-\mathrm{Exc}_{\mathrm{Sh}}+1} C_\varepsilon^{12k+36\mathrm{Exc}_{\mathrm{Sh}}} \cdot \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\right)^{k\ell}.$$

Since (3.23) is an upper bound on (3.22), rearranging the terms in the above gives:

$$(3.22) \leqslant 2n \cdot \overline{\phi}_{\max}^{|S(\mathrm{Sh})|} \cdot \left(\frac{\overline{\phi}_{\max}^{3} C_\varepsilon^{12}}{\mathcal{T}_{\min}^{12}}\right)^{k} \cdot \left(\frac{\overline{\phi}_{\max}^{12} C_\varepsilon^{36}}{\mathcal{T}_{\min}^{36} n}\right)^{\mathrm{Exc}_{\mathrm{Sh}}} \cdot \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\right)^{k\ell}.$$

Plugging this upper bound on (3.22) into (3.21) gives us:

$$\begin{aligned}
&\mathop{\mathbf{E}}_{\boldsymbol{\tau}} \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} [\boldsymbol{U}] \leqslant \\
&4nq \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\right)^{k\ell} \cdot \\
&\sum_{\mathrm{Sh}\in \mathrm{Shps}(k,\ell)} \left(\frac{2^r \overline{\phi}_{\max}^{r}}{n^{.5}}\right)^{\frac{|S(\mathrm{Sh})|}{r}} \left(\frac{C^{75\mathrm{a}_{\max}} \overline{\phi}_{\max}^{12} C_\varepsilon^{36}}{\mathcal{T}_{\min}^{36} n^{.5}}\right)^{\mathrm{Exc}_{\mathrm{Sh}}} \\
&\left(\frac{C^{25\mathrm{a}_{\max}} \overline{\phi}_{\max}^{3} C_\varepsilon^{12}}{\mathcal{T}_{\min}^{12}}\right)^{k} \left(\frac{2\overline{\phi}_{\max}}{n}\right)^{\Delta(\mathrm{Sh})}.
\end{aligned}$$

To notationally lighten the above, we choose β as a constant larger than $2\overline{\phi}_{\max}$, $\frac{C^{75\mathrm{a}_{\max}} \overline{\phi}_{\max}^{12} C_\varepsilon^{36}}{\mathcal{T}_{\min}^{36}}$ and $\frac{C^{25\mathrm{a}_{\max}} \overline{\phi}_{\max}^{3} C_\varepsilon^{12}}{\mathcal{T}_{\min}^{12}}$. Then:

$$\mathop{\mathbf{E}}_{\boldsymbol{\tau}} \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}} [\boldsymbol{U}] \leqslant 4nq \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\right)^{k\ell} \beta^{k} \cdot \sum_{\mathrm{Sh}\in \mathrm{Shps}(k,\ell)} \left(\frac{\beta^r}{n^{.5}}\right)^{\frac{|S(\mathrm{Sh})|}{r}} \left(\frac{\beta}{n^{.5}}\right)^{\mathrm{Exc}_{\mathrm{Sh}}} \left(\frac{\beta}{n}\right)^{\Delta(\mathrm{Sh})} \tag{3.24}$$

To obtain a bound on (3.24) we first bound:

$$\sum_{\mathrm{Sh}\in \mathrm{Shps}(k,\ell)} \left(\frac{\beta^r}{n^{.5}}\right)^{\frac{|S(\mathrm{Sh})|}{r}} \left(\frac{\beta}{n^{.5}}\right)^{\mathrm{Exc}_{\mathrm{Sh}}} \left(\frac{\beta}{n}\right)^{\Delta(\mathrm{Sh})} \tag{3.25}$$

We proceed by partition all $\mathrm{Sh} \in \mathrm{Shps}(k,\ell)$ into sets where each set of Sh share the same $|S(\mathrm{Sh})|$, $\mathrm{Exc}_{\mathrm{Sh}}$, and $\Delta(\mathrm{Sh})$. We bound the sum for each set with the following claim proved in Appendix 3.12.

Claim 3.7.22. Let $\mathcal{U}_{s,x,\Delta}$ denote the set of all $\mathrm{Sh} \in \mathrm{Shps}(k,\ell)$ with $|S(\mathrm{Sh})| = s$, $\mathrm{Exc}_{\mathrm{Sh}} = x$, and $\Delta(\mathrm{Sh}) = \Delta$. Then

$$\sum_{\mathrm{Sh}\in\mathcal{U}_{s,x,\Delta}} \left(\frac{\beta^r}{n^{.5}}\right)^{\frac{|S(\mathrm{Sh})|}{r}} \left(\frac{\beta}{n^{.5}}\right)^{\mathrm{Exc}_{\mathrm{Sh}}} \left(\frac{\beta}{n}\right)^{\Delta(\mathrm{Sh})} \leqslant \left(\frac{\beta^r\cdot(4k\ell)^r}{n^{.5}}\right)^{\frac{s}{r}} \left(\frac{\beta\cdot 2(k\ell)^3}{n^{.5}}\right)^{x} \left(\frac{4\beta\cdot(2k\ell)^2}{n}\right)^{\Delta} (2k\ell)^{O(k\log k\ell)}.$$

Using this claim we can derive

(3.25)

$$= \sum_{s\in[k\ell],x\in[2k\ell],\Delta\in[k\ell]} \sum_{\mathrm{Sh}\in\mathcal{U}_{s,x,\Delta}} \left(\frac{\beta^r}{n^{.5}}\right)^{\frac{s}{r}} \left(\frac{\beta}{n^{.5}}\right)^{x} \left(\frac{\beta}{n}\right)^{\Delta}$$

$$\leqslant \sum_{s\in[k\ell],x\in[2k\ell],\Delta\in[k\ell]} \left(\frac{\beta^r\cdot(4k\ell)^r}{n^{.5}}\right)^{\frac{s}{r}} \left(\frac{\beta\cdot 2(k\ell)^3}{n^{.5}}\right)^{x} \left(\frac{4\beta\cdot(2k\ell)^2}{n}\right)^{\Delta} (2k\ell)^{O(k\log k\ell)}$$

$$\leqslant (2k\ell)^{O(k\log k\ell)} k\ell\cdot 2k\ell\cdot k\ell \max_{s\in[k\ell],x\in[2k\ell],\Delta\in[k\ell]} \left(\frac{\beta^r\cdot(4k\ell)^r}{n^{.5}}\right)^{\frac{s}{r}} \left(\frac{\beta\cdot 2(k\ell)^3}{n^{.5}}\right)^{x} \left(\frac{4\beta\cdot(2k\ell)^2}{n}\right)^{\Delta}$$

We set the bicycle-free radius to $r := o(\log n^{.5}/\log 4k\ell\beta)$ so that all three terms $\frac{\beta^r\cdot(4k\ell)^r}{n^{.5}}$, $\frac{\beta\cdot 2(k\ell)^3}{n^{.5}}$, $\frac{4\beta\cdot(2k\ell)^2}{n}$ are less than 1. Then we observe that

$$(3.25) \leqslant 2k^3\ell^3(2k\ell)^{O(k\log k\ell)}.$$

Plugging this bound into (3.24) gives us:

$$\mathop{\mathbf{E}}_{\boldsymbol{\tau}} \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}}[\boldsymbol{U}] \leqslant \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\right)^{k\ell}\cdot 2nq\beta^k 2k^3\ell^3(2k\ell)^{O(k\log k\ell)}$$

Set $k\ell = O(\log n\cdot\log\log n)$, and $\ell = \omega(\log k\ell)$. Then we have the bound

$$\left(n \mathop{\mathbf{E}}_{\boldsymbol{\tau}} \mathop{\mathbf{E}}_{\boldsymbol{H}|\boldsymbol{\tau}}[\boldsymbol{U}]\right)^{1/k} \leqslant \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\cdot\left(4q\beta^k\cdot n^2\cdot(k\ell)\cdot(2k\ell)^{O(\log k\ell)/\ell}\right)^{1/k\ell}\right)^{\ell}$$

$$\leqslant \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\cdot\left(8q\beta^{k}\cdot n^{3}\right)^{1/k\ell}\right)^{\ell}$$
$$\leqslant \left((1+\varepsilon)\rho(\mathrm{Bl},\mathrm{M}^{\times})\cdot\left(1+\frac{2\log(8q\beta^{k})}{k\ell}+\frac{2\log(n^{3})}{k\ell}\right)\right)^{\ell}$$
$$\leqslant \left(\left(1+\varepsilon+O\left(\frac{1}{\log\log n}\right)\right)\rho(\mathrm{Bl},\mathrm{M}^{\times})\right)^{\ell}.$$

We now complete the proof of Theorem 3.7.4.

3.8 Weak Recovery

We begin the section by briefly describing an algorithm for weak recovery.

Weak Recovery Algorithm

1. Fix $\delta > 0$ such that $\lambda_L \geqslant (1+\delta)^4$.
2. $C = O_{\mathrm{M},\delta}(1)$ is a sufficiently large constant depending on model M and δ.
3. For $(\log\log n)^3 \leqslant t \leqslant (\log\log n)^5$, $v_t \in \mathbb{R}^{nq}$ be the eigenvector with largest eigenvalue of $\underline{A}_{G}^{(t)}$ and let Λ_t^t denote the largest eigenvalue. Compute v_t, Λ_t and $\underline{A}_{G}^{(t)}$ for all t in this range.
4. Find $\mathrm{m} \in [(\log\log n)^3, (\log\log n)^5]$ such that for all $s \in [(\log\log n)^3, \mathrm{m}]$, we have
$$\|\underline{A}_{G}^{(s)} v_{\mathrm{m}}\| \leqslant \Lambda_{\mathrm{m}}^{s}(1+\delta)^{\mathrm{m}-s}$$
(see Claim 3.8.6 for proof of existence of m)
5. For each $0 \leqslant \ell \leqslant C$, set
$$w_\ell \stackrel{\mathrm{def}}{=} \underline{A}_{G}^{(\mathrm{m}-\ell)} v_{\mathrm{m}}$$
and let $\overline{w}_\ell \stackrel{\mathrm{def}}{=} \frac{1}{\|w_\ell\|}\cdot w_\ell$.
6. Output the set of vectors $\{\overline{u}_{\ell,\beta}\}$ in $\mathbb{R}^n$ for $0 \leqslant \ell \leqslant C$, $\beta \in [q]$ defined as,
$$\overline{u}_{\ell,\beta}[i] = \overline{w}_\ell[i,\beta]$$

Theorem 3.8.1. *There exists a constant $C = O_{\mathrm{M},\delta}(1)$ depending on model M and δ such that the following holds with probability $1 - o_n(1)$: For some $\ell \in \{1, 2, \ldots, C\}$ and $\beta \in [q]$, the unit vector $\overline{u}_{\ell,\beta}$ is correlated with the coloring in the following sense: $\exists \tau \in T, \alpha \in [q]$ such that if we construct $\underline{\chi}^{\tau,\alpha} \in \mathbb{R}^n$ as*

$$\underline{\chi}^{\tau,\alpha}[i] = \mathbf{1}[\tau(i) = \tau](\mathbf{1}[\boldsymbol{c}(i) = \alpha] - \mathbb{P}_\tau(\alpha))$$

then

$$|\langle \overline{u}_{\ell,\beta}, \underline{\chi}^{\tau,\alpha} \rangle| \geqslant \Omega_{\mathrm{M}}(1) \cdot \sqrt{n}$$

In the rest of the section, we will outline the proof of correctness of the above described weak-recovery algorithm. To this end, we begin by recalling the matrix $\underline{A}_{\boldsymbol{G}} \in \mathbb{R}^{nq \times nq}$. For all $i \neq j \in [n]$,

$$\underline{A}_{\boldsymbol{G}}[i,j] \stackrel{\text{def}}{=} \sum_{e \in \mathcal{K}_n, e \ni ij} \overline{\mathbf{M}}_{e,i|j} \cdot (\mathbf{1}[e \in \boldsymbol{G}] - \underset{\mathrm{M}}{\mathbf{Pr}}[e \in \boldsymbol{G} | \boldsymbol{\tau}])$$

and $\underline{A}_{\boldsymbol{G}}[i,i] = 0$. Similarly, for all $i \neq j \in [n]$ we set

$$\underline{B}_{\boldsymbol{G}}[i,j] \stackrel{\text{def}}{=} \sum_{e \in \mathcal{K}_n, e \ni ij} \overline{\mathbf{M}}_{e,i|j} \cdot (\mathbf{1}[e \in \boldsymbol{G}] - \underset{\mathrm{M}}{\mathbf{Pr}}[e \in \boldsymbol{G} | \boldsymbol{\tau}, \boldsymbol{c}])$$

and $\underline{B}_{\boldsymbol{G}}[i,i] = 0$. Finally, let

$$\begin{aligned} \underline{R}_{\boldsymbol{G}}[i,j] &\stackrel{\text{def}}{=} \underline{A}_{\boldsymbol{G}}[i,j] - \underline{B}_{\boldsymbol{G}}[i,j] \\ &= \sum_{e \in \mathcal{K}_n, e \ni ij} \overline{\mathbf{M}}_{e,i|j} \left(\underset{\mathrm{M}}{\mathbf{Pr}}[e \in \boldsymbol{G} | \boldsymbol{\tau}, \boldsymbol{c}] - \underset{\mathrm{M}}{\mathbf{Pr}}[e \in \boldsymbol{G} | \boldsymbol{\tau}] \right) \end{aligned}$$

Let $\underline{x}_{\boldsymbol{G}} \in \mathbb{R}^{|T|}$ encode the number of variables of each type in $\boldsymbol{G}$ and $\underline{y}_{\boldsymbol{G}} \in \mathbb{R}^{q|T|}$ encode the number of variables of each type and color in $\boldsymbol{G}$. Then block $\underline{R}_{\boldsymbol{G}}[i,j]$ only depends on $\boldsymbol{\tau}(i), \boldsymbol{\tau}(j), \boldsymbol{c}(i), \boldsymbol{c}(j)$ and $\underline{y}_{\boldsymbol{G}}$. More specifically

$$\begin{aligned} &\underline{R}_{\boldsymbol{G}}[i,j] = \\ &\sum_{e \in \mathcal{K}_n, e \ni ij} \overline{\mathbf{M}}_{e,i|j} \left(\underset{\mathrm{M}}{\mathbf{Pr}}[e \in \boldsymbol{G} | \boldsymbol{\tau}(i), \boldsymbol{\tau}(j), \boldsymbol{c}(i), \boldsymbol{c}(j), \underline{y}_{\boldsymbol{G}}] - \underset{\mathrm{M}}{\mathbf{Pr}}[e \in \boldsymbol{G} | \boldsymbol{\tau}(i), \boldsymbol{\tau}(j), \underline{x}_{\boldsymbol{G}}] \right) \end{aligned}$$

Remark 3.8.2. We remark that with probability $1 - o_n(1)$ each entry of $\underline{y}_{\boldsymbol{G}}$ satisifies $\underline{y}_{\boldsymbol{G}}[\tau, \alpha] \in (1 \pm \epsilon) \cdot \mathbf{E}_{\mathrm{M}}[\underline{y}_{\boldsymbol{G}}[\tau, \alpha]]$ for some small constant ϵ. From now on we only consider $\boldsymbol{G}$ that satisifies this condition.

Next we introduce notation for the non-backtracking product of two matrices.

Definition 3.8.3. For two matrices $A, B \in \mathbb{R}^{nq \times nq}$, define $A \circ B$ to be the non-bactracking product of A and B by setting for all $i, j \in [n]$

$$(A \circ B)[i,j] = \begin{cases} \sum_k A[i,k]B[k,j] & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases}$$

Inductively define $A^{(s)} \stackrel{\text{def}}{=} A^{(s-1)} \circ A$

Suppose $v_t \in \mathbb{R}^{nq}$ be the eigenvector with largest eigenvalue of $\underline{A}_{\boldsymbol{G}}^{(t)}$, and let its eigenvalue be Λ_t^t. By Theorem 3.6.1, we know that for each s, with probability $1 - o_n(1)$,

$$\lambda_{max}(\underline{A}_{\boldsymbol{G}}^{(s)}) \geqslant \Omega_{\mathrm{M}}(1) \cdot \lambda_L^s$$

Therefore,

$$v_s \underline{A}_{\boldsymbol{G}}^{(s)} v_s \geqslant \Omega_{\mathrm{M}}(1) \cdot \lambda_L^s$$

On the other hand, by the spectral norm bound for all $s \geqslant (\log\log n)^3$ in the null model,

$$v_s \underline{B}_{\boldsymbol{G}}^{(s)} v_s \leqslant (1 + o(1))^s \left(\sqrt{\lambda_L}\right)^s$$

For simplifying notation, we will ignore the $(1 + o(1))^s$ term in the above bound, here in the rest of the section. Rewriting the difference we get,

$$v_t^T \left(\underline{A}_{\boldsymbol{G}}^{(t)} - \underline{B}_{\boldsymbol{G}}^{(t)}\right) v_t = \sum_{s=0}^{t-1} v_t^T \underline{B}_{\boldsymbol{G}}^{(s)} \circ \underline{R}_{\boldsymbol{G}} \circ \underline{A}_{\boldsymbol{G}}^{(t-s-1)} v_t$$

Now we can replace the non-backtracking product in the above expression with the usual matrix product using Lemma 3.8.4.

Lemma 3.8.4. *For all $A, B, R \in \mathbb{R}^{nq \times nq}$,*

$$\|A^{(s)} \circ R \circ B^{(t)} - A^{(s)} R B^{(t)}\| \leqslant q\|R\|_\infty \cdot \left(\|A^{(s)}\|\|B^{(t-s)}\| + \|A\|_{1\to 1}\|A^{(s-1)}\|\|B^{(t)}\| + \|B^T\|_{1\to 1}\|A^{(s)}\|\|B^{(t-1)}\|\right)$$

where $\|R\|_\infty = \max_{\ell,\ell' \in [nq]} |R_{\ell,\ell'}|$

We will postpone the proof of this Lemma to later in the section, and proceed with the argument.

Notice that under the condition in Remark 3.8.2,

$$\|\underline{R}_G\|_\infty \leqslant O_{\mathrm{M}}(1) \cdot \frac{1}{n} \tag{3.26}$$

The maximum degree of a variable in the factor graph is $O(\log n)$ with probability $1 - o_n(1)$. Therefore a naive bound on the spectral norm of $\underline{A}_G^{(s)}$ would be

$$\|\underline{A}_G^{(s)}\| \leqslant O(\log n)^s \leqslant o(n^{1/4}) \tag{3.27}$$

for $s \leqslant o(\log n / \log\log n)$. Similarly, we can bound $\|\underline{B}_G^{(t)}\| \leqslant o(n^{1/4})$. Using these bounds in Lemma 3.8.4, we can replace non-backtracking product by the usual product to conclude,

$$v_t^T \left(\underline{A}_G^{(t)} - \underline{B}_G^{(t)}\right) v_t = \sum_{s=0}^{t-1} v_t^T \underline{B}_G^{(s)} \underline{R}_G \underline{A}_G^{(t-s-1)} v_t + o_n(1)$$

We will now rewrite the matrix $\underline{R}_G$ explicitly in terms of the coloring $\boldsymbol{c}$. To this end, we make a few definitions. For types $\tau, \tau' \in T$ and colors $\alpha, \alpha' \in [q]$ define $\Gamma_{\alpha,\alpha'}^{\tau,\tau'} \in \mathbb{R}^{[q]\times[q]}$ as,

$$\Gamma_{\alpha,\alpha'}^{\tau,\tau'} \stackrel{\text{def}}{=} \sum_{e \in \mathcal{K}_n, e \ni ij} \overline{\mathbf{M}}_{e,i|j} \cdot \mathop{\mathbf{Pr}}_{\mathrm{M}}[e \in \boldsymbol{G} | \boldsymbol{\tau}(i) = \tau, \boldsymbol{\tau}(j) = \tau', \boldsymbol{c}(i) = \alpha, \boldsymbol{c}(j) = \alpha', \underline{y}_G])$$

In terms of the matrices $\{\Gamma_{\alpha,\alpha'}^{\tau,\tau'}\}$ we can write for $i \neq j$,

$$\begin{aligned}\underline{R}_G[i,j] = \sum_{\tau,\tau' \in T} \sum_{\alpha,\alpha' \in [q]} \Gamma_{\alpha,\alpha'}^{\tau,\tau'} \cdot \mathbf{1}[\boldsymbol{\tau}(i) = \tau]\mathbf{1}[\boldsymbol{\tau}(j) = \tau'] \cdot \\ \left(\mathbf{1}[\boldsymbol{c}(i) = \alpha]\mathbf{1}[\boldsymbol{c}(j) = \alpha'] - \mathbf{Pr}[\boldsymbol{c}(i) = \alpha | \boldsymbol{\tau}(i), \underline{y}_G] \mathbf{Pr}[\boldsymbol{c}(j) = \alpha' | \boldsymbol{\tau}(j), \underline{y}_G]\right)\end{aligned}$$

For every type τ and a color α, let $\underline{\chi}^{\tau,\alpha}, \boldsymbol{c}^{\tau,\alpha}, \boldsymbol{\tau}^{\tau,\alpha} \in \mathbb{R}^n$ be defined as follows:

$$\begin{aligned}\underline{\chi}^{\tau,\alpha}[i] &\stackrel{\text{def}}{=} \mathbf{1}[\tau(i) = \tau] \cdot (\mathbf{1}[\boldsymbol{c}(i) = \alpha] - \mathbf{Pr}[\boldsymbol{c}(i) = \alpha | \boldsymbol{\tau}(i) = \tau, \underline{y}_G]) \\ \chi^{\tau,\alpha}[i] &\stackrel{\text{def}}{=} \mathbf{1}[\tau(i) = \tau] \cdot (\mathbf{1}[\boldsymbol{c}(i) = \alpha]) \\ \boldsymbol{\tau}^{\tau,\alpha}[i] &\stackrel{\text{def}}{=} \mathbf{1}[\boldsymbol{\tau}(i) = \tau] \cdot \mathbf{Pr}[\boldsymbol{c}(i) = \alpha | \boldsymbol{\tau}(i) = \tau, \underline{y}_G]\end{aligned}$$

Hence for $i \neq j$ we have,

$$\underline{R}_G[i,j] = \sum_{\tau,\tau' \in T} \sum_{\alpha,\alpha' \in [q]} \Gamma^{\tau,\tau'}_{\alpha,\alpha'} \cdot \left(\chi^{\tau,\alpha}[i] \underline{\chi}^{\tau',\alpha'}[j] + \underline{\chi}^{\tau,\alpha}[i] \overline{\chi}^{\tau',\alpha'}[j] \right)$$

Define $\underline{R}[i,i]$ so that we have the equality,

$$\underline{R}_G = \sum_{\tau,\tau',\alpha,\alpha'} \Gamma^{\tau,\tau'}_{\alpha,\alpha'} \otimes \left(\chi^{\tau,\alpha} (\underline{\chi}^{\tau',\alpha'})^T + \underline{\chi}^{\tau,\alpha} (\overline{\chi}^{\tau',\alpha'})^T \right) \tag{3.28}$$

Using (3.28) for the matrix $\underline{R}$, we can write,

$$v_t^T \left(\underline{A}_G^{(t)} - \underline{B}_G^{(t)} \right) v_t =$$
$$\sum_{\tau,\tau',\alpha,\alpha'} v_t^T \sum_{s=1}^{t} \underline{B}_G^{(s)} \left(\Gamma^{\tau,\tau'}_{\alpha,\alpha'} \otimes \left(\chi^{\tau,\alpha} (\underline{\chi}^{\tau',\alpha'})^T + \underline{\chi}^{\tau,\alpha} (\overline{\chi}^{\tau',\alpha'})^T \right) \right) \underline{A}_G^{(t-s-1)} v_t + o_n(1)$$

The second term corresponding to $\underline{\chi}^{\tau,\alpha} (\overline{\chi}^{\tau',\alpha'})^T$ is negligible. Specifically, we will prove the following Lemma.

Lemma 3.8.5. *With probability* $1 - o_n(1)$ *the following holds, for all* $\tau, \alpha, \tau', \alpha'$ *For all* $1 \leqslant s, t \leqslant \sqrt{\log n}$,

$$\| \underline{B}_G^{(s)} \left(\Gamma^{\tau,\tau'}_{\alpha,\alpha'} \otimes \underline{\chi}^{\tau,\alpha} (\overline{\chi}^{\tau',\alpha'})^T \right) \underline{A}_G^{(t)} \| \leqslant \left(\sqrt{\lambda_L} \right)^{s+t} \cdot O_{\mathrm{M}}(1)$$

We postpone the proof to later in the section and proceed with the main argument.

Using Lemma 3.8.5, we can drop all terms arising from $\underline{\chi}^{\tau,\alpha} (\overline{\chi}^{\tau',\alpha'})^T$ by losing less than $O_{\mathrm{M}}(1) \sum_{s=0}^{t-1} \left(\sqrt{\lambda_L} \right)^s$. Since $\lambda_L > (1+\delta)^4$, for all t larger than a fixed constant $\Theta_M(1)$, this sum $O_{\mathrm{M}}(1) \sum_{s=0}^{t-1} \left(\sqrt{\lambda_L} \right)^s \leqslant \frac{1}{10} \lambda_L^t$. Therefore, we arrive at our inequality,

$$\sum_{\tau,\tau',\alpha,\alpha'} v_t^T \sum_{s=0}^{t-1} \underline{B}_G^{(s)} \left(\Gamma^{\tau,\tau'}_{\alpha,\alpha'} \otimes \chi^{\tau,\alpha} (\underline{\chi}^{\tau',\alpha'})^T \right) \underline{A}_G^{(t-s-1)} v_t \geqslant 0.9 \cdot \Lambda_t^t$$

Let us write the matrix $\Gamma^{\tau,\tau'}_{\alpha,\alpha'} = \sum_{\beta,\beta' \in [q]} \Gamma^{\tau,\tau'}_{\alpha,\alpha'}[\beta,\beta'] \cdot \mathbf{e}_\beta (\mathbf{e}_{\beta'})^T$ where $\mathbf{e}_\beta, \mathbf{e}'_\beta$ are standard basis vectors in $\mathbb{R}^q$. Note that $\Gamma^{\tau,\tau'}_{\alpha,\alpha'}$ has entries that are $O_{\mathrm{M}}(1)/n$. There must exist some choice of $\tau, \tau', \alpha, \alpha', \beta, \beta'$ such that,

$$v_t^T \sum_{s=0}^{t-1} \underline{B}_{\boldsymbol{G}}^{(s)} \left(\mathbf{e}_\beta \otimes \chi^{\tau,\alpha} (\mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'})^T \right) \underline{A}_{\boldsymbol{G}}^{(t-s-1)} v_t \geqslant \Omega_{\mathrm{M}}(1) \cdot \Lambda_t^t \cdot n$$

where $\Omega_{\mathrm{M}}(1)$ hides a constant depending on the model M. Rewriting the above inequality,

$$\begin{aligned}\Omega_{\mathrm{M}}(1) \cdot \Lambda_t^t \cdot n &\leqslant \sum_{s=0}^{t-1} \left\langle v_t^T, \underline{B}_{\boldsymbol{G}}^{(s)} \mathbf{e}_\beta \otimes \chi^{\tau,\alpha} \right\rangle \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(t-s-1)} v_t \right\rangle \\ &\leqslant \sum_{s=0}^{t-1} \| \underline{B}_{\boldsymbol{G}}^{(s)} \mathbf{e}_\beta \otimes \chi^{\tau,\alpha} \| \left| \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(t-s-1)} v_t \right\rangle \right|\end{aligned}$$

Using Lemma 3.8.8 on the fixed vector $\mathbf{e}_\beta \otimes \chi^{\tau,\alpha}$ and the planted distribution we get that with probability $1 - o_n(1)$,

$$\Omega_{\mathrm{M}}(1) \cdot \Lambda_t^t \cdot n \leqslant \sum_{s=0}^{t-1} \left(\sqrt{\lambda_L} \right)^s \cdot \sqrt{n} \left| \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(t-s-1)} v_t \right\rangle \right|$$

For notational convenience, let us reparametrize $s \to t - s$ and conclude,

$$\Omega_{\mathrm{M}}(1) \cdot \Lambda_t^t \cdot \sqrt{n} \leqslant \sum_{s=1}^{t} \left(\sqrt{\lambda_L} \right)^{t-s} \left| \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(s-1)} v_t \right\rangle \right|$$

With high probability, the maximum degree of a variable is $O(\log n) \ll o(\log^2 n)$, and therefore $\| \underline{A}^{(s)} \| \leqslant o(\log n)^{2s}$.

Since $\Lambda_t \geqslant \lambda_L \geqslant (1+\delta) \left(\sqrt{\lambda_L} \right)$ we can bound the terms for $s = 1, \ldots, (\log\log n)^3$ as follows,

$$\begin{aligned}&\sum_{s=1}^{(\log\log n)^3} \left(\sqrt{\lambda_L} \right)^{t-s} \left| \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(s-1)} v_t \right\rangle \right| \\ \leqslant &\sum_{s=1}^{(\log\log n)^3} \frac{\Lambda_t^{t-s}}{(1+\delta)^{t-s}} \cdot o((\log n)^{2s}) \cdot \sqrt{n} \leqslant o(1) \cdot \Lambda_t^t \cdot \sqrt{n}\end{aligned}$$

where the last inequality holds for $t > (\log\log n)^4$. Deleting terms for small s, we have the correlation inequality,

$$\Omega_{\mathrm{M}}(1) \cdot \Lambda_t^t \cdot \sqrt{n} \leqslant \sum_{s=(\log\log n)^3}^{t} \left(\sqrt{\lambda_L} \right)^{t-s} \left| \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(s-1)} v_t \right\rangle \right| \tag{3.29}$$

for $t > (\log\log n)^4$.

For each s, recall that v_s is the top eigenvector of $\underline{A}_{\boldsymbol{G}}^{(s)}$, and Λ_s^s denotes the largest eigenvalue.

Claim 3.8.6. There exists a $\mathrm{m} \in [(\log\log n)^{4.5}, (\log\log n)^5]$ such that, for all $s \in [(\log\log n)^3, \mathrm{m}]$,

$$\|A^{(s)} v_{\mathrm{m}}\| \leqslant \Lambda_{\mathrm{m}}^s (1+\delta)^{t-s}$$

Before we see the proof of above claim, let us see how it leads to an algorithm. Applying (3.29) for this choice of m, we conclude that

$$\sqrt{n} \cdot \Omega_{\mathrm{M}}(1) \cdot \Lambda_{\mathrm{m}}^{\mathrm{m}} \leqslant \sum_{s=(\log\log n)^3}^{\mathrm{m}} \left(\sqrt{\lambda_L}\right)^{\mathrm{m}-s} \left| \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(s-1)} v_{\mathrm{m}} \right\rangle \right| \tag{3.30}$$

In (3.30), we can bound the sum of all terms with $s < t^* - C$ as follows.

$$\sum_{s=(\log\log n)^3}^{\mathrm{m}-C} \left(\sqrt{\lambda_L}\right)^{\mathrm{m}-s} \left| \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(s-1)} v_t \right\rangle \right| \tag{3.31}$$

$$\leqslant \sum_{s=(\log\log n)^3}^{\mathrm{m}-C} \left(\sqrt{\lambda_L}\right)^{\mathrm{m}-s} \|\mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}\| \left\| \underline{A}_{\boldsymbol{G}}^{(s-1)} v_t \right\| \tag{3.32}$$

$$\leqslant \sum_{s=(\log\log n)^3}^{\mathrm{m}-C} \left(\sqrt{\lambda_L}\right)^{\mathrm{m}-s} \Lambda_{\mathrm{m}}^{s-1} \cdot \sqrt{n} \cdot (1+\delta)^{\mathrm{m}-s+1} \tag{3.33}$$

$$\leqslant (1+\delta)\Lambda_{\mathrm{m}}^{\mathrm{m}}\sqrt{n} \cdot \left(\sum_{s=(\log\log n)^3}^{\mathrm{m}-C} \left(\frac{\left(\sqrt{\lambda_L}\right)(1+\delta)}{\Lambda_{\mathrm{m}}} \right)^{\mathrm{m}-s} \right) \tag{3.34}$$

$$= \Lambda_{\mathrm{m}}^{\mathrm{m}} \cdot \sqrt{n} \cdot \left(\frac{1}{\delta(1+\delta)^{C-2}} \right) \tag{3.35}$$

Using (3.30) and (3.35), for sufficiently large $C = O_{\mathrm{M},\delta}(1)$, we conclude that

$$\Omega_{\mathrm{M}}(1) \cdot \Lambda_{\mathrm{m}}^{\mathrm{m}} \cdot \sqrt{n} \leqslant \sum_{s=\mathrm{m}-C+1}^{\mathrm{m}-1} \left(\sqrt{\lambda_L}\right)^{\mathrm{m}-s} \left| \left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \underline{A}_{\boldsymbol{G}}^{(s-1)} v_{\mathrm{m}} \right\rangle \right| \tag{3.36}$$

Note that there are only $C = O_{\mathrm{M},\delta}(1)$ terms in the sum, so one of them is large. In particular, there exists some $\ell \in [\mathrm{m}-C, \mathrm{m}]$ such that if we set

$$w_\ell \stackrel{\text{def}}{=} \underline{A}_{\boldsymbol{G}}^{(\ell)} v_{\mathrm{m}}$$

then,

$$\left|\left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, w_\ell \right\rangle\right| \geqslant \frac{1}{C\left(\sqrt{\lambda_L}\right)^C} \cdot \Lambda_{\mathrm{m}}^{\mathrm{m}} \cdot \Omega_{M,\delta}(1) \cdot \sqrt{n}\,,$$

But note that by the choice of m,

$$\|w_\ell\| \leqslant \Lambda_{\mathrm{m}}^{\ell} \cdot (1+\delta)^{\mathrm{m}-\ell} \leqslant \Lambda_{\mathrm{m}}^{\mathrm{m}} \cdot (1+\delta)^{C}\,.$$

So there exists some $\ell \in [\mathrm{m}-C, \mathrm{m}]$, such that the unit vector $\overline{w}_\ell = \frac{w_\ell}{\|w_\ell\|}$ satisfies,

$$\left|\left\langle \mathbf{e}_{\beta'} \otimes \underline{\chi}^{\tau',\alpha'}, \overline{w}_\ell \right\rangle\right| \geqslant \Omega_{M,\delta}(1) \cdot \sqrt{n}\,.$$

In other words, for some choice of $\ell \in [\mathrm{m}-C, \mathrm{m}]$ and $\beta' \in [q]$, if we construct $\overline{u} \in \mathbb{R}^n$ as,

$$\overline{u}[i] = \overline{w}_\ell[i, \beta']$$

then $|\langle \overline{u}, \underline{\chi}^{\tau',\alpha'} \rangle| \geqslant \Omega_{\mathrm{M},\delta}(1) \cdot \sqrt{n}$. This finishes the proof of Theorem 3.8.1.[6]

Proof. (Proof of Claim 3.8.6) The idea behind the proof is a descent/bootstrap argument to get a contradiction. Let us start with $t = (\log\log n)^5$ as the guess for m. If current value of t satisfies the condition of the claim, we are done. Otherwise, there exists $s < t$ such that,

$$\|\underline{A}_{\boldsymbol{G}}^{(s)} v_t\| \geqslant \Lambda_t^s (1+\delta)^{t-s}$$

This implies that,

$$\Lambda_s^s \geqslant \Lambda_t^s (1+\delta)^{t-s}$$

or equivalently,

$$\log \Lambda_s \geqslant \log \Lambda_t + \log(1+\delta) \cdot (t-s)/s$$

Suppose we use s as the new candidate for m and recurse. Let us suppose we iteratively construct a sequence of $t_0 = (\log\log n)^5 > \ldots > t_r$ in this manner. The value of $\log \Lambda_{t_i}$ increases along the sequence. By Fact 3.8.7, if we obtain a sequence of $t_0 = (\log\log n)^5 > \ldots > t_r = (\log\log n)^{4.5}$ then we will have,

$$\log \Lambda_r \geqslant (\log t_0 - \log t_r) \log(1+\delta) \geqslant \log(1+\delta) \cdot \Omega(\log\log\log n)$$

[6]Note that the $\underline{\chi}^{\tau',\alpha'}$ here is a bit different from the $\underline{\chi}^{\tau',\alpha'}$ defined in the theorem, but by Remark 3.8.2 they are within a multiplicative factor of $(1 \pm \epsilon)$ from each other. Thus the inequality still hold for the $\underline{\chi}^{\tau',\alpha'}$ defined in the theorem statement.

This suggests that $\|\underline{A}_G^{(t_r)}\|^{1/t_r} \geqslant \omega(1)$ for some $t_r = \Omega(\log\log n)^{4.5}$. A contradiction, since with probability $1 - o_n(1)$, we will have $\|\underline{A}_G^{(t_r)}\|^{1/t_r} = O(1)$. This follows from the fact that with probability $1 - o_n(1)$, degree of every vertex in $A_G^{(s)}$ is at most $O(D^s)$ for some constant D for all $s > (\log\log n)^2$. Therefore, the sequence terminates and we find a $t_r \in [(\log\log n)^{4.5}, (\log\log n)^5]$, implying the claim.

Fact 3.8.7. *Given a sequence of positive integers, $a_1 \geqslant a_2 \geqslant \ldots a_r$,*

$$\sum_{i=1}^{r} \frac{a_i - a_{i-1}}{a_{i-1}} \geqslant \sum_{i=1}^{r} \sum_{x=a_{i-1}}^{a_i - 1} \frac{1}{x} = \sum_{x=a_r}^{a_1} \frac{1}{x} \approx \ln(a_1) - \ln(a_r) \tag{3.37}$$

□

Proof. (Proof of Lemma 3.8.4) For $s \in \mathbb{N}$, let $\mathcal{P}_s$ be the set of length s non-backtracking walks in complete graph $\mathcal{K}_n$. So $\underline{\alpha} = (\alpha_0, \alpha_1, \ldots, \alpha_s) \in \mathcal{P}_s$ will be a non-backtracking path in $\mathcal{K}_n$ with vertices $\alpha_0, \ldots, \alpha_s$.

For $\ell, \ell' \in [n]$, we can write

$$A^{(s)} R B^{(t)}[\ell, \ell'] = \sum_{\substack{\underline{\alpha} \in \mathcal{P}_s, \underline{\beta} \in \mathcal{P}_t \\ \alpha_0 = \ell, \beta_t = \ell'}} \cdot \prod_{i=1}^{s} A_{\alpha_{i-1}\alpha_i} \cdot R_{\alpha_s, \beta_0} \cdot \prod_{j=1}^{t} B_{\beta_{j-1}\beta_j} \tag{3.38}$$

$$A^{(s)} \circ R \circ B^{(t)}[\ell, \ell'] = \sum_{\substack{\underline{\alpha} \in \mathcal{P}_s, \underline{\beta} \in \mathcal{P}_t \\ \alpha_0 = \ell, \beta_t = \ell' \\ \alpha_s \neq \beta_0, \alpha_{s-1} \neq \beta_0, \alpha_s \neq \beta_1}} \cdot \prod_{i=1}^{s} A_{\alpha_{i-1}\alpha_i} \cdot R_{\alpha_s, \beta_0} \prod_{j=1}^{t} B_{\beta_{j-1}\beta_j} \tag{3.39}$$

It is clear that the difference $A^{(s)} \circ R \circ B^{(t)} - A^{(s)} R B^{(t)}$ consists of three different terms.

Term 1: $\alpha_s = \beta_0$ Consider the block-diagonal matrix $\mathcal{D}_1 \in \mathbb{R}^{nq \times nq}$ given by,

$$\mathcal{D}_1[i, j] = \mathbf{1}[i = j] \cdot R_{ij}.$$

then we can write this term as $A^{(s)} \mathcal{D}_1 B^{(t)}$. Hence we get the following bound,

$$\|A^{(s)} \mathcal{D}_1 B^{(t)}\| \leqslant q \|R\|_\infty \cdot \|A^{(s)}\| \cdot \|B^{(t-s)}\|$$

Term 2: $\alpha_{s-1} = \beta_0$ Consider the block-diagonal matrix $\mathcal{D}_2 \in \mathbb{R}^{nq \times nq}$ given by,

$$\mathcal{D}_2[i,i] = \sum_j A_{ij} R_{ji}.$$

then we can write this term as $A^{(s-1)} \mathcal{D}_2 B^{(t)}$. We have the upper bound,

$$\|\mathcal{D}_2\| \leqslant q \|R\|_\infty \|A\|_{1\to 1}$$

which implies that

$$\|A^{(s-1)} \mathcal{D}_2 B^{(t)}\| \leqslant q \|x\|_\infty \|y\|_\infty \|A\|_{1\to 1} \|A^{(s-1)}\| \|B^{(t)}\|$$

Term 3: $\alpha_s = \beta_1$ Consider the block-diagonal matrix $\mathcal{D}_3 \in \mathbb{R}^{nq \times nq}$ given by,

$$\mathcal{D}_3[i,i] = (\sum_j R_{ij} B_{ji}).$$

then we can write this term as $A^{(s)} \mathcal{D}_3 B^{(t-1)}$. Analogous to the previous case, we get an upper bound of,

$$\|A^{(s)} \mathcal{D}_3 B^{(t-1)}\| \leqslant q \|R\|_\infty \|B^T\|_{1\to 1} \|A^{(s)}\| \|B^{(t-1)}\|$$

Adding the three terms, we have the claim of the lemma.

