Title
Spectral analysis of sparse random graphs and hypergraphs

Author
Zhu, Yizhe

Publication Date
2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Spectral analysis of sparse random graphs and hypergraphs

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Mathematics

by

Yizhe Zhu

Committee in charge:

Professor Ioana Dumitriu, Chair
Professor Mikhail Belkin
Professor Todd Kemp
Professor Rayan Saab
Professor Jason Schweinsberg

2021

The dissertation of Yizhe Zhu is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

Dedicated to my parents and grandparents.

EPIGRAPH

A mathematician is a person who can find analogies between theorems; a better mathematician is one who can see analogies between proofs and the best mathematician can notice analogies between theories. One can imagine that the ultimate mathematician is one who can see analogies between analogies.

—Stefan Banach

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Ioana Dumitriu, for her guidance and support over the years of my PhD study. She has led me to the beautiful field of random matrix theory and broadened my horizons for its extensive applications. This thesis would not have been possible without her. Her kindness and encouragement helped me become a more independent researcher. I feel so honored and lucky to have such a great mathematician and, more importantly, such a wonderful person as my advisor.

I also thank Mikhail Belkin, Todd Kemp, Rayan Saab, and Jason Schweinsberg for serving on my dissertation committee and for their insightful feedback.

Thank you to Soumik Pal for sharing his wisdom in probability theory and suggesting interesting topics on tensors and hypergraphs. I am very fortunate to work with him.

I thank my excellent collaborators: Sam Cole, Simon Coste, Yu-Chen Cheng, Kameron Decker Harris, and Zhixin Zhou. I have learned tremendously from many productive discussions with them.

I want to thank Nicholas Cook, László Erdős, Christopher Hoffman, Antti Knowles, Torben Krüger, Johannes Lederer, Marina Meila, Mark Rudelson, Roland Speicher, Roman Vershynin, Ke Wang, and Yaokun Wu for their valuable advice on my research.

I particularly want to thank my peers in probability: Lucas Benigni, Shuntao Chen, Joonyong Choi, David Clancy, Xiucai Ding, Yingjia Fu, Yukun He, Jiaoyang Huang, Yiping Hu, Yun Li, Mounmanti Podder, Jacob Richey, Yandi Shen, Ludovic Stephan, Nathan Noiry, Xuan Wu, Zhichao Wang, Fan Yang, and Sheng Yin. I am grateful for their encouragement during my PhD years.

Finally, I thank my parents, and Jingwen, for their love and support.

Chapter 2 is extracted from "Yizhe Zhu. A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Structures & Algorithms*, 56.1 (2020): 251-279". The thesis author is the author of this paper.

Chapter 3 is extracted from the preprint "Ioana Dumitriu and Yizhe Zhu. Global eigenvalue fluctuations of random biregular bipartite graphs. *arXiv:2008.11760* (2020)". The thesis author is the co-author of this paper.

Chapter 4 is extracted from "Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, (2021): 1-57". The thesis author is the co-author of this paper.

VITA

2015	B. S. in Mathematics and Applied Mathematics, Shanghai Jiao Tong University
2015–2019	Ph. D. Student in Mathematics, University of Washington
2018	M. S. in Mathematics, University of Washington
2019–2021	Graduate Teaching Assistant, University of California San Diego
2021	Ph. D. in Mathematics, University of California San Diego

PUBLICATIONS

Ioana Dumitriu and Yizhe Zhu. Sparse general Wigner-type matrices: Local law and eigenvector delocalization. *Journal of Mathematical Physics*, 60.2 (2019): 023301.

Ioana Dumitriu and Yizhe Zhu. Spectra of random regular hypergraphs. *arXiv:1905.06487* (2019).

Kameron Decker Harris and Yizhe Zhu. Deterministic tensor completion with hypergraph expanders. *arXiv:1910.10692* (2019).

Yizhe Zhu. A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Structures & Algorithms*, 56.1 (2020): 251-279.

Sam Cole and Yizhe Zhu. Exact recovery in the hypergraph stochastic block model: A spectral algorithm. *Linear Algebra and its Applications*, 593 (2020): 45-73.

Yizhe Zhu. On the second eigenvalue of random bipartite biregular graphs. *arXiv:2005.08103* (2020).

Ioana Dumitriu and Yizhe Zhu. Global eigenvalue fluctuations of random biregular bipartite graphs. *arXiv:2008.11760* (2020).

Simon Coste and Yizhe Zhu. Eigenvalues of the non-backtracking operator detached from the bulk. *Random Matrices: Theory and Applications*, (2020): 2150028.

Yu-Chen Cheng, Hong Qian, and Yizhe Zhu. Asymptotic behavior of a sequence of conditional probability distributions and the canonical ensemble. *Annales Henri Poincaré*, 22 (2021): 1561–1627.

Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, (2021): 1-57.

Zhixin Zhou and Yizhe Zhu. Sparse random tensors: Concentration, regularization and applications. *Electronic Journal of Statistics*, 15.1 (2021): 2483-2516.

ABSTRACT OF THE DISSERTATION

Spectral analysis of sparse random graphs and hypergraphs

by

Yizhe Zhu

Doctor of Philosophy in Mathematics

University of California San Diego, 2021

Professor Ioana Dumitriu, Chair

This thesis concerns the spectral and combinatorial properties of sparse random graphs and hypergraphs. We present three models, including inhomogeneous random graphs, random bipartite biregular graphs, and the hypergraph stochastic block model, emphasizing the limiting spectral distributions, eigenvalue fluctuations, and top eigenvalues and eigenvectors, respectively. We first present a graphon approach to finding the limiting spectral distribution of Wigner-type matrices, building a connection between random matrices and graph limits. For random bipartite biregular graphs, we analyze their cycle structure and compute the global eigenvalue fluctuations. Finally, we study a spectral algorithm for community detection in sparse hypergraph stochastic block models using a new matrix that counts self-avoiding walks on hypergraphs.

Chapter 1

Introduction

1.1 Sparse graph-based random matrices

The spectral analysis of random graphs is in the intersection of random matrix theory and graph theory. It combines tools from combinatorics, probability, statistics, and numerical linear algebra and has various applications in coding theory, network analysis, and machine learning.

The synergy of combinatorics and probability in the study of spectra of random graphs has been successfully developed in the past decade. From the random matrix perspective, we would like to understand how the structure of the random graphs is reflected in their spectral properties and how the universality phenomenon can be extended to sparse random graph models [82, 83, 108, 117, 27, 29, 28]. For dense graphs, the similarities with their adjacency matrices and classical random matrices are so great that one can extend the universality results over from random matrix theory; however, the dense graphs are less interesting from an application perspective. A real-world social or collaborative network is rarely dense; on the contrary, the degrees of vertices remain bounded or grow very slowly as the size of the network increases. So it makes sense to examine those graph-based matrices whose models are not dense but sparse.

The sparsity creates an immediate issue: sparse graph-based random matrices do not satisfy the bounded moments assumption in many results of the Wigner and Wishart ensembles. As such, new tools and methods have been developed to understand their spectra [39, 40, 170, 108, 78, 31, 32, 14]. Often, these new methods still necessitate concentration results that are not true in very sparse regimes (corresponding to bounded expected degree or $O(\log n)$ expected degrees), when few entries per row are nonzero [39, 81, 121, 14, 15]. In such regimes, one has to try and exploit the combinatorial properties of the model. Powerful tools from combinatorics, including graph limits [34], non-backtracking operators [167], switchings [144], etc., have been successfully applied to various sparse random matrix problems [39, 40, 37, 48, 67, 90, 66].

From an application perspective, sparse random matrix techniques have been widely used to analyze the structure of random graphs and provide the theoretical foundation for applications in statistics, theoretical computer science, and data science, including community detection [1], matrix completion [122], matrix sketching [174], error-correcting codes [114], etc. A remarkable achievement in the last decade is the success of proving the community detection threshold conjecture for the stochastic block model in different sparsity regimes [145, 146, 141, 40, 145, 2]. Many efficient algorithms that achieve the information-theoretical thresholds are spectral, and their analysis is based on random matrix theory [141, 40, 4].

1.1.1 Random graph and random hypergraph models

Many random graph and hypergraph models have been extensively studied from a random matrix perspective in the last decade. We briefly introduce several models that will be studied in this thesis. A more detailed discussion can be found in each chapter.

Inhomogeneous Erdős-Rényi random graphs

One of the most basic models for random graphs is the *Erdős-Rényi random graph*, denoted by $\mathcal{G}(n,p)$, where every edge between two vertices appears with probability p. The inhomogeneous Erdős-Rényi model $\mathcal{G}(n,(p_{ij}))$, where edges exist independently with given probabilities p_{ij}, is a generalization of the classical Erdős-Rényi model $\mathcal{G}(n,p)$.