□

We will need the following theorem about local statistics on the expectation and concentration of local statistics in order to complete the proof of Lemma 3.8.5.

Lemma 3.8.8. *Fix a vector $x \in \mathbb{R}^{nq}$, such that $\|x\|_\infty = O(1)$. With probability $1 - o_n(1)$, for all $1 \leqslant s \leqslant \sqrt{\log n}$ we have,*

$$\|\underline{A}_{\boldsymbol{G}}^{(s)} x\| \leqslant C \left(\sqrt{\lambda_L}\right)^s \cdot \sqrt{n} \tag{3.40}$$

for an absolute constant $C \geqslant 1$.

Proof sketch. This is equivalent to proving $\langle \underline{A}_{\boldsymbol{G}}^{(s)} x, \underline{A}_{\boldsymbol{G}}^{(s)} x \rangle \leqslant C \left(\sqrt{\lambda_L}\right)^{2s} n$. Indeed, this quantity can be rewritten as:

$$\langle \left(\underline{A}_{\boldsymbol{G}}^{(s)}\right)^2, x x^\top \rangle.$$

Via similar calculations to the ones done in Section 3.6, we can show that this quantity is dominated by the contribution of walks that are self-avoiding for the first s steps, and retrace the same steps taken in the next s steps, which in turn can be used to show that this quantity concentrates around the expected total weight of walks in the associated random tree that walk out s steps and walk back s steps, and hence for large enough s is at most $C\left(\sqrt{\lambda_L}\right)^{2s} n$ for an absolute constant C. □

Proof of Lemma 3.8.5. Let $\Gamma^{\tau,\tau'}_{\alpha,\alpha'} = \sum_{j\in[q]} u_j v_j^T$ be the singular decomposition of $\Gamma^{\tau,\tau'}_{\alpha,\alpha'}$.

$$\|\underline{B}_{\boldsymbol{G}}^{(s)}\Gamma^{\tau,\tau'}_{\alpha,\alpha'} \otimes \underline{\chi}^{\tau,\alpha}(^{-\tau',\alpha'})^T \underline{A}_{\boldsymbol{G}}^{(t)}\| \leqslant \sum_j \|\underline{B}_{\boldsymbol{G}}^{(s)} u_j \otimes \underline{\chi}^{\tau,\alpha}\| \|v_j \otimes (^{-\tau',\alpha'})^T \underline{A}_{\boldsymbol{G}}^{(t)}\| \tag{3.41}$$

Applying Lemma 3.8.8 to the planted model M with the fixed vector $u_j \otimes \underline{\chi}^{\tau,\alpha}$, we conclude that with probability $1 - o_n(1)$,

$$\|\underline{B}_{\boldsymbol{G}}^{(s)} u_j \otimes \underline{\chi}^{\tau,\alpha}\| \leqslant C\left(\sqrt{\lambda_L}\right)^s \cdot \|u_j \otimes \underline{\chi}^{\tau,\alpha}\| \leqslant \left(\sqrt{\lambda_L}\right)^s \cdot \|u_j\| \cdot n^{1/2} \tag{3.42}$$

Similarly, applying Lemma 3.8.8 to the null model $\mathrm{M}^{\times}$ with fixed vector $v_j \otimes \mu^{\tau',\alpha'}$, we conclude that with probability $1 - o_n(1)$, for all $1 \leqslant t \leqslant \sqrt{\log n}$,

$$\|v_j \otimes (^{-\tau',\alpha'})^T \underline{A}_{\boldsymbol{G}}^{(t)}\| \leqslant C\left(\sqrt{\lambda_L}\right)^s \cdot \|v_j \otimes (^{-\tau',\alpha'})^T\| \leqslant \left(\sqrt{\lambda_L}\right)^t \cdot \|v_j\| \cdot n^{1/2} \tag{3.43}$$

Finally, note that $\sum_j \|u_j\|\|v_j\| = \|\Gamma^{\tau,\tau'}_{\alpha,\alpha'}\|_{Fr} = O_{\mathrm{M}}(\frac{1}{n})$ where O_{M} hides a fixed constant depending on the model. Using (3.42) and (3.43) in (3.41), we conclude the proof. □

3.9 Belief propagation for M

We briefly describe the belief propagation (BP) algorithm that aims to estimate the mariginal distribution of $\boldsymbol{c}(v)$, $v \in [n]$ under the Boltzmann distribution μ with Hamiltonian $\mathbf{H}$. Define the messages $\{m_c^{v\to e}\}_{c\in[q]}$ that a variable v passes to some constraint $e \in \boldsymbol{E}_i$, and the messages $\{m_c^{e\to u}\}_{c\in[q]}$ that a constraint $e \in \boldsymbol{E}_i$ passes to a variable u. Intuitively speaking, $m_c^{v\to e}$ is an estimate of the marginal probability that v is assigned the color c when the constraint e is absent, and $m_c^{e\to u}$ is an estimate of the marginal probability that u has color c when all other constraints involving u are absent. Since the distribution of $\boldsymbol{c}(v)$ under μ depends on the

constraints that contain v, we only focus on the messages $m^{v\to e}$, $m^{e\to u}$ such that $v \in \partial e$ (i.e. e contains v) and $e \in \partial u$ (i.e. e contains u).

$$m_c^{v\to e}[t+1] = \frac{1}{Z^{v\to e}} \mathbb{P}_{\boldsymbol{\tau}(v)}(c) \prod_{f\in\partial v\setminus e} m_c^{f\to v}[t] \ , \tag{3.44}$$

where $Z^{v\to e} = \sum_{c\in[q]} \mathbb{P}_{\boldsymbol{\tau}(v)}(c) \prod_{f\in\partial v\setminus e} m_c^{f\to v}[t]$.

For factors $e \in \boldsymbol{E}_i$, the messages are defined as

$$m_c^{e\to u}[t+1] = \frac{1}{Z^{e\to u}} \sum_{\boldsymbol{c}_e|\boldsymbol{c}_e(u)=c} \phi_i(\boldsymbol{c}_e) \prod_{v\in e\setminus u} m_{\boldsymbol{c}_e(v)}^{v\to e}[t] \ , \tag{3.45}$$

where $Z^{e\to u} = \sum_{c\in[q]} \sum_{\boldsymbol{c}_e|\boldsymbol{c}_e(u)=c} \phi_i(\boldsymbol{c}_e) \prod_{v\in e\setminus u} m_{\boldsymbol{c}_e(v)}^{v\to e}[t]$.

To obtain an estimate of the marginal probability of the assignment to a variable v, apply the message update rules until reaching some fixed point $\{\hat{m}_c^{v\to e}, \hat{m}_c^{e\to u}\}_{c\in[q]}$. The estimate is called the belief and is given by

$$b_c^v = \frac{1}{Z^v} \mathbb{P}_{\boldsymbol{\tau}(v)}(c) \prod_{f\in\partial v} \hat{m}_c^{f\to v} \ ,$$

where $Z^v = \sum_{c\in[q]} \mathbb{P}_{\boldsymbol{\tau}(v)}(c) \prod_{f\in\partial v} \hat{m}_c^{f\to v}$.

3.10 Proof of Lemma 3.2.5

Recall that:

$$\overline{\phi_i} = \sum_{(c_1,\ldots,c_{\mathrm{a}(i)})\in[q]^{\mathrm{a}(i)}} \left(\prod_{k=1}^{\mathrm{a}(i)} \mathbb{P}_{\mathrm{Cl}(i)_k}(c_k) \right) \cdot \phi_i(c_1,\ldots,c_{\mathrm{a}(i)}) \tag{3.46}$$

We first explain how to solve the distinguishing problem and then explain the recovery algorithm. By definition of the BP update functions $\mathrm{Y}_{v\to e}$ (equation (3.44)) and $\mathrm{Y}_{e\to v}$ (equation (3.45)), the set of trivial messages $\overline{m}$ being a BP fixed point is equivalent to:

$$\overline{\phi_i} = \sum_{\substack{(c_1,\ldots,c_{\mathrm{a}(i)})\in[q]^{\mathrm{a}(i)}:\\ c_j=c}} \left(\prod_{k\neq j} \mathbb{P}_{\mathrm{Cl}(i)_k}(c_k) \right) \cdot \phi_i(c_1,\ldots,c_{\mathrm{a}(i)}) \quad \forall c\in[q] : \mathbb{P}_{\mathrm{Cl}(i)_j}(c) > 0 \tag{3.47}$$

If the set of trivial messages is not a fixed point of the belief propagation update rule: then there exist $c \in [q]$, $i \in [F]$, and $j \in [\mathrm{a}(i)]$ with $\mathbb{P}_{\mathrm{Cl}(i)_j}(c) > 0$ such that

$$\overline{\phi_i} \neq \sum_{\substack{(c_1,\ldots,c_{\mathrm{a}(i)}) \in [q]^{\mathrm{a}(i)}: \\ c_j = c}} \left(\prod_{k \neq j} \mathbb{P}_{\mathrm{Cl}(i)_k}(c_k) \right) \cdot \phi_i(c_1, \ldots, c_{\mathrm{a}(i)}).$$

Let $\deg_{i,j}(v)$ be the number of type-i factors with variable v in the j-th position. Via standard results for Poisson(d) approximating Binom$(n, d/n)$ we have the following:

- In $\mathrm{M}^{\times}$, for any variable v of type $\mathrm{Cl}(i)_j$ and any constant T,

  $$\mathbf{E} \deg_{i,j}(v)^T = \mathbf{E}\, \boldsymbol{X}^T \pm o_n(1)$$

  where $\boldsymbol{X} \sim \mathrm{Poisson}(\lambda)$ and $\lambda = \overline{\phi_i}$.

- On the other hand, in the planted model M:

  $$\mathbf{E} \deg_{i,j}(v)^T = \mathbf{E}\, \boldsymbol{Y}^T \pm o_n(1)$$

  where $\boldsymbol{Y}$ is distributed as the mixture of Poisson distributions $p_1\mathrm{Poisson}(\lambda_1) + \cdots + p_s\mathrm{Poisson}(\lambda_s)$ where s is the number of colors which vertex v has nonzero probability of attaining, not all λ_i are equal, and all $p_i > 0$.

By (3.46) $p_1\lambda_1 + \cdots + p_s\lambda_s = \lambda$. We first recall the following well known fact about Poisson random variables.

Fact 3.10.1. *If* $\boldsymbol{A} \sim \mathrm{Poisson}(\mu)$, *then* $\mathbf{E}\, \boldsymbol{A}^2 = \mu^2 + \mu$.

As a consequence of Fact 3.10.1: $\mathbf{E}\, \boldsymbol{X}^2 = \lambda^2 + \lambda$, and $\mathbf{E}\, \boldsymbol{Y}^2 = p_1(\lambda_1^2 + \lambda_1) + \cdots + p_s(\lambda_s^2 + \lambda_s)$.

$$\begin{aligned} &\mathbf{E}\, \boldsymbol{Y}^2 - \mathbf{E}\, \boldsymbol{X}^2 = p_1 f(\lambda_1) + \ldots p_s f(\lambda_s) - f(\lambda) \\ &= p_1 f(\lambda_1) + \ldots p_s f(\lambda_s) - f(p_1\lambda_1 + \cdots + p_s\lambda_s) \end{aligned}$$

Since not all λ_i are equal, all $p_i > 0$ and f is strictly convex, $\mathbf{E}\, \boldsymbol{Y}^2 - \mathbf{E}\, \boldsymbol{X}^2$ is equal to a constant δ strictly greater than 0. Suppose $n_{i,j,2}(\boldsymbol{G}) := \mathbf{E} \sum_{v \in [n]} \deg_{i,j}(v)^2$, then $|\mathbf{E}_{\boldsymbol{G} \sim \mathcal{N}}\, n_{i,j,2}(\boldsymbol{G}) - \mathbf{E}_{\boldsymbol{G} \sim \mathcal{P}}\, n_{i,j,2}(\boldsymbol{G})| \geqslant \Omega(n)$. Since $\mathbf{E}\, \boldsymbol{Y}^4$ and $\mathbf{E}\, \boldsymbol{X}^4$ are constants, the variance of $n_{i,j,2}(\boldsymbol{G})$ is $O(n)$ for both $\boldsymbol{G} \sim \mathcal{N}$ and $\boldsymbol{G} \sim \mathcal{P}$. This informs using the following polynomial time distinguisher:

Compute $n_{i,j,2}(\boldsymbol{G})$ and if $|n_{i,j,2}(\boldsymbol{G}) - \mathbf{E}_{\boldsymbol{G}\sim\mathcal{N}}[n_{i,j,2}(\boldsymbol{G})]| < |n_{i,j,2}(\boldsymbol{G}) - \mathbf{E}_{\boldsymbol{G}\sim\mathcal{P}}[n_{i,j,2}(\boldsymbol{G})]|$ output "null"; otherwise output "planted".

We now discuss performing recovery. Recall the inner product $\langle\cdot,\cdot\rangle_{\mathbf{H}}$ from Section 3.5 which is defined as follows: First, we define a $nq \times nq$-dimensional positive diagonal matrix $\mathbf{H}_{\boldsymbol{\tau}}$ where the (v,v) block is equal to:

$$\mathbf{H}_{\boldsymbol{\tau},(v,v)}[c,c] := \begin{cases} \mathbb{P}_{\boldsymbol{\tau}(v)}(c) & \text{if } \mathbb{P}_{\boldsymbol{\tau}(v)}(c) > 0 \\ 1 & \text{otherwise.} \end{cases}$$

The inner product on $\mathbb{R}^{nq}$ is then:

$$\langle x, y\rangle_{\mathbf{H}} := x^{\top}\mathbf{H}_{\boldsymbol{\tau}}^{-1}y.$$

And let $\|\cdot\|$ denote the norm induced by the above inner product. Let $\boldsymbol{c}$ be the hidden coloring. Our goal in recovery is to output a vector v such that $\langle v, \boldsymbol{c} - \mathbf{E}\,\boldsymbol{c}|\boldsymbol{\tau}\rangle_{\mathbf{H}} \geqslant \varepsilon \cdot \|v\| \cdot \|\boldsymbol{c} - \mathbf{E}\,\boldsymbol{c}|\boldsymbol{\tau}\|$. Let c and c' be two colors such that:

$$\begin{aligned} d_c &= \sum_{\substack{(c_1,\ldots,c_{\mathrm{a}(i)})\in[q]^{\mathrm{a}(i)}:\\ c_j=c}} \left(\prod_{k\neq j}\mathbb{P}_{\mathrm{Cl}(i)_k}(c_k)\right)\cdot\phi_i(c_1,\ldots,c_{\mathrm{a}(i)}) \\ &> \sum_{\substack{(c_1,\ldots,c_{\mathrm{a}(i)})\in[q]^{\mathrm{a}(i)}:\\ c_j=c'}} \left(\prod_{k\neq j}\mathbb{P}_{\mathrm{Cl}(i)_k}(c_k)\right)\cdot\phi_i(c_1,\ldots,c_{\mathrm{a}(i)}) = d_{c'} \end{aligned}$$

The distribution of the number of type-i factors that a color c vertex is part of is Poisson(d_c) and similarly is Poisson$(d_{c'})$ for a color c' vertex. The following algorithm can then be shown to produce a vector v meeting the aforementioned goal.

For each vertex u of type $\mathcal{T}(\mathrm{Cl}(i)_j)$, let m_u be the number of type-i factors it is part of in the j-th position. If m_u has a higher probability of being sampled from Poisson(d_c) than Poisson$(d_{c'})$ then assign the u-th block of vector v to be the indicator of color c. Otherwise assign the u-th block of vector v to be the indicator of color c'.

Since $d_c \neq d_{c'}$ there is a constant $\varepsilon > 0$ such that with high probability

$$\left(\frac{1}{2}+\varepsilon\right)\mathbb{P}_{\mathrm{Cl}(i)_j}(c)\mathcal{T}(\mathrm{Cl}(i)_j)n$$

variables of color c and type $\mathrm{Cl}(i)_j$ are assigned the correct color and also

$$\left(\frac{1}{2}+\varepsilon\right)\mathbb{P}_{\mathrm{Cl}(i)_j}(c)\mathcal{T}(\mathrm{Cl}(i)_j)n$$

variables of color c' and type $\mathrm{Cl}(i)_j$ are assigned the correct color. Consequently:

$$\langle v, \boldsymbol{c}-\mathbf{E}\,\boldsymbol{c}|\boldsymbol{\tau}\rangle_{\mathbf{H}} \geqslant \varepsilon'\cdot\|v\|\cdot\|\boldsymbol{c}-\mathbf{E}\,\boldsymbol{c}|\boldsymbol{\tau}\|$$

for some $\varepsilon'>0$.

3.11 The partial derivative matrix

Recall the BP update function Γ defined by equations (3.44) and (3.45).

We observe that by definition

$$\overline{\mathbf{M}}_{\theta(e_j),\mathrm{i}(v_j)|\mathrm{i}(v_{j+1})} = \frac{\partial\Gamma(m)^{v_j\to e_{j-1}}}{\partial m^{e_j\to v_j}}\Big|_{\overline{m}}\cdot\frac{\partial\Gamma(m)^{e_j\to v_j}}{\partial m^{v_{j+1}\to e_j}}\Big|_{\overline{m}}.$$

Thus we first compute the two derivative matrices. For any pairs of colors $c,d\in[q]$,

$$\begin{aligned}
&\frac{\partial\Gamma(m)_c^{v_j\to e_{j-1}}}{\partial m_d^{e_j\to v_j}} = \\
&\frac{1}{Z^{v_j\to e_{j-1}}}\mathbb{P}_{\boldsymbol{\tau}(v_j)}(c)\\
&\prod_{a\in\partial v_j\setminus\{e_{j-1},e_j\}} m_c^{a\to v_j}\cdot\mathbf{1}_{d=c} - \frac{m_c^{v_j\to e_{j-1}}}{Z^{v_j\to e_{j-1}}}\cdot\mathbb{P}_{\boldsymbol{\tau}(v_j)}(d)\prod_{a\in\partial v_j\setminus\{e_{j-1},e_j\}} m_d^{a\to v_j}\\
=&\frac{1}{m_d^{e_j\to v_j}}\cdot\frac{1}{Z^{v_j\to e_{j-1}}}\mathbb{P}_{\boldsymbol{\tau}(v_j)}(c)\\
&\prod_{a\in\partial v_j\setminus e_{j-1}} m_c^{a\to v_j}\cdot\mathbf{1}_{d=c} - \frac{m_c^{v_j\to e_{j-1}}}{m_d^{e_j\to v_j}}\cdot\frac{\mathbb{P}_{\boldsymbol{\tau}(v_j)}(d)}{Z^{v_j\to e_{j-1}}}\prod_{a\in\partial v_j\setminus e_{j-1}} m_d^{a\to v_j}\\
=&\frac{m_c^{v_j\to e_{j-1}}}{m_d^{e_j\to v_j}}\cdot\mathbf{1}_{d=c} - \frac{m_c^{v_j\to e_{j-1}}}{m_d^{e_j\to v_j}}\cdot m_d^{v_j\to e_{j-1}}
\end{aligned}$$

The last equality is derived from the fixed point identity

$$m_c^{v\to e} = \frac{1}{Z^{v\to e}}\mathbb{P}_{\boldsymbol{\tau}(v)}(c)\prod_{a\in\partial v\setminus e} m_c^{a\to v}.$$

Evaluating the derivative at the factorized fixed point gives the transformation matrix

$$\frac{\partial\Gamma(m)_c^{v_j\to e_{j-1}}}{\partial m_d^{e_j\to v_j}}\bigg|_{\overline{m}} = \text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})\cdot\left(\mathbb{P}_{\boldsymbol{\tau}(v_j)}(c)\cdot\mathbf{1}_{d=c} - \mathbb{P}_{\boldsymbol{\tau}(v_j)}(c)\cdot\mathbb{P}_{\boldsymbol{\tau}(v_j)}(d)\right)\ ,$$

where $\text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})$ denote the size of $\mathbb{P}_{\boldsymbol{\tau}(v_j)}$'s support. To write the matrix compactly we define $\mathbf{D}_\tau := \text{Diag}(\mathbb{P}_\tau)$, and derive from the above computation that $\frac{\partial\Gamma(m)^{v_j\to e_{j-1}}}{\partial m^{e_j\to v_j}}\Big|_{\overline{m}} = \text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})\cdot\left(\mathbf{D}_{\boldsymbol{\tau}(v_j)} - \mathbb{P}_{\boldsymbol{\tau}(v_j)}\mathbb{P}^T_{\boldsymbol{\tau}(v_j)}\right)$.

For any edge of the form $v_{j+1}\xrightarrow{e_j} v_j$ on this path where $\theta(e_j)=\phi_i$ we have,

$$\begin{aligned}
\frac{\partial\Gamma(m)_c^{e_j\to v_j}}{\partial m_d^{v_{j+1}\to e_j}} =& \frac{1}{Z^{e_j\to v_j}}\sum_{\boldsymbol{c}_{e_j}|\boldsymbol{c}_{e_j}(v_j,v_{j+1})=(c,d)}\phi_i(\boldsymbol{c}_{e_j})\prod_{w\in e_j\setminus\{v_j,v_{j+1}\}} m^{w\to e_j}_{\boldsymbol{c}_{e_j}(w)}\\
&-\frac{m_c^{e_j\to v_j}}{Z^{e_j\to v_j}}\sum_{c'\in\mathsf{C}}\sum_{\boldsymbol{c}_{e_j}|\boldsymbol{c}_{e_j}(v_j,v_{j+1})=(c',d)}\phi_i(\boldsymbol{c}_{e_j})\prod_{w\in e_j\setminus\{v_j,v_{j+1}\}} m^{w\to e_j}_{\boldsymbol{c}_{e_j}(w)}\\
=&\frac{1}{Z^{e_j\to v_j}}\cdot\frac{1}{m_d^{v_{j+1}\to e_j}}\sum_{\boldsymbol{c}_{e_j}|\boldsymbol{c}_{e_j}(v_j,v_{j+1})=(c,d)}\phi_i(\boldsymbol{c}_{e_j})\prod_{w\in e_j\setminus v_j} m^{w\to e_j}_{\boldsymbol{c}_{e_j}(w)}\\
&-\frac{m_c^{e_j\to v_j}}{Z^{e_j\to v_j}}\cdot\frac{1}{m_d^{v_{j+1}\to e_j}}\sum_{c'\in\mathsf{C}}\sum_{\boldsymbol{c}_{e_j}|\boldsymbol{c}_{e_j}(v_j,v_{j+1})=(c',d)}\phi_i(\boldsymbol{c}_{e_j})\prod_{w\in e_j\setminus v_j} m^{w\to e_j}_{\boldsymbol{c}_{e_j}(w)}
\end{aligned}$$

Recall we defined a distribution μ_i over $\boldsymbol{c}_{e_j}$ and stochastic matrices

$$\boldsymbol{\Psi}_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)}$$

before.

Evaluating the derivative at the factorized fixed point gives the transformation matrix

$$\frac{\partial\Gamma(m)^{e_j\to v_j}}{\partial m^{v_{j+1}\to e_j}}\bigg|_{\overline{m}} = \frac{1}{\mathbb{P}_{\boldsymbol{\tau}(v_{j+1})}(d)\cdot\text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})}\left(\underset{\boldsymbol{c}_{e_j}\sim D_\theta}{\mathbf{Pr}}[\boldsymbol{c}_{e_j}(v_{j+1})=d\mid\boldsymbol{c}_{e_j}(v_j)=c]\right.$$

$$- \frac{1}{\text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})} \sum_{c' \in C} \mathop{\mathbf{Pr}}_{\boldsymbol{c}_{e_j} \sim D_\theta} [\boldsymbol{c}_{e_j}(v_{j+1}) = d \mid \boldsymbol{c}_{e_j}(v_j) = c'] \Bigg) \ ,$$

Then $\left.\frac{\partial m^{e_j \to v_j}}{\partial m^{v_{j+1} \to e_j}}\right|_{(m)^{\text{fp}}} = \frac{1}{\text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})} \left(\mathbf{I} - \frac{1}{\text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})} \mathbf{1}\mathbf{1}^\top\right) \boldsymbol{\Psi}^T_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})}$ where the † in the superscript denotes the pseudoinverse of the matrix.

Then we can write the transformation matrix for the step $v_{j+1} \xrightarrow{e_j} e_{j-1}$ as

$$\begin{aligned}
&\overline{\mathbf{M}}_{\theta(e_j),\mathrm{i}(v_j)|\mathrm{i}(v_{j+1})} = \\
&\left(\mathbf{D}_{\boldsymbol{\tau}(v_j)} - \mathbb{P}_{\boldsymbol{\tau}(v_j)} \mathbb{P}^\top_{\boldsymbol{\tau}(v_j)}\right) \cdot \\
&\left(\mathbf{I} - \frac{1}{\text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})} \mathbf{1}\mathbf{1}^\top\right) \boldsymbol{\Psi}^\top_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \mathbf{D}_{\boldsymbol{\tau}(v_j)} \left(\mathbf{I} - \mathbf{1}\mathbb{P}^\top_{\boldsymbol{\tau}(v_j)}\right) \left(\mathbf{I} - \frac{1}{\text{support}(\mathbb{P}_{\boldsymbol{\tau}(v_j)})} \mathbf{1}\mathbf{1}^\top\right) \boldsymbol{\Psi}^\top_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \mathbf{D}_{\boldsymbol{\tau}(v_j)} \left(\mathbf{I} - \mathbf{1}\mathbb{P}^\top_{\boldsymbol{\tau}(v_j)}\right) \boldsymbol{\Psi}^\top_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \left(\mathbf{I} - \mathbb{P}_{\boldsymbol{\tau}(v_j)} \mathbf{1}^\top\right) \mathbf{D}_{\boldsymbol{\tau}(v_j)} \boldsymbol{\Psi}^\top_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \left(\mathbf{I} - \mathbb{P}_{\boldsymbol{\tau}(v_j)} \mathbf{1}^\top\right) \boldsymbol{\Psi}_{\theta(e_j),\mathrm{i}(v_j)|\mathrm{i}(v_{j+1})} \ .
\end{aligned}$$

This establishes the first part of Claim 3.4.1. To establish the second part, consider the following chain of equalities where the first equality is one we know from the above chain.

$$\begin{aligned}
\overline{\mathbf{M}}_{\theta(e_j),\mathrm{i}(v_j)|\mathrm{i}(v_{j+1})} &= \left(\mathbf{I} - \mathbb{P}_{\boldsymbol{\tau}(v_j)} \mathbf{1}^\top\right) \mathbf{D}_{\boldsymbol{\tau}(v_j)} \boldsymbol{\Psi}^T_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \mathbf{D}_{\boldsymbol{\tau}(v_j)} \left(\mathbf{I} - \mathbf{1}\mathbb{P}^\top_{\boldsymbol{\tau}(v_j)}\right) \boldsymbol{\Psi}^\top_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \mathbf{D}_{\boldsymbol{\tau}(v_j)} \left(\boldsymbol{\Psi}_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \left(\mathbf{I} - \mathbb{P}_{\boldsymbol{\tau}(v_j)} \mathbf{1}^\top\right)\right)^\top \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \mathbf{D}_{\boldsymbol{\tau}(v_j)} \left(\boldsymbol{\Psi}_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} - \mathbb{P}_{\boldsymbol{\tau}(v_{j+1})} \mathbf{1}^\top\right)^\top \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \mathbf{D}_{\boldsymbol{\tau}(v_j)} \left(\left(\mathbf{I} - C_{\boldsymbol{\tau}(v_{j+1})} \mathbf{1}^\top\right) \boldsymbol{\Psi}_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)}\right)^\top \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})} \\
&= \mathbf{D}_{\boldsymbol{\tau}(v_j)} \overline{\mathbf{M}}^\top_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)} \mathbf{D}^\dagger_{\boldsymbol{\tau}(v_{j+1})}
\end{aligned}$$

If $\mathbb{P}_{\boldsymbol{\tau}(v_j)}(c) = 0$, then the c-th column of $\overline{\mathbf{M}}_{\theta(e_j),\mathrm{i}(v_{j+1})|\mathrm{i}(v_j)}$ and the c-th row of $\overline{\mathbf{M}}_{\theta(e_j),\mathrm{i}(v_j)|\mathrm{i}(v_{j+1})}$ are 0, and hence the second part of Claim 3.4.1 follows as well.

3.12 Random graph lemmas

3.12.1 Proof of Lemma 3.7.8

Let $\mathrm{M}^{\times}$ be a null model, and let $\boldsymbol{H} \sim \mathrm{M}^{\times}$. We use $\mathrm{a}_{\max}$ to denote $\max_{i\in[F]} \mathrm{a}(i)$ and $\overline{\phi}_{\max}$ to denote $\max_{i\in[F]} \overline{\phi_i}$.

Define the notion of (τ, Cl)-consistent:

Definition 3.12.1. Given τ and Cl, we say a subgraph Γ of a bipartite factor graph with right vertex set $R(\Gamma)$ is (τ, Cl)-*consistent* if every $\gamma = (v_1, \ldots, v_{\mathrm{a}(i)}) \in R(\Gamma)$ satisfies $(\tau(v_1), \ldots, \tau(v_{\mathrm{a}(i)})) = \mathrm{Cl}(\theta(\gamma))$.

It is easy to see the following.

Observation 3.12.2. Suppose Γ is a subgraph of $\mathrm{Bip}(\mathcal{K}_n)$. Then the probability that Γ is a subgraph of $\boldsymbol{H}$ is equal to $\prod_{\gamma\in R(\Gamma)} \frac{\overline{\phi_{\theta(\gamma)}}}{n^{\mathrm{a}(\theta(\gamma))-1}} \mathbf{1}[\Gamma \text{ is } (\tau, \mathrm{Cl})\text{-consistent}]$.

Definition 3.12.3 (Partially labeled graph)**.** A *partially labeled graph* $\Gamma = (L, R, P, p, E)$ is given by a left vertex set L, a right vertex set R, a distinguished set of left vertices P along with an injective labeling of the distinguished vertices $p : P \to [n]$, and edge set E.

Definition 3.12.4 (Occurrence of partially labeled graph)**.** An *occurrence* of a partially labeled graph $\Gamma = (L, R, P, p, E)$ in $\mathrm{Bip}(\boldsymbol{H})$ is a pair of injective functions $f_L : L \to L(\boldsymbol{H})$ and $f_R : R \to R(\boldsymbol{H})$ such that for all $v \in P$ satisfies $f_L(v) = p(v)$, and if $\{u, v\} \in E$, then $\{f_L(u), f_R(v)\} \in E(\mathrm{Bip}(\boldsymbol{H}))$.

Given a partially labeled subgraph Γ we are interested in bounding the expected number of occurrences of Γ in $\mathrm{Bip}(\boldsymbol{H})$.

Lemma 3.12.5. *Given partially labeled graph $\Gamma = (L, R, P, p)$ with no isolated right vertices, the expected number of occurrences of Γ in $\mathrm{Bip}(\boldsymbol{H})$ is at most*

$$n^{|L|+|R|-|P|-|E|}(F\mathrm{a}_{\max}\overline{\phi}_{\max})^{|E|}.$$

Proof. There are at most $n^{|L|-|P|}$ choices for f_L. For each potential choice of f_R, we can associate $t_{f_R} : R \to [F]$ such that $t_{f_R}(r)$ is the type of $f_R(r)$. There are at most $F^{|R|}$ possible values for t_{f_R}. For each fixed choice of f_L and t, we wish to bound the

expected number of f_R such that (f_L, f_R) is an occurrence and $t_{f_R} = t$. The number of such potential f_R is bounded by

$$\prod_{i\in[R]} \mathrm{a}(t(i))^{\deg_\Gamma(i)} n^{\mathrm{a}(t(i))-\deg_\Gamma(i)}$$

and the probability that (f_L, f_R) is a valid occurrence for a given such f_R is at most

$$\prod_{i\in[R]} \frac{\overline{\phi_i}}{n^{\mathrm{a}(t(i))-1}},$$

which gives us a bound of

$$\prod_{i\in[R]} \left(\frac{\mathrm{a}(t(i))}{n}\right)^{\deg_\Gamma(i)} \cdot n \cdot \overline{\phi_i} \leqslant \left(\frac{\mathrm{a}_{\max}}{n}\right)^{|E|} n^{|R|} \overline{\phi}_{\max}^{|R|}.$$

Combining this with the bound on total number of f_L and t_{f_R} gives us a bound of:

$$n^{|L|-|P|} F^{|R|} \left(\frac{\mathrm{a}_{\max}}{n}\right)^{|E|} n^{|R|} \overline{\phi}_{\max}^{|R|} = n^{|L|+|R|-|P|-|E|} F^{|R|} \mathrm{a}_{\max}^{|E|} \overline{\phi}_{\max}^{|R|}.$$

Since there are no isolated vertices, $|E| \geqslant |R|$ and hence the above is at most

$$n^{|L|+|R|-|P|-|E|} (F \mathrm{a}_{\max} \overline{\phi}_{\max})^{|E|}.$$

□

Definition 3.12.6. For a graph Γ and a subset of its vertices S we use $B_\Gamma(S, r)$ to denote the radius-r ball around set S within Γ. We also abuse notation and use $B_{\boldsymbol{H}}(S, r)$ to mean $B_{\mathrm{Bip}(\boldsymbol{H})}(S, r)$.

Lemma 3.12.7. *Given a set of vertices S in $\mathrm{Bip}(\boldsymbol{H})$, the probability that $|E(B_{\boldsymbol{H}}(S, r))| - |V(B_{\boldsymbol{H}}(S, r))| + |S| \geqslant t$ is at most $\left(\frac{(F \mathrm{a}_{\max} \overline{\phi}_{\max})^{2(r+1)} (36 t^3 r^2)^5 |S|}{n}\right)^t$.*

In preparation to prove Lemma 3.12.7 we will need the following statement about counts of trees with a bounded number of leaves. The statement along with a proof can be found in [BMR21, Lemma 6.33].

Lemma 3.12.8. *The number of nonisomorphic trees on v vertices and L leaves is bounded by $(4Lv)^{2L+1}$.*

Proof of Lemma 3.12.7. Let us call a partially labeled subgraph $\Gamma = (L, R, P, p, E)$ a *candidate witness* if

- $p(P) = S$,
- Γ can be expressed as $F \cup B$ where F is a forest and $B = \{\{u_1, v_1\}, \ldots, \{u_t, v_t\}\}$ is a set of t additional edges,
- F has $|P|$ connected components where each connected component contains exactly one $v \in P$ and has depth r when rooted at v.

If $|E(B_{\boldsymbol{H}}(S,r))| - |V(B_{\boldsymbol{H}}(S,r))| + |S| \geqslant t$, then there must be an occurrence of some candidate witness $\Gamma = (L, R, P, p, E)$ within $\boldsymbol{H}$. We will first find a "simple" subgraph of $\Gamma = F \cup B$, which we call the *trim* of Γ. First let us augment F to $\widetilde{F}$ by adding a single vertex w and connecting it to all vertices in P – note that $\widetilde{F}$ is a tree. Now let $\mathrm{Trim}(\widetilde{F})$ be the tree obtained by only choosing vertices that lie on paths from vertices in $L := \{u_1, \ldots, u_t, v_1, \ldots, v_t\}$ to w. Since the depth of $\mathrm{Trim}(\widetilde{F})$ is $r+1$ and has at most $2t$ leaves when rooted at w, the number of vertices in $\mathrm{Trim}(\widetilde{F})$ is at most $2tr + 1$. Let $\mathrm{Trim}(\Gamma) = (L', R', P, p, E')$ be the graph obtained by deleting w from $\mathrm{Trim}(\widetilde{F})$, adding edges $\{u_1, v_1\}, \ldots, \{u_t, v_t\}$, and adding vertices in $P \setminus V(\mathrm{Trim}(\widetilde{F}))$. Since $\mathrm{Trim}(\Gamma)$ is a subgraph of Γ there must be an occurrence of $\mathrm{Trim}(\Gamma)$ in $\boldsymbol{H}$.

$\mathrm{Trim}(\Gamma)$ has at most $2tr$ vertices and

$$|E(\mathrm{Trim}(\Gamma))| - |L(\mathrm{Trim}(\Gamma))| - |R(\mathrm{Trim}(\Gamma))| + |P| \geqslant t.$$

Thus, from Lemma 3.12.5 the probability that $\mathrm{Trim}(\Gamma)$ occurs in $\boldsymbol{H}$ is bounded by $\left(\frac{(F\mathsf{a}_{\max}\overline{\phi}_{\max})^{2(r+1)}}{n}\right)^t$. Thus:

$$\begin{aligned}
&\mathbf{Pr}[|E(B_{\boldsymbol{H}}(S,r))| - |V(B_{\boldsymbol{H}}(S,r))| + |S| \geqslant t] \leqslant \\
&\mathbf{Pr}[\text{there is a candidate witness } \Gamma \text{ in } \boldsymbol{H}] \\
&\leqslant \mathbf{Pr}[\text{there is a trim of a candidate witness } \mathrm{Trim}(\Gamma) \text{ in } \boldsymbol{H}] \\
&\leqslant \sum_{\Gamma' \text{ trim of a candidate witness}} \mathbf{Pr}[\Gamma' \text{ in } \boldsymbol{H}] \\
&\leqslant \sum_{\Gamma' \text{ trim of a candidate witness}} \left(\frac{(F\mathsf{a}_{\max}\overline{\phi}_{\max})^{2(r+1)}}{n}\right)^t. \qquad (3.48)
\end{aligned}$$

Next we bound the number of terms in the above summation. Since each Γ' in the above sum can be specified by taking a tree on at most $2tr+1$ vertices and at most $2t+1$ leaves, deleting one vertex, and labeling each neighbor of this deleted vertex with an element of P, from Lemma 3.12.8 and the fact the maximum degree in a tree is bounded by the number of leaves the number of terms is at most:

$$(4(2t+1)(2tr+1))^{4t+3} \cdot (2tr+1) \cdot |P|^{2t+1} \leqslant ((36t^3r^2)^5|P|)^t.$$

Plugging this into (3.48) and using $|S| = |P|$ gives:

$$\mathbf{Pr}[|E(B_{\boldsymbol{H}}(S,r))| - |V(B_{\boldsymbol{H}}(S,r))| + |S| \geqslant t] \leqslant \left(\frac{(F\mathrm{a}_{\max}\overline{\phi}_{\max})^{2(r+1)}(36t^3r^2)^5|S|}{n}\right)^t$$

□

Corollary 3.12.9. *With probability* $1 - o_n(1)$, $\mathrm{Bip}(\boldsymbol{H})$ *is* r*-bicycle free for* $r = \frac{\log n}{\log\log n}$.

Proof. This is a simple consequence of Lemma 3.12.7. Indeed, by Lemma 3.12.7 the probability that the radius-$(r+1)$ neighborhood of a single vertex $v \in [n]$ contains more than one cycle is at most $\frac{1}{n^{2-o_n(1)}}$, and hence by a union bound over all vertices the probability of any left vertex containing more than one cycle in its radius-$r+1$ neighborhood is bounded by $\frac{1}{n^{1-o_n(1)}}$. Since every right vertex is incident to a left vertex, the statement we wish to prove follows. □

3.12.2 Proof of Lemma 3.12.11

We will need the following combinatorial lemma that appears in [FM17, Lemma A.2].

Lemma 3.12.10. *If* e *distinct edges of a graph* Γ *belong to a* r*-bicycle frame, then* $\mathrm{Exc}(\Gamma) \geqslant \frac{e}{r}$.

Our proof of the statement below follows the same strategy as the proof of a similar statement appearing in [FM17].

Lemma 3.12.11. *Suppose* S *and* L *are disjoint sets of right vertices of* $\mathcal{K}_n$ *of size at most* $\log^2 n$, $\mathbf{1}_\gamma$ *is the indicator random variable for whether* γ *is in* $\boldsymbol{H}$, μ_γ *is the probability that* γ *is in* $\boldsymbol{H}$, *and* $\boldsymbol{\mathcal{E}}$ *denotes the event that* $\boldsymbol{H}$ *is* r*-bicycle free for* $r = \frac{\log n}{\log\log n}$. *Then:*

$$\left|\mathbf{E}\left[\prod_{\gamma\in S}(\mathbf{1}_\gamma - \mu_\gamma)\prod_{\gamma\in L}\mathbf{1}_\gamma \mathbf{1}[\boldsymbol{\mathcal{E}}]\right]\right| \leqslant \prod_{\gamma\in S\cup L}\mu_\gamma \cdot 2^{|S|}\left(\frac{1}{n^{.5}}\right)^{\frac{|S|}{r} - \mathrm{Exc}(\mathrm{Clos}(S\cup L))}.$$

Proof.

$$\begin{aligned}\left|\mathbf{E}\left[\prod_{\gamma\in S}(\mathbf{1}_\gamma-\mu_\gamma)\prod_{\gamma\in L}\mathbf{1}_\gamma\mathbf{1}[\boldsymbol{\mathcal{E}}]\right]\right| &= \left|\sum_{J\subseteq S}\mathbf{E}\left[\prod_{\gamma\in J}\mathbf{1}_\gamma\cdot\prod_{\gamma\in S\setminus J}(-\mu_\gamma)\cdot\prod_{\gamma\in L}\mathbf{1}_\gamma\mathbf{1}[\boldsymbol{\mathcal{E}}]\right]\right| \\ &= \left|\sum_{J\subseteq S}(-1)^{|S|-|J|}\prod_{\gamma\in S\setminus J}\mu_\gamma\,\mathbf{E}\left[\prod_{\gamma\in J\cup L}\mathbf{1}_\gamma\mathbf{1}[\boldsymbol{\mathcal{E}}]\right]\right| \\ &= \left|\sum_{J\subseteq S}(-1)^{|J|}\prod_{\gamma\in S\cup L}\mu_\gamma\,\mathbf{Pr}\left[\boldsymbol{\mathcal{E}}|\gamma\in\boldsymbol{H}\;\forall\gamma\in J\cup L\right]\right| \\ &= \prod_{\gamma\in S\cup L}\mu_\gamma\left|\sum_{J\subseteq S}(-1)^{|J|}\mathbf{Pr}\left[\boldsymbol{\mathcal{E}}|\gamma\in\boldsymbol{H}\;\forall\gamma\in J\cup L\right]\right|\end{aligned} \tag{3.49}$$

Now we focus our attention on understanding the quantity

$$\left|\sum_{J\subseteq S}(-1)^{|J|}\mathbf{Pr}\left[\boldsymbol{\mathcal{E}}|\gamma\in\boldsymbol{H}\;\forall\gamma\in J\cup L\right]\right|.$$

Let $\boldsymbol{g}_0$ be $\boldsymbol{H}\setminus(S\cup L)$.

$$\left|\sum_{J\subseteq S}(-1)^{|J|}\mathbf{Pr}\left[\boldsymbol{\mathcal{E}}|\gamma\in\boldsymbol{H}\;\forall\gamma\in J\cup L\right]\right| = \left|\mathop{\mathbf{E}}_{\boldsymbol{g}_0}\sum_{J\subseteq S}(-1)^{|J|}\mathbf{Pr}[\boldsymbol{\mathcal{E}}|\gamma\in\boldsymbol{H}\;\forall\gamma\in J\cup L,\boldsymbol{g}_0]\right| \tag{3.50}$$

For $K\subseteq S$, define $f_{\boldsymbol{g}_0}(K)$ as 1 if $\mathrm{Clos}(\boldsymbol{g}_0\cup K\cup L)$ has no r-bicycles and 0 otherwise. Suppose there is $s\in S$ that $f_{\boldsymbol{g}_0}$ does not depend on – that is, for any $K\subseteq S$, $f_{\boldsymbol{g}_0}(K)=f_{\boldsymbol{g}_0}(K\Delta\{s\})$, then for every J which contains s:

$$\mathbf{Pr}[\boldsymbol{\mathcal{E}}|\gamma\in\boldsymbol{H}\;\forall\gamma\in J\cup L,\boldsymbol{g}_0]=\mathbf{Pr}[\boldsymbol{\mathcal{E}}|\gamma\in\boldsymbol{H}\;\forall\gamma\in J\cup L\setminus\{s\},\boldsymbol{g}_0].$$

This means (3.50) is equal to:

$$\begin{aligned}(3.50) &= \left|\mathop{\mathbf{E}}_{\boldsymbol{g}_0}\mathbf{1}[f_{\boldsymbol{g}_0}\text{ depends on every }s\in S]\sum_{J\subseteq S}(-1)^{|J|}\mathbf{Pr}\left[\boldsymbol{\mathcal{E}}|\gamma\in\boldsymbol{H}\;\forall\gamma\in J\cup L,\boldsymbol{g}_0\right]\right| \\ &\leqslant 2^{|S|}\cdot\mathop{\mathbf{Pr}}_{\boldsymbol{g}_0}[f_{\boldsymbol{g}_0}\text{ depends on every }s\in S].\end{aligned} \tag{3.51}$$

Let E_S be the set of all edges incident to S. If f_{g_0} depends on every $s \in S$, then the function h_{g_0} defined on subsets of E_S which is 1 on input $K \subseteq E_S$ if $\text{Clos}(g_0 \cup L) \cup K$ has no r-bicycles depends on at least $|S|$ edges in E_S. That means:

(3.51)

$$\leqslant 2^{|S|} \cdot \mathop{\mathbf{Pr}}_{g_0}[h_{g_0} \text{ depends on at least } |S| \text{ edges in } E_S]$$
$$\leqslant 2^{|S|} \cdot \mathop{\mathbf{Pr}}_{g_0}[\text{At least } |S| \text{ edges in } E_S \text{ part of } r\text{-bicycle frame in } \text{Clos}(g_0 \cup S \cup L)]$$

which, via Lemma 3.12.10, can be bounded by

$$\leqslant 2^{|S|} \cdot \mathop{\mathbf{Pr}}_{g_0}\left[\text{Exc}(B_{\text{Clos}(g_0 \cup S \cup L)}(\text{Clos}(S), r)) \geqslant \frac{|S|}{r}\right]$$
$$\leqslant 2^{|S|} \cdot \mathop{\mathbf{Pr}}_{g_0}\left[\text{Exc}(B_{\text{Clos}(g_0)}(\text{Clos}(S \cup L), r)) \geqslant \frac{|S|}{r} - \text{Exc}(\text{Clos}(S \cup L))\right]$$
$$\leqslant 2^{|S|} \cdot \mathop{\mathbf{Pr}}_{H}\left[\text{Exc}(B_{H}(\text{Clos}(S \cup L), r)) \geqslant \frac{|S|}{r} - \text{Exc}(\text{Clos}(S \cup L))\right]$$

By Lemma 3.12.7, the bounds on size of $|S|$ and $|L|$, and the value of r, we can conclude that the above is at most:

$$\leqslant 2^{|S|} \min\left\{\left(\frac{1}{n^{.5}}\right)^{\frac{|S|}{r} - \text{Exc}(\text{Clos}(S \cup L))}, 1\right\}$$
$$\leqslant 2^{|S|}\left(\frac{1}{n^{.5}}\right)^{\frac{|S|}{r} - \text{Exc}(\text{Clos}(S \cup L))}.$$

Plugging this back into (3.49) gives us the desired statement. □

3.12.3 Proof of Claim 3.7.22

Proof. Consider any $\text{Sh} \in \mathcal{U}_{s,x,\Delta}$. Recall that $S(\text{Sh})$ is the set of singleton vertices in $R(\text{Sh})$ and $D(\text{Sh})$ the set of duplicative vertices in $R(\text{Sh})$. $D^*(\text{Sh})$ is the maximum weight subset of $D(W)$ that makes Sh r-bicycle free, and $\Delta(\text{Sh}) = w(D(\text{Sh})) - w(D^*(\text{Sh}))$. Thus we deduce that $\Delta(\text{Sh}) \geqslant |D(\text{Sh}) \setminus D^*(\text{Sh})|$ We apply the following procedures to the walk Sh.

1. Break the walk Sh into $\leqslant s + \Delta + |D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})| + k$ segments by first removing the vertices $S(\mathrm{Sh})$ and $D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})$ from Sh and second breaking the remaining segments at endpoints of the k links in Sh. Denote the new union of walks Sh_1.

2. Since $S(\mathrm{Sh})$ and $D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})$ are removed from Sh, Sh_1 is singleton free, and the graph on Sh_1, denoted by $\boldsymbol{G}(\mathrm{Sh}_1)$, is r-bicycle free.

3. We contract the graph $\boldsymbol{G}(\mathrm{Sh}_1)$ by merging all adjacent edges that share a degree-2 vertex. We denote the resulting graph $\boldsymbol{G}(\mathrm{Sh}_1)_c$, and we note that the vertices left in $\boldsymbol{G}(\mathrm{Sh}_1)_c$ are those with degree $\geqslant 3$ in $\boldsymbol{G}(\mathrm{Sh}_1)$.

We make the following observations on the size of $\boldsymbol{G}(\mathrm{Sh}_1)$ and $\boldsymbol{G}(\mathrm{Sh}_1)_c$.

The number of vertices in $\boldsymbol{G}(\mathrm{Sh}_1)$ is $|V(\mathrm{Sh})| - |S(\mathrm{Sh})| - |D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})|$. The number of edges in $\boldsymbol{G}(\mathrm{Sh}_1)$ is $\leqslant |E(\mathrm{Sh})| - 2|S(\mathrm{Sh})| - 2|D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})|$.

To bound the number of vertices in $\boldsymbol{G}(\mathrm{Sh}_1)_c$ is we apply the following lemma from [MOP20].

Lemma 3.12.12 (Lemma 6.18 in [MOP20]). *Let C be a $(k, 2\ell)$-nonbacktracking, internally 2ℓ-bicycle-free linkage. Assume $\log k\ell = o(\ell)$. Then $\boldsymbol{G}(C)$ has at most $O(k \log k\ell)$ vertices of degree exceeding 2.*

Applying the lemma to the walk Sh_1, we obtain that the number of degree $\geqslant 3$ vertice in Sh_1 is $O(k \log k\ell)$. Thus the number of vertices in $\boldsymbol{G}(\mathrm{Sh}_1)_c$ is $O(k \log k\ell)$. The number of edges in $\boldsymbol{G}(\mathrm{Sh}_1)_c$ is

$$
\begin{aligned}
&|E(\boldsymbol{G}(\mathrm{Sh}_1))| - (|V(\boldsymbol{G}(\mathrm{Sh}_1))| - |V(\boldsymbol{G}(\mathrm{Sh}_1)_c)|) \\
&\leqslant |E(\mathrm{Sh})| - 2|S(\mathrm{Sh})| - 2|D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})| - |V(\mathrm{Sh})| + |S(\mathrm{Sh})| + \\
&|D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})| + O(k \log k\ell) \\
&= (|E(\mathrm{Sh})| - |V(\mathrm{Sh})|) - |S(\mathrm{Sh})| - |D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})| + O(k \log k\ell) \\
&\leqslant x + O(k \log k\ell).
\end{aligned}
$$

Now to count the number of distinct $\mathrm{Sh} \in \mathcal{U}_{s,x,\Delta}$, it suffices to count 1. the number of distinct sets of breaking points ($S(\mathrm{Sh})$, $D(\mathrm{Sh}) \setminus D^*(\mathrm{Sh})$, and k link endpoints), 2. the number of distinct graphs $\boldsymbol{G}(\mathrm{Sh}_1)$, 3. given the breaking points and $\boldsymbol{G}(\mathrm{Sh}_1)$, the number of distinct walk segments in $\boldsymbol{G}(\mathrm{Sh}_1)$ with those breaking points. We count each of the three quantities separately and multiply them together to obtain an upper bound on $|\mathcal{U}_{s,x,\Delta}|$.

The number of distinct sets of breaking points: these breaking points breaks Sh into at most $s + 2\Delta + k$ segments. So there are $(2k\ell)^{s+2\Delta+k}$ ways to choose these breaking points.

The number of distinct graphs $\boldsymbol{G}(\mathrm{Sh}_1)$: $\boldsymbol{G}(\mathrm{Sh}_1)$ can be contracted to a graph $\boldsymbol{G}(\mathrm{Sh}_1)_c$ on $O(k \log k\ell)$ vertices and $x + O(k \log k\ell)$ edges. Each edge in $\boldsymbol{G}(\mathrm{Sh}_1)_c$ represents a length-$\leqslant 2k\ell$ simple path in $\boldsymbol{G}(\mathrm{Sh}_1)$. Thus there are

$$O(k \log k\ell)^{2(x+O(k \log k\ell))} \cdot (2k\ell)^{x+O(k \log k\ell)}$$

distinct graphs $\boldsymbol{G}(\mathrm{Sh}_1)$.

Given the breaking points and $\boldsymbol{G}(\mathrm{Sh}_1)$, the number of distinct walk segments in $\boldsymbol{G}(\mathrm{Sh}_1)$ with those breaking points: since $\boldsymbol{G}(\mathrm{Sh}_1)$ is r-bicycle free with $r \geqslant 2\ell$ and each segment is of length $\leqslant 2\ell$, there are only 2 distinct length $\leqslant 2\ell$ walk between any two vertices in $\boldsymbol{G}(\mathrm{Sh}_1)$. Thus the number of distinct walk segments are $2^{s+2\Delta+k}$.

Combine the three bound together we obtain that

$$|\mathcal{U}_{s,x,\Delta}| \leqslant (2k\ell)^{s+2\Delta+x+O(k \log k\ell)} O(k \log k\ell)^{2(x+O(k \log k\ell))} 2^{s+2\Delta+k}.$$

From this bound we quickly derive that

$$\begin{aligned}
&\sum_{\mathrm{Sh} \in \mathcal{U}_{s,x,\Delta}} \left(\frac{\beta^r}{n^{.5}}\right)^{\frac{|S(\mathrm{Sh})|}{r}} \left(\frac{\beta}{n^{.5}}\right)^{\mathrm{Exc}_{\mathrm{Sh}}} \left(\frac{\beta}{n}\right)^{\Delta(\mathrm{Sh})} \\
&= |\mathcal{U}_{s,x,\Delta}| \cdot \left(\frac{\beta^r}{n^{.5}}\right)^{\frac{s}{r}} \left(\frac{\beta}{n^{.5}}\right)^{x} \left(\frac{\beta}{n}\right)^{\Delta} \\
&= \left(\frac{\beta^r \cdot (4k\ell)^r}{n^{.5}}\right)^{\frac{s}{r}} \left(\frac{\beta \cdot 2k\ell \cdot O(k \log k\ell)^2}{n^{.5}}\right)^{x} \left(\frac{4\beta \cdot (2k\ell)^2}{n}\right)^{\Delta} \cdot \\
&2^k (2k\ell \cdot O(k \log k\ell)^2)^{O(k \log k\ell)} \\
&\leqslant \left(\frac{\beta^r \cdot (4k\ell)^r}{n^{.5}}\right)^{\frac{s}{r}} \left(\frac{\beta \cdot 2(k\ell)^3}{n^{.5}}\right)^{x} \left(\frac{4\beta \cdot (2k\ell)^2}{n}\right)^{\Delta} (2k\ell)^{O(k \log k\ell)}
\end{aligned}$$

The last inequality follows since we pick ℓ such that $\log k\ell = o(\ell)$. □

Chapter 4

Explicit near-Ramanujan graphs

This chapter has been adapted from [MOP20], a paper co-authored by the author of this thesis, Ryan O'Donnell, and Pedro Paredes.

For every constant $d \geqslant 3$ and $\varepsilon > 0$, we give a deterministic $\mathrm{poly}(n)$-time algorithm that outputs a d-regular graph on $\Theta(n)$ vertices that is ε-near-Ramanujan; i.e., its eigenvalues are bounded in magnitude by $2\sqrt{d-1}+\varepsilon$ (excluding the single trivial eigenvalue of d).

4.1 Introduction

In this work, we obtain explicit d-regular ε-near-Ramanujan graphs for every $d \geqslant 3$ and every $\varepsilon > 0$. As an example, we give the first explicit family of 7-regular graphs with $\lambda_2(G), |\lambda_n(G)| \leqslant 2\sqrt{6}+\varepsilon$. Our main result is the following:

Theorem 4.1.1. *For any $d \geqslant 3$ and any $\varepsilon > 0$, there is an explicit (deterministic polynomial-time computable) infinite family of d-regular graphs G with*

$$\max\{\lambda_2(G), |\lambda_n(G)|\} \leqslant 2\sqrt{d-1}+\varepsilon.$$

The key technical result that we prove in service of this is the following:

Theorem 4.1.2. *Let G be an arbitrary d-regular n-vertex graph. Assume that the r-neighborhood of every vertex contains at most one cycle, where $r \gg (\log\log n)^2$. Then a random edge-signing of G has all its eigenvalues bounded in magnitude by $2\sqrt{d-1}+o_n(1)$, with high probability.*

See Section 4.1.4 for a comparison of Theorem 4.1.2 with a similar theorem of Bilu and Linial [BL06], which has an alternate hypothesis and a weaker conclusion.

4.1.1 On near-Ramanujan graphs

Let us put our results into context. Loosely speaking, *expander graphs* are sparse graphs in which every small set of vertices has many edges on its boundary. For an early paper working out relationships between various possible definitions, see Alon [Alo86]. For a thorough reference describing expanders' myriad applications and connections to various parts of computer science and mathematics, see the survey of Hoory, Linial, and Wigderson [HLW06].

A good way to quantify the definition of expansion is through the eigenvalues of the graph.

Definition 4.1.3 (Graph eigenvalues)**.** Let G be an n-vertex d-regular multigraph. We write $\lambda_i = \lambda_i(G)$ for the *eigenvalues* of its adjacency matrix A, and we always assume they are ordered with $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. A basic fact is that $\lambda_1 = d$ always; this is called the *trivial* eigenvalue. We also write $\lambda = \lambda(G) = \max\{\lambda_2, |\lambda_n|\}$.

The extent to which a d-regular graph G is "expanding" is governed by the magnitude of its nontrivial eigenvalues; in particular, by λ_2 and (to a lesser extent) $|\lambda_n|$. Together these are captured by the parameter $\lambda(G)$. The smaller $\lambda(G)/d$ is, the better G's expansion; typically, a graph is called expanding when this ratio is bounded away from 1.

Definition 4.1.4 (Spectral expanders)**.** An infinite sequence of d-regular graphs (G_n) is said to be a *family of expanders* if there is a constant $\delta > 0$ such that $\lambda(G_n) \leqslant (1-\delta)d$ for all n.

Pinsker [Pin73] introduced the English terminology and showed that random (bipartite) graphs have positive expansion properties with high probability (see also [BK67, Mar73a]). Indeed, it can be shown [Alo86] that a uniformly random d-regular graph has $\lambda_2(G) < (1-\delta)d$ with high probability, for some universal $\delta > 0$ (see Theorem 4.1.8 below for a much stronger result). However, for almost all of the numerous practical applications of expanders in theoretical computer science (error correcting codes, derandomization, complexity theory, cryptography, metric embeddings, etc.) it is important for the graphs to be *explicit* — i.e., constructible

by a deterministic polynomial-time algorithm. Indeed, it is even better if they are *strongly explicit*, meaning that their adjacency list is computable in poly $\log n$ time.

4.1.2 Review of Ramanujan and near-Ramanujan families

Margulis [Mar73b] was the first to provide an explicit expander family; a slight variant of it, which is 8-regular, was shown [GG81] to have $\lambda \leqslant 5\sqrt{2} \approx 7.1$ (see [HLW06]). A natural question then is to provide explicit d-regular expanders, for various values of d, with λ as small as possible as a function of d. The well-known *Alon–Boppana bound* shows that $2\sqrt{d-1}$ is essentially a lower bound:

Theorem 4.1.5. *([Alo86, Nil91, Fri93].) Let G be an n-vertex d-regular multigraph. Then $\lambda_2(G) \geqslant 2\sqrt{d-1} - O(1/\log^2 n)$.*

On the other hand, using the resolution of the Ramanujan–Petersson conjectures in various number-theoretic settings, it is possible to construct d-regular expander families that meet the bound $\lambda(G) \leqslant 2\sqrt{d-1}$ for some values of d. Lubotzky–Phillips–Sarnak [LPS88] dubbed such graphs *Ramanujan*.

Definition 4.1.6 (Ramanujan graphs)**.** A d-regular (multi)graph G is called *(two-sided) Ramanujan* whenever $\lambda(G) \leqslant 2\sqrt{d-1}$. When we merely have $\lambda_2(G) \leqslant 2\sqrt{d-1}$, we call G *one-sided Ramanujan*; if G is bipartite this implies that $|\lambda_i| \leqslant 2\sqrt{d-1}$ for all $i \neq 1, n$, with $\lambda_n(G) = -d$ being inevitable.

We remark that some expander properties (e.g., edge-expansion for small sets) only need a one-sided eigenvalue bound, whereas others (e.g., the Expander Mixing Lemma) need a two-sided bound.

Regarding the explicit construction of d-regular Ramanujan graphs using number theory, the case when $d-1$ is an odd prime is due to Ihara [Iha66] (implicitly) and to Lubotzky–Phillips–Sarnak [LPS88] and Margulis [Mar88] (independently); the $d-1=2$ case is by Chiu [Chi92]; and, the general prime power case mentioned below is due to Morgenstern [Mor94]. For extensions to general d where the eigenvalue bound depends on the number of distinct prime divisors of $d-1$, see [Piz90, Cla06].

Theorem 4.1.7. *([Mor94].) For any $d \geqslant 3$ with $d-1$ a prime power, there is a strongly explicit family of d-regular Ramanujan graphs.*

For all other values of d — e.g., for $d = 7$ — it is unknown if infinite families of d-regular Ramanujan graphs exist (but see Theorem 4.1.12 below for the one-sided bipartite case). However, it is known that *near*-Ramanujan graph families exist for every d. Alon [Alo86] conjectured that a random n-vertex d-regular graph $\boldsymbol{G}$ has $\lambda(\boldsymbol{G}) \leqslant 2\sqrt{d-1} + o_n(1)$ with high probability, and this was proven two decades later by Friedman [Fri08]. Bordenave [Bor19] has recently given a simpler proof, and our paper will involve modifying and derandomizing Bordenave's work.

Theorem 4.1.8. *([Fri08].) Fix any $d \geqslant 3$ and $\varepsilon > 0$ and let $\boldsymbol{G}$ be a uniformly random d-regular graph. Then*

$$\mathbf{Pr}\left[\lambda(\boldsymbol{G}) \leqslant 2\sqrt{d-1} + \varepsilon\right] \geqslant 1 - o_n(1).$$

In fact [Bor19], $\boldsymbol{G}$ achieves the subconstant $\varepsilon = \widetilde{O}(1/\log^2 n)$ with probability at least $1 - 1/n^{.99}$.

A natural question then is whether, for every d, one can achieve *explicit* graph families that are "ε-near-Ramanujan" as above. In their work introducing the *zig-zag product*, Reingold–Vadhan–Wigderson [RVW02] asked whether explicit families could at least reach a bound of $O(\sqrt{d})$; towards this, their work gave strongly explicit families with $\lambda(G) \leqslant O(d^{2/3})$. By extending their approach, Ben-Aroya and Ta-Shma reached $d^{1/2+o(1)}$:

Theorem 4.1.9. *([RVW02, BT11].) There are strongly explicit families of d-regular multigraphs G satisfying the bound $\lambda(G) \leqslant \sqrt{d} \cdot 2^{O(\sqrt{\log d})}$.*

Bilu and Linial [BL06] got even closer to $O(\sqrt{d})$, using a new approach based on random *lifts* that will prove important in our paper. Their graph families are not strongly explicit, although Bilu–Linial point out they are at least "probabilistically strongly explicit" (q.v. Theorem 4.1.13).

Theorem 4.1.10. *([BL06].) There are explicit families of d-regular multigraphs G satisfying the bound $\lambda(G) \leqslant \sqrt{d} \cdot O(\log^{1.5} d)$.*

Due to their asymptotic-in-d nature, neither of Theorems 4.1.9 and 4.1.10 gives much help for specific small values of d not covered by Morgenstern, such as $d = 7$. In such cases, one can use a simple idea due to Cioabă and Murty [CM08] (cf. [dlHM06]): take a prime (or prime power) $q < d - 1$, form a $(q+1)$-regular

Ramanujan graph, and then add in $d-q-1$ arbitrary perfect matchings. It is shown in [CM08] that each perfect matching increases $\lambda(G)$ by at most 1. Hence:

Theorem 4.1.11. *([CM08].) For any $d \geqslant 3$, there is a strongly explicit family of d-regular multigraphs with $\lambda(G) \leqslant 2\sqrt{d-1} + gap(d)$, where $gap(d)$ denotes the least value g such that $d-1-g$ is a prime (power). One can bound $gap(d)$ by $O(\log^2 d)$ under Cramér's conjecture, by $O(\sqrt{d}\log d)$ under the Riemann Hypothesis, or by $O(d^{.525})$ unconditionally.*

For example, this gives strongly explicit 7-regular multigraphs with $\lambda(G) \leqslant 2\sqrt{5}+1 < 5.5$. For comparison, the Ramanujan bound is $2\sqrt{6} < 4.9$.

Finally, Marcus–Spielman–Srivastava [MSS15a, MSS15b] recently introduced the *Interlacing Polynomials Method* and used it to show that *one-sided bipartite* Ramanujan graphs exist for all $d \geqslant 3$ and all even n. Their proof was merely existential, but Cohen [Coh16] was able to make it explicit (though not strongly so):

Theorem 4.1.12. *([MSS15a, MSS15b, Coh16].) For any $d \geqslant 3$, there is an explicit family of one-sided bipartite, d-regular, Ramanujan multigraphs.*

As mentioned, this theorem gives an n-vertex graph for every even n, which is slightly better than all other results mentioned in this section, which merely give graphs for a dense sequence of n's (typically, a sequence n_j with $n_{j+1} - n_j = o(n_j)$). Also, as pointed out to us by Nikhil Srivastava, pairing left and right vertices in the construction from Theorem 4.1.12 and merging them gives "twice-Ramanujan" graphs of every even degree; i.e., $2d$-regular graphs for all $d \geqslant 3$ with $\lambda(G) \leqslant 4\sqrt{d-1}$.[1] One can then obtain $(2d+1)$-regular graphs with $\lambda(G) \leqslant 4\sqrt{d-1}+1$ by adding an arbitrary perfect matching via the result of [CM08].

Our results. As mentioned, our Theorem 4.1.1 gives $\text{poly}(n)$-time deterministically computable n-vertex d-regular graphs G with $\lambda(G) \leqslant 2\sqrt{d-1}+\varepsilon$, for any

[1] We include a short proof here: let $\widetilde{A} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$ be the adjacency matrix of a d-regular bipartite Ramanujan graph. Then $A + A^T$ is the adjacency matrix of the merged graph. For any x orthogonal to $\underline{1}$, $(A+A^T)x = \begin{bmatrix} \mathbb{1} & \mathbb{1} \end{bmatrix} \widetilde{A} \begin{bmatrix} x \\ x \end{bmatrix}$. Thus $\|(A+A^T)x\| \leqslant \sqrt{2} \cdot \left\| \widetilde{A} \begin{bmatrix} x \\ x \end{bmatrix} \right\|$. Since $\begin{bmatrix} x \\ x \end{bmatrix}$ is orthogonal to both $\begin{bmatrix} \underline{1} \\ \underline{1} \end{bmatrix}$ and $\begin{bmatrix} \underline{1} \\ -\underline{1} \end{bmatrix}$, we have $\left\| \widetilde{A} \begin{bmatrix} x \\ x \end{bmatrix} \right\| \leqslant 2\sqrt{d-1}\sqrt{2}\|x\|$. One can then conclude that $\|(A+A^T)x\| \leqslant 4\sqrt{d-1}\|x\|$.

$d \geqslant 3$ and $\varepsilon > 0$. To be more precise, the running time of our algorithm is $n^{f(d,\varepsilon)}$ where $f(d,\varepsilon) = O(d^{1/4}\log(d)/\sqrt{\varepsilon})$. Although our graphs are not strongly explicit, they are "probabilistically poly log n-time computable", a relaxation of a notion defined by [BL06]. Essentially, this means we show there *exist* near-Ramanujan graphs whose adjacency lists are computable in poly log n time, and furthermore there is a poly $\log(n)$-time randomized algorithm for finding them with high probability. More precisely, the following statement holds:

Theorem 4.1.13. *There is a deterministic polynomial-time algorithm with the following properties:*

- *It takes as input N, $d \geqslant 3$, and $\varepsilon > 0$ written as binary strings.*
- *It also takes as input a "seed" $s \in \{0,1\}^{O(\log^2 N)}$ (the $O(\cdot)$ hides a factor of $O(d^{1/4}\log(d)/\sqrt{\varepsilon})$).*
- *It outputs a Boolean circuit C that implements the "adjacency list" of a d-regular graph G on $N' \sim N$ vertices in* poly $\log(N)$ *time. (This means that on input $u \in [N']$ and $i \in [d]$, both expressed in binary, $C(u,i)$ outputs the $v \in [N']$ that is the ith neighbor of u in G.)*
- *With high probability over the choice of seed s, the resulting graph G satisfies the bound $\lambda(G) \leqslant 2\sqrt{d-1} + \varepsilon$.*

The difference betweeen our notion and that of [BL06] is that we use a seed of length-$O(\log^2 N)$, whereas the notion in [BL06] requires the seed length to be $O(\log N)$.

4.1.3 On Bordenave's theorem with random edge-signs

Since our result may be viewed as a derandomization of the Friedman/Bordenave theorem (Theorem 4.1.8), let us take some time to describe this result. Friedman's original proof is notably quite involved (100 pages). Bordenave's proof is certainly simpler (more like 30 pages), although it is by no means easy. However, Bordenave's proof can become still simpler if one is willing consider a variant: when $\boldsymbol{G}$ is not just a random d-regular graph, but rather a *randomly edge-signed* random d-regular graph.

Let us say a few words about why this makes things simpler. First, it turns out that in this case one need not worry about the "trivial eigenvalue" of d; it no longer exists, and the statement to be proven is simply that $\rho(\boldsymbol{G}) \leqslant 2\sqrt{d-1}+\varepsilon$ with high probability, where $\rho(\boldsymbol{G})$ is the spectral radius (largest eigenvalue-magnitude) of the (signed) adjacency matrix of $\boldsymbol{G}$. Second, with random edge-signs, each entry of $\boldsymbol{G}$'s adjacency matrix becomes a symmetric random variable, and it is always more pleasant in probability theory when one's random variables naturally have mean zero.

In fact, there are scenarios in which one might actually *want* to consider random edge-signed d-regular graphs. For example when studying the Max-Cut problem, the setting of sparse random graphs is a very natural and challenging one, and many algorithms/complexity results depend on eigenvalue bounds for such graphs. Having random edge-signs simply means studying the equally natural 2XOR (aka 2Lin) problem, one that has a long history in theoretical computer science as well [Hås84].

Undoubtedly experts would know that including random edge-signs should make Bordenave's proof simpler, but it doesn't appear to have been directly explored until the recent work of Deshpande et al. [DMO+19]. That paper proved the analogue of Friedman/Bordenave for random edge-signings of random (c,d)-biregular graphs. The case when $c=d$ is essentially the same as the d-regular random graph case, but the nature of the proof simplification is perhaps obscured, particularly because [DMO+19] directly cited several lemmas from Bordenave [Bor19]. A similar situation occurred in a subsequent work [MOP19], which has random edge-signs within an even more complicated random graph model.

In fact, a side motivation we had for this paper was to carefully set out a self-contained proof — as simple as possible — of "Alon's Conjecture" for randomly edge-signed graphs. A reader not interested in derandomization may nevertheless find our proof of the below theorem of interest, particularly since it contains a substantial portion of Bordenave's proof of Friedman's theorem.

Theorem 4.1.14. *Let $d \geqslant 3$ and $\varepsilon > 0$. If $\boldsymbol{G}$ is a random edge-signed d-regular n-vertex graph, then*

$$\mathbf{Pr}\left[\rho(\boldsymbol{G}) \leqslant 2\sqrt{d-1}+\varepsilon\right] \geqslant 1 - o_n(1).$$

In the course of proving this theorem, we are able to observe that in fact Theorem 4.1.2 holds. That is, Theorem 4.1.14 does not thoroughly rely on having a

random edge-signing of a *random* d-regular graph. Instead, it works for a random edge-signing of *any* d-regular graph that has one particular property: namely, every vertex-neighborhood of radius $O((\log\log n)^2)$ should have at most one cycle. This property — called tangle-freeness by Bordenave (simplifying Friedman's notion of "tangles") — is a property that random d-regular graphs have with high probability, even for neighborhoods of the much larger radius $\Theta(\log_{d-1} n)$.

With Theorem 4.1.2 in hand, we are in a position rather like that of Bilu–Linial, who similarly showed [BL06, Cor. 3.1] that a random edge-signing of any sufficiently good small-set expander has spectral radius at most $\sqrt{d} \cdot O(\log^{1.5} d)$ (with high probability). As in Bilu–Linial, it is also fairly straightforward to see that Theorem 4.1.2 can be derandomized effectively using almost-k-wise independent binary random variables.

We next describe how this derandomized result on edge-signings leads to our main Theorem 4.1.1.

4.1.4 Explicit near-Ramanujan graphs via repeated 2-lifts

Let $G = (V, E)$ be an n-vertex d-regular graph, and let $\widetilde{G}$ be the edge-signed version of it associated to edge-signing $w : E \to \{\pm 1\}$. As observed by Bilu and Linial [BL06], this edge-signing is in a sense equivalent to the "2-lift" $G_2 = (V_2, E_2)$ of G defined by

$$V_2 = V \times \{\pm 1\}, \qquad E_2 = \Big\{ \{(u,\sigma),(v,\sigma \cdot w(u,v))\} : (u,v) \in E \Big\}.$$

This G_2 is a $2n$-vertex d-regular graph, and the equivalence is that G_2's eigenvalues are precisely the multiset-union of G's eigenvalues and $\widetilde{G}$'s eigenvalues. (The latter refers to the eigenvalues of $\widetilde{G}$'s signed adjacency matrix, whose nonzero entries are $w(u,v)$ for each $\{u,v\} \in E$.) In particular, if all the eigenvalues of G and $\widetilde{G}$ have magnitude at most $2\sqrt{d-1} + \varepsilon$ (excluding G's trivial eigenvalue of d), then the same is true of G_2 (excluding *its* trivial eigenvalue). Thus Theorem 4.1.2 can provide us with a (derandomizable) way of doubling the number of vertices in an ε-near-Ramanujan graph. It is not hard to see (Theorem 4.2.12) that if G is "r-bicycle-free" — meaning that every radius-r vertex neighborhood in G has at most one cycle — then G_2 will also be r-bicycle-free. Thus we may repeatedly double the number of vertices in an ε-near-Ramanujan graph, so long as the parameter r remains $\omega((\log\log|V|)^2)$, where $|V|$ is the "current" number of vertices. (Unfortunately,

we do not see an obvious way to get the parameter r to increase as we perform 2-lifts.) This is roughly the same strategy employed in [BL06].

As a consequence, to obtain a final d-regular ε-near-Ramanujan graph with $\Theta(N)$ vertices, all we need to get started is some d-regular ε-near-Ramanujan graph H on a smaller number of vertices, n, which is $O((\log\log N)^2)$-bicycle-free. Thanks to Friedman/Bordenave, we know that a *random* d-regular n-vertex graph is (with high probability) near-Ramanujan, and it's not hard to show it's $\Theta(\log n)$-bicycle-free. Thus we could get started with H being a random d-regular graph on, say, $n = 2^{\sqrt{\log N}}$ vertices, or even something smaller like $n = \text{quasipoly}(\log\log N)$.

Of course, to get a construction which is overall explicit, we need to derandomize the Friedman/Bordenave analysis for this base graph H. The advantage is we now have $\text{poly}(N)$ time to spend on constructing a graph with $n \ll N$ vertices. A trivial exponential-time derandomization won't work, but nor do we need a polynomial-time derandomization; a quasipolynomial-time derandomization is more than sufficient. And as we will see in Section 4.4, it is possible to derandomize Bordenave's proof in deterministic $n^{O(\log n)}$ time using $O(\log n)$-wise uniform permutations. The proof of this is not completely straightforward because Bordenave's proof uses a twist on the Trace Method (since the plain Trace Method provably fails).

4.2 Preliminaries

4.2.1 Standard derandomization tools

Throughout we use **boldface** to denote random variables.

Definition 4.2.1 (**(δ, k)-wise uniform bits**)**.** Let $\delta \in [0,1]$ and $k \in \mathbb{N}^+$. A sequence of Boolean random variables $\boldsymbol{y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_n) \in \{\pm 1\}^n$ is said to be (δ, k)-*wise uniform*[2] if, for every $S \subseteq [n]$ with $0 < |S| \leqslant k$, it holds that $|\mathbf{E}[\prod_{i \in S} \boldsymbol{y}_i]| \leqslant \delta$. When $\delta = 0$, we simply say that the sequence is *(truly) k-wise uniform*; indeed, in this case the bits are individually uniformly distributed and are k-wise independent.

A classic result of Naor and Naor [NN93] shows that (δ, k)-wise uniform bits can be constructed efficiently and deterministically from a truly random seed

[2]Frequently called (δ, k)-wise *independent* in the literature.

of length $O(\log k + \log\log n + \log(1/\delta))$. Indeed, these bits can be generated "strongly explicitly" (using [Sho90]; cf. [AGHP92]):

Theorem 4.2.2. *([NN93].) There is a deterministic algorithm that, given δ, k, and N, runs in time $\mathrm{poly}(N/\delta)$ and outputs a multiset $Y \subseteq \{\pm 1\}^N$ of cardinality $S = \mathrm{poly}(k\log(N)/\delta)$ (a power of 2) such that, for $\boldsymbol{y} \sim Y$ chosen uniformly at random, the sequence $\boldsymbol{y}$ is (δ, k)-wise uniform. Indeed, if the algorithm is additionally given $1 \leqslant s \leqslant S$ and $1 \leqslant i \leqslant N$ (written in binary), it can output the ith bit of the sth string in Y in deterministic time $\mathrm{poly}\log(N/\delta)$.*

We will make use of the fact that the parameters in this theorem have excellent dependence on N and k. We now discuss the analogous concept for random permutations, where it is not known if the parameter dependence can be as strong.

Definition 4.2.3 ((δ, k)-wise uniform permutations)**.** Let $\delta \in [0,1]$ and $k \in \mathbb{N}^+$. Let $[n]_k$ denote the set of all sequences of k distinct indices from $[n]$. A random permutation $\boldsymbol{\pi} \in S_n$ is said to be (δ, k)-*wise uniform* if, for every sequence $(i_1, \ldots, i_k) \in [n]_k$, the distribution of $(\boldsymbol{\pi}(i_1), \ldots, \boldsymbol{\pi}(i_k))$ is δ-close in total variation distance from the uniform distribution on $[n]_k$. When $\delta = 0$, we simply say that the permutation is *(truly) k-wise uniform*.

Kassabov [Kas07] and Kaplan–Naor–Reingold [KNR09] independently obtained a deterministic construction of (δ, k)-wise uniform permutations with seed length $O(k\log n + \log(1/\delta))$. Again, the construction is even "strongly explicit":

Theorem 4.2.4. *([KNR09, Kas07].) There is a deterministic algorithm that, given δ, k, and n, runs in time $\mathrm{poly}(n^k/\delta)$ and outputs a multiset $\Pi \subseteq S_n$ (closed under inverses) of cardinality $S = \mathrm{poly}(n^k/\delta)$ (a power of 2) such that, for $\boldsymbol{\pi} \sim \Pi$ chosen uniformly at random, $\boldsymbol{\pi}$ is a (δ, k)-wise uniform permutation. Indeed, if the algorithm is additionally given $1 \leqslant s \leqslant S$ and $1 \leqslant i \leqslant n$ (written in binary), it can output $\pi_s(i)$ and $\pi_s^{-1}(i)$ (where π_s is the sth permutation in Π) in deterministic time $\mathrm{poly}(k\log(n/\delta))$.*

We will also use a convenient theorem of Alon and Lovett [AL13]:

Theorem 4.2.5. *([AL13].) Let $\boldsymbol{\pi} \in S_n$ be a (δ, k)-wise uniform permutation. Then one can define a (truly) k-wise uniform permutation $\boldsymbol{\pi}' \in S_n$ such that the total variation distance between $\boldsymbol{\pi}$ and $\boldsymbol{\pi}'$ is $O(\delta n^{4k})$.*

Combining the previous two results yields the following:

Corollary 4.2.6. *([KNR09, Kas07, AL13]) There is a deterministic algorithm that, given k and n, runs in time $\mathrm{poly}(n^k)$ and outputs a multiset $\Pi \subseteq S_n$ (closed under inverses) such that, when $\boldsymbol{\pi} \sim \Pi$ is chosen uniformly at random, $\boldsymbol{\pi}$ is n^{-100k}-close in total variation distance to a (truly) k-wise uniform permutation. (And the final "indeed" statement from Theorem 4.2.4 also holds.)*

4.2.2 Elementary graph theory

4.2.2.1 Random d-regular graphs

We will be concerned with d-regular (multi)graphs. We start by describing the standard way to generate random d-regular graphs: the *configuration model*, see [BC78, Bol80, Bol01].