Many popular graph models arise as special cases of $\mathcal{G}(n,(p_{ij}))$ such as random graphs with given expected degrees [64], stochastic block models [113], and W-random graphs [136, 42]. It is a popular topic attracting attention from different areas: the limiting spectral distribution and generalization of the universality phenomenon beyond Wigner matrices in random matrix theory [178, 56, 8]; the study of graph limits and inhomogeneous random graph models beyond the Erdős-Rényi graphs in random graph theory [135, 36]; community detection and network analysis in statistics and machine learning [1, 97].

Stochastic block models

The *Stochastic Block Model* (SBM) represents a generalization of Erdős-Rényi graphs to allow for more heterogeneity. This model is designed to produce graphs containing communities and to serve as a benchmark for clustering algorithms. Specifically, let A be the adjacency matrix of an SBM. Suppose we have a partition of $[n] = V_1 \cup V_2 \cup \ldots \cup V_d$ for some integer d, and that $|V_i| = n_i$ for $i = 1, \ldots, d$. For any pair $(k,l) \in [d] \times [d]$, there is a $p_{kl} \in [0,1]$ such that for any $i \in V_k$, $j \in V_l$,

$$a_{ij} = \begin{cases} 1, & \text{with probability } p_{kl}, \\ 0, & \text{otherwise.} \end{cases}$$

The task for community detection is to find the unknown partition of a random graph sampled from the SBM. In the last decade, there has been considerable activity [129, 40, 1, 31, 32] in understanding the spectral properties of matrices associated with the SBM and other generalized graph models, in particular in connection to spectral clustering methods.

Random regular and bipartite biregular graphs

An expander graph is a sparse graph with connectivity properties which exhibits rapid mixing [63]. Expander graphs play an important role in computer science, including sampling, complexity theory, and the design of error-correcting codes (see [114]).

When a graph is *d-regular*, i.e., each vertex has degree d, quantification of expansion is possible based on the eigenvalues of the adjacency matrix. By random d-regular graph on n vertices, we mean a random graph chosen uniformly from the space of all simple d-regular graphs on n vertices. The adjacency matrix of a random d-regular graph is a random matrix model with dependent entries. This has been extensively analyzed in the last decade with respect to global and local statistics and universality [77, 170, 76, 27, 29, 28, 117], and the spectral gap [49, 37, 67, 168].

In many applications, one would like to construct bipartite expander graphs with two unbalanced disjoint vertex sets, among which bipartite biregular graphs are of particular interest. An (n,m,d_1,d_2)*-bipartite biregular graph* is a bipartite graph $G=(V_1,V_2,E)$ where $|V_1|=n, |V_2|=m$ and every vertex in V_1 has degree d_1 and every vertex in V_2 has degree d_2. Note that we must have $nd_1=md_2=|E|$. The spectra of random bipartite biregular graphs have been studied in [75, 169, 175, 48, 179, 80].

Random hypergraphs

A *hypergraph* H consists of a set V of vertices and a set E of hyperedges such that each hyperedge is a nonempty set of V. A hypergraph H is *k-uniform* for an integer $k \geq 2$ if every hyperedge $e \in E$ contains exactly k vertices. The degree of i, denoted $\deg(i)$, is the number of hyperedges incident to i. A hypergraph is *d-regular* if all of its vertices have degree d. A hypergraph is (d,k)*-regular* if it is both d-regular and k-uniform.

Many clustering methods are based on graphs, which represent pairwise relationships among objects. However, in many real-world problems, pairwise relations are not sufficient, while higher-order relations between objects cannot be fully described as edges on graphs. Hypergraphs can be used to represent more complex relationships among data, and they have been shown empirically to have advantages over graphs; see [176, 152]. Thus, it is of practical interest to develop algorithms based on hypergraphs, and much work has already been done to that end; see, for example, [176, 130, 172, 100, 50, 109, 11]. There are two natural generalizations of random graph models to random hypergraphs: *Erdős-Rényi random hypergraphs* and *random regular hypergraphs*.

A k-uniform Erdős-Rényi random hypergraph $\mathcal{H}_k(n,p)$ is a random hypergraph on n vertices, where each hyperedge is of size k and appears independently with probability p. The spectra of the Laplacian, adjacency matrix, and the adjacency tensor for Erdős-Rényi random hypergraphs have been studied in [137, 177]. Inhomogeneous random hypergraphs and hypergraph stochastic block models are popular in the fields of network modeling and clustering [131, 124, 98, 20, 65, 151].

Random regular hypergraphs, where each hyperedge has the same size and each vertex has the same degree, serve as a natural model for the study of hypergraphs with regularity and dependency [74, 127]. They have been used to study the average behavior of optimization algorithms on hypergraphs [154, 72]. The spectra of such random hypergraph models have been analyzed through their adjacency matrix [91, 132, 79] and adjacency tensor [95].

1.1.2 Spectral statistics

In this section, we define the spectral statistics that will be discussed in this thesis.

Empirical spectral distribution

For any $n \times n$ Hermitian matrix A with eigenvalues $\lambda_1, \dots, \lambda_n$, the *empirical spectral distribution* (ESD) of A is defined by

$$\mu_A(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i}(x). \tag{1.1.1}$$

The limiting spectral distribution, or the global law, describes the limit shape of the spectrum. A Wigner matrix is a Hermitian random matrix whose entries are i.i.d. random variables up to the symmetry constraint and have zero expectation and variance 1. As has been known since Wigner's seminal paper [173] in various formats, for Wigner matrices, the empirical spectral distribution converges almost surely to the semicircle law, with a density function

$$\mu(x) = \frac{1}{2\pi}\sqrt{4-x^2}\mathbf{1}_{[-2,2]}(x).$$

Global eigenvalue fluctuations

A *linear statistic* of an $n \times n$ matrix A with eigenvalues $\lambda_1, \ldots, \lambda_n$ is a functional of the form

$$\mathcal{L}(f) = \sum_{i=1}^{n} f(\lambda_i),$$

where f is a function belonging to a certain class. When f is a suitable test function, the first order behavior of $\mathcal{L}(f)$, is given by

$$\frac{1}{n}\mathcal{L}(f) = \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i) \to \int_{\mathbb{R}} f(x)d\mu_A(x),$$

where μ_A the limiting spectral distribution of A. The global fluctuation for linear statistics, is another spectral statistic of interest, which examines the second order behavior of $\mathcal{L}(f)$, given by

$$X_f = \mathcal{L}(f) - \mathbb{E}\mathcal{L}(f).$$

Under a suitable scaling, one would like to prove that X_f converges in distribution to a certain random variable whose variance depends on f. The term "global" in "global fluctuation" refers to the fact that all eigenvalues contribute similarly to $\mathcal{L}(f)$.

Spectral gap

Let A be the adjacency matrix of a d-regular graph. The first eigenvalue $\lambda_1(A)$ is always d. The second eigenvalue in absolute value $\lambda(A) = \max\{\lambda_2(A), -\lambda_n(A)\}$ is of particular interest, since the difference between d and λ, also known as the *spectral gap*, provides an estimate on the expansion property of the graph [12, 63, 114].

For a bipartite biregular graph, its first eigenvalue is $\sqrt{d_1 d_2}$. The difference between $\sqrt{d_1 d_2}$ and $\lambda_2(A)$ is called the spectral gap. The spectral gap of bipartite biregular graphs has found applications in error-correcting codes, matrix completion and community detection, see for example [165, 160, 96, 48, 51].

For inhomogeneous Erdős-Rényi graphs, the spectral gap is referred to separation between the few largest eigenvalues outside the bulk of the spectrum (the outliers) and the edge of the bulk spectrum [40, 57, 55, 14].

1.2 The moment method

The three different results we present in the thesis share the same philosophy: the spectral statistics of a random matrix can be studied through the analysis of certain combinatorial objects in the random graph or random matrix model. An important tool of the analysis is the moment method.

Compared to the Stieltjes transform method in the study of local statistics, the moment method has some advantages in the study of global statistics and the edge behavior of the spectrum. For example, in the very sparse regime, including Erdős-Rényi graphs with bounded expected degrees [40, 38], random regular graphs and quasi-regular graphs with fixed degrees [37, 47, 48], many variants of the classical moment method provides stronger results and more precise information of the spectrum compared to the Stieltjes transform method. The advantage is that it incorporates the information of the combinatorial structures even when concentration tools fail.