Definition 4.2.7 (Configuration model)**.** Given integers $n > d > 0$ with nd even, the *configuration model* produces a random n-vertex, d-regular undirected multigraph (with loops) $\boldsymbol{G}$. This multigraph is induced by a uniformly random matching $\boldsymbol{M}$ on the set of "half-edges", $[n] \times [d] \cong [nd]$ (where $(v,i) \in [n] \times [d]$ is thought of as half of the ith edge emanating from vertex v). We identify $\boldsymbol{M}$ with a symmetric matrix in $\{0,1\}^{nd \times nd}$ having 1's precisely in the entries corresponding to matched pairs $\{(v,i),(v',i')\}$. We may think of $\boldsymbol{M}$ being generated as follows: First a uniformly random permutation $\boldsymbol{\pi} \in S_{nd}$ is chosen; then we set $\boldsymbol{M}_{\boldsymbol{\pi}(j),\boldsymbol{\pi}(j+1)} = \boldsymbol{M}_{\boldsymbol{\pi}(j+1),\boldsymbol{\pi}(j)} = 1$ for each odd $j \in [nd]$.

Given $\boldsymbol{M}$, the multigraph $\boldsymbol{G}$ is formed by "attaching" the matched half-edges. More formally, the (v,v')-entry of $\boldsymbol{G}$'s adjacency matrix $\boldsymbol{A}$ is the sum, over all $i,i' \in [d]$, of $\boldsymbol{M}_{(v,i),(v',i')}$. Hence

$$\begin{aligned}&\boldsymbol{A}_{v,v'} = \\ &\sum_{i,i'=1}^{d} \sum_{\substack{\text{odd} \\ j \in [nd]}} (1[\boldsymbol{\pi}(j) = (v,i)] \cdot 1[\boldsymbol{\pi}(j+1) = (v',i')] \\ &+ 1[\boldsymbol{\pi}(j) = (v',i')] \cdot 1[\boldsymbol{\pi}(j+1) = (v,i)]).\end{aligned}$$

Note that $\boldsymbol{A}_{v,v}$ will always be even; a self-loop is considered to contribute degree 2.

It is well known that a graph $\boldsymbol{G}$ drawn from the configuration model is simple — i.e., has no cycles of length 1 or 2 — with probability $\Omega_d(1)$. As it is pleasant to

work with simple graphs, we will show in Section 4.6 that this continues to hold for *pseudorandom* d-regular graphs, when an $O(d^2)$-wise uniform permutation is used in the configuration model. We also record the well known fact that for $\boldsymbol{G}$ drawn from the configuration model, when $\boldsymbol{G}$ is conditioned on being simple, its conditional distribution is uniformly random among all d-regular graphs.

Although the configuration model is the most natural way to generate large random d-regular graphs, the fact that it does not produce simple graphs with high probability is mildly annoying. (In particular, this causes a slight technical hitch for establishing our "probabilistically strongly explicit" construction.) To sidestep this, we will also consider the *random lift* model for producing random d-regular graphs.

Definition 4.2.8 (Lift model)**.** Fix a (simple) *base* graph $\underline{G} = (\underline{V}, \underline{E})$ on $\underline{n}$ vertices. Then for $n \in \mathbb{N}^+$, an *n-lift of $\underline{G}$* is graph G defined by a collection of permutations $\pi_{uv} \in S_n$, one for each edge $(u, v) \in \underline{E}$, under the constraint that $\pi_{uv} = \pi_{vu}^{-1}$. The vertex set of G is $\underline{V} \times [n]$, and the edges of G are given by all pairs $(u, i), (v, j)$ satisfying $(u, v) \in \underline{E}$ and $\pi_{uv}(i) = j$. When the permutations π_{uv} are independent and uniformly random, we call the associated graph $\boldsymbol{G}$ a *(uniformly) random n-lift of $\underline{G}$*. Observe that if $\underline{G}$ is a d-regular graph, then $\boldsymbol{G}$ is always a d-regular (simple) graph on $\underline{n}n$ vertices.

Bordenave [Bor19] also confirmed Theorem 4.1.8 (the Alon Conjecture) in the case that $\boldsymbol{G}$ is a random n-lift of any fixed d-regular Ramanujan base graph $\underline{G}$. The simplest case is $\underline{G} = K_{d+1}$, the complete graph on $d + 1$ vertices. This gives a way to randomly construct arbitrarily large d-regular near-Ramanujan graphs that are always simple. We will also derandomize this result, as it will be convenient for our "probabilistically strongly explicit" construction to have guaranteed simplicity.

4.2.2.2 Bicycle-freeness

It is well known that a d-regular random graph is likely to have at most one cycle in any neighborhood of radius $c \log_{d-1} n$, for a certain universal $c > 0$. (This holds in either the configuration or the random lift model.) Let us make some definitions to codify this.

Definition 4.2.9 (Excess)**.** Given a multigraph $H = (V, E)$, its *excess* is $\mathrm{exc}(H) = |E| - |V|$.

Definition 4.2.10 (A/uni/bi-cyclic)**.** A connected multigraph H with $\mathrm{exc}(H) = -1$, 0, 1 (respectively) is said to be *acyclic*, *unicyclic*, *bicyclic* (respectively). In either of the first two cases, we call H *bicycle-free* (or *at most unicyclic*).

Definition 4.2.11 (Bicycle-free at radius r)**.** We say a multigraph is *bicycle-free at radius r* if the distance-r neighborhood of every vertex is bicycle-free. Another way to say this is that a breadth-first search of depth r, started at any vertex, encounters at most one "back-edge". We remark that this notion was termed *r-tangle-free* by Bordenave [Bor19].

Proposition 4.2.12. *If G is bicycle-free at radius r, and G_2 is a 2-lift of G, then G_2 is bicycle-free at radius r.*

Proof. Let (v, i) be any vertex in G_2. Let H be the distance-r neighborhood of v in G and let H_2 be the subgraph of G_2 induced by $V(H) \times [2]$. Observe that the distance-r neighborhood of (v, i) is contained in H_2, and that $\mathrm{exc}(H_2) \leqslant 0$ since $\mathrm{exc}(H) \leqslant 0$. If H_2 is disconnected it is isomorphic to a disjoint union of two copies of H and thus the distance-r neighborhood of (v, i) is then isomorphic to H. Otherwise, if H_2 is connected, $\mathrm{exc}(H_2) \leqslant 0$ implies that it has at most one cycle. □

It is easy to see that any n-vertex, d-regular graph that is bicycle-free at radius r must have $r \lesssim \log_{d-1} n$. On the other hand, as mentioned earlier, a random d-regular graph achieves this bound up to a constant factor, and we will derandomize the proof of this fact, within the $O(\log n)$-wise uniform configuration/lift model, in Section 4.4.1.

In a graph that is bicycle-free at radius r, by definition we have $\mathrm{exc}(H) \leqslant 0$ for all subgraphs H contained in a single distance-r neighborhood. In fact, this property is enough to guarantee that $\mathrm{exc}(H)$ is small for *any* subgraph H with at most $\exp(r)$ vertices, regardless of whether it's contained in a single distance-r neighborhood:

Theorem 4.2.13. *Let H be a v-vertex graph that is bicycle-free at radius r. Assume $r \geqslant 10 \ln v$. Then $\mathrm{exc}(H) \leqslant \frac{\ln(ev)}{r} v$.*

The rest of this subsection is devoted to the proof of the above theorem of elementary graph theory.

Definition 4.2.14 ($\mathrm{Cyc}_{\leqslant g}(G)$ and girth)**.** Given a graph G, let $\mathrm{Cyc}_{\leqslant g}(G)$ denote the collection of all cycles in G of length at most g. Recall that if $\mathrm{Cyc}_{\leqslant g}(G)$ is empty then G is said to have *girth* exceeding g.

The following fact is essentially immediate from the definitions:

Fact 4.2.15. *Suppose G is bicycle-free at radius r. Then the cycles in $\mathrm{Cyc}_{\leqslant 2r}(G)$ are vertex-disjoint.*

Indeed, more generally:

Proposition 4.2.16. *Suppose G is bicycle-free at radius r. For each $C \in \mathrm{Cyc}_{\leqslant 2r}(G)$, let C^+ denote the collection of vertices within distance $r - \mathrm{len}(C)/2$ of C. Then the sets $\{C^+ : C \in \mathrm{Cyc}_{\leqslant 2r}(G)\}$ are pairwise disjoint.*

Proof. If $u \in C_1^+ \cap C_2^+$, the distance-r neighborhood of u is enough to include both C_1 and C_2. □

Next, let us now recall the "Moore bound for irregular graphs". Suppose H is a graph with v vertices and $\mathrm{exc}(H) = \varepsilon v$; hence H has average degree $2 + 2\varepsilon$. If we build a breadth-first search tree from some vertex, then after depth t we would "expect" to encounter at least $(1 + 2\varepsilon)^t$ vertices. If this exceeds v — roughly, if $t \geqslant (\ln v)/(2\varepsilon)$ — then the breadth-first search must encounter a cycle. Thus we have a heuristic argument that $\mathrm{girth}(H) \lesssim (\ln v)/\varepsilon$; i.e., $\varepsilon \lesssim (\ln v)/\mathrm{girth}(H)$. Indeed, Alon–Hoory–Linial have precisely established this kind of result; we quote their theorem in a slightly simplified form:

Theorem 4.2.17. *([AHL02].) Let H be a graph with v vertices, $\mathrm{exc}(H) = \varepsilon v$ (for $\varepsilon \geqslant 0$), and girth g. Then $v \geqslant (1 + 2\varepsilon)^{g/2 - 3/2}$.*

Corollary 4.2.18. *Let H be a graph with $v \geqslant 3$ vertices and girth $g \geqslant 20 \ln v$. Then $\mathrm{exc}(H) \leqslant ((2 \ln v)/g)v$.*

We can now prove Theorem 4.2.13, which replaces "girth" with "bicycle-free radius" in the above with only a small loss in parameters.

Proof of Theorem 4.2.13. We will show the theorem assuming H is connected (the only case we'll need). It is an exercise to extend it to the general case by considering H's connected components.

Let $c = |\mathrm{Cyc}_{2r}(H)|$. By deleting at most c edges from H we can obtain a v-vertex graph $\widetilde{H}$ with girth at least (in fact, exceeding) $2r$. Applying Theorem 4.2.18 to $\widetilde{H}$, we conclude that $\mathrm{exc}(H) \leqslant \frac{\ln v}{r} v + c$. Thus it remains to show $c \leqslant v/r$. This is trivial if $c = 0$, and if $c = 1$ then it can only fail if $r > v$ — but then H is unicyclic and hence has excess 0. Assuming then that $c \geqslant 2$, choose paths in H to minimally connect the c cycles of $\mathrm{Cyc}_{2r}(H)$. Now for each $C \in \mathrm{Cyc}_{2r}(H)$, if we "charge" to it the $r - \mathrm{len}(C)/2$ closest path-vertices, then no vertex is charged to multiple cycles, by virtue of Theorem 4.2.16. If we also charge the vertices of C to itself, then for each $C \in \mathrm{Cyc}_{2r}(H)$ we have charged a batch of $\mathrm{len}(C) + (r - \mathrm{len}(C)/2) > r$ vertices, and these batches are disjoint. Thus $cr \leqslant v$, i.e. $c \leqslant v/r$, as required. □

4.2.3 Non-backtracking walks and the Ihara–Bass formula

The Friedman/Bordenave theorem ultimately uses the Trace Method to analyze the eigenvalues of random d-regular graphs; this involves counting closed walks in them. As observed in [Fri08, Bor19], it is much easier to count *non-backtracking* walks, and luckily the *Ihara–Bass formula* gives an easy translation between eigenvalues of the adjacency matrix of a graph and the eigenvalues of its *non-backtracking matrix*.

Definition 4.2.19 (Non-backtracking matrix [Has89])**.** Let $G = (V, E)$ be a multigraph with adjacency matrix A. Let $\underline{E}$ denote the (multi)set of all directed edges formed by replacing each undirected edge in E with two opposing directed edges. Then G's *non-backtracking* matrix B has rows and columns indexed by $\underline{E}$, with

$$B_{(u_1,v_1),(u_2,v_2)} = \begin{cases} 1 & \text{if } v_1 = u_2 \text{ and } v_2 \neq u_1, \\ 0 & \text{otherwise.} \end{cases}$$

(Note that this matrix is not symmetric in general.) In case G is an edge-signed graph, the entry 1 above should be replaced by A_{u_2,v_2}, the sign of G on edge $\{u_2, v_2\}$.

In a number-theoretic context, Ihara [Iha66] implicitly showed a relationship between the eigenvalues of A and B when G is regular. Serre [Ser77] and several others suggested the translation to graph theory, and Bass [Bas92] (following [Has89]) explicitly established:

Theorem 4.2.20. *(Ihara–Bass formula.) Let G be a d-regular (multi)graph and write $q = d - 1$. Then*

$$\det(\mathbb{1} - zB) = (1 - z^2)^{\mathrm{exc}(G)} \det((1 + qz^2)\mathbb{1} - zA),$$

where $\mathbb{1}$ denotes the identity matrix (of appropriate dimension).

This theorem has been given many proofs, and it can be generalized to irregular graphs, edge-weighted graphs, and infinite graphs. We will use the following result, which is immediate from the edge-weighted generalization [WF09] when all weights are ± 1:

Theorem 4.2.21. *([WF09].) The Ihara–Bass formula holds as stated above for edge-signed graphs.*

The utility of Ihara–Bass is that it gives a direct correspondence between the spectra of A and B. To see this, consider the zeroes of the polynomials (in z) on the left- and right-hand sides. We have that z is a zero of the left-hand side precisely if z^{-1} is an eigenvalue of B. On the other hand, z is a zero of the right-hand side precisely if $z^{-1} = \pm 1$ or if z^{-1} is such that $z^{-1} + q/z^{-1}$ is an eigenvalue of A. Thus if we want to deduce, say, the eigenvalues of B from the eigenvalues of A, we have the following:

Proposition 4.2.22. *(Consequence of Ihara–Bass.) Let $G = (V, E)$ be a $(q+1)$-regular edge-signed graph with adjacency matrix A and non-backtracking matrix B. Let $\lambda \neq 0, \pm 1$ be a number such that $\lambda + q/\lambda$ is an eigenvalue of A. Then λ is an eigenvalue of B.*

In fact, Theorem 4.2.22 is the only consequence of Ihara–Bass we will need in this paper, and for the convenience of the reader we give a self-contained proof (inspired by [AFH15]):

Proof. Let $f : V \to \mathbb{C}$ be an eigenvector for A with eigenvalue $\lambda + q/\lambda$. Define $g : \underline{E} \to \mathbb{C}$ by $g_{vw} = A_{vw} f_v - \lambda f_w$. We claim that $Bg = \lambda g$. It then follows that λ is an eigenvalue of B, given that $g \not\equiv 0$ (a consequence of $f \not\equiv 0$: choose $\{v, w\} \in E$ with f_v, f_w not both 0, and then $g_{vw} = 0 = g_{wv}$ is impossible because $\lambda \neq \pm 1$). To verify the claim, for any $uv \in \underline{E}$ we have

$$(Bg)_{uv} = \sum_{\substack{w \sim v \\ w \neq u}} A_{vw} g_{vw}$$

$$= \sum_{w \sim v} A_{vw}(A_{vw}f_v - \lambda f_w) - A_{vu}(A_{vu}f_v - \lambda f_u)$$
$$= -\lambda \sum_{w \sim v} A_{vw}f_w + qf_v + \lambda A_{vu}f_u.$$

But $\sum_{w\sim v} A_{vw}f_w = (Af)_v = (\lambda + q/\lambda)f_v$. Thus $(Bg)_{uv} = -\lambda^2 f_v + \lambda A_{vu}f_u = \lambda g_{uv}$, as needed. □

When G is unsigned, A has a "trivial" eigenvalue of $d = q+1$, corresponding to $\lambda = q$; this yields the "trivial" eigenvalue of $q = d-1$ for B. For general edge-signed G, if $\lambda = \pm\sqrt{q} = \pm\sqrt{d-1}$ in Theorem 4.2.22, then $\lambda + q/\lambda = \pm 2\sqrt{q} = \pm 2\sqrt{d-1}$. Thus the Ramanujan eigenvalue bound of $2\sqrt{d-1}$ for A is equivalent to the bound $\sqrt{d-1}$ for B. As for the "$+\varepsilon$", a simple calculation (appearing in [Bor19]) shows:

Corollary 4.2.23. *Let $G = (V, E)$ be a d-regular edge-signed graph ($d \geqslant 3$) with adjacency matrix A and non-backtracking matrix B. If A has an eigenvalue of magnitude $2\sqrt{d-1} + \varepsilon$ (for $\varepsilon \geqslant 0$) then B has an eigenvalue of magnitude $\sqrt{d-1} + \sqrt{\varepsilon}\sqrt{\sqrt{q} + \varepsilon/4} + \varepsilon/2$ (which is $\sqrt{d-1} + \Theta(d^{1/4}\sqrt{\varepsilon})$ for fixed d and $\varepsilon \to 0$).*

4.3 On random edge-signings of fixed base graphs

In this section we will prove Theorem 4.1.2. In fact, we will prove the following refined version:

Theorem 4.3.1. *Let $G = (V, E)$ be an arbitrary d-regular n-vertex graph, where $d \leqslant \operatorname{poly}\log n$. Assume that G is bicycle-free at radius $r \gg (\log\log n)^2$. Then for $\boldsymbol{G}$ a uniformly random edge-signing of G, except with probability at most n^{-100} the non-backtracking matrix $\boldsymbol{B}$ of $\boldsymbol{G}$ satisfies the spectral radius bound*

$$\rho(\boldsymbol{B}) \leqslant \sqrt{d-1} \cdot \left(1 + O\left(\frac{(\log\log n)^2}{r}\right)\right),$$

and hence (by Theorem 4.2.23) the signed adjacency matrix $\boldsymbol{A}$ of $\boldsymbol{G}$ satisfies the bound

$$\rho(\boldsymbol{A}) \leqslant 2\sqrt{d-1} \cdot \left(1 + O\left(\frac{(\log\log n)^4}{r^2}\right)\right).$$

Furthermore, let $C = C(n)$ satisfy $1 \leqslant C \leqslant \operatorname{poly}\log n$ and suppose we merely assume that the random edge-signs are (δ, k)-wise uniform for $\delta \leqslant n^{-O(C\log d)}$ and $k \geqslant 2C\log n$.

Then the above bounds continue hold, with an additional additive $O(\sqrt{d}/C)$ in the $\rho(\boldsymbol{B})$ bound and $O(\sqrt{d}/C^2)$ in the $\rho(\boldsymbol{A})$ bound.

As in [Fri08, Bor19], the proof of Theorem 4.3.1 will use the Trace Method. In preparation for this, we make some definitions:

Definition 4.3.2 (Hikes)**.** Let $G = (V, E)$ be an undirected graph. For $\ell \in \mathbb{N}$, we define an *ℓ-hike* $\mathcal{H}$ to be a closed walk in G of exactly 2ℓ steps which is non-backtracking except possibly between the ℓth and $(\ell+1)$th step. Given an edge-signing $w : E \to \{\pm 1\}$ we write $w(\mathcal{H})$ for the product of the edge-signs that $\mathcal{H}$ traverses, counted with multiplicity. Finally, we call a hike *even* (respectively, *singleton-free*) if each undirected edge traversed by $\mathcal{H}$ is traversed an even number of times (respectively, at least twice).

A straightforward use of the Trace Method will now imply:

Proposition 4.3.3. *Let $\ell \in \mathbb{N}^+$ and define $\boldsymbol{T} = \mathrm{tr}\big(\boldsymbol{B}^\ell (\boldsymbol{B}^\top)^\ell\big)$ (which is an upper bound on $\rho(\boldsymbol{B})^{2\ell}$). Then for a uniformly random edge-signing $\boldsymbol{w} : E \to \{\pm 1\}$,*

$$\begin{aligned}\mathbf{E}[\boldsymbol{T}] &\leqslant d^2 \cdot \#\{\text{even } (\ell-1)\text{-hikes } \mathcal{H} \text{ in } G\} \\ &\leqslant d^2 \cdot \#\{\text{singleton-free } (\ell-1)\text{-hikes } \mathcal{H} \text{ in } G\}.\end{aligned}$$

Furthermore, if $\boldsymbol{w}$ is merely $(\delta, 2\ell)$-wise uniform, the bound holds up to an additive $\delta n d^{2\ell+2}$.

Proof. We have

$$\boldsymbol{T} = \sum_{\underline{e}_0, \underline{e}_1, \dots, \underline{e}_{2\ell-1}, \underline{e}_{2\ell} = \underline{e}_0 \in \underline{E}} \boldsymbol{B}_{\underline{e}_0, \underline{e}_1} \boldsymbol{B}_{\underline{e}_1, \underline{e}_2} \cdots \boldsymbol{B}_{\underline{e}_{\ell-1}, \underline{e}_\ell} \boldsymbol{B}_{\underline{e}_{\ell+1}, \underline{e}_\ell} \boldsymbol{B}_{\underline{e}_{\ell+2}, \underline{e}_{\ell+1}} \cdots \boldsymbol{B}_{\underline{e}_{2\ell}, \underline{e}_{2\ell-1}}. \tag{4.1}$$

Recalling the definition of $\boldsymbol{B}$, one immediately sees that $\boldsymbol{T}$ is "something like" the sum of $\boldsymbol{w}(\mathcal{H})$ over all ℓ-hikes in G. But being careful, one sees we precisely have the following:

> $\boldsymbol{T}$ is equal to the sum of $\boldsymbol{w}(\mathcal{H})$ over all "special" $(\ell+1)$-hikes in G, where we call an $(\ell+1)$-hike *special* if its $(\ell+2)$th step is the reverse of its $(\ell+1)$th step, and the last step is the reverse of the first step.[3]

Next, we employ the following easy fact:

[3]The astute reader will note that the sign of the first/last edge in $\mathcal{H}$ is never counted in Equation (4.1); however it is okay to count it twice, as $\boldsymbol{w}(\mathcal{H})$ does, since $(\pm 1)^2 = 1$.

Fact 4.3.4. *If $\boldsymbol{w} : E \to \{\pm 1\}$ is a fully uniformly random edge-signing, then $\mathbf{E}[\boldsymbol{w}(\mathcal{H})]$ will be 1 if $\mathcal{H}$ is an even hike, and will be 0 otherwise.*

Thus

$$\mathop{\mathbf{E}}_{\boldsymbol{w}:E\to\{\pm 1\}}[\boldsymbol{T}] = \#\{\text{even, special } (\ell+1)\text{-hikes } \mathcal{H} \text{ in } G\}. \tag{4.2}$$

Since an $(\ell+1)$-hike involves at most 2ℓ undirected edges, a crude upper bound on the number of all $(\ell+1)$-hikes in G is $nd^{2\ell}$. Thus for an edge-signing $\boldsymbol{w}$ that is merely $(\delta, 2\ell)$-wise uniform, Equation (4.2) holds up to an additive $\delta nd^{2\ell}$. Finally, every even special $(\ell+1)$-hike $\mathcal{H}$ can be formed from an even $(\ell-1)$-hike $\mathcal{H}'$ by: (i) attaching a step and its reverse to the beginning/end of $\mathcal{H}$; (ii) attaching a step and its reverse to the midpoint of $\mathcal{H}$. As there are at most $(d-1)^2 \leqslant d^2$ choices for how to perform (i) and (ii), the inequality in the proposition's statement follows. □

At this point, edge-signs are out of the way and we are reduced to counting singleton-free hikes. In aid of this, we borrow some terminology from [MOP19]:

Definition 4.3.5. Given an $(\ell-1)$-hike $\mathcal{H}$ in graph G, we write $G_{\mathcal{H}} = (V_{\mathcal{H}}, E_{\mathcal{H}})$ for the subgraph of G formed by the union of the edges visited by $\mathcal{H}$. We think of $G_{\mathcal{H}}$ as being "revealed" as the $2(\ell-1)$ steps of $\mathcal{H}$ are taken in order. We classify each step of $\mathcal{H}$ as either *stale*, *fresh*, or *boundary*. If a step of $\mathcal{H}$ traverses a previously-explored edge in $G_{\mathcal{H}}$ (in either direction), we call the step *stale*; otherwise, if it steps to a previously-unvisited vertex, we call the step *fresh*; otherwise, we call it *boundary*. For the purposes of this definition, at the beginning of $\mathcal{H}$ the initial vertex is considered to be "previously visited".

We now put bounds on the different kinds of steps. For the fresh steps, we only need the singleton-free property:

Proposition 4.3.6. *In a singleton-free $(\ell-1)$-hike, at least half of all steps must be stale. Thus there are fewer than ℓ fresh steps.*

For the boundary steps of $\mathcal{H}$, it is easy to see that there are exactly $\mathrm{exc}(G_{\mathcal{H}})+1$ of them. Thus we can bound them using only the bicycle-free property. Together with the simple bound $|V_{\mathcal{H}}| \leqslant 2\ell$, Theorem 4.2.13 implies

Proposition 4.3.7. *If $\mathcal{H}$ is an $(\ell-1)$-hike in a graph G which is bicycle-free at radius $r \geqslant 10\ln(2\ell)$, then $\mathcal{H}$ has at most $O(\frac{\log \ell}{r}) \cdot \ell$ boundary steps.*

Finally, to handle the stale steps we group them into "stretches".

Proposition 4.3.8. *In an $(\ell-1)$-hike $\mathcal{H}$, the stale steps may be partitioned into at most $O(\frac{\log \ell}{r}) \cdot \ell$ stretches of consecutive stale steps, each stretch having length at most r, and none straddling the "turnaround" at step ℓ.*

Proof. We begin by partitioning the stale steps into maximal contiguous stretches. It is easy to see that each of these must be preceded in $\mathcal{H}$ by a boundary step (with a single possible exception of the "turnaround" at step ℓ). Thus Theorem 4.3.7 implies that there are at most $O(\frac{\log \ell}{r}) \cdot \ell$ maximal stretches of stale steps. If a maximal stretch straddles the turnaround, we can split it in two. Finally, if necessary we now subdivide the stretches into length at most r. Since there are fewer than 2ℓ stale steps, this subdivision can be done without increasing the number of stretches by more than $2\ell/r \leqslant O(\frac{\log \ell}{r}) \cdot \ell$. □

We may now make our final estimate:

Theorem 4.3.9. *In a d-regular graph G that is bicycle-free at radius $r \geqslant 10\ln(2\ell)$, the number of singleton-free $(\ell-1)$-hikes $\mathcal{H}$ is at most $O(\ell^3 n) \cdot (d-1)^\ell \cdot (dr\ell)^{O(\frac{\log \ell}{r}) \cdot \ell}$.*

Proof. Following [Bor19], we use an encoding argument. To each $\mathcal{H}$ we associate a string $\textsc{struct}(\mathcal{H})$ over the alphabet $\{\mathtt{F}, \mathtt{B}, \mathtt{S}\}$, where we replace each fresh step with an F, each boundary step with a B, and each stale stretch with an S. Our goal will be to show:

Claim 4.3.10. For any string σ with c_f, c_b, c_s occurrences of F, B, S (respectively), there are no more than $2n \cdot (d-1)^{c_f+c_b} \cdot (2r\ell)^{c_s}$ singleton-free $(\ell-1)$-hikes $\mathcal{H}$ with $\textsc{struct}(\mathcal{H}) = \sigma$.

Let us complete the proof of the theorem assuming this claim. By Theorems 4.3.6 to 4.3.8, we have the bounds

$$c_f < \ell, \qquad c_b, c_s < m := O(\tfrac{\log \ell}{r}) \cdot \ell.$$

Crudely, there are at most $O(\ell^3)$ possibilities for the triple (c_f, c_b, c_s). Also, the following two quantities are increasing in c_f, c_b, c_s:

$$2n \cdot (d-1)^{c_f+c_b} \cdot (2r\ell)^{c_s}, \qquad \Sigma_{c_f,c_b,c_s} := \#\text{ strings of } c_f \text{ F's, } c_b \text{ B's, } c_s \text{ S's.}$$

Thus we can upper-bound the number of all singleton-free $(\ell-1)$-hikes by

$$O(\ell^3 n)\cdot(d-1)^{\ell+m}\cdot(2r\ell)^m\cdot\Sigma_{\ell,m,m}\leqslant O(\ell^3 n)\cdot(d-1)^{\ell}\cdot(dr\ell)^{O(m)},$$

as needed, where we used the simple bound $\Sigma_{\ell,m,m}\leqslant \ell^{O(m)}$.

It remains to prove the claim. Let σ be as given. We may recover all possible associated $\mathcal{H}$, in a vertex-by-vertex fashion, by first specifying the initial vertex (n choices) and then proceeding through the symbols of σ in order. If we are at an F or a B symbol, we can recover the next vertex by specifying one of $d-1$ neighbors of the current vertex; there are only $d-1$ possibilities, since $\mathcal{H}$ is non-backtracking. (Exception: there are d choices at the very beginning of the hike; we compensated for this with the factor $2>\frac{d}{d-1}$.) To complete the proof of the claim, we need to show that for each stale stretch, there are at most $2r\ell$ possibilities. Recall that a stale stretch beginning from a vertex v consists of walking in non-backtracking fashion for at most r steps over the previously seen portion K of $G_{\mathcal{H}}$. This subgraph K has at most 2ℓ vertices, and by the bicycle-free property, this walk is confined to a subgraph of K that is at most unicyclic. It is easy to see this walk is determined by specifying its final vertex (at most 2ℓ possibilities), the number of times the cycle in v's distance-r neighborhood (should it exist) is traversed (fewer than $r/2$ possibilities), and the direction in which the cycle is traversed (2 possibilities). Thus indeed each stale stretch can be completely determined by specifying one of at most $2\ell\cdot(r/2)\cdot 2=2r\ell$ possibilities. □

Combining this with Theorem 4.3.3 now yields:

Corollary 4.3.11. *Let $G=(V,E)$ be an arbitrary d-regular n-vertex graph. Assume that G is bicycle-free at radius r. Let $\ell\in\mathbb{N}^+$ and $0<\eta<1$ be parameters. Then for $\boldsymbol{G}$ a uniformly random edge-signing of G, except with probability at most η the non-backtracking matrix $\boldsymbol{B}$ of $\boldsymbol{G}$ has spectral radius bound*

$$\rho(\boldsymbol{B})\leqslant\sqrt{d-1}\cdot(1+O(\varepsilon_1)+O(\varepsilon_2)),\tag{4.3}$$

where

$$\varepsilon_1:=\frac{\log(n/\eta)}{\ell},\qquad \varepsilon_2:=\frac{\log(d\ell)\log(\ell)}{r},$$

provided $\varepsilon_1,\varepsilon_2\leqslant 1$.

Furthermore, if the random edge-signs of $\boldsymbol{G}$ are merely $(\delta,2\ell)$-wise uniform, the bound holds up to an additional additive $(\delta n/\eta)^{\frac{1}{2\ell}}\cdot O(d)$.

Proof. We have obtained that, for a uniformly random edge-signing $\boldsymbol{w} : E \to \{\pm 1\}$,

$$\mathbf{E}[\boldsymbol{T}] \leqslant O(d^2\ell^3 n) \cdot (d-1)^{\ell} \cdot (dr\ell)^{O(\frac{\log \ell}{r})\cdot \ell}.$$

Note that $r \lesssim \log_{d-1} n$ always holds, and hence we must have $\ell \leqslant n$ (else $\varepsilon_2 > 1$). Also we must have $\ell \geqslant \log n$ (else $\varepsilon_1 > 1$). Thus we may coarsen $O(d^2\ell^3 n)$ in the above to $O(n^5)$, and coarsen $(dr\ell)^{O(\cdot)}$ to $(d\ell)^{O(\cdot)}$. Now since $\boldsymbol{T}$ is a nonnegative random variable, Markov's inequality implies that except with probability at most η,

$$\boldsymbol{T} \leqslant O(n^5/\eta) \cdot (d-1)^{\ell} \cdot (d\ell)^{O(\frac{\log \ell}{r})\cdot \ell},$$

and hence

$$\rho(\boldsymbol{B}) \leqslant \boldsymbol{T}^{\frac{1}{2\ell}} \leqslant O(n^5/\eta)^{\frac{1}{2\ell}} \cdot \sqrt{d-1} \cdot (d\ell)^{O(\frac{\log \ell}{r})},$$

which directly implies Inequality (4.3).

Finally, in the $(\delta, 2\ell)$-wise uniform case, we get an additional additive $\delta n d^{2\ell+2}$ in the bound on $\mathbf{E}[\boldsymbol{T}]$; this gets a factor of $1/\eta$ after the application of Markov, and becomes $(\delta n/\eta)^{\frac{1}{2\ell}} \cdot O(d)$ after taking 2ℓth roots. □

Finally, the reader may verify that Theorem 4.3.1 follows from Theorem 4.3.11 in the fully uniform case by taking $\ell = \Theta(r \log(n)/\log\log n)$, and in the derandomized case by taking $\ell = \Theta(C \log(n/\eta))$.

Remark 4.3.12. Alternatively, by taking $\eta = \exp(-\exp(r^{.49}))$ and $\ell = \exp(r^{.49})$ in Theorem 4.3.11, we may conclude that $\rho(\boldsymbol{B}) \leqslant \sqrt{d-1} \cdot (1 + o_r(1))$ holds in the fully uniform case except with probability at most $\exp(-\exp(r^{.49}))$.

4.4 Weakly derandomizing Bordenave's theorem

In this section we give a weak derandomization of Bordenave's proof of Theorem 4.1.8, using "off-the-shelf" tools; the derandomization is "weak" in the sense that it only yields a quasipoly(n)-time deterministic construction. As discussed in Section 4.2.2.1, we will derandomize both the configuration model version and the random lift version. Specifically, we show the conclusion of Theorem 4.1.8 holds even for the "almost k-wise uniform" versions of these models, $k = O(\log n)$.

Definition 4.4.1 ((δ, k)-wise uniform configuration/lift models)**.** When the permutation $\boldsymbol{\pi} \in S_{nd}$ used in the configuration model is not uniformly random but is

merely (δ, k)-wise uniform, we will say that $\boldsymbol{G}$ is drawn from the *(δ, k)-wise uniform configuration model*. Similarly, when the $\boldsymbol{\pi}_{uv} \in S_n$ used in the random lift model are independent but merely (δ, k)-wise uniform, we will say that $\boldsymbol{G}$ is a *(δ, k)-wise uniform random n-lift of base graph $\underline{G}$*.

(For simplicity, in the lift model we will henceforth only concern ourselves with $\underline{G} = K_{d+1}$.)

We will not fully recap Bordenave's proof of Theorem 4.1.8 in this work, although the reader unfamiliar with it will get some insight knowing that our proof of Theorem 4.3.1 is modeled on it. Bordenave employs two twists on the Trace Method to show that a random d-regular graph $\boldsymbol{G}$ has spectral radius at most $2\sqrt{d-1} + \varepsilon$ (when the trivial eigenvalue of d is ignored). The less important (but still challenging) twist involves replacing the non-backtracking matrix $\boldsymbol{B}$ by a centered variant, $\underline{\boldsymbol{B}}$, that enables one to ignore the trivial eigenvalue. The more conceptually important twist comes from the fact, originally recognized by Friedman, that even after passing to $\underline{\boldsymbol{B}}$, the Trace Method still fails. The reason, in brief, is as follows: A successful use of the Trace Method would have to consider walks of length ℓ for ℓ at least a large multiple of $\log n$, in order to overcome the factor of n arising from the n different walk starting points (cf. the error term ε_1 just after Inequality (4.3)). But for walks of this long length, one can show that the expected trace of $\underline{\boldsymbol{B}}^{\ell}(\underline{\boldsymbol{B}}^{\top})^{\ell}$ is simply too large — much larger than the target $\mathrm{poly}(n) \cdot (d-1)^{\ell}$ needed to get the "correct" final bound.

However, as first demonstrated by Friedman, the expectation is too large only because of certain low-probability events. Bordenave's way of handling things is to show that: (i) a random d-regular graph $\boldsymbol{G}$ is, with high probability, bicycle-free at large radius r; (ii) when $\boldsymbol{G}$ is so bicycle-free, the rth power of its non-backtracking matrix, $\boldsymbol{B}^r$, coincides with a certain "bicycle-discarding" variant $\boldsymbol{B}^{(r)}$; (iii) the usual Trace Method *can* be successfully applied to $\boldsymbol{B}^{(r)}$; i.e., the expected trace of powers of $\underline{\boldsymbol{B}}^{(r)}$ *is* suitably small.

Thus our weak derandomization of Bordenave's proof has two ingredients, corresponding to (i) and (iii) above. In Section 4.4.1 we derandomize a standard proof that a random d-regular graph is bicycle-free at large radius (in either the configuration model or the random lift model). In Section 4.4.2 we examine the key probabilistic ingredient in Bordenave's use of the Trace Method, [Bor19, Prop. 11], which encapsulates the fact that for a centered version $\underline{\boldsymbol{M}}$ of the configuration model matching matrix, the random variables $\underline{\boldsymbol{M}}_{(v,i),(v',i')}$ are close to k-wise independent

for $k \ll \sqrt{dn}$.

(In Section 4.6, we also show a derandomization of the most basic fact about the configuration model, that $\boldsymbol{G}$ is simple with probability $\Omega_d(1)$. This is just a "bonus" for the reader who prefers the configuration model; it will be more convenient to use the random lift model for our explicit near-Ramanujan graphs, due to its guaranteed simplicity.)

4.4.1 Bicycle-freeness

The following relatively straightforward fact about d-regular n-vertex graphs is crucial for Bordenave's proof: with high probability they are bicycle-free at radius r, provided $r \lesssim c \log_{d-1} n$ for some constant $c < 1/4$. This fact is proved for completeness by Bordenave [Bor19, Lem. 9] (and in [Bor19, Lem. 27] for random lifts); another proof appears earlier in, e.g., [LS10, Lem. 2.1]. We would like a derandomized version of this fact for the k-wise uniform configuration model, $k = O(r)$. This motivates looking for a moments-based proof, such as the one suggested by Wormald [Wor99, Lem. 2.7] and carried out for Erdős–Rényi $\mathcal{G}(n, m)$ graphs in [JŁR00, Thm. 5.5]. The essential point will be that minimal witnesses to failure have only $O(r)$ edges.

Definition 4.4.2 (Minimal bicycle)**.** We say a connected multigraph is a *minimal bicycle* if it is bicyclic but has no proper subgraph that is bicyclic. It is easy to see (cf. [JŁR00, Proof of Thm. 5.5]) that any minimal bicycle is either a "handcuffs graph" (two cycles joined by a path), a "figure-eight graph" (two cycles attached at a vertex), or a "theta graph" (a cycle with a "diagonal").

We now prove:

Proposition 4.4.3. *Fix $d \geqslant 3$ and $k \geqslant 1$. Let $\boldsymbol{G}$ be drawn from the d-regular n-vertex configuration model using a $2k$-wise uniform permutation. Then $\boldsymbol{G}$ is bicycle-free at radius $k/4$, except with probability at most $O(k^3(d-1)^k/n)$.*

As a corollary, the failure probability is at most $1/n^{.99}$ provided $k < c \log_{d-1} n$ for a certain universal $c > 0$. This statement remains true if $\boldsymbol{G}$ is instead a $2k$-wise uniform random n-lift of K_{d+1}. Finally, by Theorem 4.2.5, these statements remain true in the $(\delta, 2k)$-wise uniform versions of the models, $\delta \leqslant 1/n^{8k+2}$.

Proof. We first consider the configuration model. Fix a minimal bicycle H with h vertices and hence $h+1$ edges, where $h < k$. Let the random variable $\boldsymbol{X}_H$ denote

the number of times that H appears in $\boldsymbol{G}$. This is a polynomial of degree at most $h+1 \leqslant k$ in the entries of $\boldsymbol{G}$'s adjacency matrix and hence a polynomial of degree at most $2k$ in the permutation indicators $1[\boldsymbol{\pi}(j) = (v,i)]$. Thus to compute $\mathbf{E}[\boldsymbol{X}_H]$ we may assume $\boldsymbol{G}$ is drawn from the usual configuration model (with a truly random permutation). In this case, it is elementary to compute an exact formula for $\mathbf{E}[\boldsymbol{X}_H]$; as per [Bor16, eqn. (2.4)], it is

$$\mathbf{E}[\boldsymbol{X}_H] = \frac{1}{bc} \frac{n(n-1)(n-2)\cdots(n-h+1)}{(nd-1)(nd-3)(nd-5)\cdots(nd-2h-1)} \tag{4.4}$$

$$\prod_{u \in V(H)} d(d-1)\cdots(d-\deg_H(u)+1), \tag{4.5}$$

where b (respectively, c) is the number of edge- (respectively, vertex-)isomorphisms of H. For any minimal bicycle H we have $b \geqslant 1$, $c \geqslant 2$, and $\deg_H(u) \geqslant 2$ for all $u \in V(H)$. The last of these facts implies the product on the right in Equation (4.4) is at most $(d(d-1))^{h+1}$. Also, the large fraction in the middle is asymptotic to $(d^{h+1}n)^{-1}$, and it is not hard to check it is always at most twice that. Hence we conclude $\mathbf{E}[\boldsymbol{X}_H] \leqslant (d-1)^{h+1}/n \leqslant (d-1)^k/n$. Finally, it is easy to see that, up to isomorphism, the number of minimal bicycles with fewer than k vertices is at most $O(k^3)$. Thus by Markov's inequality we conclude that the probability of having any minimal bicycle on fewer than k vertices is at most $k^3(d-1)^k/n$. The claim about the configuration model now follows because any bicyclic radius-$k/4$ vertex neighborhood in $\boldsymbol{G}$ must contain a minimal bicycle with fewer than k vertices. (The "worst case" is a figure-eight graph.)