We give a quick illustration of the moment method in the study of Wigner matrices (see [18] for more details). Let W_n be a random Hermitian matrix with i.i.d. entries of mean zero and variance one, and we assume all moments are finite for simplicity. Denote $A_n = \frac{1}{\sqrt{n}} W_n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. Denote the empirical spectral distribution of A_n by μ_n. Then the k-th moment of μ_{A_n} satisfies

$$\int x^k \mu_{A_n}(x) dx = \frac{1}{n} \mathrm{tr}(A_n^k). \tag{1.2.1}$$

The moment method to show μ_{A_n} has a limit is by first proving the convergence of the expected trace of A_n^k for each fixed k:

$$\frac{1}{n} \mathbb{E}[\mathrm{tr} A_n^k] = \frac{1}{n} \sum_{i_1, \ldots, i_k \in [n]} \mathbb{E}[(A_n)_{i_1 i_2} \cdots (A_n)_{i_k i_1}]. \tag{1.2.2}$$

When k is even, the leading order of the above expression is given by the number of closed walks on rooted planar trees, which matches the moment of a semicircle distribution, i.e., the Catalan number.

For global eigenvalue fluctuations, the goal is to find the limiting law for the linear statistic of a suitable test function f, given by

$$\sum_{i=1}^{n} f(\lambda_i) - \sum_{i=1}^{n} \mathbb{E} f(\lambda_i). \tag{1.2.3}$$

The moment method can be used to examine the fluctuation of linear statistics for a test function $f(x) = x^k$ with a fixed k. The linear statistic of f can be written as

$$X_k := \mathrm{tr}(A_n^k) - \mathbb{E}\mathrm{tr}(A_n^k). \tag{1.2.4}$$

The moments of X_k can be interpreted as counting cycles in a certain graph, and they match the moments of a Gaussian random variable with an explicit variance. This calculation implies a central limit theorem. Then one can extend the CLT to polynomial test functions.

The moment method is also a powerful tool to estimate the spectral norm, defined by

$$\|A_n\| = \sup_{x:\|x\|_2=1} \|A_n x\|_2.$$

Consider the following inequality:

$$\mathbb{E}\|A_n\| \leq \mathbb{E}[\operatorname{tr}(A_n)^{2k}]^{\frac{1}{2k}} \leq n^{\frac{1}{2k}}\mathbb{E}\|A_n\|.$$

When k is on the order of $\log n$, the three quantities are of the same order. Therefore we can bound the spectral norm $\|A\|$ by counting a certain type of closed walk of length $k = O(\log n)$.

We have developed and generalized the moment method to three different random graph and random hypergraph models in order to analyze the three different types of spectral statistics mentioned above. The new challenges, beyond the Wigner matrix example, are inhomogeneity, edge dependence, and sparsity.

In Chapter 2, we consider random matrices $A_n = \frac{1}{\sqrt{n}}W_n$, where entries w_{ij} are independent, wtih mean zero and variance s_{ij}. This type of random matrix is called a Wigner-type matrix with a variance profile [8]. We give a new formula for the moments of the limiting spectral distribution for Wigner-type matrices, in terms of homomorphism densities, a quantity that is well studied in the dense graph limit (i.e., the graphon theory [135]). This is an inhomogeneous generalization of the Catalan number expression for the moments of Wigner matrices.

In Chapter 3, we study the global eigenvalue fluctuation for random bipartite biregular graphs. In this case, the linear eigenvalue statistic $\mathcal{L}(f)$ has a closed-form expression when f is a Chebyshev polynomial, which counts the number of cyclically non-backtracking closed walks in a bipartite biregular graph. Such closed walks can be analyzed through the cycle counts in the graph. In this random graph model with edge dependence, we are able to analyze the law of cycle counts through switching operations [143, 144], a tool from random graph theory. The fluctuations of cycle counts essentially determine the global eigenvalue fluctuations.

In Chapter 4, we develop the moment method for sparse random hypergraphs, to solve a community detection problem in the hypergraph block model [20, 101]. We apply this method to a new matrix that counts self-avoiding walks on hypergraphs, whose spectral norm is analyzed by counting concatenations of self-avoiding walks of length $O(\log n)$.

1.3 Contribution of this thesis

We summarize the main results in each chapter as follows. More detailed introductions will be given in each chapter.

Chapter 2: A graphon approach to the limiting spectral distributions of Wigner-type matrices

We analyze the limiting spectral distributions of general Wigner-type matrices. Such random matrices have independent entries up to symmetry, but with different variances. This approach determines the moments of the limiting measures and the equations of their Stieltjes transforms explicitly with weaker assumptions on the convergence of variance profiles than previous results in [157, 19]. As applications, we determine the limiting spectral distributions for three sparse inhomogeneous random graph models with sparsity $\omega(1/n)$: inhomogeneous random graphs with roughly equal expected degrees, W-random graphs, and stochastic block models with a growing number of blocks. Our theorems can also be applied to study random Gram matrices with a variance profile for which we can find the limiting spectral distributions under weaker assumptions than previous results in [105].

Chapter 3: Global eigenvalue fluctuations for random bipartite biregular graphs

We compute the global eigenvalue fluctuations for the spectral statistic $\mathcal{L}(f)$ of uniformly distributed random biregular bipartite graphs with fixed and growing degrees, for a large class of analytic functions f. As a key step in the proof, we obtain a total variation distance bound for the Poisson approximation of the number of cycles and cyclically non-backtracking walks in random biregular bipartite graphs. As an application, we translate the results to adjacency matrices of uniformly distributed random regular hypergraphs.

Chapter 4: Community detection in the sparse hypergraph stochastic block model

We consider the community detection problem in sparse random hypergraphs. Angelini et al. in [20] conjectured the existence of a sharp threshold on model parameters for community detection in the hypergraph stochastic block model. We settled the positive part of the conjecture for the case of two blocks: above the threshold, there is a spectral algorithm that asymptotically almost surely constructs a partition of the hypergraph which correlated with the true partition. Our method is a generalization to sparse random hypergraphs of the method developed by Massoulié in [141] for sparse random graphs.

Chapter 2

A graphon approach to limiting spectral distributions of Wigner-type matrices

2.1 Introduction

Eigenvalue Statistics of Random Matrices

Classically, as has been known since Wigner's seminal paper [173] in various formats, the empirical spectral distribution for Wigner matrices converges almost surely to the semicircle law. The i.i.d. requirement and the constant variance condition are not essential for proving the semicircle law, as can be seen from the fact that generalized Wigner matrices, whose entries have different variances but each column of the variance profile is stochastic, turned out to obey the semicircle law [19, 86, 102], under various conditions as well. Beyond the semicircle law, the Wigner matrices exhibit universality [85, 166], a phenomenon that has been recently shown to hold for other models, including generalized Wigner matrices [86], adjacency matrices of Erdős-Rényi random graphs [82, 83, 170, 116] and general Wigner-type matrices [8].

A slightly different direction of research is to investigate structured random matrix models whose limiting spectral distribution is not the semicircle law. One such example is random block matrices, whose limiting spectral distribution has been found in [157, 89] using free probability. Ding [73] used moment methods to derive the limiting spectral distribution of random block matrices for a fixed number of blocks (a claim in [73] that the method extends to the growing number of blocks case is unfortunately incorrect). Recently Alt et al. [17] provided a unified way to study the global law for a general class of non-Hermitian random block matrices including Wigner-type matrices.

Graphons and Convergence of Graph Sequences

Understanding large networks is a fundamental problem in modern graph theory and to properly define a limit object, an important issue is to have good definitions of convergence for graph sequences. Graphons, introduced in 2006 by Lovász and Szegedy [136] as limits of dense graph sequences, aim to provide a solution to this question. Roughly speaking, the set of finite graphs endowed with the cut metric (See Definition 2.2.3) gives rise to a metric space, and the completion of this space is the space of graphons. These objects may be realized as symmetric, Lebesgue measurable functions from $[0,1]^2$ to $\mathbb{R}$. They also characterize the convergence of graph sequences based on graph homomorphism densities [44, 45]. Recently, graphon theory has been generalized for sparse graph sequences [42, 43, 93, 126].

The most relevant results for our endeavor are the connections between two types of convergences: left convergence in the sense of homomorphism densities and convergence in cut metric. In our approach, for the general Wigner-type matrices, we will regard the variance profile matrices S_n as a graphon sequence. The convergence of empirical spectral distributions is connected to the convergence of this graphon sequence associated with S_n in either left convergence sense or in cut metric.

For inhomogeneous random graphs with bounded expected degree introduced by Bollobás, Janson and Riordan [36], their graphon limits will be 0 and our main result will not cover this regime. This is because the graphon limit is only suitable for graph sequences with unbounded degrees. Instead, the spectrum of random graphs with bounded expected degrees was studied in [39] by local weak convergence [34, 10], a graph limit theory for graph sequences with bounded degrees.

Contribution of this chapter

We obtained a formula to compute the moments of limiting spectral distributions of general Wigner-type matrices from graph homomorphism densities, and we derived quadratic vector equations as in [7] from this formula. Previous approaches to the problem require the variance profiles to converge to a function whose set of discontinuities has measure zero [157, 19, 105], we make no such requirement here. The method in [157] is based on free probability theory, and it is assumed that all entries of the matrix are Gaussian, while our Theorem 2.3.2 and Theorem 2.3.4 work for non-i.i.d. entries with general distributions. Especially, we cover a variety of sparse matrix models (see Section 2.4-2.7). The argument in [19] is based on a sophisticated moment method for band matrix models, and our moment method proof based on graphon theory is much simpler and can be applied to many different models including random Gram matrices. For random Gram matrices, in [105], it is assumed that all entries have $(4+\varepsilon)$ moments and the variance profile is continuous. The continuity assumption is used to show the Stieltjes transform of the empirical measure converges to the Stieltjes transform of the limiting measure. We remove the technical higher moments and the continuity assumptions since our combinatorial approach requires less regularity.

All the previous results above assume the limiting variance profile exists and is continuous. This assumption is used to have an error control under L^∞-norm between the n-step variance profile and the limiting variance profile, which will guarantee that either the moments of the empirical measure converge or the Stieltjes transform the empirical measure converges. However, this L^∞-convergence is only a stronger sufficient condition compared to our condition in Theorem 2.3.2 and Theorem 2.3.4. The key observation in our approach is that permuting a random matrix does not change its spectrum, but the continuity of the variance is destroyed. The cut metric in the graphon theory is a suitable tool to exploit the permutation invariant property of the spectrum (see Theorem 2.3.4).

Moreover, we realize that to make the moments of the empirical measure converge, we don't need to assume the moments of the limiting measure is an integral in terms of the limiting variance profile. All we need is the convergence of homomorphism density from trees. We show two examples in Section 2.4 where we don't have a limiting variance profile but the moments of the empirical measure still converge: generalized Wigner matrices and inhomogeneous random graphs with roughly equal expected degrees.

Besides, if the limiting distribution is not the semicircle law, previous results only implicitly characterize the Stieltjes transform of the limiting measure by the quadratic vector equations (see (2.3.2), (2.3.3)), which are not easy to solve. Our combinatorial approach explicitly determines the moments of the limiting distributions in terms of sums of graphon integrals. Our convergence condition (see Theorem 2.3.2 (1)) is the weakest so far for the existence of limiting spectral distributions and covers a variety of models like generalized Wigner matrices, adjacency matrices of sparse stochastic block models with a growing number of blocks, and random Gram matrices.

The organization of this chapter is as follows: In Section 2.2, we introduce definitions and facts that will be used in our proofs. In Section 2.3, we state and prove the main theorems for general Wigner-type matrices and then specialize our results to different models in Section 2.4-2.7. In Section 2.8, we extend our results to random Gram matrices with a variance profile.

2.2 Preliminaries

Our main task in this chapter is to investigate the convergence of the sequence of empirical spectral distribution to the limiting spectral distribution for a given sequence of structured random matrices. A useful tool to study the convergence of measure is the Stieltjes transform. Let μ be a probability measure on $\mathbb{R}$. The *Stieltjes transform* of μ is a function $s(z)$ defined on the upper half plane $\mathbb{C}^+$ by the formula:

$$s(z) = \int_{\mathbb{R}} \frac{1}{z-x} d\mu(x), \quad z \in \mathbb{C}^+.$$

Suppose that μ is compactly supported, and denote $r := \sup\{|t| \mid t \in \operatorname{supp}(\mu)\}$. We then have a power series expansion

$$s(z) = \sum_{k=0}^{\infty} \frac{\beta_k}{z^{k+1}}, \quad |z| \geq r, \tag{2.2.1}$$

where $\beta_k := \int_{\mathbb{R}} x^k d\mu(x)$ is the k-th moment of μ for $k \geq 0$.

Definition 2.2.1. The *rooted planar tree* is a planar graph with no cycles, with one distinguished vertex as a root, and with a choice of ordering at each vertex. The ordering defines a way to explore the tree starting at the root. *Depth-first search* is an algorithm for traversing rooted planar trees. One starts at the root and explores as far as possible along each branch before backtracking. An enumeration of the vertices of a tree is said to have *depth-first search order* if it is the output of the depth-first search.

The Dyck paths of length $2k$ are bijective to rooted planar trees of $k+1$ vertices by the depth-first search (see Lemma 2.1.6 in [18]). Hence the number of rooted planar trees with $k+1$ vertices is the k-th Catalan number $C_k := \frac{1}{k+1}\binom{2k}{k}$.

We introduce definitions from graphon theory. For more details, see [135].

Definition 2.2.2. A *graphon* is a symmetric, integrable function $W : [0,1]^2 \to \mathbb{R}$.

Here symmetric means $W(x,y) = W(y,x)$ for all $x, y \in [0,1]$. Every weighted graph G has an associated graphon W^G constructed as follows. First divide the interval $[0,1]$ into intervals

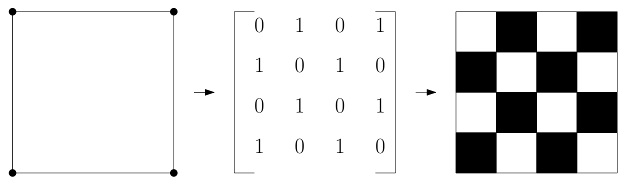

Figure 2.1: A graphon representation of a cycle of length 4

$I_1, \ldots, I_{|V(G)|}$ of length $\frac{1}{|V(G)|}$, then give the edge weight β_{ij} on $I_i \times I_j$, for all $i, j \in V(G)$. In this way, every finite weighted graph gives rise to a graphon (see Figure 2.1).

The most important metric on the space of graphons is the cut metric. The space that contains all graphons taking values in $[0,1]$ endowed with the cut metric is a compact metric space.

Definition 2.2.3. For a graphon $W : [0,1]^2 \to \mathbb{R}$, the *cut norm* is defined by

$$\|W\|_\square := \sup_{S,T \subseteq [0,1]} \left| \int_{S \times T} W(x,y) dxdy \right|,$$

where S, T range over all measurable subsets of $[0,1]$. Given two graphons $W, W' : [0,1]^2 \to \mathbb{R}$, define $d_\square(W,W') := \|W - W'\|_\square$ and the *cut metric* $\delta_\square$ is defined by

$$\delta_\square(W,W') := \inf_\sigma d_\square(W^\sigma, W'),$$

where σ ranges over all measure-preserving bijections $[0,1] \to [0,1]$ and

$$W^\sigma(x,y) := W(\sigma(x), \sigma(y)).$$

Using the cut metric, we can compare two graphs with different sizes and measure their similarity, which defines a type of convergence of graph sequences whose limiting object is the graphon we introduced. Another way of defining the convergence of graphs is to consider graph homomorphisms.

Definition 2.2.4. For any graphon W and multigraph $F = (V,E)$ (without loops), define the *homomorphism density* from F to W as

$$t(F,W) := \int_{[0,1]^{|V|}} \prod_{ij \in E} W(x_i,x_j) \prod_{i \in V} dx_i.$$

One may define *homomorphism density from partially labeled graphs to graphons*, as follows.

Definition 2.2.5. Let $F = (V,E)$ be a k-labeled multigraph. Let $V_0 = V \setminus [k]$ be the set of unlabeled vertices. For any graphon W, and $x_1,\dots,x_k \in [0,1]$, define

$$t_{x_1,\dots,x_k}(F,W) := \int_{x \in [0,1]^{|V_0|}} \prod_{ij \in E} W(x_i,x_j) \prod_{i \in V_0} dx_i. \tag{2.2.2}$$

This is a function of $x_1,\dots,x_k$.

It is natural to think two graphons W and W' are similar if they have similar homomorphism densities from any finite graph G. This leads to the following definition of left convergence.

Definition 2.2.6. Let W_n be a sequence of graphons. We say W_n is *convergent from the left* if $t(F,W_n)$ converges for any finite simple (no loops, no multi-edges, no directions) graph F.

The importance of homomorphism densities is that they characterize convergence under the cut metric. Let $\mathcal{W}_0$ be the set of all graphons such that $0 \le W \le 1$. The following is a characterization of convergence in the space $\mathcal{W}_0$, known as Theorem 11.5 in [135].

Theorem 2.2.7. *Let $\{W_n\}$ be a sequence of graphons in $\mathcal{W}_0$ and let $W \in \mathcal{W}_0$. Then $t(F,W_n) \to t(F,W)$ for all finite simple graphs if and only if $\delta_\square(W_n,W) \to 0$.*

2.3 Main results for general Wigner-type matrices

2.3.1 Set-up and main results

Let A_n be a Hermitian random matrix whose entries above and on the diagonal of A_n are independent. Assume a *general Wigner-type matrix* A_n with a variance profile matrix S_n satisfies the following conditions:

1. $\mathbb{E}a_{ij} = 0, \mathbb{E}|a_{ij}|^2 = s_{ij}$.

2. (Lindeberg's condition) for any constant $\eta > 0$,

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{1\leq,i,j\leq n} \mathbb{E}[|a_{ij}|^2 \mathbf{1}(|a_{ij}| \geq \eta\sqrt{n})] = 0. \tag{2.3.1}$$

3. $\sup_{ij} s_{ij} \leq C$ for some constant $C \geq 0$.