As for the model where $\boldsymbol{G}$ is a $2k$-wise uniformly random n-lift of K_{d+1}, the proof is nearly identical. The only difference arises in the computation of $\mathbf{E}[\boldsymbol{X}_H]$ — instead of using an exact closed form expression for the quantity, one can elementarily upper bound $\mathbf{E}[\boldsymbol{X}_H]$ by $O((d+1)^h/n)$ (assuming, say, $k \leqslant \sqrt{n}$). From this slightly weaker bound, one can still draw the same conclusion that the failure probability is at most $1/n^{.99}$ for $k < c \log_{d-1} n$ (possibly with slightly smaller c). □

4.4.2 Bordenave's key probabilistic proposition

In this section we examine the last place in Bordenave's argument that uses randomness of the underlying graph $\boldsymbol{G}$; namely, [Bor19, Prop. 11] for the configuration model and [Bor19, Prop. 28] for the random lift model. These nearly-identical propositions give an upper bound on a certain moment arising in his use of the

Trace Method. Unfortunately, the propositions are not as self-contained as the ones covered in Section 4.4.1. Rather than trying to give a complete summary of how Bordenave's argument works, we will proceed in a "black-box" fashion, only giving the bare minimum needed to verify derandomizability. We refer the reader to [Bor19] for the complete picture. As in [Bor19], we will focus on the configuration model, and then describe the modifications necessary for the random lift model.

Here is the key probabilistic proposition (which can be viewed as a far more sophisticated version of Theorem 4.3.4):

Proposition 4.4.4. *([Bor19, Prop. 11].) Let $\underline{E} = [n] \times [d]$, and let $\boldsymbol{M}$ be a uniformly random matching on $\underline{E}$ as in the configuration model Theorem 4.2.7. Also let $\underline{\boldsymbol{M}}$ be the matrix obtained from $\boldsymbol{M}$ by subtracting $\frac{1}{n'}$ from each entry, where $n' := dn$. Then for any $\gamma \in \underline{E}^{2k}$ with $1 \leqslant k \leqslant \sqrt{m}$ and any $0 \leqslant k_0 \leqslant k$, we have*

$$\left| \mathbf{E} \left[\prod_{t=1}^{k_0} \underline{\boldsymbol{M}}_{\gamma_{2t-1},\gamma_{2t}} \prod_{t=k_0+1}^{k} \boldsymbol{M}_{\gamma_{2t-1},\gamma_{2t}} \right] \right| \leqslant O\Big(2^b \cdot (\tfrac{1}{n'})^a \cdot (\tfrac{3k}{\sqrt{n'}})^{a_1} \Big). \tag{4.6}$$

Here a, b, and a_1 on the right-hand side of Inequality (4.6) are certain quantities relating to the multiplicities of half-edges in γ and to k_0. We omit these definitions here, as they won't be relevant for us.

Note that when $\boldsymbol{M}$ is formed from a random permutation $\boldsymbol{\pi}$ on $[nd]$ as in Theorem 4.2.7, each entry $\boldsymbol{M}_{e,f}$ is a polynomial of degree 2 in the indicators $1[\boldsymbol{\pi}(j) = (v,i)]$. It follows that the quantity inside the expectation in Inequality (4.6) is a polynomial of degree at most $2k$ in these indicators. We conclude:

Corollary 4.4.5. *Let $\boldsymbol{G}$ be drawn from the d-regular n-vertex configuration model using a $2k$-wise uniform permutation, and write $\boldsymbol{M}$ for the matching matrix inducing $\boldsymbol{G}$. Then Inequality (4.6) continues to hold.*

Bordenave also proved an analogue of Theorem 4.4.4 for the random lift model. The statement is extremely similar to Theorem 4.4.4, with "n'" being n, and with the rows/columns of "$\boldsymbol{M}$" being the potential "half-edges" in the lifted graph; for the exact statement we refer the reader to [Bor19, Prop. 28]. Further, Theorem 4.4.5 is true when $\boldsymbol{G}$ is drawn from a $2k$-wise uniform lift model.

With Theorem 4.4.4 in hand, Bordenave does some intricate — but entirely non-probabilistic — path-counting to complete his use of the Trace Method. (This

is like a much more sophisticated version of the part of Section 4.3 beginning with Theorem 4.3.2.) This part of his proof involves considering paths of length $2\ell m$, where "ℓ" and "m" are parameters he selects (with ℓ being at least the bicycle-free radius, and m being large enough so that $\ell m \gg \log n$). The crucial observation for us is that Bordenave *only* employs Theorem 4.4.4 with its parameter "k" set to $2\ell m$ (and the same is true in the random lift model).

Bordenave directly sets $\ell = \Theta(\log_{d-1}(n))$ and $m = \Theta(\log(n)/\log\log(n))$ to obtain best parameters, but we will work more generally, since we may be interested in minimizing $k = 2\ell m$ to save on random bits. Carefully examining [Bor19, Proofs of Prop. 14, 18], one may extract the below proposition. The random matrices $\underline{\boldsymbol{B}}^{(\ell)}$ and $\boldsymbol{R}_1^{(\ell)}, \ldots, \boldsymbol{R}_\ell^{(\ell)}$ mentioned in it are derived from the randomness of the configuration model; again, see [Bor19] for details.

Proposition 4.4.6. *Assuming d, ℓ, m satisfy $\mathrm{poly}(d\ell m)^m \ll n$, it holds that*

$$\mathbf{E}\left[\|\underline{\boldsymbol{B}}^{(\ell)}\|^{2m}\right] \leqslant \mathrm{poly}(n)\cdot(d-1)^{\ell m}, \qquad \mathbf{E}\left[\sum_{i=1}^{\ell}\|\underline{\boldsymbol{R}}_i^{(\ell)}\|^{2m}\right] \leqslant \mathrm{poly}(d\ell m)^m\cdot(d-1)^{2\ell m},$$

Furthermore, this only relies on Inequality (4.6) with $k = 2\ell m$, and therefore by Theorem 4.4.5 it continues to hold even in the $4\ell m$-wise independent configuration model. Thus in this model, Markov's inequality implies that except with probability at most n^{-100},

$$\|\underline{\boldsymbol{B}}^{(\ell)}\| \leqslant \mathrm{poly}(n)^{\frac{1}{2m}}\cdot\sqrt{d-1}^{\ell}, \qquad \sum_{i=1}^{\ell}\|\underline{\boldsymbol{R}}_i^{(\ell)}\| \leqslant \mathrm{poly}(n)^{\frac{1}{2m}}\cdot(d-1)^{\ell}.$$

This proposition holds just the same in the random lift model with base graph $\underline{G} = K_{d+1}$ (indeed, with any d-regular base graph). One simply has to follow through the analogous propositions, [Bor19, Proofs of Prop. 29, 33], in the same way.[4]

Finally, [Bor19, Prop. 8] is the following:

[4]Bordenave carries these propositions out for not-necessarily-regular base graphs of maximum degree d. His computations depend on the base graph through the Perron eigenvalue ρ_1 of its non-backtracking operator B, which in the d-regular case is just $d-1$. In [Bor19, (67)] Bordenave selects $\rho > \rho_1$ and $c_\rho \geqslant 1$ such that $\|(B^\top)^k 1_e\|_1 \leqslant c_\rho \rho^k$ holds for all k and all edges in the base graph. In our d-regular case, we can simply take $c_\rho = 1$ and $\rho = \rho_1 = d-1$ when carrying through his computations.

Proposition 4.4.7. *Suppose $\boldsymbol{G}$ drawn from the d-regular configuration model is bicycle-free at radius ℓ. Let $n' = dn$. Then the largest magnitude eigenvalue of the associated non-backtracking matrix $\boldsymbol{B}$, excluding the trivial eigenvalue of d, is at most*

$$\left(\|\underline{\boldsymbol{B}}^{(\ell)}\| + \frac{1}{n'} \cdot \sum_{i=1}^{\ell} \|\underline{\boldsymbol{R}}_i^{(\ell)}\|\right)^{1/\ell}.$$

Again, Bordenave has a very similar analogue [Bor19, Prop. 26] in the random lift model, with "n'" equal to n, and with the quantity bounding the largest-in-magnitude "new" eigenvalue of the lifted graph (which is precisely what one needs to bound to show the near-Ramanujan property, assuming the base graph is itself d-regular Ramanujan).

We can now finish the proof as Bordenave does (in either the configuration or random lift model), combining Theorem 4.2.23, Theorems 4.4.3, 4.4.6 and 4.4.7, and also Theorem 4.6.1 (if desired). Using the parameter settings $\ell = c \log_{d-1} n$ and $m = (C/c)\log(d-1)/\sqrt{\varepsilon}$ where c is the constant from Theorem 4.4.3 and C is a large enough universal constant, we get the following:

Theorem 4.4.8. *Fix $3 \leqslant d \leqslant C^{-1}\sqrt{\log n}$ and let $\varepsilon \leqslant 1$ and k satisfy*

$$\varepsilon \geqslant C^3 \cdot \left(\frac{\log\log n}{\log_{d-1} n}\right)^2, \qquad k \geqslant C\log(n)/\sqrt{\varepsilon}.$$

Let $\boldsymbol{G}$ be chosen from the d-regular n-vertex k-wise uniform configuration model, or as a k-wise uniform random n-lift of K_{d+1}. Then except with probability at most $1/n^{.99}$, the following hold:

- *$\boldsymbol{G}$ is bicycle-free at radius $c\log_{d-1} n$;*
- *$\lambda(\boldsymbol{G}) \leqslant 2\sqrt{d-1} \cdot (1+\varepsilon)$.*

Additionally, in the configuration model case, $\boldsymbol{G}$ is simple with probability at least $e^{-(d-1)^2/4}/2$. Finally, by Theorem 4.2.5, these statements remains true in the (δ, k)-wise uniform configuration model, $\delta \leqslant 1/n^{8k+1}$.

4.5 Explicit near-Ramanujan graphs

With the tools developed in Section 4.3 and Section 4.4 we are now ready to establish our explicit near-Ramanujan graph constructions. For ease of reading, in

this section we will merely prove Theorem 4.1.1, the deterministic polynomial-time ("weakly explicit") construction, with d and ε assumed to be constant. We leave the slightly more technical proof of the "probabilistically poly $\log n$-time computable" construction (Theorem 4.1.13), with worked out dependence on $d = d(n)$ and $\varepsilon = \varepsilon(n)$, for Section 4.7.

Recall we want to show there is a deterministic algorithm that on input $N, d \geqslant 3$ and $\varepsilon > 0$, outputs in $\mathrm{poly}(N)$-time a d-regular graph G on $N' \sim N$ vertices with $\lambda(G) \leqslant 2\sqrt{d-1} + \varepsilon$.

Before getting into the details, we recap the construction as outlined in Section 4.1.4:

1. Using Theorem 4.4.8 we construct a d-regular simple graph G_0 on some "small" number of vertices $n_0 = n_0(N)$, which is bicycle-free at radius $\Omega(\log n_0)$ and has $\lambda(G_0) \leqslant 2\sqrt{d-1} + \epsilon$. The quantity n_0 should satisfy
$$2^{\omega((\log\log N)^2)} \leqslant n_0 \leqslant 2^{O(\sqrt{\log N})},$$
the left inequality so that G_0 is sufficiently bicycle-free for Step 2 below, and the right inequality so that G_0 is constructible in deterministic $\mathrm{poly}(N)$ time. We have a wide range of allowable possibilities here; for concreteness we will take n_0 near the upper limit to allow for slightly better dependence on non-constant d, ε in Section 4.7.

2. Next we repeatedly use Theorem 4.3.1 (roughly $\log(N/n_0) \sim \log N$ times) to double the number of vertices in our construction from Step 1, while keeping $\lambda \leqslant 2\sqrt{d-1} + \epsilon$ and also retaining that the graph is bicycle-free at radius $\Omega(\log n_0)$ (Theorem 4.2.12). Importantly, since Theorem 4.3.1 is a high-probability result, we will be able to reuse the seed for each of the $\log N$ pseudorandom edge-signings.

Step 1 details. Here the algorithm will select n_0 to be an even integer on the order of $2^{\Theta(\sqrt{\log N})}$. Theorem 4.4.8 tells us that for a sufficiently large $k = O(\log n_0) = O(\sqrt{\log N})$, and for sufficiently small $\delta = n_0^{-\Theta(k)} = 1/\mathrm{poly}(N)$, a random d-regular n_0-vertex graph $\boldsymbol{G}_0$ chosen from the (δ, k)-wise uniform configuration or random-lift-of-K_{d+1} model will with high probability satisfy:

$$\boldsymbol{G}_0 \text{ is bicycle-free at radius } \Omega(\log n_0) = \Omega(\sqrt{\log N}); \qquad \lambda(\boldsymbol{G}_0) \leqslant 2\sqrt{d-1} + \varepsilon. \tag{4.7}$$

(Recall we are treating d and ε as constant here.) $\boldsymbol{G}_0$ will also be simple with $\Omega(1)$ probability in the configuration model case, and with probability 1 in the random lift case. In the former case, we need a (δ, k)-wise permutation in S_{nd}; in the latter case, we need $\binom{d+1}{2}$ independent (δ, k)-wise permutations in S_n. Either way, Theorem 4.2.4 tells us that a deterministic algorithm can enumerate all possibilities for $\boldsymbol{G}_0$ in $\mathrm{poly}(N)$ time and pick out any fixed simple one G_0 satisfying (4.7).

Step 2 details. Here the algorithm will be applying Theorem 4.3.1 some $t \sim \log_2 N$ times, starting with G_0, and each time interpreting the edge-signing produced as a 2-lift as discussed in Section 4.1.4. This produces a sequence of pseudorandom d-regular simple graphs $\boldsymbol{G}_1, \dots, \boldsymbol{G}_t$, where $\boldsymbol{G}_i$ has $n_0 2^t$ vertices. The parameter t is chosen to be least possible such that the final number of vertices, $N' = n_0 2^t$, is as at least N. It is not hard to check that by adjusting n_0 by a factor of at most 2, we can ensure that $N'/N = 1 + o_N(1)$, where the $o_N(1)$ term is $O(1/n_0) = 1/2^{\Theta(\sqrt{\log N})}$.

For simplicity, we will use the same values for the parameters r, k, and δ in each application of Theorem 4.3.1; only the value of n will change (ranging from n_0 up to N'). We may take $r = \Omega(\sqrt{\log N})$, the bicycle-free radius from Equation (4.7) (observe that the bicycle-free radius cannot decrease for *any* 2-lift of a graph). Note that the failure probability of any single 2-lift is at most $1/2^{\Theta(\sqrt{\log N})}$, and hence a union bound tells us that the probability of *any* of the 2-lifts "failing" is low, $\frac{\log N}{2^{\Theta(\sqrt{\log N})}}$. We take the parameter "k" to be $\Theta\left(\frac{\log N}{\sqrt{\varepsilon}}\right)$ (the hidden constant sufficiently large depending on d). Finally, we take $\delta = 1/N^{\Theta(1/\sqrt{\varepsilon})}$ (again with the hidden constant sufficiently large depending on d). By plugging these parameters into Theorem 4.3.1 we conclude that with high probability, all "new" eigenvalues arising in the 2-lifted adjacency matrices $\boldsymbol{A}_1, \dots, \boldsymbol{A}_t$ are at most $2\sqrt{d-1} + \varepsilon$ in magnitude, and hence $\boldsymbol{G}_t$ is ε-near Ramanujan.

It remains to observe that with these parameter settings, using Theorem 4.2.2, a deterministic algorithm can in $\mathrm{poly}(N/\delta) = \mathrm{poly}(N)$ time do the following: First, produce a single $(\delta, 2\ell)$-wise uniform multiset of strings $Y \subseteq \{\pm 1\}^{N'd/4}$; here $N'd/4$ bits are sufficient to edge-sign/2-lift any of the graphs $\boldsymbol{G}_i$. Then, for $i = 1, \dots, t$ the algorithm can search Y for a "good" string $y_i \in Y$, meaning one with the property that using it to do an edge-signing/2-lift of G_i yields graph G_{i+1} which is ε-near Ramanujan. As argued in the previous paragraph, a $1 - O\left(\frac{\log N}{2^{\Theta(\sqrt{\log N})}}\right)$

fraction of strings in Y have this property. We can check the goodness of any string y in $\mathrm{poly}(N)$ time using the following fact.

Fact 4.5.1. *For any rational approximation ρ of $2\sqrt{d-1}+\varepsilon$, one can decide in $\mathrm{poly}(n)$ time whether $\lambda(G) \leqslant \rho$.*

This concludes the proof of Theorem 4.1.1.

4.6 Simplicity

In the fully uniform configuration model, the probability of $\boldsymbol{G}$ being simple (i.e., being an ordinary graph with no self-loops or parallel edges) is known [BC78, Bol80] to tend to the constant $\exp(-(d^2-1)/4)$, as $n \to \infty$. We establish that the $O(d^2)$-wise uniform configuration model suffices for this:

Proposition 4.6.1. *Let $3 \leqslant d \ll \sqrt{\log n}$ and let $k \geqslant Cd^2$, where C is a certain universal constant. Let $\boldsymbol{G}$ be drawn from the d-regular n-vertex configuration model using a k-wise uniform permutation. Then*

$$\mathbf{Pr}[\boldsymbol{G} \text{ is simple}] = e^{-(d^2-1)/4}(1 \pm e^{-100d^2}).$$

By Theorem 4.2.5, this remains true if the permutation is merely (δ, k)-wise uniform, $\delta \leqslant n^{-C'd^2}$.

The proof is a straightforward derandomization of Bollobás's original analysis of simplicity in the configuration model [Bol80]. Unlike several later refinements that used the Chen–Stein method, Bollobás's proof uses the method of moments, making it particularly convenient to derandomize using k-wise uniform permutations.

Proof of Theorem 4.6.1. Let us recap Bollobás's proof concerning an n-vertex d-regular configuration model graph $\boldsymbol{G}$ formed from a truly random permutation $\boldsymbol{\pi} \sim S_{nd}$. He defines $\boldsymbol{X}_1$ to be the number of self-loops in $\boldsymbol{G}$ (i.e., $\frac{1}{2}\operatorname{tr}(\boldsymbol{A})$), $\boldsymbol{X}_2$ to be the number of 2-cycles (i.e., $\sum_{v<v'} \binom{\boldsymbol{A}_{v,v'}}{2}$), and $\boldsymbol{X} = \boldsymbol{X}_1 + \boldsymbol{X}_2$. Note that $\boldsymbol{G}$ is simple if and only if $\boldsymbol{X} = 0$. The idea of the proof is that it is nearly the case that $\boldsymbol{X}_1, \boldsymbol{X}_2$ are independent Poisson random variables with respective means

$$\lambda_1 = \lambda \cdot \frac{nd}{nd-1} = \lambda \cdot (1 \pm O(1/n))$$

$$\lambda_2 = \lambda^2 \cdot \frac{nd \cdot (nd-d)}{(nd-1)\cdot(nd-3)} = \lambda^2 \cdot (1 \pm O(1/n)),$$

where $\lambda := (d-1)/2$. Thus $\boldsymbol{X}$ should be nearly Poisson with mean $\lambda_1 + \lambda_2 \sim \lambda + \lambda^2 = (d^2-1)/4$, and hence we should have $\mathbf{Pr}[\boldsymbol{X} = 0] \sim e^{-(d^2-1)/4}$.

More precisely, Bollobás first establishes [Bol80, ineq. (11)] the following estimate for all integers $0 \leqslant r \leqslant 8 \log n$:

$$E_r := \mathbf{E}\left[\binom{\boldsymbol{X}}{r}\right] \text{ satisfies } \left|E_r - \frac{(\lambda_1+\lambda_2)^r}{r!}\right| \leqslant \frac{(\lambda_1+\lambda_2)^r}{r!} \cdot O(r^2/n). \tag{4.8}$$

(Actually, Bollobás has $O((\log n)^2/n)$ on the right-hand side rather than $O(r^2/n)$, but inspection of his proof confirms the above.) The key point for our proof of Theorem 4.6.1 is that Inequality (4.8) continues holds when the permutation $\boldsymbol{\pi} \in S_{nd}$ defining $\boldsymbol{G}$ is merely $4r$-wise uniform. This is simply because $\binom{\boldsymbol{X}}{r}$ is a polynomial of degree at most $4r$ in the indicators $1[\boldsymbol{\pi}(j) = (v,i)]$. Thus to complete the proof, it suffices to derive the conclusion

$$\mathbf{Pr}[\boldsymbol{X} = 0] = e^{-(d^2-1)/4}(1 \pm e^{-100d^2}) \tag{4.9}$$

from the estimates in Inequality (4.8) with $r = O(d^2)$. This can be done exactly as in Bollobás's work. He uses the following inclusion-exclusion-type inequality, which holds (for any $u \in \mathbb{N}$) due to $\boldsymbol{X}$ being $\mathbb{N}$-valued:

$$\sum_{r=0}^{2u+1} (-1)^r E_r \leqslant \mathbf{Pr}[\boldsymbol{X} = 0] \leqslant \sum_{r=0}^{2u} (-1)^r E_r. \tag{4.10}$$

Notice that $E_r \approx \frac{(\lambda_1+\lambda_2)^r}{r!}$, and

$$\sum_{r=0}^{\infty} (-1)^r \frac{(\lambda_1+\lambda_2)^r}{r!} = e^{-(\lambda_1+\lambda_2)} = e^{-(\lambda+\lambda^2)\cdot(1\pm O(1/n))} = e^{-(d^2-1)/4} \cdot (1 \pm O(d^2/n)); \tag{4.11}$$

also, $O(d^2/n) \ll e^{-100d^2}$ since $d \ll \sqrt{\log n}$. Thus we can establish Equation (4.9) by bounding the two errors distinguishing the infinite sum in Equation (4.11) from the sums on the left- and right-hand side of Inequality (4.10). The two distinctions are: the error in $E_r \approx \frac{(\lambda_1+\lambda_2)^r}{r!}$, boundable using Inequality (4.8); and, the tail of the infinite sum from $2u$ or $2u+1$ onward. In absolute value, these two errors are

boundable by:

$$O(1/n) \cdot \sum_{r=0}^{2u \text{ or } 2u+1} \frac{(\lambda_1+\lambda_2)^r}{r!} \cdot r^2, \quad \text{and} \quad \sum_{r=2u+1 \text{ or } 2u+2}^{\infty} \frac{(\lambda_1+\lambda_2)^r}{r!}.$$

The first quantity above can be bounded by $O(u^2/n) \cdot e^{\lambda_1+\lambda_2}$, and the second quantity can be bounded by $O((\lambda_1+\lambda_2)^{2u+1}/(2u+1)!)$ provided $u \geqslant \lambda_1+\lambda_2$. Recalling $\lambda_1+\lambda_2 = \Theta(d^2)$ and $d \ll \sqrt{\log n}$, we see that by taking $u = O(d^2)$ sufficiently large, both errors can be made much smaller than e^{-100d^2}, and we obtain Equation (4.9) with $r = O(d^2)$ as needed. □

4.7 The probabilistically poly log n-time computable construction

We now walk through the steps of Section 4.5 giving precise parameter details along the way, and extract a probabilistically poly log n-time computable construction of near-Ramanujan graphs.

Assume we are given N, $3 \leqslant d \leqslant \frac{(\log N)^{1/8}}{C}$ and $\epsilon \gg \frac{(\log\log N)^4}{\log N} \cdot \sqrt{d}$ where C is the constant from the statement of Theorem 4.4.8.

Revisiting Step 1. Choose parameters as follows: $\alpha = 1/\sqrt{\binom{d+1}{2}}$; n_0 as the largest multiple of $d+1$ smaller than $2^{\alpha\sqrt{\log N}}$; $k = C\alpha\sqrt{\log N} \cdot d^{1/4}/\sqrt{\epsilon}$ (which is $\approx \log n_0$); and $\delta = 1/N^{8k+1}$. Recall that the key result used in this step is that by Theorem 4.4.8, $\boldsymbol{G}_0$ drawn from the n_0-vertex (δ,k)-wise random-lift-of-K_{d+1} model is a simple graph that with high probability satisfies:

$$\boldsymbol{G}_0 \text{ is bicycle-free at radius } \Omega\left(\frac{\alpha\sqrt{\log N}}{\log(d-1)}\right); \qquad \lambda(\boldsymbol{G}_0) \leqslant 2\sqrt{d-1}+\varepsilon. \tag{4.12}$$

As an upshot of Theorem 4.2.4, $\boldsymbol{G}_0$ can be sampled using $\boldsymbol{s}$, a uniform binary string of length $O\left(\frac{\log N \cdot d^{1/4}}{\sqrt{\epsilon}}\right)$ as a seed. In particular, $\boldsymbol{s}$ is divided into $\binom{d+1}{2}$ disjoint substrings $\boldsymbol{s}_{e_1}, \ldots, \boldsymbol{s}_{e_{\binom{d+1}{2}}}$ each of length $\ell_1 = O\left(\frac{\alpha^2 \log N \cdot d^{1/4}}{\sqrt{\epsilon}}\right)$ indexed by edges of K_{d+1}; the (δ,k)-wise uniform permutation $\boldsymbol{\pi}_{uv}$ corresponding to edge (u,v) is taken to be the $\boldsymbol{s}_{uv}$th permutation in the multiset of permutations Π from the statement

of Theorem 4.2.4. Additionally, given $\boldsymbol{s}$ and a vertex $(u, i) \in V(\boldsymbol{G}_0)$, it is possible to return a list of its neighbors in time $T_1 = O\left(d \cdot \text{poly}\left(\frac{\alpha^2 \log N \cdot d^{1/4}}{\sqrt{\epsilon}}\right)\right)$.

Revisiting Step 2. Let $t = \left\lceil \log\left(\frac{N}{n_0}\right) \right\rceil$; let β be a large enough constant; let $k = \frac{2\beta d^{1/4}}{\sqrt{\epsilon}} \log N$; and let $\delta = N^{-O(\beta d^{1/4} \log d/\sqrt{\epsilon})}$. The main result used in Step 2 is that from Theorem 4.3.1 the graphs $\boldsymbol{G}_1, \ldots, \boldsymbol{G}_t$ where $\boldsymbol{G}_i$ is obtained via a 2-lift of $\boldsymbol{G}_{i-1}$ induced by a (δ, k)-wise uniform signing have their nontrivial eigenvalues bounded by $2\sqrt{d-1} + \epsilon$ in magnitude, except with probability $O(t/n_0^{100})$. From Theorem 4.2.2, a (δ, k)-wise uniform signing of any $\boldsymbol{G}_i$ can be obtained by first sampling a random binary string $\boldsymbol{s}'$ of length $\ell_2 = O\left(\frac{d^{1/4} \log d \cdot \log N}{\sqrt{\epsilon}}\right)$ and choosing the $\boldsymbol{s}$th string in the multiset of signings Y from the theorem statement. In fact, given $\boldsymbol{s}'$ and edge $e \in \boldsymbol{G}_i$ one can also output the sign assigned to edge e in time $T_2 = \text{poly}\left(\beta d^{1/4} \log d \log N/\sqrt{\epsilon}\right)$. Finally, by the union bound, the bound of $O(t/n_0^{100})$ on the probability that $\boldsymbol{G}_t$ is not ϵ-near Ramanujan holds if we use independently chosen seeds $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_t$ to perform the 2-lifts. Note that $t < \log N$ and $n_0 \geqslant 2^{(\log N)^{1/4}}$ and hence the failure probability is $o_N(1)$.

Probabilistically strongly explicit near-Ramanujan graphs. Given a uniform binary string $\boldsymbol{s}$ of length $\ell_1 + t \cdot \ell_2$ as a random seed, call the substring given by the first ℓ_1 bits $\boldsymbol{s}_1$ and the substring given by the next $t \cdot \ell_2$ bits $\boldsymbol{s}_2$. Let $\boldsymbol{G}_0$ be sampled from $\boldsymbol{s}_1$ as described in Step 1, and let $\boldsymbol{G}_t$ be the "final graph" obtained by the sequence of 2-lifts in Step 2 from $\boldsymbol{s}_2$. Each vertex in $\boldsymbol{G}_i$ can be naturally identified with a tuple $(v, a, x) \in [d] \times [n_0] \times \{0,1\}^i$. Let x be a string in $\{0,1\}^t$, let $x^{\leqslant i}$ denote its i-bit prefix. Given a vertex (v, a, x) in $\boldsymbol{G}_t$ and seeds $\boldsymbol{s}_1$ and $\boldsymbol{s}_2$, we describe an algorithm to output a list of its d neighbors in $\widetilde{O}(T_1 + dT_2)$-time where the $\widetilde{O}(\cdot)$ hides factors of poly log N. From Step 1, we know that there is an T_1-time algorithm to output a list of d neighbors of $(v, a, x^{\leqslant 0})$ in $\boldsymbol{G}_0$.

Next, given a list of neighbors of $(v, a, x^{\leqslant i-1})$ in $\boldsymbol{G}_{i-1}$ it is possible to output a list of neighbors of $(v, a, x^{\leqslant i})$ in $\boldsymbol{G}_i$ in $\widetilde{O}(dT_2)$-time in the following way. Let (w, b, y) be a neighbor of $(v, a, x^{\leqslant i-1})$. Then exactly one of $(w, b, y \wedge 0)$ and $(w, b, y \wedge 1)$ is a neighbor of $(v, a, x^{\leqslant i})$ where $\wedge$ denotes concatenation. It is possible to obtain the sign on edge $\{(v, a, x^{\leqslant i-1}), (w, b, y)\}$ in the 2-lift from $\boldsymbol{G}_{i-1}$ to $\boldsymbol{G}_i$ in T_2 time from $\boldsymbol{s}_2$. If the sign is a -1, then $(w, b, y \wedge (1 - x_i))$ is a neighbor of $(v, a, x^{\leqslant i})$; otherwise

$(w, b, y \wedge x_i)$ is a neighbor. Thus, in $\widetilde{O}(dT_2)$ time, we can obtain a length-d (and hence complete) list of neighbors of $(v, a, x^{\leqslant i})$.

As a result, after spending T_1 time generating a list of neighbors of $(v, a, x^{\leqslant 0})$, we can use the above routine t times to obtain a list of neighbors of (v, a, x) in $\boldsymbol{G}_t$ in $T_1 + t \cdot \widetilde{O}(dT_2) \leqslant \widetilde{O}(T_1 + dT_2)$. From the upper and lower bounds on d and ϵ, this quantity is always $O(\operatorname{poly}\log N)$.

To summarize, we have an algorithm that takes in a random seed of length $O\left(\frac{d^{1/4}\log d \cdot \log^2 N}{\sqrt{\epsilon}}\right)$ and implements the adjacency matrix of a corresponding random graph $\boldsymbol{G}$ such that:

- Given any vertex v of $\boldsymbol{G}$, its list of neighbors can be generated in $O(\operatorname{poly}\log N)$ time.
- $\boldsymbol{G}$ is ϵ-near Ramanujan with probability $1 - o_N(1)$.

This yields the conclusion of Theorem 4.1.13.

Chapter 5

Girth-density tradeoffs in hypergraphs

This chapter is adapted from [HKM23], co-authored by the author of this thesis, Jun-Ting (Tim) Hsieh, and Pravesh Kothari.

The hypergraph Moore bound is an elegant statement that characterizes the extremal trade-off between the girth — the number of hyperedges in the smallest cycle or *even cover* (a subhypergraph with all degrees even) and size — the number of hyperedges in a hypergraph. For graphs (i.e., 2-uniform hypergraphs), a bound tight up to the leading constant was proven in a classical work of Alon, Hoory and Linial [AHL02]. For hypergraphs of uniformity $k > 2$, an appropriate generalization was conjectured by Feige [Fei08]. The conjecture was settled up to an additional $\log^{4k+1} n$ factor in the size in a recent work of Guruswami, Kothari and Manohar [GKM21]. Their argument relies on a connection between the existence of short even covers and the spectrum of a certain randomly signed *Kikuchi* matrix. Their analysis, especially for the case of odd k, is significantly complicated.

In this work, we present a substantially simpler and shorter proof of the hypergraph Moore bound. Our key idea is the use of a new *reweighted* Kikuchi matrix and an *edge deletion* step that allows us to drop several involved steps in [GKM21]'s analysis such as combinatorial bucketing of rows of the Kikuchi matrix and the use of the Schudy–Sviridenko polynomial concentration. Our simpler proof also obtains tighter parameters: in particular, the argument gives a new proof of the classical Moore bound of [AHL02] with no loss (the proof in [GKM21] loses a $\log^3 n$ factor), and loses only a single logarithmic factor for all $k > 2$-uniform

hypergraphs.

As in [GKM21], our ideas naturally extend to yield a simpler proof of the full trade-off for strongly refuting smoothed instances of constraint satisfaction problems with similarly improved parameters.

5.1 Introduction

What is the maximum girth of a graph on n vertices and average degree d? For d-regular graphs, a simple "ball growing" argument shows that the graph must have a cycle of length at most $2\log_{d-1} n + 2$. This threshold is called the *Moore bound* [Wik22] (see Page 180 of [Big93]) and graphs achieving it are called Moore graphs. In a classical paper that resolved a question of Bollobás [Bol78], Alon, Hoory and Linial [AHL02] proved that the same upper bound holds even for irregular graphs. Later on, Hoory [Hoo02] obtained a better bound for bipartite graphs and Babu and Radhakrishnan [BR14] found an elegant proof based on the entropy of random walks.

Girth-density trade-offs for hypergraphs. This work is about a natural and well-studied generalization of the Moore bound to $k > 2$-uniform hypergraphs. A cycle[1] in a hypergraph, more descriptively called an *even cover*, is a collection of hyperedges such that every vertex participates in an even number of them. The girth of a hypergraph is the smallest size of an even cover in it. When specialized to graphs, an even cover is simply a union of cycles and thus, this formulation naturally generalizes the standard notion of girth in graphs.

Analogously to the Moore bound, understanding the maximum number of hyperedges that one can pack in a hypergraph while avoiding an even cover of a given length is a basic *hypergraph Turán* problem. Hypergraph Turán problems are typically significantly more difficult than their counterparts in graphs. Indeed, even the original hypergraph Turán conjecture from the 1940s that studies an appropriate analog of triangle free hypergraphs is still open. We direct the reader to the recent survey of Keevash [Kee11] for an overview of hypergraph Turán theory.

[1] There are several well-studied combinatorial notions of cycles in contrast to the more linear algebraic notion of even covers.

Applications of girth-density trade-offs. Like the graph Moore bound, girth-density trade-offs for hypergraphs have foundational connections to several research directions in theoretical computer science. One source of such applications is the observation that for a collection of linear equations in $\mathbb{F}_2$ on n variables, if we associate each equation to the set indicated by its coefficient vector, then the girth of the resulting hypergraph on $[n]$ is the same as the size of the smallest linearly dependent subset of equations. As a consequence, rate vs distance trade-offs for *low density parity check* (LDPC) codes are equivalent to the girth vs size trade-offs for k-uniform hypergraphs with hyperedges corresponding to the columns of the parity check matrix. As a result, there is an extensive line of work that studies the girth-density trade-offs for hypergraphs (see e.g. [BKHL99, BMS08, AF09]).

Naor and Verstraëte [NV08] started a systematic study of hypergraph girth density trade-offs. They were explicitly motivated by mapping the rate-distance trade-offs for LDPC codes and computing product representations of square integers arises as a step in sub-exponential time algorithms for integer factoring. In particular, they showed that every k-uniform hypergraph on n vertices and $O(n^{k/2} \log n)$ hyperedges must have an even cover of size $O(\log n)$. Improving the bounds of [NV08] for $k = 3$, Feige [Fei08] proved that every 3-uniform hypergraph on n vertices and $O(n^{3/2}) \log \log n$ hyperedges has an even cover of length $O(\log n)$. Feige's motivation was a connection, via the connection to linear equations modulo 2 discussed above, to refuting random 3SAT formulas and generalizations. In particular, by exploiting this improved bound, Feige derived a weak refutation algorithm for smoothed 3SAT formulas with $O(n^{1.5} \log \log n)$ constraints.

Hypergraph Moore bound. Feige's result leaves open the uncharted territory of hypergraph sizes between $m \sim n$ and $m \sim n^{k/2}$ — a polynomially large multiplicative interval when $k > 2$. The work of Feige, Kim and Ofek [FKO06] found an intriguing connection between the girth bounds in this interesting regime and the foundational average-case problem of refuting random 3SAT formulas [Fei02b]. They observed that *random* hypergraphs with $m \gtrsim n^{1.4}$ hyperedges[2] must have an even cover of length $O(n^{0.2})$ and used a tour de force argument based on the second moment method to establish that at the same density, random hypergraphs should contain $n^{1.4}$ different almost disjoint even covers of size $n^{0.2}$. As a conse-

[2]Throughout this work, we will use the notation $f \gtrsim g$ to stand for "there exists a constant $C > 0$ such that $f \geqslant Cg$".

quence, they obtained their celebrated result on the existence of polynomial size witnesses of unsatisfiability for random 3SAT formulas with $O(n^{1.4})$ constraints — a threshold that is $n^{0.1}$ factor smaller than the $m \gtrsim n^{1.5}$ bound for the best known efficient refutation algorithms. Motivated both by whether FKO witnesses could be efficiently constructed (and potentially refute a strong form of Feige's Random 3SAT hypothesis [Fei02b]) and investigating whether such certificates exist in semi-random and smoothed 3SAT formulas, Feige [Fei08] conjectured the following *hypergraph Moore bound*.

Conjecture 5.1.1 (Hypergraph Moore Bound (Feige's conjecture), Conjecture 1.2 of [Fei08])**.** *For every $k \in \mathbb{N}$ and $1 \leqslant r \leqslant n$, every hypergraph with n vertices and $m \gtrsim n(\frac{n}{r})^{\frac{k}{2}-1}$ hyperedges has an even cover of size $O(r \log n)$.*

In addition to a complete rate-distance profile for LDPC codes, Feige's conjecture implies (see Section 9 in [GKM21] for an exposition) a significantly simpler and 2nd-moment-method-free proof of the existence of the FKO [FKO06] refutation witnesses below the spectral threshold for random 3SAT (and other CSPs) that also generalizes to semirandom and smoothed instances[3].

Feige's conjecture was recently settled by Guruswami, Kothari and Manohar [GKM21] up to an additional $\log^{4k+1} n$ multiplicative factor in the density m. Their proof goes via a new connection between the existence of small even covers in k-uniform hypergraphs and sub-exponential size spectral refutations of semirandom k-XOR formulas via a certain *Kikuchi* matrix.

While [GKM21] begins with an elegant and simple observation, their technical analysis especially for odd k (the "hard" case in all algorithms and certificates for refutation) is quite complicated and involves manipulating the Kikuchi matrix via "row bucketing" and "row pruning" in various steps and invoking the Schudy–Sviridenko concentration inequality [SS12] (that extends the breakthrough work of Kim and Vu [KV00]) for polynomials with combinatorial structure in the monomials. As a consequence, even for the simplest case of $k = 2$ (i.e., recovering the classical Moore bound), their proof incurs an additional $\log^3 n$ factor.

[3] A smoothed Boolean CSP instance is obtained by starting from a worst-case instance and perturbing the literal patterns by independently flipping each with some small constant probability (with probability $1/2$ in the special case of the semirandom model). In particular, in contrast to random CSPs where the variables in every clause are generated uniformly at random, smoothed and semirandom CSP instances have a worst-case clause structure.

5.1.1 Our results

The main result of this work is a simple and short proof of the hypergraph Moore bound that is *almost* tight up to a single logarithmic factor.

Theorem 5.1.2. *For every $k \in \mathbb{N}$ and $1 \leqslant r \leqslant n$, every hypergraph on n vertices and $m \gtrsim n \log n \cdot (\frac{n}{r})^{\frac{k}{2}-1}$ hyperedges has an even cover of size $O(r \log n)$.*

In Section 5.2.1 and Appendix 5.5, as evidence of the power of our proof strategy, we obtain yet another proof of the classical Moore bound [AHL02] with the same leading constant. Independently of our work, David Munhá Correia and Benny Sudakov [CS22] informed us that they have found a simple, combinatorial argument for analyzing the Kikuchi matrix to prove a hypergraph Moore bound for even arity k that also loses only a single logarithmic factor.