Remark 2.3.1. If we assume entries of A_n are of the form $a_{ij} = s_{ij}\xi_{ij}$ where the ξ_{ij}'s have mean 0, variance 1 and are i.i.d. up to symmetry, then the Lindeberg's condition (2.3.1) holds by the Dominated Convergence Theorem.

To begin with, we associate a graphon W_n to the matrix S_n in the following way. Consider S_n as the adjacency matrix of a weighted graph G_n on $[n]$ such that the weight of the edge (i, j) is s_{ij}, then W_n is defined as the corresponding graphon to G_n. We say W_n is a *graphon representation* of S_n. We define $M_n := \frac{1}{\sqrt{n}}A_n$ and denote all rooted planar tree with $k+1$ vertices as $T_j^{k+1}, 1 \leq j \leq C_k$. Now we are ready to state our main results for the limiting spectral distributions of general Wigner-type matrices.

Theorem 2.3.2. *Let A_n be a general Wigner-type matrix and W_n be the corresponding graphon of S_n. The following holds:*

1. *If for any finite tree T, $t(T,W_n)$ converges as $n \to \infty$, the empirical spectral distribution of M_n converges almost surely to a probability measure μ such that for $k \geq 0$,*

$$\int x^{2k} d\mu = \sum_{j=1}^{C_k} \lim_{n\to\infty} t(T_j^{k+1}, W_n), \quad \int x^{2k+1} d\mu = 0.$$

2. *If $\delta_\square(W_n, W) \to 0$ for some graphon W as $n \to \infty$, then for all $k \geq 0$,*

$$\int x^{2k} d\mu = \sum_{j=1}^{C_k} t(T_j^{k+1}, W), \quad \int x^{2k+1} d\mu = 0.$$

Remark 2.3.3. Similar moment formulas appear in the study of traffic distributions in free probability theory [138, 139].

Using the connection between the moments of the limiting spectral distribution and its Stieltjes transform described in (2.2.1), we can derive the equations for the Stieltjes transform of the limiting measure by the following theorem.

Theorem 2.3.4. *Let A_n be a general Wigner-type matrix and W_n be the corresponding graphon of S_n. If $\delta_\square(W_n, W) \to 0$ for some graphon W, then the empirical spectral distribution of $M_n := \frac{A_n}{\sqrt{n}}$ converges almost surely to a probability measure μ whose Stieltjes transform $s(z)$ is an analytic solution defined on $\mathbb{C}^+$ by the following equations:*

$$s(z) = \int_0^1 a(z,x)dx, \tag{2.3.2}$$

$$a(z,x)^{-1} = z - \int_0^1 W(x,y)a(z,y)dy, \quad x \in [0,1], \tag{2.3.3}$$

where $a(z,x)$ is the unique analytic solution of (2.3.3) *defined on $\mathbb{C}^+ \times [0,1]$.*

Moreover, for $|z| > 2\|W\|_\infty^{1/2}$,

$$a(z,x) = \sum_{k=0}^{\infty} \frac{\beta_{2k}(x)}{z^{2k+1}}, \quad \beta_{2k}(x) := \sum_{j=1}^{C_k} t_x(T_j^{k+1}, W), \tag{2.3.4}$$

$$\textit{where} \quad t_{x_1}(T_j^{k+1}, W) := \int_{[0,1]^k} \prod_{uv \in E(T_j^{k+1})} W(x_u, x_v) \prod_{i=2}^{k+1} dx_i. \tag{2.3.5}$$

Remark 2.3.5. In (2.3.5), $t_{x_1}(T_j^{k+1}, W)$ is a function of x_1, and in (2.3.4) $t_x(T_j^{k+1}, W)$ is the function evaluated at $x_1 = x$.

Theorem 2.3.4 holds under a stronger condition compared to Theorem 2.3.2. We provide two examples in Section 2.4 to show that it's possible to have tree densities converge but the empirical graphon does not converge under the cut metric. We show that the limiting spectral distribution can still exist. However, to have the equations (2.3.2) and (2.3.3), we need a well-defined measurable function W that W_n converges to, therefore we need the condition of graphon convergence under the cut metric.

(2.3.2) and (2.3.3) have been known as *quadratic vector equations* in [7, 9], where the properties of the solution are discussed under more assumptions on variance profiles to prove local law and universality. A similar expansion as (2.3.4) and (2.3.5) has been derived in [84]. The central role of (2.3.3) in the context of random matrices has been recognized by many authors, see [103, 157, 110].

Wigner-type matrices is a special case for the Kronecker random matrices introduced in [17], and the global law has been proved in Theorem 2.7 of [17], which states the following: let H_n be a Kronecker random matrix and μ_n^H be its empirical spectral distribution, then there exists a deterministic sequence of probability measure μ_n such that $\mu_n^H - \mu_n$ converges weakly in probability to the zero measure as $n \to \infty$. In particular, for Wigner-type matrices, the global law holds under the assumptions of bounded variances and bounded moments. Our Theorem 2.3.2 and Theorem 2.3.4 give a moment method proof of the global law in [17] for Wigner-type matrices under bounded variances and Lindeberg's condition. Our new contribution is a weaker condition for the convergence of the empirical spectral distribution μ_n^M of M_n.

In Section 2.3.2 and Section 2.3.3 we provide the proofs for Theorem 2.3.2 and Theorem 2.3.4 respectively. We briefly summarize the proof ideas here. In the proof of Theorem 2.3.2, we revisit the standard path-counting moment method proof for the semicircle law (see for example [23]). Since our matrix model has a variance profile, we encode different variances as weights on the paths and represent the moments of the empirical measure as a sum of homomorphism densities. Then if the tree homomorphism densities converge, the limiting spectral distribution exists.

For the proof of Theorem 2.3.4, since we assume that the variance profile convergences under the cut norm, we can obtain a limiting graphon W. To obtain (2.3.3) We expand $a(z,x)$ in (2.3.3) as a power series of homomorphism density from partially labeled trees to graphon W denoted by $\beta_{2k}(x)$ in (2.3.4). Then we prove a graphon version of the Catalan number recursion formula for $\beta_{2k}(x)$ in (2.3.11) and show that this essentially implies the quadratic vector equations (2.3.2) and (2.3.3). This recursion formula (2.3.11) for tree homomorphism densities to a graphon could be of independent interest.

2.3.2 Proof of Theorem 2.3.2

Using the truncation argument as in [23, 73], we can first apply moment methods to a general Wigner-type matrix with bounded entries in the following lemma.

Lemma 2.3.6. *Assume a Hermitian random matrix A_n with a variance profile S_n satisfies*

1. $\mathbb{E}a_{ij} = 0, \mathbb{E}|a_{ij}|^2 = s_{ij}$. $\{a_{ij}\}_{1\leq i,j\leq n}$ *are independent up to symmetry.*

2. $|a_{ij}| \leq \eta_n\sqrt{n}$ *for some positive decreasing sequence* η_n *such that* $\eta_n \to 0$.

3. $\sup_{ij} s_{ij} \leq C$ *for a constant* $C \geq 0$.

Let W_n be the graphon representation of S_n. Then for every fixed integer $k \geq 0$, we have the following asymptotic formulas:

$$\frac{1}{n}\mathbb{E}[\mathrm{tr} M_n^{2k}] = \sum_{j=1}^{C_k} t(T_j^{k+1}, W_n) + o(1), \tag{2.3.6}$$

$$\frac{1}{n}\mathbb{E}[\mathrm{tr} M_n^{2k+1}] = o(1), \tag{2.3.7}$$

where $\{T_j^{k+1}, 1 \leq j \leq C_k\}$ are all rooted planar trees of $k+1$ vertices.

Proof. We start with expanding the expected normalized trace. For any integer $h \geq 0$,

$$\frac{1}{n}\mathbb{E}[\mathrm{tr} M_n^h] = \frac{1}{n^{h/2+1}}\mathbb{E}\mathrm{tr}(A_n^h) = \frac{1}{n^{h/2+1}} \sum_{1 \leq i_1, \ldots, i_h \leq n} \mathbb{E}[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_h i_1}].$$

Each term in the above sum corresponds to a closed walk (with possible self-loops) $(i_1, i_2, \ldots, i_h)$ of length h in the complete graph K_n on vertices $\{1, \ldots, n\}$. Any closed walk can be classified into one of the following three categories.

- $\mathcal{C}_1$: All closed walks such that each edge appears exactly twice.
- $\mathcal{C}_2$: All closed walks that have at least one edge which appears only once.
- $\mathcal{C}_3$: All other closed walks.

By independence, it's easy to see that every term corresponding to a walk in $\mathcal{C}_2$ is zero. We call a walk that is not in $\mathcal{C}_2$ a *good walk*. Consider a good walk that uses p different edges $e_1, \ldots, e_p$ with corresponding multiplicity $t_1, \ldots, t_p$ and each $t_i \geq 2$, such that $t_1 + \cdots + t_p = h$. Now the term corresponding to a good walk has the form $\mathbb{E}[a_{e_1}^{t_1} \cdots a_{e_p}^{t_p}]$. Such a walk uses at most $p+1$ vertices and an upper bound for the number of good walks of this type is $n^{p+1} p^h$. Since $|a_{ij}| \leq \eta_n \sqrt{n}$, and $\sup_{ij} \mathrm{Var}(a_{ij}) = \sup_{ij} s_{ij} \leq C$, we have

$$\mathbb{E} a_{e_1}^{t_1} \cdots a_{e_p}^{t_p} \leq \mathbb{E}[a_{e_1}^2] \cdots \mathbb{E}[a_{e_p}^2](\eta_n \sqrt{n})^{t_1 + \cdots + t_p - 2p} \leq \eta_n^{h-2p} n^{h/2-p} C^p.$$

When $h = 2k+1$, we have

$$\begin{aligned}
\frac{1}{n}\mathbb{E}[\mathrm{tr}M_n^{2k+1}] &= \frac{1}{n^{h/2+1}}\sum_{p=1}^{k}\sum_{\text{good walks of } p \text{ edges}}\mathbb{E}[a_{e_1}^{t_1}\cdots a_{e_p}^{t_p}] \\
&\leq \frac{1}{n^{k+3/2}}\sum_{p=1}^{k} n^{p+1}p^h(\eta_n^{h-2p}n^{h/2-p})C^p \\
&= \sum_{p=1}^{k} p^h\eta_n^{h-2p}C^p = O(\eta_n) = o(1).
\end{aligned}$$

When $h = 2k$, let S_i denote the sum of all terms in $\mathcal{C}_i, 1 \leq i \leq 3$. By independence, we have $S_2 = 0$. Each walk in $\mathcal{C}_3$ uses p different edges with $p \leq k-1$. We then have

$$\begin{aligned}
S_3 &= \frac{1}{n^{h/2+1}}\sum_{p=1}^{k-1}\sum_{\text{good walk of } p \text{ edges}}\mathbb{E}a_{e_1}^{t_1}\cdots a_{e_p}^{t_p} \\
&\leq \frac{1}{n^{k+1}}\sum_{p=1}^{k-1} n^{p+1}p^h\left(\eta_n^{h-2p}n^{h/2-p}\right)\left(\sup_{ij} s_{ij}\right)^p \\
&= \sum_{p=1}^{k-1} p^h\eta_n^{h-2p}C^p = o(1).
\end{aligned}$$

Now it remains to compute S_1. For the closed walk that contains a self-loop, the number of distinct vertices is at most k, which implies the total contribution of such closed walks is $O(n^k)$, hence such terms are negligible in the limit of S_1. We only need to consider closed walks that use $k+1$ distinct vertices. Each closed walk in $\mathcal{C}_1$ with $k+1$ distinct vertices in $\{1,\ldots n\}$ is a closed walk on a tree of $k+1$ vertices that visits each edge twice. Given an unlabeled rooted planar tree T and a depth-first search closed walk with vertices chosen from $[n]$, there is a one-to-one correspondence between such walk and a labeling of T (See Figure 2.2). There are C_k many rooted planar trees with $k+1$ vertices and for each rooted planar tree T_j^{k+1}, the ordering of the vertices from 1 to $k+1$ is fixed by its depth-first search. Let $T_{l,j}^{k+1}$ be any labeled tree with the unlabeled rooted tree T_j^{k+1} and a labeling $l = (l_1,\ldots,l_{k+1}), 1 \leq l_i \leq n, 1 \leq i \leq k+1$ for its vertices from 1 to $k+1$. For terms in $\mathcal{C}_1$, any possible labeling l must satisfy that $l_1,\ldots,l_{k+1}$ are

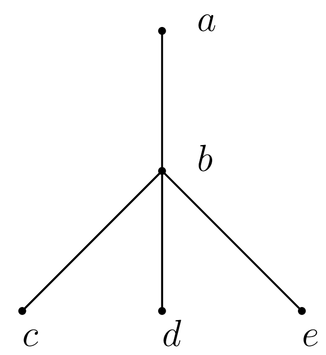

Figure 2.2: A closed walk *abcbdbeba* corresponds to a labeling of the rooted planar tree.

distinct. Let $E(T_{l,j}^{k+1})$ be the edge set of $T_{l,j}^{k+1}$. Then S_1 can be written as

$$S_1 = \frac{1}{n^{k+1}} \sum_{j=1}^{C_k} \sum_{l=(l_1,\ldots,l_{k+1})} \mathbb{E} \prod_{e \in E(T_{l,j}^{k+1})} a_e^2 = \sum_{j=1}^{C_k} \frac{1}{n^{k+1}} \sum_{l=(l_1,\ldots,l_{k+1})} \prod_{e \in E(T_{l,j}^{k+1})} s_e. \tag{2.3.8}$$

Consider

$$S_1' := \sum_{j=1}^{C_k} \frac{1}{n^{k+1}} \sum_{1 \le l_1,\ldots,l_{k+1} \le n} \prod_{e \in E(T_{l,j}^{k+1})} s_e,$$

where l now stands for every possible labelling which allows some of $l_1, \ldots l_{k+1}$ to coincide, then we have

$$|S_1 - S_1'| \le \frac{1}{n^{k+1}} C_k (k+1) n^k (\sup_{ij} s_{ij})^k = O\left(\frac{1}{n}\right).$$

On the other hand,

$$\begin{aligned} t(T_j^{k+1}, W_n) &= \int_{[0,1]^{k+1}} \prod_{uv \in E(T_j^{k+1})} W_n(x_u, x_v) dx_1 \ldots dx_{k+1} \\ &= \frac{1}{n^{k+1}} \sum_{1 \le l_1,\ldots,l_{k+1} \le n} \prod_{uv \in E(T_{l,j}^{k+1})} s_{l_u l_v} = \frac{1}{n^{k+1}} \sum_{1 \le l_1,\ldots,l_{k+1} \le n} \prod_{e \in E(T_{l,j}^{k+1})} s_e. \end{aligned} \tag{2.3.9}$$

Note that $S_1' = \sum_{j=1}^{C_k} t(T_j^{k+1}, W_n)$. From (2.3.8) and (2.3.9), we get $S_1 = \sum_{j=1}^{C_k} t(T_j^{k+1}, W_n) + o(1)$.

Combining the estimates of S_1, S_2 and S_3, the conclusion of Lemma 2.3.6 follows. $\square$

Lemma 2.3.6 connects the moments of the trace of M_n to homomorphism densities from trees to the graphon W_n. To proceed with the proof of Theorem 2.3.2, we need the following lemma.

Lemma 2.3.7. *In order to prove the conclusion of Theorem 2.3.2, it suffices to prove it under the following conditions:*

1. $\mathbb{E}a_{ij} = 0$, $\mathbb{E}|a_{ij}|^2 = s_{ij}$ *and* $\{a_{ij}\}_{1\le i,j\le n}$ *are independent up to symmetry.*

2. $|a_{ij}| \le \eta_n\sqrt{n}$ *for some positive decreasing sequence* η_n *such that* $\eta_n \to 0$.

3. $\sup_{ij} s_{ij} \le C$. *for some constant* $C \ge 0$.

The proof of Lemma 2.3.7 follows verbatim as the proof of Theorem 2.9 in [23], so we do not give it here. The followings are two results that are used in the proof and will be used elsewhere in the paper, so we give them here. See Section A in [23] for further details.

Lemma 2.3.8 (Rank Inequality)**.** *Let* A_n, B_n *be two* $n \times n$ *Hermitian matrices. Let* F^{A_n}, F^{B_n} *be the empirical spectral distributions of* A_n *and* B_n*, then*

$$\|F^{A_n} - F^{B_n}\| \le \frac{\operatorname{rank}(A_n - B_n)}{n},$$

where $\|\cdot\|$ *is the* L^∞*-norm.*

Lemma 2.3.9 (Lévy Distance Bound)**.** *Let* L *be the Lévy distance between two distribution functions, we have for any* $n \times n$ *Hermitian matrices* A_n *and* B_n*,*

$$L^3(F^{A_n}, F^{B_n}) \le \frac{1}{n}\operatorname{tr}[(A_n - B_n)(A_n - B_n)^*].$$

With Lemma 2.3.7, we will prove Theorem 2.3.2 under assumptions in Lemma 2.3.7.

Proof of Theorem 2.3.2. By Lemma 2.3.7, it suffices to prove Theorem 2.3.2 under the conditions (1)-(3) in Lemma 2.3.7. We now assume these conditions hold. Then (2.3.6) and (2.3.7) in Lemma 2.3.6 can be applied here.