Our techniques extend to give a simple and tighter proof of a sub-exponential time strong refutation algorithm for semirandom k-XOR formulas when the number of constraints is below the "spectral threshold" $n^{k/2}$, which is spelled out in Section 5.4. Via the standard XOR trick (see, for example, [AOW15]), this recovers a tighter trade-off for refuting smoothed Boolean constraint satisfaction problems as in [GKM21]. Prior to our work, a bound tight up to $\log n$ factors was not known even for the (significantly) easier setting of *fully random* k-XOR refutation for odd k (the argument of [WAM19] obtains such a result for even k) where the best known bound due to [RRS17] loses a $\log^{2k} n$ factor.

Theorem 5.1.3 (Informal)**.** *Fix $k \in \mathbb{N}$ and $r \leqslant n$, there is an $n^{O(r)}$-time algorithm such that given a semirandom k-XOR instance ψ with n variables and $m \gtrsim n \log n \cdot (\frac{n}{r})^{\frac{k}{2}-1}$ constraints, it certifies that ψ is not $(1/2 + 0.01)$-satisfiable.*

We believe that the last remaining logarithmic factor in the theorems above is also unnecessary. However, removing it seems related to certain technical difficulties that arise in beating the logarithmic factor incurred in spectral norm bounds for the matrix Rademacher series [Tro15]. In particular, the tightest known proof for the closely related problem of refuting fully random k-XOR formulas below the spectral threshold also loses a $\log n$ factor for even k [WAM19] and a $\log^{2k} n$-factor for odd k [RRS17]. For some easier settings such as refutation in the polynomial time regime [dT22] and understanding the SDP value of random NAE-3SAT and generalizations [FM17, DMO+19, MOP20], recent works manage to circumvent this difficulty by application of powerful tools such as the Ihara–Bass

formula and the largest eigenvalue of non-backtracking walk matrices. Our proof suggests a natural and more elementary route to removing this final logarithmic factor but requires resolving the count of certain "walks" that arise in our analysis.

Key ideas. The abstract strategy employed by [GKM21] is to construct a so called *Kikuchi* matrix $A_{\mathcal{H}}$ (first introduced in the work of [WAM19] on Gaussian tensor PCA) associated with our hypergraph $\mathcal{H}$ where:

(i) $\mathbb{1}^\top A_{\mathcal{H}} \mathbb{1}$ is a surrogate for the number of hyperedges in $\mathcal{H}$.

(ii) The lack of short even covers in $\mathcal{H}$ can be turned into a certificate that $\mathbb{1}^\top A_{\mathcal{H}} \mathbb{1}$ is small, which then translates to a bound on $|\mathcal{H}|$.

The certificate used by [GKM21] is $\|A_{\mathcal{H}}\|_{\infty \to 1}$, which they control by bucketing the rows by weight, bounding the spectral norm of each submatrix, and stitching these norms together.

Our key insight is in the style of certificate we provide — we give a matrix Q such that $Q \succeq A_{\mathcal{H}}$. Such a certificate implies a bound of $\mathrm{tr}(Q)$ on our surrogate for $|\mathcal{H}|$. The inequality $Q \succeq A_{\mathcal{H}}$ is equivalent to proving $\left\|Q^{-1/2} A_{\mathcal{H}} Q^{-1/2}\right\|_2 \leqslant 1$, which can be done via the trace moment method with relative ease. Our reweighting strategy is akin to constructing a "diagonal weighted" *dual solution* for certifying upper bounds on the value of the basic SDP relaxation for quadratic optimization problems on the hypercube such as Max-Cut and the Grothendieck problems. Our *reweighting* strategy simplifies the analysis, removes the need for the "row bucketing" step in [GKM21], and lets us obtain a sharper result.

Our proof for odd k requires combining our reweighted Kikuchi matrix with a new "edge deletion" operation that controls the "heavy rows" in the Kikuchi matrix. At a high level, our strategy involves deleting an appropriately chosen set of entries of the Kikuchi matrix in comparison to the "row pruning" strategy of [GKM21] which involves deleting entire rows (vertices in the Kikuchi graph). This seemingly technical change leads to a great deal of simplification and in particular allows replacing the use of the Schudy–Sviridenko inequality [SS12] and the carefully introduced logarithmic factors in the hypergraph regularity decomposition in [GKM21].

Organization. The rest of this paper is organized as follows. In Section 5.2, we give a complete (and sharper) proof of the hypergraph Moore bound for the even

arity case, starting with a proof of a weak version (that loses additional constant factors) of the classical Moore bound using our ideas. We will include detailed commentary for the sake of exposition and a short overview of the additional ideas (including our new edge deletion trick) to handle the odd arity case. In Section 5.3, we will give a proof of the hypergraph Moore bound for odd arity. Finally, in Section 5.4, we will extend our techniques to obtain strong refutation algorithms for semirandom and smoothed Boolean CSPs.

5.2 Warm-up: hypergraph Moore bound in the even arity case

In this section, we will give a proof of the Moore bound for hypergraphs of even arity with the goal of providing an exposition of our main ideas. As an illustration of the power of our reweighting idea, in Section 5.2.1 we will give a simple proof of the classical Moore bound [AHL02] that is tight up to an absolute constant factor (as opposed to the $\log^3 n$ loss incurred by the strategy of [GKM21]).[4] In Section 5.2.2, we will generalize the reweighting idea to prove hypergraph Moore bound for all even arities. Finally in Section 5.2.3, we will discuss the key new idea of *edge deletions* that is crucial for our simpler and tighter proof for the case of odd k.

5.2.1 Weak Moore bound for graphs

In this section, we prove a weak Moore bound for graphs to illustrate our reweighting strategy in a simple setting. The resulting bound is weak in the sense that it incurs a constant factor loss when compared to [AHL02]. In Appendix 5.5, we implement this strategy (in a way that is less generalizable to hypergraphs) to recover the tight $2\log_{d-1} n$ bound.

We note that [GKM21] also proved a weaker Moore bound (Proposition 2.3 of [GKM21]) to illustrate their "row bucketing" strategy that partitions the vertices into $O(\log n)$ buckets, each of which has vertices with degrees within a multiplicative constant factor of each other. This strategy splits the adjacency matrix A into

[4] In Appendix 5.5, we present a proof that uses one additional tool to recover the classical Moore bound for irregular graphs with the same leading constant.

$O(\log^2 n)$ pieces and ends up requiring an average degree $d \gtrsim \log^3 n$ in order to contain a cycle of length $O(\log n)$.

This simple exercise will show how our reweighting handles different degrees automatically, avoiding the lossy row bucketing step completely.

Proposition 5.2.1 (Weak Moore bound for irregular graphs)**.** *Every graph with n vertices and average degree $d > 16$ has a cycle of length at most $2\lceil \log_{(d/16)} n \rceil$.*

The core of the proof of Proposition 5.2.1 is the following spectral norm bound on the reweighted adjacency matrix.

Claim 5.2.2. Let G be a graph with n vertices and average degree $d > 1$ that has no cycle of length $\leqslant \ell$ for some even $\ell \in \mathbb{N}$. Let A be the $\{0,1\}$ adjacency matrix of G, and let $\Gamma = D + d\mathbb{1}$ be the diagonal matrix such that $D_{uu} = d_u$ where d_u is the degree of vertex u. Then, $\left\|\Gamma^{-1/2} A \Gamma^{-1/2}\right\|_2 < \frac{2n^{1/\ell}}{\sqrt{d}}$.

We now complete the proof of Proposition 5.2.1.

Proof of Proposition 5.2.1 by Claim 5.2.2. Suppose G has no cycle of length $\leqslant \ell$, then Claim 5.2.2 implies that $A \prec \frac{2n^{1/\ell}}{\sqrt{d}}\Gamma$. Then, the quadratic form $\mathbb{1}^\top A \mathbb{1} < \frac{2n^{1/\ell}}{\sqrt{d}} \operatorname{tr}(\Gamma)$ since $\mathbb{1}^\top \Gamma \mathbb{1} = \operatorname{tr}(\Gamma)$. By definition, $\mathbb{1}^\top A \mathbb{1} = nd$ and $\operatorname{tr}(\Gamma) = \sum_{u=1}^n (d_u + d) = 2nd$. Thus, $n^{1/\ell} > \sqrt{d}/4$, and taking logs, we get

$$\frac{1}{\ell} \log n > \frac{1}{2} \log(d/16) \Rightarrow \frac{\ell}{2} < \log_{d/16} n \,.$$

ℓ is even, so we have $\ell < 2\lceil \log_{d/16} n \rceil$. Thus, by the contrapositive, G must contain a cycle of length $2\lceil \log_{d/16} n \rceil$. This completes the proof. □

We now prove Claim 5.2.2 using the well-known trace moment method, which reduces to counting weighted closed walks in the graph. In the analysis, we will see exactly how the choice of the reweighting matrix Γ accounts for different vertex degrees.

Proof of Claim 5.2.2. Let $\widetilde{A} = \Gamma^{-1/2} A \Gamma^{-1/2}$. For even $\ell \in \mathbb{N}$, the trace moment method states that $\|\widetilde{A}\|_2^\ell \leqslant \operatorname{tr}(\widetilde{A}^\ell) = \operatorname{tr}((\Gamma^{-1}A)^\ell)$, which is a summation of all (weighted) closed walks of length ℓ in G. Since there is no cycle of length $\leqslant \ell$, the only closed walks are the ones that *backtrack* to the original vertex, meaning that there can be at most $\ell/2$ "new" edges and at least $\ell/2$ "old" edges in the walk. We encode each closed walk $u_1 \to u_2 \to \cdots \to u_\ell \to u_1$ as follows,

- Choose a starting vertex $u_1 \in [n]$.
- One bit $b_i \in \{0,1\}$ at each step i to encode whether this step uses a new edge or an old one.
  - If $b_i = 0$ (new edge), select one of u_i's neighbors as u_{i+1}.
  - If $b_i = 1$ (old edge), we must backtrack to the previous vertex u_{i-1}.

For $b \in \{0,1\}$ and $u \in [n]$, let $N_b(u) \subseteq [n]$ be the possible next steps in the walk from u. Then, simply expanding $\mathrm{tr}((\Gamma^{-1}A)^\ell)$, we get

$$\mathrm{tr}((\Gamma^{-1}A)^\ell) = \sum_{b\in\{0,1\}^\ell} \sum_{u_1\in[n]} \sum_{u_2\in N_{b_1}(u_1)} \Gamma^{-1}_{u_1u_1} \sum_{u_3\in N_{b_2}(u_2)} \Gamma^{-1}_{u_2u_2} \cdots \sum_{u_{\ell+1}\in N_{b_\ell}(u_\ell)} \Gamma^{-1}_{u_\ell u_\ell} \cdot \mathbb{1}(u_{\ell+1} = u_1)\,.$$

As we can see, each step $u_i \to u_{i+1}$ gets a factor $\Gamma^{-1}_{u_iu_i} = \frac{1}{d_{u_i}+d}$. We can now bound the above by observing that if $b_i = 0$ (new edge), then $|N_0(u_i)| \leqslant d_{u_i}$ and

$$\sum_{u_{i+1}\in N_0(u_i)} \Gamma^{-1}_{u_iu_i} \leqslant \frac{d_{u_i}}{d_{u_i}+d} < 1\,,$$

and if $b_i = 0$ (old edge), then $|N_1(u_i)| = 1$ (the previous step) and

$$\sum_{u_{i+1}\in N_1(u_i)} \Gamma^{-1}_{u_iu_i} \leqslant \frac{1}{d_{u_i}+d} < \frac{1}{d}\,.$$

Finally, considering $b \in \{0,1\}^\ell$, $u_1 \in [n]$, and there are at least $\ell/2$ old edges, we have

$$\mathrm{tr}((\Gamma^{-1}A)^\ell) < 2^\ell n \left(\frac{1}{d}\right)^{\ell/2}\,,$$

and taking the ℓ-th root completes the proof. □

5.2.2 The case of even arity hypergraphs

In this section, we prove the existence of small even covers in even arity hypergraphs.

Theorem 5.2.3 (Theorem 5.1.2, even k)**.** *For even $k \in \mathbb{N}$ and any $r \in \mathbb{N}$ with $k \leqslant r \leqslant n/8$, any k-uniform hypergraph $\mathcal{H}$ with n vertices and $m \geqslant 128n \log n \cdot (\frac{n}{r})^{k/2-1}$ hyperedges has an even cover of size at most $\lceil r \log_2 n \rceil + 1$.*

The proof is simple and almost identical to the proof of the weak Moore bound (Proposition 5.2.1) but with A being the adjacency matrix of the *Kikuchi graph* which we define below.

Definition 5.2.4 (Kikuchi graph)**.** Let $\mathcal{H}$ be a k-uniform hypergraph on vertex set $[n]$ for even k. For an integer parameter r, define the *Kikuchi graph* K_r associated to $\mathcal{H}$ is a graph on vertex set $\binom{[n]}{r}$ such that a pair of vertices $S, T \in \binom{[n]}{r}$ have an edge between them if the symmetric difference $S \oplus T \in \mathcal{H}$. For such an edge, we write $S \overset{C}{\leftrightarrow} T$ and think of the edge as "colored" by $C \in \mathcal{H}$ where $C = S \oplus T$. We call the adjacency matrix A of K_r the *Kikuchi matrix.*

The key insight of [GKM21] (and also our starting point) is relating even covers in $\mathcal{H}$ to cycles in the associated Kikuchi graph. For sets $R_1, R_2, \ldots, R_\ell \subseteq [n]$ let $\oplus_{i \leqslant \ell} R_i$ denote the set of elements of $[n]$ that appear in an odd number of R_is (i.e., the sum modulo 2 of the indicator vectors of R_is).

Observation 5.2.5 (Closed walks in the Kikuchi graph)**.** Let $\mathcal{H}$ be a k-uniform hypergraph on $[n]$ for even k and let $S_1 \to S_2 \to \cdots S_\ell \to S_1$ be a closed walk on vertices in K_r such that for every $i \leqslant \ell$, $S_i \overset{C_i}{\leftrightarrow} S_{i+1}$ for $C_1, C_2, \ldots, C_\ell \in \mathcal{H}$ (denoting $S_{\ell+1} = S_1$). Then, $\oplus_{i \leqslant \ell} C_i = 0$. Further, if $\mathcal{H}$ has no even cover of length ℓ, then every hyperedge in $\mathcal{H}$ appears an even number of times in the multiset $\{C_1, C_2, \ldots, C_\ell\}$. We will call such walks in K_r *trivial*.

Proof. Note that $S_i \oplus S_{i+1} = C_i$ for every $i \leqslant \ell$. If we add both sides of all ℓ such equalities then each S_i occurs in exactly two of the equations so the LHS must be 0. Thus, $\oplus_{i \leqslant \ell} C_i = 0$.

Next, we repeatedly remove hyperedges that occur an even number of times in the multiset $\{C_1, C_2, \ldots, C_\ell\}$ to obtain a collection of $\ell' \leqslant \ell$ distinct hyperedges of $\mathcal{H}$. The sum (modulo 2) of the remaining hyperedge should still be 0 as we removed hyperedges in pairs. The resulting ℓ' must be 0 as otherwise the remaining hyperedges form an even cover of length $\ell' \leqslant \ell$. □

Consider a hypergraph $\mathcal{H}$ with n vertices and m hyperedges, and its associated Kikuchi graph (V, E) with parameter r. Each $C \in \mathcal{H}$ introduces $\frac{1}{2}\binom{k}{k/2}\binom{n-k}{r-k/2}$ edges

in the Kikuchi graph (select $k/2$ vertices from C and select $r - k/2$ vertices from $[n] \setminus C$ to complete S), thus the total edges $|E| = \frac{1}{2}\binom{k}{k/2}\binom{n-k}{r-k/2} \cdot m$. Let d_S be the degree of $S \in V$, and let d denote the average degree, then simple calculations show that

$$d = \frac{\binom{k}{k/2}\binom{n-k}{r-k/2}m}{\binom{n}{r}} \geqslant \left(\frac{r}{n}\right)^{k/2} m \cdot \binom{k}{k/2}\left(1 - \frac{2r}{n}\right)^{k/2}\left(1 - \frac{k}{2r}\right)^{k/2} \geqslant \frac{1}{2}\left(\frac{r}{n}\right)^{k/2} m \tag{5.1}$$

when $k \leqslant r \leqslant n/8$.

We will follow the reweighting strategy with $\Gamma = D + d\mathbb{1}$ to bound the spectral norm of the reweighted Kikuchi matrix. The following lemma is analogous to Claim 5.2.2.

Lemma 5.2.6. *Let $k, r, n \in \mathbb{N}$ such that $k \leqslant r \leqslant n$, and let $\ell \in \mathbb{N}$ be even. Let A be the Kikuchi matrix with parameter r of a k-uniform hypergraph $\mathcal{H}$ on n vertices, and let $\Gamma = D + d\mathbb{1}$ where D is the degree matrix and d is the average degree of the Kikuchi graph. Suppose there is no even cover of size at most ℓ in $\mathcal{H}$, then*

$$\left\|\Gamma^{-1/2} A \Gamma^{-1/2}\right\|_2 < 2n^{r/\ell}\sqrt{\frac{\ell}{d}}\,.$$

We can immediately complete the proof of Theorem 5.2.3.

Proof of Theorem 5.2.3 by Lemma 5.2.6. Suppose that there is no even cover of size $\leqslant \ell := \lceil r \log_2 n \rceil$ (assume this is even, otherwise add 1). Then, $n^{r/\ell} \leqslant 2$ and Lemma 5.2.6 states that the Kikuchi graph (V, E) satisfies $A \prec 4\sqrt{\ell/d} \cdot \Gamma$ where $\Gamma = D + d\mathbb{1}$. Then,

$$\mathbb{1}^\top A \mathbb{1} < 4\sqrt{\frac{\ell}{d}} \cdot \operatorname{tr}(\Gamma) = 4\sqrt{\frac{\ell}{d}} \cdot \sum_{S \in V}(d_S + d) = 8\sqrt{\frac{\ell}{d}} \cdot |V| d\,.$$

On the other hand, $\mathbb{1}^\top A \mathbb{1} = 2|E| = |V|d$. Thus, we have $d < 64\ell$. By (5.1) we have $d \geqslant \frac{1}{2}(\frac{r}{n})^{k/2} m$ when $k \leqslant r \leqslant n/8$. Thus, if there is no even cover of size $\leqslant \ell$, then $m < 128 n \log n \cdot (\frac{n}{r})^{k/2-1}$, completing the proof. □

Now, we prove Lemma 5.2.6 by counting weighted closed walks in the Kikuchi graph, essentially the same way we prove Claim 5.2.2.

Proof of Lemma 5.2.6. Let $\widetilde{A} = \Gamma^{-1/2} A \Gamma^{-1/2}$. We use the trace power method:

$$\|\widetilde{A}\|_2^\ell \leqslant \operatorname{tr}(\widetilde{A}^\ell) = \operatorname{tr}((\Gamma^{-1}A)^\ell)\,.$$

We upper bound $\operatorname{tr}((\Gamma^{-1}A)^\ell)$ by counting (weighted) closed walks of length ℓ in the Kikuchi graph. Note that each edge (S,T) of the Kikuchi graph corresponds to a hyperedge $S \oplus T \in \mathcal{H}$. Since there is no even covers of size at most ℓ, any closed walk must contain an even number of each hyperedge in $\mathcal{H}$.

We can encode a closed walk $S_1 \to S_2 \to \cdots \to S_\ell \to S_1$ as follows:

- Choose a starting vertex $S_1 \in V$.
- One bit $b_i \in \{0,1\}$ at each step i to encode whether this step uses a new hyperedge or an old one.
  - If $b_i = 0$ (new hyperedge), select one of S_i's neighbors as S_{i+1}.
  - If $b_i = 1$ (old hyperedge), select an old hyperedge C from the previous steps, and set $S_{i+1} = S_i \oplus C$.

Note that there are at most $\ell/2$ new hyperedges and at least $\ell/2$ old hyperedges since each hyperedge must occur an even number of times. For $b \in \{0,1\}$ and $S \in V$, let $N_b(S) \subseteq V$ be the possible next steps in the walk from S (according to b). Each step $S_i \to S_{i+1}$ gets a factor $(\Gamma^{-1}A)_{S_i,S_{i+1}} = \Gamma^{-1}_{S_i,S_i} = \frac{1}{d_{S_i}+d}$. Thus,

$$\operatorname{tr}((\Gamma^{-1}A)^\ell) = \sum_{b\in\{0,1\}^\ell} \sum_{S_1\in V} \sum_{S_2\in N_{b_1}(S_1)} \frac{1}{d_{S_1}+d} \sum_{S_3\in N_{b_2}(S_2)} \frac{1}{d_{S_2}+d} \cdots \sum_{S_{\ell+1}\in N_{b_\ell}(S_\ell)} \frac{\mathbb{1}(S_{\ell+1}=S_1)}{d_{S_\ell}+d}\,.$$

We can upper bound the above as follows. If $b = 0$, then $|N_0(S_i)| \leqslant d_{S_i}$ and $\sum_{S_{i+1}\in N_0(S_i)} \Gamma^{-1}_{S_iS_i} \leqslant \frac{d_{S_i}}{d_{S_i}+d} < 1$. If $b = 1$, then $|N_1(S_i)| \leqslant \ell$ as there are only ℓ options to choose one of the previous steps, and $\sum_{S_{i+1}\in N_1(S_i)} \Gamma^{-1}_{S_iS_i} \leqslant \frac{\ell}{d_{S_i}+d} < \frac{\ell}{d}$. Furthermore, we can assume that $\ell \leqslant d$, otherwise we can simply treat all steps as new hyperedges.

Finally, $b \in \{0,1\}^\ell$, there are $|V| = \binom{n}{r}$ choices for the starting vertex S_1, and there are at least $\ell/2$ old hyperedges. Thus, we have

$$\operatorname{tr}((\Gamma^{-1}A)^\ell) < 2^\ell \binom{n}{r} \left(\frac{\ell}{d}\right)^{\ell/2} \leqslant 2^\ell n^r \left(\frac{\ell}{d}\right)^{\ell/2}\,.$$

Taking the ℓ-th root completes the proof. □

Summary. As the above short and simple arguments illustrate, reweighted Kikuchi matrix appears to be a clean and simple way to handle irregularities in the degree of graphs (Kikuchi or otherwise) in spectral double counting. For $k > 2$, we suspect that the extra $\log n$ factor incurred in our analysis can likely be removed by a better counting of the weighted closed walks.

5.2.3 Overview of the odd arity case

As in many previous works on refuting constraint satisfaction problems, the odd arity case requires significantly more work. Indeed, even the definition of the Kikuchi graph (Definition 5.2.4) only makes sense when k is even. We present the proof of the odd arity case in Section 5.3, and here we outline some of our key ideas.

Bipartite hypergraph. The main insight is to transform the hypergraph $\mathcal{H}$ to a "bipartite" hypergraph (this abstraction is closely related to the Cauchy-Schwarz trick in the context of odd-arity CSP refutation). First, we partition the hyperedges of $\mathcal{H}$ into $\mathcal{H}_1, \dots, \mathcal{H}_p$ such that for each $\mathcal{H}_i$, all hyperedges in $\mathcal{H}_i$ contains a "center" vertex $u_i \in [n]$. We denote $\widetilde{\mathcal{H}}_i$ to be $\{C \setminus \{u_i\} : C \in \mathcal{H}_i\}$, i.e. removing the center vertex, and denote $\widetilde{C} := C \setminus \{u_i\}$. Then, we construct a $2(k-1)$-uniform hypergraph as follows: for each $i \in [p]$ and each distinct pair $C, C' \in \mathcal{H}_i$, we add a hyperedge $\widetilde{C} \oplus \widetilde{C}'$ (let's assume $\widetilde{C}, \widetilde{C}'$ are disjoint for now).

Let's make some quick calculations. Suppose that $\mathcal{H}$ has m hyperedges, and suppose that there are roughly $p \approx n$ partitions and each partition size is $\approx m/n$. Then, the new $2(k-1)$-uniform hypergraphs will contain roughly $n \cdot (m/n)^2 = \frac{m^2}{n}$ hyperedges. Now, since $2(k-1)$ is even, we can apply our bound for the even case: there is an even cover of size $r \log n$ when $\frac{m^2}{n} \geqslant \widetilde{O}(n)(\frac{n}{r})^{(k-1)-1}$, meaning $m \geqslant \widetilde{O}(n)(\frac{n}{r})^{\frac{k}{2}-1}$, the correct bound!

The issue is that this new hypergraph has small even covers for trivial reasons: any 3 pairs (C_1, C_2), (C_2, C_3) and (C_3, C_1) from $\mathcal{H}_i$ form an even cover of size 3. Nevertheless, we can proceed to analyze the Kikuchi matrix of the new hypergraph (Definition 5.3.6), assuming that there is no small even cover in the *original* hypergraph. Note that now an edge (S, T) is associated with 2 hyperedges C, C' from the

same $\mathcal{H}_i$, which we denote as $S \overset{C,C'}{\longleftrightarrow} T$. Assuming that there is no even cover of size $\leqslant 2\ell$, we bound the number of "trivial" closed walks where each hyperedge is used an even number of times.

Encoding a closed walk. The standard technique of bounding counts of closed walks in the trace moment method is to give a small *encoding* of a walk. In our case, in a length-ℓ closed walk of the Kikuchi graph, each step is associated with two hyperedges, and we have two types of steps:

1. a step using 2 *new* hyperedges, and
2. a step using at least 1 *old* hyperedge.

The first type is bounded exactly the same way as the even arity case by our weight matrix Γ, the trouble is the second type: while we can easily encode one edge in the step, we need too many bits to encode the other edge.

Deleting bad edges of the Kikuchi graph. The main insight is that in the end, we only care about bounding $\mathbb{1}^\top A \mathbb{1}$. Again, let d be the average degree of the Kikuchi graph, $A \in \mathbb{R}^{N \times N}$ be the Kikuchi matrix, and $\Gamma = D + d\mathbb{1}$ be our diagonal weight matrix. If we delete (say) half of the edges of the Kikuchi graph such that we have $\|\Gamma^{-1/2} A' \Gamma^{-1/2}\|_2 \leqslant \lambda(d)$, where A' is the modified Kikuchi matrix, for some "good enough" $\lambda(d)$, then we will have $\frac{Nd}{2} \leqslant \mathbb{1}^\top A' \mathbb{1} \leqslant \lambda \operatorname{tr}(\Gamma) = \lambda \cdot 2Nd$, essentially only losing a constant factor in the density.

We define an appropriate edge deletion process, prove that the fraction of edges removed is small (Claim 5.3.10), and show that the resulting subgraph has combinatorial properties that let us encode steps of the second type efficiently (Lemma 5.3.9).

Improving the row pruning step of [GKM21]. The analysis of [GKM21] also requires reducing the Kikuchi graph to obtain certain combinatorial properties. However, instead of deleting "bad" edges, they delete "bad" vertices, which they defined as vertices that are bad for *some* i in the bipartite hypergraph (they call this row pruning as each row of the Kikuchi matrix corresponds to a vertex). Crucially, doing so requires a union bound over i, hence they need a strong bound on fraction of bad vertices for each i. Furthermore, they proved their bound using tail

inequalities for low-degree polynomials by Schudy and Sviridenko [SS12], which is a powerful black-box concentration inequality but loses log factors and requires involved analysis. All this combined with the row bucketing step introduces several log factors.

Hyperedges with large intersections. It turns out that the fraction of bad edges highly depends on *large intersections* of hyperedges in $\mathcal{H}$. To bound the fraction of edges deleted, we require our hypergraph to be somewhat "regular" – that is, no small subset appears in more than an appropriately chosen threshold of hyperedges in $\mathcal{H}$. To this end, we invoke the hypergraph regularity decomposition of [GKM21] (with more transparently chosen thresholds that do not involve carefully chosen logarithmic factors) to decompose the hypergraph into at most k subhypergraphs such that each piece satisfies the required regularity conditions (see Algorithm 5.3.2 and Observation 5.3.4 & 5.3.3). Then, there must be one subhypergraph $\mathcal{H}^{(i)}$ with at least m/k hyperedges, and we will show that there exists an even cover within $\mathcal{H}^{(i)}$.

5.3 Hypergraph Moore bound for odd arity hypergraphs

In this section we prove the hypergraph Moore bound for k-uniform hypergraphs when k is odd.

Theorem 5.3.1 (Theorem 5.1.2, odd k)**.** *There is a universal constant B such that for any odd $k \in \mathbb{N}$, and any $r \in \mathbb{N}$ satisfying $2k \leqslant r \leqslant \frac{n}{B^k}$, any k-uniform hypergraph $\mathcal{H}$ with n vertices and $m \geqslant B^k n \log n \cdot \left(\frac{n}{r}\right)^{k/2-1}$ hyperedges has an even cover of size at most $r \log_2 n$.*

Our proof strategy broadly involves the following steps.

- **Hypergraph decomposition.** We partition $\mathcal{H}$ into subhypergraphs $\mathcal{H}^{(0)}, \ldots, \mathcal{H}^{(k-1)}$ with the property that every size-$(i+1)$ set in $\mathcal{H}^{(i)}$ is contained in only a small number of clauses, and every clause in $\mathcal{H}^{(i)}$ intersects many other clauses at a size-i set. One of the $\mathcal{H}^{(i)}$ must contain at least m/k clauses, and we find an even cover in that $\mathcal{H}^{(i)}$.

- **Large i.** When $i \geqslant \frac{k+1}{2}$, we give a direct reduction to the hypergraph Moore bound for even arity hypergraphs and apply Theorem 5.2.3.
- **Kikuchi graph.** To handle the remaining values of i, we show the existence of an even cover by proving the contrapositive — a hypergraph with no small even covers has a bounded number of hyperedges. To achieve this, we appropriately define the Kikuchi graph for odd arity hypergraphs, and show that the adjacency matrix $\widehat{A}$ of some suitably chosen subgraph (via the "edge deletion process" described below) satisfies $\widehat{A} \preceq Q$ for some diagonal matrix Q. Then the resulting inequality $\mathbb{1}^\top \widehat{A} \mathbb{1} \leqslant \mathrm{tr}(Q)$ can be rearranged to bound the number of hyperedges.
- **Trace method.** The way we prove $\widehat{A} \preceq Q$ is by using the trace moment method to show $\left\| Q^{-1/2} \widehat{A} Q^{-1/2} \right\|_2 \leqslant 1$. Bounding a high trace power of $Q^{-1/2} \widehat{A} Q^{-1/2}$ corresponds to bounding the total weight of closed walks that use every hyperedge an even number of times in the Kikuchi graph.
- **Edge deletion process.** We delete a small fraction of the edges in K_r with the guarantee that in the resulting subgraph any clause participates in only a small number of incident edges to every vertex.

Hypergraph decomposition. We describe our algorithm to partition our hypergraph.

Algorithm 5.3.2. We partition $\mathcal{H}$ into hypergraphs $\mathcal{H}^{(0)}, \ldots, \mathcal{H}^{(k-1)}$ via the following algorithm.

1. Set $t = k - 1$ and $\mathcal{H}_{\text{current}} := \mathcal{H}$.
2. Set counter $s = 1$. While there is $U \subseteq [n]$ such that $|U| = t$ and
$$|\{C \in \mathcal{H}_{\text{current}} : U \subseteq C\}| \geqslant \max\left\{2, \left(\frac{n}{r}\right)^{\frac{k}{2}-t}\right\}:$$
   a) Choose U satisfying the condition and let $\mathcal{H}_s^{(t)}$ be a subset of $\{C \in \mathcal{H}_{\text{current}} : U \subseteq C\}$ of size $\max\left\{2, \left(\frac{n}{r}\right)^{\frac{k}{2}-t}\right\}$.
   b) Add all clauses in $\mathcal{H}_s^{(t)}$ to $\mathcal{H}^{(t)}$.

c) Delete all clauses in $\mathcal{H}_s^{(t)}$ to $\mathcal{H}_{\text{current}}$.

d) Increment s by 1.

3. Decrement t by 1. If $t > 0$, go back to step 2; otherwise take the remaining clauses in $\mathcal{H}_{\text{current}}$ and add them to $\mathcal{H}^{(0)}$.

First, observe that the largest subhypergraph $\mathcal{H}^{(i)}$ in the partition produced by our algorithm must have at least $\frac{m}{k}$ hyperedges. Next, observe that $i \neq 0$ because if $|\mathcal{H}^{(0)}| \geqslant m/k$, then there must be a $j \in [n]$ such that $\left|\{C \in \mathcal{H}^{(0)} : j \in C\}\right| \geqslant \frac{m}{nk} \gg (\frac{n}{r})^{k/2-1}$, which would have been added to $\mathcal{H}^{(1)}$. Our goal in the rest of the proof is to find a small even cover in $\mathcal{H}^{(i)}$. The following observations articulate the properties of $\mathcal{H}^{(i)}$ we need that are guaranteed by the algorithm.

Observation 5.3.3. $\mathcal{H}^{(i)}$ can be partitioned into $\mathcal{H}_1^{(i)}, \dots, \mathcal{H}_p^{(i)}$ where for each $j \in [p]$, there is a set U_j of size i such that every $C \in \mathcal{H}_j^{(i)}$ contains U_j, and $|\mathcal{H}_j^{(i)}| \geqslant \left(\frac{n}{r}\right)^{\frac{k}{2}-i}$ and $p \leqslant m \cdot \left(\frac{r}{n}\right)^{\frac{k}{2}-i}$.

Observation 5.3.4. For $s \geqslant 1$ and any $U \subseteq [n]$ such that $|U| = i + s$, the number of hyperedges in $\mathcal{H}^{(i)}$ containing U is at most $\max\left\{1, \left(\frac{n}{r}\right)^{\frac{k}{2}-s-i}\right\}$, otherwise they would have been added to $\mathcal{H}^{(i+s)}$.

Reduction to even arity case when $i \geqslant \frac{k+1}{2}$. In this case, by Observation 5.3.4, each pair $C \neq C'$ in any $\mathcal{H}_j^{(i)}$ must satisfy $C \cap C' = U_j$. The following makes the reduction from finding even covers in $\mathcal{H}^{(i)}$ if $i \geqslant \frac{k+1}{2}$ to the even arity case concrete.

Lemma 5.3.5. *Let $\mathcal{H}$ be a k-uniform hypergraph on n vertices with no even cover of size $r \log_2 n$. Fix $1 \leqslant i \leqslant k-1$. Suppose $\mathcal{H}_1, \dots, \mathcal{H}_p$ are disjoint subsets of $\mathcal{H}$ such that for each $j \in [p]$, $|\mathcal{H}_j| \geqslant 2$ and all pairs of hyperedges $C \neq C' \in \mathcal{H}_j$ satisfy $C \cap C' = U_j$ for some $U_j \subseteq [n]$ of size i. Then,*

$$\sum_{j=1}^{p} |\mathcal{H}_j| \leqslant O(n \log n) \left(\frac{2n}{r}\right)^{k-i-1}.$$

In particular, when $i \geqslant \frac{k+1}{2}$ the above is at most $O(n \log n) \cdot \left(\frac{n}{r}\right)^{k/2-1}$.

Proof. Given such disjoint subsets $\mathcal{H}_1, \dots, \mathcal{H}_p$, we can construct a $2(k-i)$-uniform hypergraph $\widehat{\mathcal{H}}$ by the following: for each $j \in [p]$, arbitrarily order the edges: $\mathcal{H}_j = (C_1, \dots, C_{|\mathcal{H}_j|})$. Then, add the hyperedge $C_s \oplus C_{s+1}$ to $\widehat{\mathcal{H}}$ for $s = 1, \dots, |\mathcal{H}_j| - 1$. By assumption $|C_s \cap C_{s+1}| = |U_j| = i$, thus $|C_s \oplus C_{s+1}| = 2(k-i)$. The resulting $\widehat{\mathcal{H}}$ has

$$|\widehat{\mathcal{H}}| = \sum_{j=1}^{p} |\mathcal{H}_j| - 1 \geqslant \frac{1}{2} \sum_{j=1}^{p} |\mathcal{H}_j|$$

hyperedges, since $|\mathcal{H}_j| \geqslant 2$ for all $j \in [p]$.

We claim that $\widehat{\mathcal{H}}$ cannot have an even cover of size at most $\frac{r}{2} \log_2 n$. First, if $\widehat{\mathcal{H}}$ has repeated hyperedges, then there must exist $j \neq j' \in [p]$ and $C_1, C_2 \in \mathcal{H}_j$, $C_1', C_2' \in \mathcal{H}_{j'}$ such that $C_1 \oplus C_2 = C_1' \oplus C_2'$, but then $\{C_1, C_2, C_1', C_2'\}$ would be an even cover of size 4 in $\mathcal{H}$. Now, suppose $\widehat{\mathcal{H}}$ has no repeated edges but has an even cover of size ℓ. Then, for any $\widehat{C}$ in the even cover, we can uniquely identify $j \in [p]$ and $s \leqslant |\mathcal{H}_j| - 1$ such that $C_s, C_{s+1} \in \mathcal{H}_j$ and $\widehat{C} = C_s \oplus C_{s+1}$. Furthermore, by construction there must be at least two $C_s, C_{s'} \in \mathcal{H}_j$ that each occurs only once. Therefore, these edges must form an even cover of size at most 2ℓ in $\mathcal{H}$.

Since $2(k-i)$ is even and $\widehat{\mathcal{H}}$ has no even cover of size $\frac{r}{2} \log_2 n$, we can apply Theorem 5.2.3 to show that

$$|\widehat{\mathcal{H}}| \leqslant O(n \log n) \left(\frac{2n}{r}\right)^{k-i-1}.$$

This completes the proof. □

Henceforth, we assume $i \leqslant \frac{k-1}{2}$, which is the case we need an appropriate Kikuchi graph for odd arity hypergraphs.

Kikuchi matrix for odd arity hypergraphs. The following is the same Kikuchi graph defined in [GKM21, Definition 6.2].

Definition 5.3.6 (Colored Kikuchi graphs and subgraphs)**.** Fix $r \in \mathbb{N}$ and $t \in \{1, \dots, k-1\}$ such that $2k \leqslant r \leqslant n$. Let $\mathcal{H}_1, \dots, \mathcal{H}_p$ be p disjoint sets of hyperedges such that for each $i \in [p]$, all hyperedges in $\mathcal{H}_i$ have a common subset $U_i \subset [n]$ where $|U_i| = t$. For each $C \in \mathcal{H}_i$, denote $\widetilde{C} \coloneqq C \setminus U_i$, and denote $\widetilde{\mathcal{H}}_i \coloneqq \{\widetilde{C} : C \in \mathcal{H}_i\}$ which can be viewed as a $(k-t)$-uniform hypergraph. We define the *colored* Kikuchi graph K_r as follows.

The vertex set $V(K_r)$ consists of subsets of $[n] \times [2]$ of size r, where $S \in V$ is viewed as $(S^{(1)}, S^{(2)})$ where $S^{(1)}, S^{(2)} \subseteq [n]$ are colored *green* and *blue* respectively. For each $i \in [p]$ and each $C \neq C' \in \mathcal{H}_i$, let $\widetilde{C}^{(1)}$ be $\widetilde{C}$ colored green and $\widetilde{C}'^{(2)}$ be $\widetilde{C}'$ colored blue, and we add an edge between $S, T \in V$, denoted $S \overset{C,C'}{\longleftrightarrow} T$, if $S \oplus T = \widetilde{C}^{(1)} \oplus \widetilde{C}'^{(2)}$ and if one of the following holds,

- $|\widetilde{C} \cap S^{(1)}| = |\widetilde{C}' \cap T^{(2)}| = \left\lceil \frac{k-t}{2} \right\rceil$ and $|\widetilde{C}' \cap S^{(2)}| = |\widetilde{C} \cap T^{(1)}| = \left\lfloor \frac{k-t}{2} \right\rfloor$, or
- $|\widetilde{C} \cap S^{(1)}| = |\widetilde{C}' \cap T^{(2)}| = \left\lfloor \frac{k-t}{2} \right\rfloor$ and $|\widetilde{C}' \cap S^{(2)}| = |\widetilde{C} \cap T^{(1)}| = \left\lceil \frac{k-t}{2} \right\rceil$, or

Figure 5.1 shows an example of two edges $C, C' \in \mathcal{H}_i$ forming an edge (S, T) in the Kikuchi graph.

We say that the edge (S, T) is type-i, and for $S \in V$, we define the type-i degree as

$$d_{S,i} \coloneqq \left| \left\{ C \in \mathcal{H}_i : \left| \widetilde{C} \cap S^{(1)} \right| \text{ or } |\widetilde{C} \cap S^{(2)}| \in \left\{ \left\lceil \frac{k-t}{2} \right\rceil, \left\lfloor \frac{k-t}{2} \right\rfloor \right\} \right\} \right| .$$

We call any subgraph of the colored Kikuchi graph as a *colored Kikuchi subgraph*.

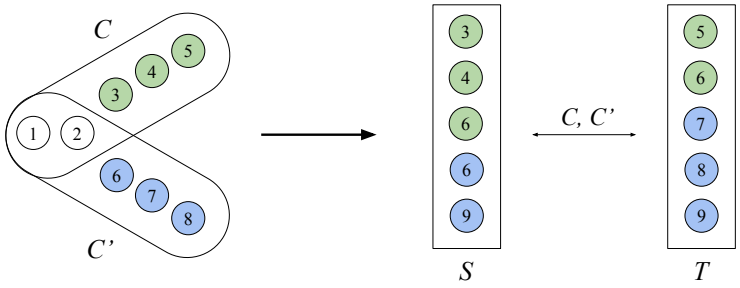


Figure 5.1: An example of Definition 5.3.6 with $k = 5$ and $t = 2$. On the left are two 5-uniform hyperedges in $\mathcal{H}_i$ with common intersection $U_i = \{1, 2\}$ and $\widetilde{C} = \{3, 4, 5\}$, $\widetilde{C}' = \{6, 7, 8\}$. On the right, S and T are vertices in the Kikuchi graph where $S^{(1)} = \{3, 4, 6\}$, $T^{(1)} = \{5, 6\}$ are colored green, and $S^{(2)} = \{6, 9\}$, $T^{(2)} = \{7, 8, 9\}$ are colored blue. C and C' form an edge between S, T because $|\widetilde{C} \cap S^{(1)}| = 2$, $|\widetilde{C} \cap T^{(1)}| = 1$, $|\widetilde{C}' \cap S^{(2)}| = 1$, and $|\widetilde{C}' \cap T^{(2)}| = 2$.

Remark 5.3.7 (Purpose of coloring)**.** The coloring in Definition 5.3.6 is needed because $C \neq C' \in \mathcal{H}_i$ may have intersection larger than t, meaning $|C \oplus C'| =$

$|\widetilde{C} \oplus \widetilde{C}'| < 2(k-t)$, making the analysis complicated. Coloring $\widetilde{C}, \widetilde{C}'$ with different colors automatically makes $\widetilde{C}^{(1)}, \widetilde{C}'^{(2)}$ disjoint, i.e. $|S \oplus T| = |\widetilde{C}^{(1)} \oplus \widetilde{C}'^{(2)}| = 2(k-t)$. Note also that a vertex $S \subseteq [n] \times [2]$ may contain two copies of some element in $[n]$ with different colors, as shown in Figure 5.1.

Observation 5.3.8 (Parameters of the Kikuchi graph)**.** The Kikuchi graph (V, E) defined in Definition 5.3.6 has $|V| = \binom{2n}{r}$, and each distinct pair $C, C' \in \mathcal{H}_i$ contributes a collection of edges $E_{C,C'}$ in E, where

$$|E_{C,C'}| = \alpha_t := \binom{k-t}{\lfloor \frac{k-t}{2} \rfloor} \binom{k-t}{\lceil \frac{k-t}{2} \rceil} \binom{2n-2(k-t)}{r-(k-t)} \cdot 2^{\mathbb{1}(k-t \text{ is odd})}$$

by first choosing $\widetilde{C} \cap S^{(1)}$, $\widetilde{C}' \cap S^{(2)}$ (or $\widetilde{C} \cap S^{(2)}$, $\widetilde{C}' \cap S^{(1)}$) and completing S's remaining $r-(k-t)$ elements. Thus, $|E| = \sum_{i=1}^{p} \binom{|\mathcal{H}_i|}{2} \cdot \alpha_t$, and standard calculations show that when $2k \leqslant r \leqslant n/8$, the average degree $d = \frac{2|E|}{|V|}$ satisfies

$$\left(\frac{r}{2n}\right)^{k-t} \sum_{i=1}^{p} \binom{|\mathcal{H}_i|}{2} \leqslant d \leqslant 2^{2k} \left(\frac{r}{2n}\right)^{k-t} \sum_{i=1}^{p} \binom{|\mathcal{H}_i|}{2}.$$

Our ideal hope is that the adjacency matrix A of the Kikuchi graph, constructed from $\mathcal{H}^{(i)} = (\mathcal{H}_1^{(i)}, \ldots, \mathcal{H}_p^{(i)})$, is bounded in the PSD order by some low-trace diagonal matrix Q. To achieve this, we prove the following lemma analogous to Lemma 5.2.6, but with the additional requirement that $d_{S,i}$ is small for all $S \in V(K_r)$ and $i \in [p]$. The proof is almost identical to the proof of Lemma 5.2.6 but the encoding for an "old hyperedge" step is different.

Lemma 5.3.9. *Let $r \geqslant 2k$. Given disjoint hyperedges $\mathcal{H}_1, \ldots, \mathcal{H}_p$, let $\widehat{A}$ be the adjacency matrix of any colored Kikuchi subgraph $\widehat{K}_r$ as defined in Definition 5.3.6, and let $\Gamma = D + d\mathbb{1}$ where D is the degree matrix and d is the average degree of G. Fix $\eta \in \mathbb{R}$ and let $\ell \in \mathbb{N}$ be even. Suppose there is no even cover of size at most ℓ, and suppose $d_{S,i} \leqslant \eta$ for all $S \in V$ and $i \in [p]$. Then,*

$$\left\| \Gamma^{-1/2} \widehat{A} \Gamma^{-1/2} \right\|_2 \leqslant 2n^{r/\ell} \sqrt{\frac{2\eta\ell}{d}}.$$

Proof. Let $\widetilde{A} = \Gamma^{-1/2} \widehat{A} \Gamma^{-1/2}$. We again use the trace power method:

$$\|\widetilde{A}\|_2^{\ell} \leqslant \operatorname{tr}(\widetilde{A}^{\ell}) = \operatorname{tr}((\Gamma^{-1}A)^{\ell}).$$

Note that each edge (S, T) in $\widehat{A}$ corresponds to two hyperedges of the same type (both from some $\mathcal{H}_i$), one green and one blue, and since there is no even covers of size at most ℓ, any closed walk must contain an even number of each hyperedge.

We encode a closed walk $S_1 \to S_2 \to \cdots \to S_\ell \to S_1$ as follows:

- Starting vertex $S_1 \in V$.
- One bit $b_i \in \{0, 1\}$ at step i to encode whether this step uses two new hyperedges or one (or more) old hyperedge.
  - If $b_i = 0$ (two new hyperedges), select one of S_i's neighbors as S_{i+1}.
  - If $b_i = 1$ (old hyperedge), select an old green (or blue) hyperedge C from the previous steps, and select a blue (or green) hyperedge C' incident to S_i.

Recall that for $b \in \{0, 1\}$, we write $N_b(S)$ as the possible next steps in the walk from S. Using the same analysis as the proof of Lemma 5.2.6, for $b = 0$,

$$\sum_{S_{i+1} \in N_0(S_i)} \frac{1}{d_{S_i} + d} \leqslant 1 ,$$

and for $b = 1$, suppose the old edge is of type $j \in [p]$, then $|N_1(S_i)| \leqslant 2\ell d_{S_i,j}$ (one previous step, 2 colors), thus

$$\sum_{S_{i+1} \in N_b(S_i)} \frac{1}{d_{S_i} + d} \leqslant \frac{2\ell d_{S_i,j}}{d_{S_i} + d} \leqslant \frac{2\eta\ell}{d} .$$

We can assume that $2\eta\ell \leqslant d$, otherwise we can simply treat all steps as new hyperedges.

There are $\binom{2n}{r} \leqslant (\frac{2en}{r})^r \leqslant n^r$ (since $r \geqslant 2k$ and $k \geqslant 3$) choices to pick the starting vertex S_1. Furthermore, there can be at most $\ell/2$ steps that use two new hyperedges, i.e. $|b| \geqslant \ell/2$, thus

$$\operatorname{tr}((\Gamma^{-1}\widehat{A})^\ell) \leqslant n^r \sum_{b \in \{0,1\}^\ell} \left(\frac{2\eta\ell}{d}\right)^{|b|} \leqslant 2^\ell n^r \left(\frac{2\eta\ell}{d}\right)^{\ell/2} .$$

Taking the ℓ-th root completes the proof. □

Construction of colored Kikuchi subgraph. Unfortunately, the requirement for all $d_{S,i}$ to be bounded by a small η prohibits us from obtaining a good bound on the adjacency matrix of the full colored Kikuchi graph K_r using Lemma 5.3.9. This motivates dropping a small number of edges from K_r, and bounding the adjacency matrix $\widehat{A}$ of the resulting subgraph $\widehat{K}_r$ instead. Thus, we proceed with identifying a suitable colored Kikuchi subgraph $\widehat{K}_r$ of $\mathcal{H}^{(i)}$ with adjacency matrix $\widehat{A}$ via the following *edge deletion process*:

> Start with the colored Kikuchi graph K_r, and delete every edge $\{S, T\}$ caused by a pair of clauses C, C' such that S or T has strictly more than 1 edge that C or C' participates in.

To obtain a handle on the average degree of $\widehat{K}_r$, we first show that the number of edges of K_r we delete to obtain $\widehat{K}_r$ is only a small fraction of the total number of edges, and then the desired lower bound follows from a lower bound on $|E(K_r)|$.

Analyzing the edge deletion process. We find it convenient to think of the fraction of deleted edges as the *probability that a uniformly random edge in K_r is absent in $\widehat{K}_r$*. With this probabilistic interpretation in hand, observe that a uniformly random edge in K_r is the same as choosing a uniformly random pair of clauses (C, C') such that C and C' both belong to the same $\mathcal{H}^{(i)}_j$ and then choosing a random edge $\{S, T\}$ in $E_{C,C'}$, the collection of edges adorned by (C, C'). We will use the notation $C'' \to_C S$ to mean $|\widetilde{C}'' \cap S| = |\widetilde{C} \cap S|$, where we recall from Definition 5.3.6 that $\widetilde{C} := C \setminus U_j$ with U_j being the size-i common intersection of $\mathcal{H}^{(i)}_j$. We then show the following.

Claim 5.3.10 (Deletion probability)**.** For every pair of clauses (C, C') such that C and C' belong to the same $\mathcal{H}^{(i)}_j$ for some $j \in [p]$,

$$\Pr_{\{S,T\} \sim E_{C,C'}}[\{S, T\} \text{ deleted}] \leqslant k \cdot 4^{k+1} \sqrt{\frac{r}{n}}.$$

Proof. Recall that we defined $\widetilde{C} = C \setminus U_j$ and $\widetilde{C}' = C' \setminus U_j$. The distribution of $S = (S^{(1)}, S^{(2)})$ (the green and blue vertices) is uniform on all sets such that:

- $|\widetilde{C} \cap S^{(1)}| = \left\lceil \frac{k-i}{2} \right\rceil$, $|\widetilde{C}' \cap S^{(2)}| = \left\lfloor \frac{k-i}{2} \right\rfloor$, or

- $|\widetilde{C} \cap S^{(1)}| = \left\lceil \frac{k-i}{2} \right\rceil, |\widetilde{C}' \cap S^{(2)}| = \left\lfloor \frac{k-i}{2} \right\rfloor$.

Then, by union bound,

$$\begin{aligned}
\mathop{\mathbf{Pr}}_{\{S,T\}\sim E_{C,C'}}[\{S,T\} \text{ deleted}] \leqslant & \mathop{\mathbf{Pr}}_{\{S,T\}\sim E_{C,C'}}\left[\exists C'' \to_C S^{(1)} : C'' \in \mathcal{H}_j^{(i)}, C'' \neq C\right] + \\
& \mathop{\mathbf{Pr}}_{\{S,T\}\sim E_{C,C'}}\left[\exists C'' \to_{C'} S^{(2)} : C'' \in \mathcal{H}_j^{(i)}, C'' \neq C'\right] + \\
& \mathop{\mathbf{Pr}}_{\{S,T\}\sim E_{C,C'}}\left[\exists C'' \to_C T^{(1)} : C'' \in \mathcal{H}_j^{(i)}, C'' \neq C\right] + \\
& \mathop{\mathbf{Pr}}_{\{S,T\}\sim E_{C,C'}}\left[\exists C'' \to_{C'} T^{(2)} : C'' \in \mathcal{H}_j^{(i)}, C'' \neq C'\right] \\
= 4 & \mathop{\mathbf{Pr}}_{\{S,T\}\sim E_{C,C'}}\left[\exists C'' \to_C S^{(1)} : C'' \in \mathcal{H}_j^{(i)}, C'' \neq C\right]
\end{aligned}$$

then by Markov's inequality,

$$\begin{aligned}
& \leqslant 4 \mathop{\mathbf{E}}_{\{S,T\}\sim E_{C,C'}}\left|C'' : C'' \to_C S^{(1)}, C'' \in \mathcal{H}_j^{(i)}, C'' \neq C\right| \\
& = 4 \sum_{\substack{C'':C''\in\mathcal{H}_j^{(i)} \\ C''\neq C}} \mathop{\mathbf{Pr}}_{\{S,T\}\sim E_{C,C'}}\left[C'' \to_C S^{(1)}\right] \qquad (5.2)
\end{aligned}$$

Once the intersection of S with $\widetilde{C}$ and $\widetilde{C}'$ is chosen, the remaining elements are selected uniformly at random without replacement. For fixed $C'' \neq C \in \mathcal{H}_j^{(i)}$, since they contain U_j of size i, $|\widetilde{C}'' \cap \widetilde{C}| = |C'' \cap C| - i$, and S must include $\lfloor \frac{k-i}{2} \rfloor - (|C'' \cap C| - i)$ additional elements from $\widetilde{C}'' \setminus \widetilde{C}$ for $C'' \to_C S^{(1)}$ to hold. Thus,

$$\mathop{\mathbf{Pr}}_{\{S,T\}\sim E_{C,C'}}\left[C'' \to_C S^{(1)}\right] \leqslant 2^k \left(\frac{r}{n}\right)^{\left\lfloor \frac{k-i}{2} \right\rfloor - |C''\cap C| + i}.$$

Thus, we can prove:

$$(5.2) \leqslant 4 \cdot 2^k \sum_{s=i}^{k-1} \sum_{\substack{U\subseteq C \\ |U|=s}} \sum_{\substack{C'':C''\in\mathcal{H}_j^{(i)} \\ C''\neq C \\ C''\cap C = U}} \left(\frac{r}{n}\right)^{\left\lfloor \frac{k-i}{2} \right\rfloor - s + i} \qquad (5.3)$$

By Observation 5.3.4, we can bound the above as

$$\leqslant 4 \cdot 2^k \sum_{s=i}^{k-1} \sum_{\substack{U \subseteq C \\ |U|=s}} \left(\frac{n}{r}\right)^{\frac{k}{2}-s} \left(\frac{r}{n}\right)^{\frac{k-i}{2}-\frac{\mathbf{1}[k-i \text{ odd}]}{2}-s+i}$$

$$\leqslant k \cdot 4^{k+1} \sqrt{\frac{r}{n}},$$

as $\frac{i}{2} - \frac{\mathbf{1}[k-i \text{ odd}]}{2} \geqslant \frac{1}{2}$ for all $i \geqslant 1$ when k is odd. □

Lower bound on average degree in A. By choosing B large enough, the upper bound on r, and Claim 5.3.10, the fraction of edges we delete from the original colored Kikuchi graph K_r to obtain $\widehat{K}_r$ is at most .5 and hence $d(\widehat{K}_r) \geqslant .5d(K_r)$ where $d(K_r)$ and $d(\widehat{K}_r)$ are the average degrees in K_r and $\widehat{K}_r$ respectively. Thus, we know:

$$d(K_r) \geqslant \left(\frac{r}{2n}\right)^{k-i} \sum_{j=1}^{p} \binom{|\mathcal{H}_j^{(i)}|}{2} \geqslant \left(\frac{r}{2n}\right)^{k-i} \cdot p \cdot \binom{m/kp}{2} \geqslant \left(\frac{r}{2n}\right)^{k-i} \cdot \frac{m^2}{4k^2 p}$$

where the first inequality uses Observation 5.3.8, and the second inequality is due to Jensen's inequality.

By the upper bound $p \leqslant m \cdot \left(\frac{r}{n}\right)^{\frac{k}{2}-i}$ as noted in Observation 5.3.3:

$$d(K_r) \geqslant \frac{1}{4k^2 2^k} \cdot \left(\frac{r}{n}\right)^{k-i} \cdot \left(\frac{n}{r}\right)^{\frac{k}{2}-i} \cdot m = \frac{1}{4k^2 2^k} \cdot \left(\frac{r}{n}\right)^{\frac{k}{2}} \cdot m.$$

As an upshot, we know:

Claim 5.3.11. $d(\widehat{K}_r) \geqslant \frac{1}{8k^2 2^k} \cdot \left(\frac{r}{n}\right)^{k/2} \cdot m.$

Spectral double counting. With a lower bound on $d(\widehat{K}_r)$ in hand, we are now ready to perform our weighted spectral double counting argument to complete the proof of Theorem 5.3.1.

Proof of Theorem 5.3.1. Recall that our goal is to prove that there is a small even cover in $\mathcal{H}^{(i)}$, the largest piece obtained from the decomposition, and also recall

that if $i \geqslant \frac{k+1}{2}$, then we are done by Lemma 5.3.5. Hence, we assume $i \leqslant \frac{k-1}{2}$ for the rest of the proof.

Suppose there are no even covers in $\mathcal{H}$ of size $\ell = r \log n$, then there are also none in $\mathcal{H}^{(i)}$ from Lemma 5.3.9 we get:

$$\left\| \Gamma^{-1/2} \widehat{A} \Gamma^{-1/2} \right\|_2 \leqslant 4\sqrt{\frac{2\ell}{d(\widehat{K}_r)}}.$$

Thus, $\widehat{A} \preceq 4\sqrt{\frac{2\ell}{d(\widehat{K}_r)}}\Gamma$, and by taking the quadratic form with the all-ones vector, we get:

$$2|E(\widehat{K}_r)| = \mathbb{1}^\top \widehat{A} \mathbb{1} \leqslant 4\sqrt{\frac{2\ell}{d(\widehat{K}_r)}} \cdot \operatorname{tr}(\Gamma) = 16\sqrt{\frac{2\ell}{d(\widehat{K}_r)}} \cdot |E(\widehat{K}_r)|,$$

which implies

$$d(\widehat{K}_r) \leqslant 128\ell,$$

and by our lower bound on $d(\widehat{K}_r)$ from Claim 5.3.11, we get

$$\frac{1}{8k^2 2^k} \cdot \left(\frac{r}{n}\right)^{\frac{k}{2}} \cdot m \leqslant 128 r \log n,$$

which we can rearrange as

$$m \leqslant B^k n \log n \cdot \left(\frac{n}{r}\right)^{k/2-1}.$$

for some large enough constant B. Thus, if m is lower bounded as in the theorem statement, there must be an even cover of size $\ell \log n$. □

5.4 Strong refutation of semirandom k-XOR

In this section, we show that our reweighted Kikuchi matrix and edge deletion process yield a significantly simpler analysis of strong refutation algorithms for semirandom k-XOR formulas and lose only a single $\log n$ factor in the density. Combined with Feige's "XOR principle" [Fei02b, AOW15], we also obtain refutation algorithms for all *smoothed* Boolean CSPs. We will omit such reduction in this work and direct the reader to [GKM21] for a detailed exposition.

Theorem 5.4.1 (Semirandom k-XOR refutation)**.** *Fix $k \in \mathbb{N}$. There is an algorithm with parameter $r \in \mathbb{N}$, $2k \leqslant r \leqslant n/8$ that takes as input a semirandom k-XOR instance*

$$\psi(x) = \frac{1}{m} \sum_{C \in \mathcal{H}} b_C x_C$$

where $\mathcal{H}$ is a k-uniform hypergraph with n vertices and m hyperedges, and each $b_C \in \{\pm 1\}$ is chosen uniformly at random. The algorithm has the following guarantee: there is a universal constant C such that if $m \geqslant C^k n \log n \cdot (\frac{n}{r})^{\frac{k}{2}-1} \varepsilon^{-4}$ for $\varepsilon \in (0, 1/2)$, then with probability over $1 - \frac{1}{\mathrm{poly}(n)}$ over $\{b_C\}_{C \in \mathcal{H}}$, the algorithm runs in time $n^{O(r)}$ and certifies that $\psi(x) \leqslant \varepsilon$.

Remark 5.4.2 (Refutation strength: dependence on ε)**.** For the even arity case, we actually obtain a stronger guarantee (weaker requirement) of $m \geqslant O(n \log n) \cdot (\frac{n}{r})^{\frac{k}{2}-1} \varepsilon^{-2}$. For the odd arity case however, our analysis incurs a (likely suboptimal) dependence of $1/\varepsilon^4$ on the refutation strength (i.e., the upper bound on the value of the input k-XOR instance), though improving the $1/\varepsilon^5$ dependence of [GKM21, Theorem 5.1]. In contrast, a $1/\varepsilon^2$ dependence is known to hold for fully random k-XOR instances [RRS17]. Apart from a somewhat unsatisfying deficiency, this suboptimality turns out to be consequential – in particular, it changes the threshold at which efficient FKO refutation witnesses exist for semirandom k-SAT (and other CSPs) by a polynomial factor in n. Finding the "right" dependence of $1/\varepsilon^2$ (for the odd case) is an interesting open problem.

Our refutation algorithm will utilize the same Kikuchi graphs from Definition 5.2.4 and Definition 5.3.6 but with signs added to the edges in the natural way.

Definition 5.4.3 (Signed Kikuchi graph)**.** Let $\mathcal{H}$ be a k-uniform hypergraph associated with $\{\pm 1\}$ signs $\{b_C\}_{C \in \mathcal{H}}$. For the even arity case, let A_b be the signed adjacency matrix of the Kikuchi graph from Definition 5.2.4 where each edge $S \overset{C}{\longleftrightarrow} T$ has a sign b_C. For the odd arity case, let A_b be the signed adjacency matrix of the Kikuchi graph from Definition 5.3.6 where each edge $S \overset{C,C'}{\longleftrightarrow} T$ has a sign $b_C b_{C'}$.

5.4.1 Refuting semirandom even arity XOR

In this section, we prove Theorem 5.4.1 when k is even. As we will see in the short proof, our idea of the reweighted Kikuchi matrix from the hypergraph Moore bound naturally applies here, and in fact, we obtain the "right" $1/\varepsilon^2$ dependence in this case, i.e., we can certify that $\psi(x) \leqslant \varepsilon$ when $m \geqslant O(n \log n) \cdot (\frac{n}{r})^{\frac{k}{2}-1}\varepsilon^{-2}$.

Recall that in the Kikuchi graph (V, E), each $C \in \mathcal{H}$ contributes $\alpha := \frac{1}{2}\binom{k}{k/2}\binom{n-k}{r-k/2}$ edges in E, hence $|E| = \frac{1}{2}|V|d = m\alpha$. Thus, it is clear that

$$\psi(x) = \frac{1}{m} \cdot \frac{1}{\alpha} \sum_{(S,T)\in E} b_{S\oplus T} x_{S\oplus T} = \frac{1}{\binom{n}{r} d} (x^{\odot r})^\top A_b x^{\odot r} \tag{5.4}$$

where $x^{\odot r} \in \{\pm 1\}^{\binom{n}{r}}$ and the S-entry of $x^{\odot r}$ is x_S for $S \subseteq [n]$, $|S| = r$.

We now follow the same reweighting strategy: with $\Gamma = D + d\mathbb{1}$, we bound the spectral norm of the reweighted Kikuchi matrix $\left\|\Gamma^{-1/2} A_b \Gamma^{-1/2}\right\|_2$ with an almost identical proof as Lemma 5.2.6.

Lemma 5.4.4. *Let k be even and $r \in \mathbb{N}$. Let A_b be the signed Kikuchi graph with random $\{\pm 1\}$ coefficients $\{b_C\}_{C\in\mathcal{H}}$, and let $\Gamma = D + d\mathbb{1}$ where D is the degree matrix and d is the average degree of the Kikuchi graph. Then, with probability at least $1 - \frac{1}{\mathrm{poly}(n)}$ over the randomness of $\{b_C\}_{C\in\mathcal{H}}$,*

$$\left\|\Gamma^{-1/2} A_b \Gamma^{-1/2}\right\|_2 \leqslant O\left(\sqrt{\frac{r \log n}{d}}\right).$$

Proof. Let $\widetilde{A}_b = \Gamma^{-1/2} A_b \Gamma^{-1/2}$. We again use the trace power method $\|\widetilde{A}_b\|_2^\ell \leqslant \mathrm{tr}((\Gamma^{-1}A_b)^\ell)$ where we choose an even $\ell = 2\lceil r \log_2 n \rceil$. Observe that in expectation, $\mathbf{E}_b \mathrm{tr}((\Gamma^{-1}A_b)^\ell)$ counts the closed walks that use each hyperedge an even number of times. This is exactly the same as Lemma 5.2.6 where we count closed walks in an unsigned Kikuchi graph assuming there is no even cover of size $\leqslant \ell$. Thus, Lemma 5.2.6 shows that

$$\mathop{\mathbf{E}}_b \mathrm{tr}((\Gamma^{-1}A_b)^\ell) \leqslant 2^\ell n^r \left(\frac{\ell}{d}\right)^{\ell/2} \leqslant O\left(\frac{\ell}{d}\right)^{\ell/2}$$

when $\ell \geqslant r \log_2 n$. Then, by Markov's inequality, for any $\lambda > 0$,

$$\mathop{\mathbf{Pr}}_b\left[\|\widetilde{A}_b\|_2 \geqslant \lambda\right] = \mathop{\mathbf{Pr}}_b\left[\|\widetilde{A}_b\|_2^\ell \geqslant \lambda^\ell\right] \leqslant \lambda^{-\ell} \cdot \mathop{\mathbf{E}}_b \mathrm{tr}((\Gamma^{-1}A_b)^\ell) \leqslant O\left(\frac{\ell}{\lambda^2 d}\right)^{\ell/2}$$

Choosing $\lambda = O(\sqrt{\ell/d})$ completes the proof. □

We can complete the proof of Theorem 5.4.1 for even k.

Proof of Theorem 5.4.1 for even k. Let A_b be the signed Kikuchi graph with signs $\{b_C\}_{C\in\mathcal{H}}$, let $\Gamma = D + d\mathbb{1}$ where D is the degree matrix and d is the average degree of the Kikuchi graph, and let $\widetilde{A}_b = \Gamma^{-1/2}A_b\Gamma^{-1/2}$. The certification algorithm is simply to compute $\|\widetilde{A}_b\|_2$. Since $A_b \preceq \|\widetilde{A}_b\|_2 \cdot \Gamma$, and $\mathrm{tr}(\Gamma) = 2\binom{n}{r}d$, by Lemma 5.4.4,

$$\psi(x) = (5.4) \leqslant \frac{1}{\binom{n}{r}d}\|\widetilde{A}_b\|_2 \cdot \mathrm{tr}(\Gamma) \leqslant O\left(\sqrt{\frac{r\log n}{d}}\right)$$

using the fact that $x^{\odot r} \in \{\pm 1\}^{\binom{n}{r}}$ and $(x^{\odot r})^\top \Gamma x^{\odot r} = \mathrm{tr}(\Gamma)$. There is some constant C such that when $m \geqslant Cn\log n \cdot (\frac{n}{r})^{\frac{k}{2}-1}\varepsilon^{-2}$, by (5.1) the average degree $d \geqslant \frac{1}{2}(\frac{r}{n})^{k/2}m = \frac{C}{2}r\log n \cdot \varepsilon^{-2}$, thus giving us $\psi(x) \leqslant \varepsilon$. This completes the proof. □

5.4.2 Refuting semirandom odd arity XOR

Our proof of Theorem 5.4.1 for the odd arity case closely mimics the steps taken in proving the hypergraph Moore bound for odd arity hypergraphs (Theorem 5.3.1). Given a semirandom k-XOR instance ψ on hypergraph $\mathcal{H}$ with random signs $\{b_C\}_{C\in\mathcal{H}}$, we first apply the following hypergraph decomposition algorithm (a variant of Algorithm 5.3.2) to decompose the hypergraph into subhypergraphs $\mathcal{H}^{(1)}, \ldots, \mathcal{H}^{(k-1)}$. The main difference compared to Algorithm 5.3.2 is that in the final step, we add the "leftover" hyperedges to $\mathcal{H}^{(1)}$ instead of an extra $\mathcal{H}^{(0)}$.

Algorithm 5.4.5 (Hypergraph decomposition)**.** Given a k-uniform hypergraph $\mathcal{H}$ on n vertices and m hyperedges, and thresholds $\tau_1, \ldots, \tau_{k-1} \geqslant 2$, we partition $\mathcal{H}$ into hypergraphs $\mathcal{H}^{(1)}, \ldots, \mathcal{H}^{(k-1)}$ via the following algorithm.

1. Set $t = k - 1$ and $\mathcal{H}_{\text{current}} := \mathcal{H}$.
2. Set counter $s = 1$. While there is $T \subseteq [n]$ such that $|T| = t$ and
$$|\{C \in \mathcal{H}_{\text{current}} : T \subseteq C\}| \geqslant \tau_t :$$
   a) Choose T satisfying the condition and let $\mathcal{H}_s^{(t)}$ be a subset of $\{C \in \mathcal{H}_{\text{current}} : T \subseteq C\}$ of size τ_t.

b) Add all clauses in $\mathcal{H}_s^{(t)}$ to $\mathcal{H}^{(t)}$.

c) Delete all clauses in $\mathcal{H}_s^{(t)}$ to $\mathcal{H}_{\text{current}}$.

d) Increment s by 1.

3. Decrement t by 1. If $t > 0$, go back to step 2; otherwise take the remaining clauses in $\mathcal{H}_{\text{current}}$ and partition them into n parts $F_1, \dots, F_n$ where each clause C goes to some F_i such that $i \in C$. Add $F_1, \dots, F_n$ to $\mathcal{H}^{(1)}$ and terminate.

Notations and parameters. Throughout this section we will use the following notations.

- In Algorithm 5.4.5, we set thresholds $\tau_t = \max\left\{1, \left(\frac{n}{r}\right)^{\frac{k}{2}-t}\right\} \cdot 4k\varepsilon^{-2}$.
- In the decomposition, each $\mathcal{H}^{(t)}$ contains p_t groups $\mathcal{H}_1^{(t)}, \dots, \mathcal{H}_{p_t}^{(t)}$ where group $\mathcal{H}_i^{(t)}$ has a center $T_i^{(t)}$ of size t, and for each $C \in \mathcal{H}_i^{(t)}$, we write $\widetilde{C} = C \setminus T_i^{(t)}$.
- Each $|\mathcal{H}_i^{(t)}| = \tau_t$, with the exception that $|\mathcal{H}_i^{(1)}| \leqslant \tau_1$ may have different sizes (the leftover hyperedges in Algorithm 5.4.5). Let $m_t := \sum_{i=1}^{p_t} |\mathcal{H}_i^{(t)}|$ be the total number of hyperedges in $\mathcal{H}^{(t)}$.
- When $t = 1$ and $m \geqslant C^k n \log n \cdot \left(\frac{n}{r}\right)^{\frac{k}{2}-1} \varepsilon^{-4}$ for a large enough constant C, we have $m \geqslant n\tau_1$, hence $p_1 \leqslant \frac{m}{\tau_1} + n \leqslant \frac{2m}{\tau_1}$. Thus, we will use $p_t \tau_t \leqslant 2m$ for all $t \in [k-1]$.
- For each $t \in [k-1]$, the colored Kikuchi graph (V, E) obtained from $\mathcal{H}^{(t)} = (\mathcal{H}_1^{(t)}, \dots, \mathcal{H}_{p_t}^{(t)})$ (from Definition 5.3.6) has edges $|E| = \alpha_t \sum_{i=1}^{p_t} \binom{|\mathcal{H}_i^{(t)}|}{2} \leqslant \frac{1}{2}\alpha_t m_t \tau_t$, where $\alpha_t \approx \left(\frac{2n}{r}\right)^{r-(k-t)}$ is the number of edges contributed by each distinct pair $C, C' \in \mathcal{H}_i$ (see Observation 5.3.8).

With these notations and parameters in mind, we can write $\psi(x)$ as

$$\psi(x) = \frac{1}{m} \sum_{t=1}^{k-1} \sum_{C \in \mathcal{H}^{(t)}} b_C x_C = \frac{1}{k} \sum_{t=1}^{k-1} \psi_t(x)$$

$$\text{where} \quad \psi_t(x) := \frac{k}{m}\sum_{i=1}^{p_t}\sum_{C\in\mathcal{H}_i^{(t)}} b_C x_C = \frac{k}{m}\sum_{i=1}^{p_t} x_{T_i}\sum_{C\in\mathcal{H}_i^{(t)}} b_C x_{\widetilde{C}}\,. \tag{5.5}$$

Essentially, each ψ_t is the sub-instance of ψ restricted to the partition $\mathcal{H}^{(t)}$. Recall that for the purpose of showing existence of even covers, we only need to focus on one $\mathcal{H}^{(t)}$. For refutation however, we need to certify a bound on $\psi_t(x)$ for all $t \in [k-1]$.

Lemma 5.4.6 (Refuting each ψ_t)**.** *Fix an odd $k \in \mathbb{N}$, $t \in [k-1]$, and let $2k \leqslant r \leqslant n/8$. There is a constant C such that given a semirandom k-XOR instance ψ with n variables and $m \geqslant C^k n \log n (\frac{n}{r})^{\frac{k}{2}-1}\varepsilon^{-4}$ clauses for $\varepsilon \in (0,1/2)$, and suppose ψ_t is the subinstance from (5.5) obtained by the hypergraph decomposition algorithm (Algorithm 5.4.5), then with probability $1 - \frac{1}{\mathrm{poly}(n)}$ over the random signs, we can certify that $\psi_t(x) \leqslant \varepsilon$ in $n^{O(r)}$ time.*

Lemma 5.4.6 immediately completes the proof of Theorem 5.4.1 for odd k.

Proof of Theorem 5.4.1 by Lemma 5.4.6. Given the hypergraph $\mathcal{H}$, we apply the hypergraph decomposition algorithm (Algorithm 5.4.5) with thresholds $\tau_1, \dots, \tau_{k-1}$ and obtain subinstances $\psi_1, \dots, \psi_{k-1}$ as in (5.5). For each $t \in [k-1]$, we can certify that $\psi_t(x) \leqslant \varepsilon$ by Lemma 5.4.6 with high probability, which immediately implies the desired bound $\psi(x) \leqslant \varepsilon$. □

Edge deletion process. The proof of Lemma 5.4.6 requires deleting the "bad" edges from the signed Kikuchi matrix $A_b^{(t)}$ via a similar deletion process as the one used in the proof of Theorem 5.3.1, but with some parameter $\eta > 1$ instead of 1 and an additional *equalizing* step:

> Start with the colored Kikuchi graph, and delete every edge $\{S, T\}$ caused by a pair of clauses $C, C' \in \mathcal{H}_i^{(t)}$ such that S or T has more than η edges that C or C' participates in.
>
> Suppose $\rho < 1$ is the maximum fraction of edges deleted among all pairs of clauses. Then, for every $i \in [p_t]$ and every distinct pair $C, C' \in \mathcal{H}_i^{(t)}$, we delete (additional) edges caused by C, C' arbitrarily such that exactly ρ fraction of edges are deleted.

Observation 5.4.7 (Uniform deletion)**.** The final step in the above edge deletion process ensures that every pair C, C' contributes the *same* number of edges ($(1-\rho)\alpha_t$ to be exact) in the Kikuchi graph.

Mirroring the proof of Claim 5.3.10 yields the following generalization.

Lemma 5.4.8 (Deletion rate)**.** *Suppose a subhypergraph $\mathcal{H}^{(i)}$ satisfies that for any $s \geqslant i$ and any $T \subseteq [n]$ with $|T| = s$, the number of hyperedges in $\mathcal{H}^{(i)}$ containing T is at most τ_s, then the deletion process with parameter $\eta \geqslant 1$ satisfies*

$$\Pr_{\{S,T\}\sim E_{C,C'}}[\{S,T\} \text{ deleted}] \leqslant \frac{4^k}{\eta} \cdot \sum_{s=i}^{\lfloor \frac{k+i}{2} \rfloor} \tau_s \left(\frac{r}{n}\right)^{\lfloor \frac{k+i}{2} \rfloor - s}.$$

Proof. The proof is identical to the proof of Claim 5.3.10. Eq. (5.2) holds with an additional $1/\eta$ factor due to Markov's inequality. The lemma statement then follows immediately from (5.3). □

Proof of Lemma 5.4.6 via the Cauchy-Schwarz trick and the deletion process.

Proof of Lemma 5.4.6. We apply the Cauchy-Schwarz trick to ψ_t from (5.5):

$$\begin{aligned}\psi_t(x)^2 &\leqslant \frac{k}{m^2} \sum_{i=1}^{p_t} x_{T_i}^2 \cdot \sum_{i=1}^{p_t} \left(\sum_{C \in \mathcal{H}_i^{(t)}} b_C x_{\widetilde{C}} \right)^2 \leqslant \frac{kp_t}{m^2} \sum_{i=1}^{p_t} \sum_{C,C' \in \mathcal{H}_i^{(t)}} b_C b_{C'} x_{\widetilde{C}} x_{\widetilde{C}'} \\ &\leqslant \frac{kp_t m_t}{m^2} + \frac{kp_t}{m^2} \sum_{i=1}^{p_t} \sum_{C \neq C' \in \mathcal{H}_i^{(t)}} b_C b_{C'} x_{C \oplus C'} \end{aligned} \tag{5.6}$$

since $x \in \{\pm 1\}^n$, $b_C \in \{\pm 1\}$ and $\sum_{i=1}^{p_t} |\mathcal{H}_i^{(t)}| = m_t$. For the first term, since for all $t \in [k-1]$, we set $\tau_t \geqslant 4k\varepsilon^{-2}$ and $p_t \leqslant 2m/\tau_t \leqslant \frac{m\varepsilon^2}{2k}$, thus

$$\frac{kp_t m_t}{m^2} \leqslant \frac{\varepsilon^2}{2}. \tag{5.7}$$

We can now focus our attention on the second term in (5.6).

Given $\mathcal{H}^{(t)}$ and its partitions $\mathcal{H}_1^{(t)}, \ldots, \mathcal{H}_{p_t}^{(t)}$ of size τ_t, and signs $\{b_C\}_{C \in \mathcal{H}^{(t)}}$, let $A_b^{(t)}$ be the signed Kikuchi matrix defined in Definition 5.4.3, which is the signed version of the colored Kikuchi graph (V, E) from Definition 5.3.6. Recall from

Observation 5.3.8 that each distinct pair $C, C' \in \mathcal{H}_i^{(t)}$ contributes $\alpha_t \approx (\frac{2n}{r})^{r-(k-t)}$ edges in the graph. Thus, similar to (5.4) in the even case, we can write the second term of (5.6) as a quadratic form:

$$f_t(x) := \frac{kp_t}{m^2} \sum_{i=1}^{p_t} \sum_{C \neq C' \in \mathcal{H}_i^{(t)}} b_C b_{C'} x_{C \oplus C'} = \frac{kp_t}{2\alpha_t m^2} (x^{\odot r})^\top A_b^{(t)} x^{\odot r} \tag{5.8}$$

where $x^{\odot r} \in \{\pm 1\}^{\binom{2n}{r}}$ such that for $S \in [n] \times [2]$ with $S = (S^{(1)}, S^{(2)})$ (green and blue elements), the S-entry of $x^{\odot r}$ is $x_{S^{(1)} \oplus S^{(2)}}$.

We proceed to certify an upper bound on $f_t(x)$. Given the signed Kikuchi matrix $A_b^{(t)}$, we first apply the deletion process with parameter $\eta = B^k \varepsilon^{-2}$ for some large enough constant B. With the chosen thresholds τ_s, Lemma 5.4.8 states that the deletion probability ρ is at most

$$\rho \leqslant \frac{4^k}{\eta} \cdot \sum_{s=t}^{\lfloor \frac{k+t}{2} \rfloor} 4k\varepsilon^{-2} \cdot \max\left\{1, \left(\frac{n}{r}\right)^{\frac{k}{2}-s}\right\} \cdot \left(\frac{r}{n}\right)^{\lfloor \frac{k+t}{2} \rfloor - s} \leqslant \frac{1}{2},$$

since $s \leqslant \lfloor \frac{k+t}{2} \rfloor$ in the summation and $\lfloor \frac{k+t}{2} \rfloor \geqslant \frac{k+1}{2}$ for all $t \geqslant 1$.