(1) Since for any finite tree T, $t(T, W_n)$ converges as $n \to \infty$, we can define

$$\beta_{2k} := \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\mathrm{tr} M_n^{2k}] = \lim_{n\to\infty} \sum_{j=1}^{C_k} t(T_j^{k+1}, W_n), \quad \beta_{2k+1} := \lim_{n\to\infty} \frac{1}{n}\mathbb{E}[\mathrm{tr} M_n^{2k+1}] = 0.$$

With Carleman's Lemma (Lemma B.1 and Lemma B.3 in [23]), in order to to show the limiting spectral distribution of M_n is uniquely determined by the moments, it suffices to show that for each integer $k \geq 0$, almost surely we have

$$\lim_{n\to\infty} \frac{1}{n}\mathrm{tr} M_n^k = \beta_k, \quad \text{and} \quad \liminf_{k\to\infty} \frac{1}{k}\beta_{2k}^{1/2k} < \infty.$$

The remaining of the proof is similar to proof of Theorem 2.9 in [23], and we include it here for completeness. Let $G(\mathbf{i})$ be the graph induced by the closed walk $\mathbf{i} = (i_1, \dots i_k)$. Define $A(G(\mathbf{i})) := a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_k i_1}$. Then

$$\mathbb{E}\left|\frac{1}{n}\mathrm{tr} M_n^k - \frac{1}{n}\mathbb{E}[\mathrm{tr} M_n^k]\right|^4 = \frac{1}{n^{4+2k}} \sum_{\mathbf{i}_j, 1\leq j\leq 4} \mathbb{E}\prod_{j=1}^{4} [A(G(\mathbf{i}_j)) - \mathbb{E}A(G(\mathbf{i}_j))]$$

Consider a quadruple closed walk $\mathbf{i}_j, 1 \leq j \leq 4$. By independence, for the nonzero term, the graph $\cup_{j=1}^4 G(\mathbf{i}_j)$ has at most two connected components. Assume there are q edges in $\cup_{j=1}^4 G(\mathbf{i}_j)$ with multiplicity $v_1, \dots, v_q$, then $v_1 + \cdots + v_q = 4k$. The number of vertices in $\cup_{j=1}^4 G(\mathbf{i}_j)$ is at most $q+2$.

To make every term in the expansion of $\mathbb{E}\prod_{j=1}^4 \left(A(G(\mathbf{i}_j)) - \mathbb{E}A(G(\mathbf{i}_j))\right)$ nonzero, the multiplicity of each edge is at least 2, so $q \leq 2k$ and the corresponding term satisfies

$$\mathbb{E}\prod_{j=1}^{4}[A(G(\mathbf{i}_j)) - \mathbb{E}A(G(\mathbf{i}_j))] \leq C^q(\eta_n\sqrt{n})^{4k-2q}. \tag{2.3.10}$$

If $q = 2k$, we have $v_1 = \cdots = v_q = 2$. Since the graph $\cup_{j=1}^4 G(\mathbf{i}_j)$ has at most two connected components with at most $2k+1$ vertices, there must be a cycle in $\cup_{j=1}^4 G(\mathbf{i}_j)$. So the number of such graphs is at most n^{2k+1}. Therefore from (2.3.10),

$$\begin{aligned}\mathbb{E}\left|\frac{1}{n}\mathrm{tr}M_n^k - \frac{1}{n}\mathbb{E}[\mathrm{tr}M_n^k]\right|^4 &= \frac{1}{n^{4+2k}}\sum_{\mathbf{i}_j, 1\leq j\leq 4}\mathbb{E}\prod_{j=1}^{4}[A(G(\mathbf{i}_j)) - \mathbb{E}A(G(\mathbf{i}_j))]\\ &\leq \frac{1}{n^{4+2k}}\left(C^{2k}n^{2k+1} + \sum_{q<2k}C^q n^{q+2}(\eta_n\sqrt{n})^{4k-2q}\right) = o\left(\frac{1}{n^2}\right).\end{aligned}$$

Then by Borel-Cantelli Lemma,

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}M_n^k = \beta_k \quad a.s.$$

Moreover, since we have

$$\beta_{2k} = \lim_{n\to\infty}\sum_{j=1}^{C_k} t(T_j^{k+1}, W_n) \leq C_k C^k,$$

which implies $\liminf_{k\to\infty}\frac{1}{k}\beta_{2k}^{1/2k} = 0$.

(2) Since $\delta_{\square}(W_n, W) \to 0$, by Theorem 2.2.7, we have

$$\lim_{n\to\infty} t(T_j^{k+1}, W_n) = t(T_j^{k+1}, W)$$

for any rooted planar tree T_j^{k+1} with $k \geq 1, 1 \leq j \leq C_k$. Therefore for all $k \geq 0$,

$$\lim_{n\to\infty}\frac{1}{n}\mathrm{tr}M_n^{2k} = \sum_{j=1}^{C_k} t(T_j^{k+1}, W), \quad \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}M_n^{2k+1} = 0 \quad a.s.$$

This completes the proof. $\square$

2.3.3 Proof of Theorem 2.3.4

Proof. Since

$$\limsup_{k\to\infty}(\beta_{2k}(x))^{1/(2k+1)} \le 2\|W\|_\infty^{1/2}$$

for all $x \in [0,1]$, we have for $|z| > 2\|W\|_\infty^{1/2}$, $a(z,x) = \sum_{k=0}^{\infty} \frac{\beta_{2k}(x)}{z^{2k+1}}$ converges. Note that

$$\int_0^1 \beta_{2k}(x)dx = \sum_{j=1}^{C_k} \int_0^1 t_x(T_j^{k+1}, W)dx = \sum_{j=1}^{C_k} t(T_j^{k+1}, W) = \beta_{2k},$$

which implies for $|z| > 2\|W\|_\infty^{1/2}$, $s(z) = \sum_{k=0}^{\infty} \frac{\beta_{2k}}{z^{2k+1}} = \int_0^1 a(z,x)dx$.

Next we show (2.3.3) holds for $|z| > 2\|W\|_\infty^{1/2}$, which is equivalent to show

$$a(z,x)\int_0^1 W(x,y)a(z,y)dy = za(z,x) - 1, \quad \forall x \in [0,1]. \tag{2.3.11}$$

We order the vertices in each rooted planar tree T_j^{k+1} from 1 to $k+1$ by depth-first search order (the root for each T_j^{k+1} is always denoted by 1). Define a function

$$f_{j,k}(x_1, x_2, \ldots, x_{k+1}) =: \prod_{uv\in E(T_j^{k+1})} W(x_u, x_v).$$

Now we expand $a(z,x)$ as follows

$$a(z,x) = \sum_{k=0}^{\infty} \frac{1}{z^{2k+1}} \sum_{j=1}^{C_k} t_x(T_j^{k+1}, W) = \sum_{k=0}^{\infty} \frac{1}{z^{2k+1}} \sum_{j=1}^{C_k} \int_{[0,1]^k} f_{j,k}(x, x_2, \ldots, x_{k+1}) \prod_{i=2}^{k+1} dx_i.$$

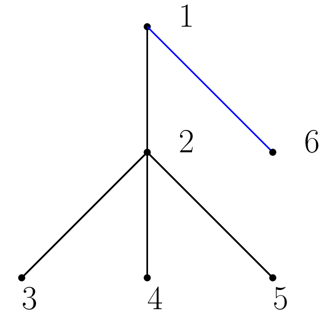

Figure 2.3: A rooted planar tree with a new edge attached with a new vertex labeled 6

Then we can write $\int_0^1 W(x,y)a(z,y)dy$ as

$$\sum_{k=0}^{\infty} \frac{1}{z^{2k+1}} \sum_{j=1}^{C_k} \int_{[0,1]^{k+1}} W(x,y) f_{j,k}(y,x_2,\dots,x_{k+1}) dy \prod_{i=2}^{k+1} dx_i. \tag{2.3.12}$$

Denote

$$B_{j,k}(x) := \int_{[0,1]^{k+1}} W(x,y) f_{j,k}(y,x_2,\dots,x_{k+1}) dy \prod_{i=2}^{k+1} dx_i.$$

Let T_j^{k+1*} be the rooted planar tree T_j^{k+1} with a new edge attached to the root and the new vertex ordered $k+2$ (See Figure 2.3). Let $t_x(T_j^{k+1*},W)$ be the homomorphism density from partially labeled graph T_j^{k+1*} to W with the new vertex labeled x. With this notation, $B_{j,k}(x_{k+2})$ can be written as

$$\begin{aligned} &\int_{[0,1]^{k+1}} W(x_{k+2},x_1) f_{j,k}(x_1,x_2\dots,x_{k+1}) \prod_{i=1}^{k+1} dx_i \\ =&\int_{[0,1]^{k+1}} \prod_{uv\in E(T_j^{k+1*})} W(x_u,x_v) \prod_{i=1}^{k+1} dx_i = t_{x_{k+2}}(T_j^{k+1*},W). \end{aligned} \tag{2.3.13}$$

So (2.3.12) and (2.3.13) implies $\int_0^1 W(x,y)a(z,y)dy = \sum_{k=0}^{\infty} \frac{1}{z^{2k+1}} \sum_{j=1}^{C_k} t_x(T_j^{k+1*},W)$.