Let $\widehat{A}_b^{(t)}$ be the Kikuchi matrix after the deletion process. By Observation 5.4.7, each distinct pair $C, C' \in \mathcal{H}_i^{(t)}$ contributes exactly $(1-\rho)$ fraction of the original edges. Thus, we have

$$(x^{\odot r})^\top \widehat{A}_b^{(t)} x^{\odot r} = (1-\rho) \cdot (x^{\odot r})^\top A_b^{(t)} x^{\odot r}. \tag{5.9}$$

Next, we follow the same argument as the proof of Lemma 5.4.4 to analyze $\widehat{A}_b^{(t)}$, using the norm bound of Lemma 5.3.9. Let $\Gamma = D + d\mathbb{1}$ where D is the degree matrix and d is the average degree, and let $\widetilde{A}_b = \Gamma^{-1/2} \widehat{A}_b^{(t)} \Gamma^{-1/2}$. To bound $\|\widetilde{A}_b\|_2$, we again use the trace power method $\|\widetilde{A}_b\|_2^\ell \leqslant \operatorname{tr}((\Gamma^{-1}\widehat{A}_b^{(t)})^\ell)$ where we choose an even $\ell = 2\lceil r \log_2 n \rceil$. Observe that in expectation, $\mathbf{E}_b \operatorname{tr}((\Gamma^{-1}A_b)^\ell)$ counts the closed walks that use each hyperedge an even number of times. This is exactly the same as Lemma 5.3.9 where we count closed walks in an unsigned Kikuchi graph assuming there is no even cover of size $\leqslant \ell$. Furthermore, $d_{S,i} \leqslant \eta$ is automatically satisfied after the deletion process. Thus, we can directly apply Lemma 5.3.9 and show that

$$\mathop{\mathbf{E}}_b \operatorname{tr}\left((\Gamma^{-1}\widehat{A}_b^{(t)})^\ell\right) \leqslant 2^\ell n^r \left(\frac{2\eta\ell}{d}\right)^{\ell/2} \leqslant O\left(\frac{\eta\ell}{d}\right)^{\ell/2}$$

when $\ell \geqslant r \log_2 n$. Then, by Markov's inequality, we have that

$$\Pr_b \left[\|\widetilde{A}_b\|_2 \geqslant O\left(\sqrt{\frac{\eta\ell}{d}}\right) \right] \leqslant \frac{1}{\mathrm{poly}(n)}.$$

Thus, with high probability we have $\widehat{A}_b^{(t)} \preceq O\left(\sqrt{\frac{\eta\ell}{d}}\right) \cdot \Gamma$, then since $\mathrm{tr}(\Gamma) = 4|E|$,

$$(x^{\odot r})^\top \widehat{A}_b^{(t)} x^{\odot r} \leqslant O\left(\sqrt{\frac{\eta\ell}{d}}\right) \cdot \mathrm{tr}(\Gamma) = O\left(\sqrt{\frac{\eta\ell}{d}}\right) \cdot |E|\,.$$

Next, let $\widehat{f}_t(x) = \frac{kp_t}{2\alpha_t m^2}(x^{\odot r})^\top \widehat{A}_b^{(t)} x^{\odot r}$. By Observation 5.3.8, we have $d \geqslant (\frac{r}{2n})^{k-t} \sum_{i=1}^{p_t} \binom{|\mathcal{H}_i^{(t)}|}{2}$ when $2k \leqslant r \leqslant n/8$. Plugging in parameters

$$|E| = \alpha_t \sum_{i=1}^{p_t} \binom{|\mathcal{H}_i^{(t)}|}{2},$$

$p_t\tau_t \leqslant 2m$, $\eta = B^k\varepsilon^{-2}$, and $\ell = 2\lceil r \log_2 n \rceil$, standard calculations show that

$$\begin{aligned}
\widehat{f}_t(x) &\leqslant O(1)\frac{kp_t}{\alpha_t m^2}\sqrt{\frac{\eta\ell}{d}}|E| \\
&\leqslant O(1)\frac{kp_t}{m^2}\sqrt{\eta\ell\left(\frac{2n}{r}\right)^{k-t}\sum_{i=1}^{p_t}\binom{|\mathcal{H}_i^{(t)}|}{2}} \\
&\leqslant O(1)\sqrt{\frac{\eta r \log n}{m\tau_t}\left(\frac{2n}{r}\right)^{k-t}}.
\end{aligned}$$

Suppose $m \geqslant C^k n \log n \cdot (\frac{n}{r})^{\frac{k}{2}-1}\varepsilon^{-4}$ for some large enough constant C. We split into cases:

1. For $t \leqslant \frac{k-1}{2}$, we set $\tau_t = (\frac{n}{r})^{\frac{k}{2}-t} \cdot 4k\varepsilon^{-2}$, thus $\widehat{f}_t(x) \leqslant \frac{\varepsilon^2}{4}$.
2. For $t \geqslant \frac{k+1}{2}$, we set $\tau_t = 4k\varepsilon^{-2}$, thus $\widehat{f}_t(x) \leqslant \frac{\varepsilon^2}{4}(\frac{n}{r})^{\frac{k}{4}-\frac{t}{2}} < \frac{\varepsilon^2}{4}$.

Therefore, by calculating $\|\widetilde{A}_b\|_2$, which can be done in $n^{O(r)}$ time, we can certify that $\widehat{f}_t(x) \leqslant \frac{\varepsilon^2}{4}$. Combined with (5.9) and the bound of $\rho \leqslant 1/2$, we can certify that

$$f_t(x) \leqslant \frac{1}{1-\rho} \cdot \widehat{f}_t(x) \leqslant \frac{\varepsilon^2}{2},$$

and with (5.7), we can certify an upper bound on (5.6):

$$\psi_t(x)^2 \leqslant (5.7) + (5.8) \leqslant \frac{\varepsilon^2}{2} + f_t(x) \leqslant \varepsilon^2 ,$$

completing the proof. □

5.5 Alternative proof of the Moore bound for irregular graphs

We proved the weak Moore bound (Proposition 5.2.1) by showing that if there is no cycle of length $\leqslant \ell$, then $A \prec \frac{2n^{1/\ell}}{\sqrt{d}}(D + d\mathbb{1})$ (Claim 5.2.2) where D is the diagonal degree matrix and d is the average degree, which then gives us a bound of $2\lceil \log_{d/16} n \rceil$. In this section, we prove that using a more carefully chosen diagonal matrix Γ', such a strategy can recover the exact Moore bound $2\log_{d-1} n$. This provides an alternative proof of the Moore bound in addition to the existing proofs by [AHL02] and [BR14][5].

Theorem 5.5.1 (Moore bound for irregular graphs)**.** *Suppose G is a graph on n vertices with average degree $d > 2$. Then G has a cycle of length $2(\lfloor \log_{d-1} n \rfloor + 1)$.*

The following lemma shows what the "correct" diagonal matrix should be to recover the exact Moore bound.

Lemma 5.5.2. *Let G be a graph with n vertices and degree matrix D that has no cycle of length $\leqslant \ell$ for some even $\ell \in \mathbb{N}$. Then, the adjacency matrix A satisfies*

$$A \preceq n^{2/\ell}\mathbb{1} + n^{-2/\ell}(D - \mathbb{1}) .$$

Proof of Theorem 5.5.1 by Lemma 5.5.2. Assuming there is no cycle of length $\leqslant \ell$, Lemma 5.5.2 implies that

$$\underline{1}^\top A \underline{1} = nd \leqslant n \cdot (n^{2/\ell} + n^{-2/\ell}(d-1)) .$$

Let $x = n^{2/\ell}$, then we have $x^2 - dx + (d-1) \geqslant 0$, which implies that $x \geqslant d - 1$ (as $x \leqslant 1$ is not valid). Taking logs, we get

$$\frac{2}{\ell}\log n \geqslant \log(d-1) \Longrightarrow \ell \leqslant 2\log_{d-1} n .$$

[5] [AHL02] and [BR14] actually obtained a slightly more precise bound depending on whether the girth of the graph is odd or even.

ℓ is even, so $\ell < 2(\lfloor \log_{d-1} n \rfloor + 1)$. This completes the proof. □

The proof of Lemma 5.5.2 is based on *non-backtracking walks*, which are walks such that no edge is the inverse of its preceding edge. We note that both proofs of [AHL02] and [BR14] also analyze non-backtracking walks. For a graph G on n vertices with adjacency matrix A, we define $A^{(s)}$ to be the $n \times n$ matrix whose (u, v) entry counts the number of length-s non-backtracking walks between vertices u and v in G. The following is a standard fact.

Fact 5.5.3 (Recurrence and generating function of $A^{(s)}$)**.** *The non-backtracking matrices $A^{(s)}$ satisfy the following recurrence:*

$$\begin{aligned} A^{(0)} &= \mathbb{1}\,, \\ A^{(1)} &= A\,, \\ A^{(2)} &= A^2 - D\,, \\ A^{(s)} &= A^{(s-1)}A - A^{(s-2)}(D - \mathbb{1})\,, \quad s > 2\,. \end{aligned}$$

The recurrences imply that these matrices have a generating function:

$$J(t) := \sum_{s=0}^{\infty} A^{(s)} t^s = (1 - t^2) \cdot H(t)^{-1}\,, \text{ where } H(t) := \mathbb{1} - At + (D - \mathbb{1})t^2$$

for $t \in [0, 1)$ whenever the series converges.

We first prove the following lemma,

Lemma 5.5.4. *Let $s, k \in \mathbb{N}$, $s \geqslant k$, and let q, r be the quotient and remainder of s divided by k, i.e. $s = qk + r$. Then,*

$$\operatorname{tr}(A^{(s)}) \leqslant \sqrt{n} \cdot \|A^{(k)}\|_2^q \cdot \|A^{(r)}\|_F.$$

Proof. $\operatorname{tr}(A^{(s)})$ counts the number of closed non-backtracking walks of length s in the graph. Now, consider the set of closed walks of length $s = qk + r$ such that after every k non-backtracking steps, we can "forget the previous step", i.e. we are allowed to backtrack at step ik for every $i = 0, \ldots, q$. The number of such walks is $\operatorname{tr}((A^{(k)})^q A^{(r)})$. The set of closed non-backtracking walk is clearly a subset of such walks, thus we have

$$\operatorname{tr}(A^{(s)}) \leqslant \operatorname{tr}((A^{(k)})^q A^{(r)}) \leqslant \left\|(A^{(k)})^q\right\|_F \cdot \left\|A^{(r)}\right\|_F.$$

Let $\lambda_1, \dots, \lambda_n$ be the eigenvalues of $A^{(k)}$ and $\lambda_{\max} = \|A^{(k)}\|_2$. Then,

$$\left\|(A^{(k)})^q\right\|_F = \sqrt{\sum_{i=1}^{n} \lambda_i^{2q}} \leqslant \sqrt{n}(\lambda_{\max})^q.$$

This completes the proof. □

With Fact 5.5.3 and Lemma 5.5.4, we now prove Lemma 5.5.2 by analyzing the convergence of $J(t)$ as t increases from 0.

Proof of Lemma 5.5.2. Let A be the adjacency matrix of G with average degree $d > 2$, and let D be the diagonal degree matrix G. Recall the definitions $J(t) = \sum_{s=0}^{\infty} A^{(s)} t^s$ and $H(t) = \mathbb{1} - At + (D - \mathbb{1})t^2$ from Fact 5.5.3. We will analyze the convergence of $\operatorname{tr}(J(t))$ as t increase from 0.

Observe that $J(0) = H(0) = \mathbb{1}$, and since $J(t)$ and $H(t)$ are both symmetric matrices, their eigenvalues move continuously on the real line as t increases from 0. Thus, suppose there is some $t^* \in (0, 1)$ such that $\operatorname{tr}(J(t)) < \infty$ for all $t \in [0, t^*)$, then $H(t) \succ 0$ for all $t \in [0, t^*)$. This is easy to see because if not, then there must be some $t' \in [0, t^*)$ such that $H(t') \succeq 0$ but has a zero eigenvalue, and $\operatorname{tr}(J(t'))$ will not converge.

We next show that we can take $t^* = n^{-2/\ell}$ assuming that G has no cycle of length $\leqslant \ell = 2k$. First, observe that every entry of $A^{(k)}$ must be either 0 or 1, otherwise if $A^{(k)}[i, j] > 1$ then there are two distinct length-k paths from i to j, meaning there is a cycle of length at most $2k = \ell$, a contradiction. Therefore, the L_1 norm of each row of $A^{(k)}$ is at most n, hence $\|A^{(k)}\|_2 \leqslant n$. Next, observe that for each $s \in \mathbb{N}$ we can write $s = qk + r$, and

$$J(t) = \sum_{s=0}^{\infty} A^{(s)} t^s \leqslant \sum_{r=0}^{k-1} \sum_{q=0}^{\infty} A^{(qk+r)} t^{qk+r}.$$

By Lemma 5.5.4, we have

$$\operatorname{tr}(J(t)) \leqslant \sum_{r=0}^{k-1} t^r \sqrt{n} \|A^{(r)}\|_F \sum_{q=0}^{\infty} \|A^{(k)}\|_2^q \cdot t^{qk} \leqslant \sum_{r=0}^{k-1} t^r \sqrt{n} \|A^{(r)}\|_F \sum_{q=0}^{\infty} (nt^k)^q.$$

Thus, if $t < n^{-1/k} < 1$, then $\operatorname{tr}(J(t)) < \infty$. Therefore, we have $H(t) \succ 0$ for all $t \in [0, n^{-1/k})$, and by continuity $H(n^{-1/k}) \succeq 0$, which means that

$$\mathbb{1} - n^{-1/k} A + n^{-2/k}(D - \mathbb{1}) \succeq 0 \Longrightarrow A \preceq n^{2/\ell}\mathbb{1} + n^{-2/\ell}(D - \mathbb{1})$$

as $\ell = 2k$. This completes the proof. □

Bibliography

[Abb17] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017. 14

[ABH16] Emmanuel Abbe, Afonso S Bandeira, and Georgina Hall. Exact recovery in the stochastic block model. *IEEE Transactions on Information Theory*, 62(1):471–487, 2016. 14, 19

[ABLS07] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. Non-backtracking random walks mix faster. *Communications in Contemporary Mathematics*, 9(04):585–603, 2007. 29

[AC08] D. Achlioptas and A. Coja-Oghlan. Algorithmic barriers from phase transitions. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 793–802, 2008. 121

[ACKZ15] Maria Chiara Angelini, Francesco Caltagirone, Florent Krzakala, and Lenka Zdeborová. Spectral detection on sparse hypergraphs. In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 66–73. IEEE, 2015. 120

[AF09] Noga Alon and Uriel Feige. On the power of two, three and four probes. In *Proceedings of the twentieth annual ACM-SIAM symposium on Discrete algorithms*, pages 346–354. SIAM, 2009. 245

[AFH15] Omer Angel, Joel Friedman, and Shlomo Hoory. The non-backtracking spectrum of the universal cover of a graph. *Transactions of the American Mathematical Society*, 367(6):4287–4318, 2015. 223

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple constructions of almost k-wise independent random variables. *Random Structures & Algorithms*, 3(3):289–304, 1992. 217

[AHL02] Noga Alon, Shlomo Hoory, and Nathan Linial. The moore bound for irregular graphs. *Graphs and Combinatorics*, 18(1):53–57, 2002. 11, 65, 221, 243, 244, 247, 249, 276, 277

[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. 24

[AL13] Noga Alon and Shachar Lovett. Almost k-wise vs. k-wise independent permutations, and uniformity for general group actions. *Theory of Computing*, 9:559–577, 2013. 217, 218

[Alo86] Noga Alon. Eigenvalues and expanders. *Combinatorica*, 6(2):83–96, 1986. 209, 210, 211

[AM09] Per Austrin and Elchanan Mossel. Approximation resistant predicates from pairwise independence. *Computational Complexity*, 18(2):249–271, 2009. 117

[AM13] Emmanuel Abbe and Andrea Montanari. Conditional random fields, planted constraint satisfaction and entropy concentration. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 332–346. Springer, 2013. 120

[AOW15] Sarah R Allen, Ryan O'Donnell, and David Witmer. How to refute a random CSP. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 689–708. IEEE, 2015. 2, 116, 247, 267

[AS12] Pranjal Awasthi and Or Sheffet. Improved spectral-norm bounds for clustering. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 37–49. Springer, 2012. 20

[AS15] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery. In *2015 IEEE 56th Annual Symposium on*

Foundations of Computer Science, pages 670–688. IEEE, 2015. 19, 117, 121, 136

[Bas92] Hyman Bass. The Ihara–Selberg zeta function of a tree lattice. *International Journal of Mathematics*, 3(6):717–797, 1992. 222

[BB20] Matthew Brennan and Guy Bresler. Reducibility and Statistical-Computational Gaps from Secret Leakage. *arXiv preprint arXiv:2005.08099*, 2020. 127

[BBH18] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. *arXiv preprint arXiv:1806.07508*, 2018. 127

[BBH+20] Matthew Brennan, Guy Bresler, Samuel B Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low-degree tests are almost equivalent. *arXiv preprint arXiv:2009.06107*, 2020. 127

[BBK+20] Afonso S Bandeira, Jess Banks, Dmitriy Kunisky, Cristopher Moore, and Alexander S Wein. Spectral planting and the hardness of refuting cuts, colorability, and communities in random graphs. *arXiv preprint arXiv:2008.12237*, 2020. 92, 93

[BC78] Edward Bender and Rodney Canfield. The asymptotic number of labeled graphs with given degree sequences. *Journal of Combinatorial Theory. Series A*, 24(3):296–307, 1978. 218, 238

[BC19] Charles Bordenave and Benoît Collins. Eigenvalues of random lifts and polynomials of random permutation matrices. *Annals of Mathematics*, 190(3):811–875, 2019. 122, 124, 125, 126, 143, 147, 152, 167, 173

[BDG+16] Gerandy Brito, Ioana Dumitriu, Shirshendu Ganguly, Christopher Hoffman, and Linh V Tran. Recovery and rigidity in a regular stochastic block model. In *Proceedings of the twenty-seventh annual ACM-SIAM symposium on Discrete algorithms*, pages 1589–1601. Society for Industrial and Applied Mathematics, 2016. 18

[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(2):687–735, 2019. 9, 23

[Big93] Norman Biggs. *Algebraic graph theory*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, second edition, 1993. 244

[BK67] Jānis Bārzdiņš and Andrey Kolmogorov. On the realization of networks in three-dimensional space. *Problemy Kibernetiki*, 19:261–268, 1967. 209

[BKHL99] Claudia Bertram-Kretzberg, Thomas Hofmeister, and Hanno Lefmann. Sparse 0- 1 matrices and forbidden hypergraphs. *Combinatorics, Probability and Computing*, 8(5):417–427, 1999. 245

[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(12):5451–5460, 2019. 24

[BKW19] Afonso S Bandeira, Dmitriy Kunisky, and Alexander S Wein. Computational hardness of certifying bounds on constrained pca problems. *arXiv preprint arXiv:1902.07324*, 2019. 89

[BL06] Yonatan Bilu and Nathan Linial. Lifts, discrepancy and nearly optimal spectral gap. *Combinatorica*, 26(5):495–519, 2006. 209, 211, 213, 215, 216

[BLM15] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. In *2015 IEEE 56th Annual Symposium on Foundations of Computer Science*, pages 1347–1357. IEEE, 2015. 14, 15, 16, 114, 117, 118, 120, 121, 122, 125, 136, 146, 167

[BMR21] Jess Banks, Sidhanth Mohanty, and Prasad Raghavendra. Local statistics, semidefinite programming, and community detection. pages 1298–1316, 2021. 13, 127, 146, 171, 175, 201

[BMS08] Louay Bazzi, Mohammad Mahdian, and Daniel A Spielman. The minimum distance of turbo-like codes. *IEEE Transactions on Information Theory*, 55(1):6–15, 2008. 245

[Bol78] Béla Bollobás. *Extremal graph theory*, volume 11 of *London Mathematical Society Monographs*. Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], London-New York, 1978. 244

[Bol80] Béla Bollobás. A probabilistic proof of an asymptotic formula for the number of labelled regular graphs. *European Journal of Combinatorics*, 1(4):311–316, 1980. 218, 238, 239

[Bol01] Béla Bollobás. *Random Graphs*. Cambridge University Press, second edition edition, 2001. 218

[Bor16] Charles Bordenave. Lecture notes on random graphs and probabilistic combinatorial optimization. Retrieved from `https://www.math.univ-toulouse.fr/~bordenave/coursRG.pdf`, 2016. 232

[Bor19] Charles Bordenave. A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts. In *Annales scientifiques de l'Ecole normale supérieure*, 2019. 11, 147, 167, 211, 214, 219, 220, 222, 224, 225, 227, 230, 231, 232, 233, 234, 235

[BR14] S Ajesh Babu and Jaikumar Radhakrishnan. An entropy-based proof for the Moore bound for irregular graphs. In *Perspectives in Computational Complexity*, pages 173–181. Springer, 2014. 244, 276, 277

[BS95] Avrim Blum and Joel Spencer. Coloring random and semi-random k-colorable graphs. *Journal of Algorithms*, 19(2):204–234, 1995. 19

[BSW01] Eli Ben-Sasson and Avi Wigderson. Short proofs are narrow—resolution made simple. *Journal of the ACM (JACM)*, 48(2):149–169, 2001. 15

[BT11] Avraham Ben-Aroya and Amnon Ta-Shma. A combinatorial construction of almost-Ramanujan graphs using the zig-zag product. *SIAM Journal on Computing*, 40(2):267–290, 2011. 211

[Cas06] Chad Casarotto. Graph theory and Cayley's formula. 2006. 69

[Chi92] Patrick Chiu. Cubic Ramanujan graphs. *Combinatorica*, 12(3):275–285, 1992. 210

[CJSX14] Yudong Chen, Ali Jalali, Sujay Sanghavi, and Huan Xu. Clustering partially observed graphs via convex optimization. *The Journal of Machine Learning Research*, 15(1):2213–2238, 2014. 19

[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(3):1027–1059, 2015. 20

[Cla06] Pete Clark. Ramanujan graphs and Shimura curves. Retrieved from `http://alpha.math.uga.edu/~pete/ramanujanrevisited.pdf`, 2006. 210

[CM08] Sebastian M. Cioabă and M. Ram Murty. Expander graphs and gaps between primes. *Forum Mathematicum*, 20(4):745–756, 2008. 211, 212

[CO04] Amin Coja-Oghlan. Coloring semirandom graphs optimally. In *International Colloquium on Automata, Languages, and Programming*, pages 383–395. Springer, 2004. 19

[CO07] Amin Coja-Oghlan. Solving np-hard semirandom graph problems in polynomial expected time. *Journal of Algorithms*, 62(1):19–46, 2007. 19

[Coh16] Michael Cohen. Ramanujan graphs in polynomial time. In *Proceedings of the 57th Annual IEEE Symposium on Foundations of Computer Science*, pages 276–281, 2016. 11, 212

[COHKL+20] Amin Coja-Oghlan, Max Hahn-Klimroth, Philipp Loick, Noela Müller, Konstantinos Panagiotou, and Matija Pasch. Inference and mutual information on random factor graphs. *arXiv preprint arXiv:2007.07494*, 2020. 120

[CS22] David Munhá Correia and Benny Sudakov. Personal communication. 2022. 247

[CSV17] Moses Charikar, Jacob Steinhardt, and Gregory Valiant. Learning from untrusted data. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 47–60. ACM, 2017. 15

[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(6):066106, 2011. 17, 117, 121, 122, 123, 136

[DKMZ11b] Aurelien Decelle, Florent Krzakala, Cristopher Moore, and Lenka Zdeborová. Inference and phase transitions in the detection of modules in sparse networks. *Physical Review Letters*, 107(6):065701, 2011. 14, 117, 121

[dlHM06] Pierre de la Harpe and Antoine Musitelli. Expanding graphs, Ramanujan graphs, and 1-factor perturbations. *Bulletin of the Belgian Mathematical Society — Simon Stevin*, 13(4):673–680, 2006. 211

[DMO$^+$19] Yash Deshpande, Andrea Montanari, Ryan O'Donnell, Tselil Schramm, and Subhabrata Sen. The threshold for SDP-refutation of random regular NAE-3SAT. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2305–2321. SIAM, 2019. 214, 247

[DSMM18] Yash Deshpande, Subhabrata Sen, Andrea Montanari, and Elchanan Mossel. Contextual stochastic block models. In *Advances in Neural Information Processing Systems*, pages 8581–8593, 2018. 130

[dT22] Tommaso d'Orsi and Luca Trevisan. A Ihara-Bass Formula for Non-Boolean Matrices and Strong Refutations of Random CSPs. *arXiv preprint arXiv:2204.10881*, 2022. 247

[Fei02a] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 534–543. ACM, 2002. 15

[Fei02b] Uriel Feige. Relations between average case complexity and approximation complexity. In *Proceedings of the 34th Annual ACM Symposium on Theory of Computing*, pages 543–543, 2002. 245, 246, 267

[Fei08] Uriel Feige. Small linear dependencies for binary vectors of low weight. In *Building bridges*, pages 283–307. Springer, 2008. 12, 243, 245, 246

[FGR+17] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2):1–37, 2017. 127

[FK00] Uriel Feige and Robert Krauthgamer. Finding and certifying a large hidden clique in a semirandom graph. *Random Structures & Algorithms*, 16(2):195–208, 2000. 19

[FK01] Uriel Feige and Joe Kilian. Heuristics for semirandom graph problems. *Journal of Computer and System Sciences*, 63(4):639–671, 2001. 19

[FKO06] Uriel Feige, Jeong Han Kim, and Eran Ofek. Witnesses for non-satisfiability of dense random 3CNF formulas. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 497–508. IEEE, 2006. 245, 246

[FM17] Zhou Fan and Andrea Montanari. How well do local algorithms solve semidefinite programs? In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 604–614, 2017. 25, 80, 83, 99, 100, 106, 146, 203, 247

[Fri93] Joel Friedman. Some geometric aspects of graphs and their eigenfunctions. *Duke Mathematical Journal*, 69(3):487–525, 1993. 210

[Fri03a] Joel Friedman. A proof of Alon's second eigenvalue conjecture. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 720–724. ACM, 2003. 25

[Fri03b] Joel Friedman. A proof of Alon's second eigenvalue conjecture. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 720–724, 2003. 147, 167

[Fri08] Joel Friedman. A proof of Alon's second eigenvalue conjecture and related problems. *Memoirs of the American Mathematical Society*, 195(910):viii+100, 2008. 11, 31, 211, 222, 225

[GG81] Ofer Gabber and Zvi Galil. Explicit constructions of linear-sized superconcentrators. *Journal of Computer and System Sciences*, 22(3):407–420, 1981. Special issued dedicated to Michael Machtey. 210

[GKM21] Venkatesan Guruswami, Pravesh K Kothari, and Peter Manohar. Algorithms and Certificates for Boolean CSP Refutation:" Smoothed is no harder than Random". *arXiv preprint arXiv:2109.04415*, 2021. 12, 243, 244, 246, 247, 248, 249, 252, 256, 257, 260, 267, 268

[GMPS18] Sainyam Galhotra, Arya Mazumdar, Soumyabrata Pal, and Barna Saha. The geometric block model, 2018. 131

[Gol00] Oded Goldreich. Candidate One-Way Functions Based on Expander Graphs. *IACR Cryptol. ePrint Arch.*, 2000:63, 2000. 127

[Gol11] Oded Goldreich. Candidate one-way functions based on expander graphs. In *Studies in Complexity and Cryptography. Miscellanea on the Interplay between Randomness and Computation*, pages 76–87. Springer, 2011. 116

[Gri01] Dima Grigoriev. Linear lower bound on degrees of positivstellensatz calculus proofs for the parity. *Theoretical Computer Science*, 259(1):613–622, 2001. 15

[GV16] Olivier Guédon and Roman Vershynin. Community detection in sparse networks via grothendieck's inequality. *Probability Theory and Related Fields*, 165(3-4):1025–1049, 2016. 15, 19, 20

[Hås84] Johan Håstad. An NP-complete problem – some aspects of its solution and some possible applications. Master's thesis, Uppsala University, 1984. 214

[Has89] Ki-ichiro Hashimoto. Zeta functions of finite graphs and representations of p-adic groups. In *Automorphic forms and geometry of arithmetic varieties*, volume 15 of *Advanced Studies in Pure Mathematics*, pages 211–280. Elsevier, 1989. 222

[HKM23] Jun-Ting Hsieh, Pravesh K Kothari, and Sidhanth Mohanty. A simple and sharper proof of the hypergraph Moore bound. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 2324–2344. SIAM, 2023. 243

[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. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017. 23

[HLM12] Simon Heimlicher, Marc Lelarge, and Laurent Massoulié. Community detection in the labelled stochastic block model. *arXiv preprint arXiv:1209.2910*, 2012. 130

[HLW06] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *American Mathematical Society Bulletin*, 43(4):439–561, 2006. 10, 209, 210

[Hoo02] Shlomo Hoory. The size of bipartite graphs with a given girth. *Journal of Combinatorial Theory, Series B*, 86(2):215–220, 2002. 244

[HS17] Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: community detection and related problems. In *Foundations of Computer Science (FOCS), 2017 IEEE 58th Annual Symposium on*, pages 379–390. IEEE, 2017. 9, 16, 23, 127

[HWX16] Bruce Hajek, Yihong Wu, and Jiaming Xu. Achieving exact cluster recovery threshold via semidefinite programming. *IEEE Transactions on Information Theory*, 62(5):2788–2797, 2016. 19

[Iha66] Yasutaka Ihara. On discrete subgroups of the two by two projective linear group over p-adic fields. *Journal of the Mathematical Society of Japan*, 18:219–235, 1966. 210, 222

[JŁR00] Svante Janson, Tomasz Łuczak, and Andrzej Rucinski. *Random graphs*. John Wiley & Sons, 2000. 231

[Kas07] Martin Kassabov. Symmetric groups and expander graphs. *Inventiones Mathematicae*, 170(2):327–354, 2007. 217, 218

[Kee11] Peter Keevash. Hypergraph turan problems. *Surveys in combinatorics*, 392:83–140, 2011. 244

[KGR11] Ulugbek Kamilov, Vivek K Goyal, and Sundeep Rangan. Optimal quantization for compressive sensing under message passing reconstruction. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 459–463. IEEE, 2011. 24

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the thiry-fourth annual ACM symposium on Theory of computing*, pages 767–775, 2002. 3, 116

[KK10] Amit Kumar and Ravindran Kannan. Clustering with spectral norm and the k-means algorithm. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pages 299–308. IEEE, 2010. 20

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. Optimal inapproximability results for max-cut and other 2-variable csps? *SIAM Journal on Computing*, 37(1):319–357, 2007. 3, 116

[KMM+13] Florent Krzakala, Cristopher Moore, Elchanan Mossel, Joe Neeman, Allan Sly, Lenka Zdeborová, and Pan Zhang. Spectral redemption in clustering sparse networks. *Proceedings of the National Academy of Sciences*, 110(52):20935–20940, 2013. 114, 118, 120

[KMOW17] Pravesh K Kothari, Ryuhei Mori, Ryan O'Donnell, and David Witmer. Sum of squares lower bounds for refuting any CSP. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 132–145, 2017. 116

[KMZ12] Florent Krzakala, Marc Mézard, and Lenka Zdeborová. Reweighted belief propagation and quiet planting for random k-sat. *Journal on Satisfiability, Boolean Modeling and Computation*, 8(3-4):149–171, 2012. 121

[KNR09] Eyal Kaplan, Moni Naor, and Omer Reingold. Derandomized constructions of k-wise (almost) independent permutations. *Algorithmica. An International Journal in Computer Science*, 55(1):113–133, 2009. 217, 218

[KV00] Jeong Han Kim and Van H Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000. 246

[KV06] Michael Krivelevich and Dan Vilenchik. Semirandom models as benchmarks for coloring algorithms. In *2006 Proceedings of the Third Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 211–221. SIAM, 2006. 19

[KV12] Bernhard Korte and Jens Vygen. *Combinatorial optimization*, volume 2. Springer, 2012. 65

[KZ09] Florent Krzakala and Lenka Zdeborova. Hiding quiet solutions in random constraint satisfaction problems. *Physical review letters*, 102:238701, 07 2009. 117, 119, 121, 126, 136

[LMP01] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001. 120

[LMR22] Siqi Liu, Sidhanth Mohanty, and Prasad Raghavendra. On statistical inference when fixed points of belief propagation are unstable. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 395–405. IEEE, 2022. 113

[LPS88] Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261—-277, 1988. 11, 210

[LS10] Eyal Lubetzky and Allan Sly. Cutoff phenomena for random walks on random regular graphs. *Duke Mathematical Journal*, 153(3):475–510, 2010. 231

[Mar73a] Grigory Margulis. Complexity of an optimum nonblocking switching network without reconnections. *Problemy Peredachi Informatsii*, 9(1):84–87, 1973. 209

[Mar73b] Grigory Margulis. Explicit construction of concentrators. *Problemy Peredachi Informatsii*, 94(4):71–80, 1973. 210

[Mar88] Grigory Margulis. Explicit group-theoretic constructions of combinatorial schemes and their applications in the construction of expanders and concentrators. *Problemy Peredachi Informatsii*, 24(1):51–60, 1988. 11, 210

[Mas14a] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703. ACM, 2014. 14

[Mas14b] Laurent Massoulié. Community detection thresholds and the weak ramanujan property. In *Proceedings of the forty-sixth annual ACM symposium on Theory of computing*, pages 694–703, 2014. 117, 121, 136

[MM09a] Marc Mezard and Andrea Montanari. Information, physics, and computation. 2009. 117, 134

[MM09b] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009. 119

[MMV12] Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan. Approximation algorithms for semi-random partitioning problems. In *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*, pages 367–384. ACM, 2012. 19

[MMV16] Konstantin Makarychev, Yury Makarychev, and Aravindan Vijayaraghavan. Learning communities in the presence of errors. In *Conference on Learning Theory*, pages 1258–1291, 2016. 15, 20

[MNS14] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. In *Conference on Learning Theory*, pages 356–370, 2014. 129

[MNS18] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018. 14, 117, 121, 136

[Moi12] Ankur Moitra. A singly-exponential time algorithm for computing nonnegative rank. *arXiv preprint arXiv:1205.0044*, 2012. 15

[Mon08] Andrea Montanari. Estimating random variables from random sparse observations. *European Transactions on Telecommunications*, 19(4):385–403, 2008. 114, 120, 130

[MOP19] Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. Explicit near-ramanujan graphs of every degree. *arXiv preprint arXiv:1909.06988*, 2019. 65, 214, 226

[MOP20] Sidhanth Mohanty, Ryan O'Donnell, and Pedro Paredes. The SDP Value for Random Two-Eigenvalue CSPs. In *37th International Symposium on Theoretical Aspects of Computer Science (STACS 2020)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik, 2020. 59, 147, 167, 206, 208, 247

[Mor94] Moshe Morgenstern. Existence and explicit constructions of $q + 1$ regular Ramanujan graphs for every prime power q. *Journal of Combinatorial Theory. Series B*, 62(1):44–62, 1994. 11, 210

[MS15] Andrea Montanari and Subhabrata Sen. Semidefinite programs on sparse random graphs and their application to community detection. *arXiv preprint arXiv:1504.05910*, 2015. 15, 19, 20

[MSS15a] Adam Marcus, Daniel Spielman, and Nikhil Srivastava. Interlacing families I: Bipartite Ramanujan graphs of all degrees. *Annals of Mathematics. Second Series*, 182(1):307–325, 2015. 11, 212

[MSS15b] Adam Marcus, Daniel Spielman, and Nikhil Srivastava. Interlacing families IV: Bipartite Ramanujan graphs of all sizes. In *Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science*, pages 1358–1377, 2015. 11, 212

[MV17] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *arXiv preprint arXiv:1711.01682*, 2017. 128

[MZ02] M Mézard and R Zecchina. The random k-satisfiability problem: from an analytic solution to an efficient algorithm, 2002 phys. *Rev. E*, 66:056126, 2002. 2

[Nil91] A. Nilli. On the second eigenvalue of a graph. *Discrete Mathematics*, 91(2):207–210, 1991. 10, 210

[Nis01] Hidetoshi Nishimori. *Statistical physics of spin glasses and information processing: an introduction*. Number 111. Clarendon Press, 2001. 119

[NM14] MEJ Newman and Travis Martin. Equitable random graphs. *Physical Review E*, 90(5):052824, 2014. 18

[NN93] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. *SIAM Journal on Computing*, 22(4):838–856, 1993. 216, 217

[NV08] Assaf Naor and Jacques Verstraëte. Parity check matrices and product representations of squares. *Combinatorica*, 28(2):163–185, 2008. 12, 245

[OW20] Ryan O'Donnell and Xinyu Wu. Explicit near-fully X-Ramanujan graphs. *arXiv preprint arXiv:2009.02595*, 2020. 122, 147, 173

[Pin73] Mark Pinsker. On the complexity of a concentrator. In *Proceedings of the 7th International Teletraffic Congress*, pages 318/1–318/4, 1973. 209

[Piz90] Arnold Pizer. Ramanujan graphs and Hecke operators. *American Mathematical Society. Bulletin. New Series*, 23(1):127–137, 1990. 210

[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. 24

[Rag08] Prasad Raghavendra. Optimal algorithms and inapproximability results for every csp? In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 245–254, 2008. 3, 116

[Ran11] Sundeep Rangan. Generalized approximate message passing for estimation with random linear mixing. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 2168–2172. IEEE, 2011. 24

[RRS17] Prasad Raghavendra, Satish Rao, and Tselil Schramm. Strongly refuting random CSPs below the spectral threshold. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 121–131, 2017. 116, 247, 268

[RTSZ19] Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Typology of phase transitions in bayesian inference problems. *Physical Review E*, 99(4):042109, 2019. 119, 121

[RVW02] Omer Reingold, Salil Vadhan, and Avi Wigderson. Entropy waves, the zig-zag graph product, and new constant-degree expanders. *Annals of Mathematics*, 155(1):157–187, 2002. 211

[Sch78] Thomas J Schaefer. The complexity of satisfiability problems. In *Proceedings of the tenth annual ACM symposium on Theory of computing*, pages 216–226, 1978. 3, 116

[Ser77] Jean-Pierre Serre. *Arbres, amalgames,* SL_2. Société Mathématique de France, Paris, 1977. Avec un sommaire anglais, Rédigé avec la collaboration de Hyman Bass, Astérisque, No. 46. 222

[Sho90] Victor Shoup. New algorithms for finding irreducible polynomials over finite fields. *Mathematics of Computation*, 54(189):435–447, 1990. 217

[SKZ17] Alaa Saade, Florent Krzakala, and Lenka Zdeborová. Spectral bounds for the ising ferromagnet on an arbitrary given graph. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(5):053403, 2017. 121

[SLKZ15] Alaa Saade, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Spectral detection in the censored block model. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1184–1188. IEEE, 2015. 120

[Sly09] Allan Sly. Reconstruction for the potts model. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 581–590, 2009. 121

[SM20] Ludovic Stephan and Laurent Massoulié. Non-backtracking spectra of weighted inhomogeneous random graphs. *arXiv preprint arXiv:2004.07408*, 2020. 122, 125

[SS12] Warren Schudy and Maxim Sviridenko. Concentration and moment inequalities for polynomials of independent random variables. In *Proceedings of the twenty-third annual ACM-SIAM symposium on Discrete Algorithms*, pages 437–446. SIAM, 2012. 246, 248, 257

[SVC16] Jacob Steinhardt, Gregory Valiant, and Moses Charikar. Avoiding imposters and delinquents: Adversarial crowdsourcing and peer prediction. In *Advances in Neural Information Processing Systems*, pages 4439–4447, 2016. 15

[Tro15] Joel A Tropp. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015. 247

[WAM19] Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore. The Kikuchi hierarchy and tensor PCA. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1446–1468. IEEE, 2019. 247, 248

[WF09] Yusuke Watanabe and Kenji Fukumizu. Graph zeta function in the Bethe free energy and loopy belief propagation. In *Proceedings of the 23rd Annual Conference on Neural Information Processing Systems*, pages 2017–2025, 2009. 223

[WF11] Yusuke Watanabe and Kenji Fukumizu. Loopy belief propagation, bethe free energy and graph zeta function. *arXiv preprint arXiv:1103.0605*, 2011. 80, 83

[Wik22] Wikipedia contributors. Moore graph — Wikipedia, the free encyclopedia, 2022. [Online; accessed 12-July-2022]. 244

[Wor99] Nicholas Wormald. Models of random regular graphs. In *Surveys in combinatorics, 1999 (Canterbury)*, volume 267 of *London Mathematical Society Lecture Note Series*, pages 239–298. Cambridge Univ. Press, Cambridge, 1999. 231

[Zhu20] Dmitriy Zhuk. A proof of the csp dichotomy conjecture. *Journal of the ACM (JACM)*, 67(5):1–78, 2020. 3, 116

[ZK11] Lenka Zdeborová and Florent Krzakala. Quiet planting in the locked constraint satisfaction problems. *SIAM Journal on Discrete Mathematics*, 25(2):750–770, 2011. 121

[ZK16] Lenka Zdeborová and Florent Krzakala. Statistical physics of inference: Thresholds and algorithms. *Advances in Physics*, 65(5):453–552, 2016. 24, 117, 119

[ZMZ14] Pan Zhang, Cristopher Moore, and Lenka Zdeborová. Phase transitions in semisupervised clustering of sparse networks. *Physical Review E*, 90(5), Nov 2014. 130, 136