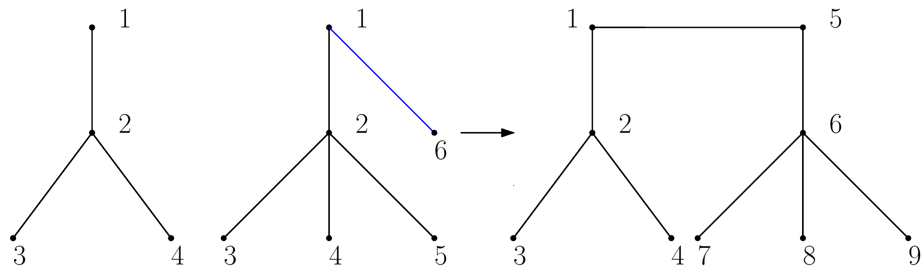

Figure 2.4: Combining T_i^{k+1} with T_j^{l+1*} yields a new rooted planar tree of $k+l+2$ vertices.

Therefore

$$a(z,x)\int_0^1 W(x,y)a(z,y)dy = \left(\sum_{k=0}^{\infty}\frac{1}{z^{2k+1}}\sum_{i=1}^{C_k} t_x(T_i^{k+1},W)\right)\left(\sum_{l=0}^{\infty}\frac{1}{z^{2l+1}}\sum_{j=1}^{C_l} t_x(T_j^{l+1*},W)\right)$$
$$= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\frac{1}{z^{2(k+l)+2}}\sum_{i=1}^{C_k}\sum_{j=1}^{C_l} t_x(T_i^{k+1},W)t_x(T_j^{l+1*},W). \tag{2.3.14}$$

Let $\{T_{i,j}^{k+l+2}, 1 \le i \le C_k, 1 \le j \le C_l\}$ be all rooted planar trees with $k+l+2$ vertices generated by combining T_i^{k+1} and T_j^{l+1*} in the following way.

1. First of all, by attaching the new labeled vertex of T_j^{l+1*} to the root of T_i^{k+1}, we get a new tree T of $k+l+2$ vertices.

2. Choose the root of T to be the root of T_i^{k+1}. Order all vertices coming from T_i^{k+1} with $1,2,\ldots,k+1$ and order vertices coming from T_j^{l+1} with $k+2,k+3,\ldots,k+l+2$ both in depth-first search order. Then T becomes a rooted planar tree $T_{i,j}^{k+l+2}$ of $k+l+2$ vertices (See Figure 2.4).

Let $t_x(T_{i,j}^{k+l+2},W)$ be the homomorphism density from partially labeled tree $T_{i,j}^{k+l+2}$ to W with the root labeled x. Using our notation, we have

$$t_x(T_i^{k+1},W)t_x(T_j^{l+1*},W) = t_x(T_{i,j}^{k+l+2},W).$$

Now let $s = k + l + 1$, then (2.3.14) can be written as

$$\sum_{s=1}^{\infty} \frac{1}{z^{2s}} \sum_{\substack{k+l+1=s \\ k,l\geq 0}} \sum_{i=1}^{C_k} \sum_{j=1}^{C_l} t_x(T_{i,j}^{s+1}, W). \tag{2.3.15}$$

Since all rooted planar trees in the set $\{T_{i,j}^{s+1}\ 1 \leq i \leq C_l, 1 \leq j \leq C_k\}$ are different, from the Catalan number recurrence, there are

$$\sum_{\substack{k+l=s-1 \\ k,l\geq 0}} C_k C_l = \sum_{k=0}^{s-1} C_k C_{s-1-k} = C_s$$

many, which implies $\{T_{i,j}^{s+1}\ 1 \leq i \leq C_l, 1 \leq j \leq C_k\}$ are all rooted planar trees of $s+1$ vertices. Now (2.3.15) can be written as

$$\sum_{s=1}^{\infty} \frac{1}{z^{2s}} \sum_{i=1}^{C_s} t_x(T_i^{s+1}, W) = za(z,x) - 1.$$

Therefore (2.3.11) holds for $|z| > 2\|W\|_{\infty}^{1/2}$. Since (2.3.11) has a unique analytic solution on $\mathbb{C}^+$ (see Theorem 2.1 in [7]), by analytic continuation, $a(z,x)$ has a unique extension on $\mathbb{C}^+ \times [0,1]$ such that (2.3.11) holds for all $z \in \mathbb{C}^+$. This completes the proof. $\square$

2.4 Generalized Wigner matrices

The semicircle law for generalized Wigner matrices whose variance profile is doubly stochastic and comes from discretizing a function with zero-measure discontinuities was proved in [149, 19]. The local semicircle law and universality of generalized Wigner matrices have been studied in [86, 87] with a lower bound on the variance profile and conditions on the distributions of entries. With Theorem 2.3.2, we can have a quick proof of the semicircle law for generalized Wigner matrices under Lindeberg's condition. Compared to [149, 19], where the L^{∞}-convergence of the variance profile is assumed, we don't even need to assume the variance profile converges under the cut metric. We will only need the weaker condition: the convergence of $t(T,W_n)$ for any finite tree T. In this section, we will show that the condition in Theorem 2.3.2, the convergence of tree integrals, is indeed a weaker condition than the convergence of the variance profile under the cut metric. Below we provide two examples where assumptions in [19, 157] fail, but our Theorem 2.3.2 holds.

We make the following assumptions for our *generalized Wigner matrices*. Let A_n be a random Hermitian matrix such that entries are independent up to symmetry, and satisfies the following conditions:

1. $\mathbb{E}[a_{ij}] = 0, \mathbb{E}\left[|a_{ij}|^2\right] = s_{ij}$,
2. $\frac{1}{n}\sum_{j=1}^{n} s_{ij} = 1 + o(1)$ for all $1 \le i \le n$.
3. for any constant $\eta > 0$, $\lim_{n\to\infty} \frac{1}{n^2} \sum_{1\le,i,j\le n} \mathbb{E}\left[|a_{ij}|^2 \mathbf{1}(|a_{ij}| \ge \eta\sqrt{n})\right] = 0$.
4. $\sup_{ij} s_{ij} \le C$ for a constant $C > 0$.

We use our general formula in Theorem 2.3.2 to get the semicircle law. An important observation is, when the variance profile is almost stochastic, the homomorphism densities in Theorem 2.3.2 are easy to compute, as shown in the following lemma. The main idea is that we can start computing the homomorphism density integral from leaves on the tree.

Lemma 2.4.1. *Let $\{W_n\}_{n\ge1}$ be any sequence of graphons such that $0 \le W_n(x,y) \le C$ almost everywhere for some constant $C > 0$. If for $x \in [0,1]$ almost everywhere,*

$$\lim_{n\to\infty} \int_0^1 W_n(x,y)dy = 1,$$

then $\lim_{n\to\infty} t(T,W_n) = 1$ for any finite tree T.

Proof. We induct on the number of vertices of a tree. Let $k = |V|$. For $k = 2$, by Dominated Convergence Theorem,

$$\lim_{n\to\infty} t(T, W_n) = \int_0^1 W_n(x,y)dxdy = 1. \tag{2.4.1}$$

Assume for any trees with $k-1$ vertices the statement holds. For any tree T with k vertices, we order the vertices in T by depth-first search. Then the vertex with label k is a leaf. Note that

$$\begin{aligned}
t(T, W_n) &= \int_{[0,1]^k} \prod_{ij\in E} W_n(x_i, x_j)dx_1 \ldots dx_k \\
&= \int_{[0,1]^k} W_n(x_{k-1}, x_k) \prod_{ij\in E\setminus\{k-1,k\}} W_n(x_i, x_j)dx_1 \ldots dx_k \\
&= \int_{[0,1]^{k-1}} \left(\int_{[0,1]} W_n(x_{k-1}, x_k)dx_k \right) \prod_{ij\in E\setminus\{k-1,k\}} W_n(x_i, x_j)dx_1 \ldots dx_{k-1}
\end{aligned}$$

Let T' be the tree T with the edge $\{k-1,k\}$ removed, then we have

$$\begin{aligned}
t(T', W_n) &= \int_{[0,1]^{k-1}} \prod_{ij\in E\setminus\{k-1,k\}} W_n(x_i, x_j)dx_1 \ldots dx_{k-1}, \\
t(T, W_n) - t(T', W_n) &= \int_{[0,1]^{k-1}} \left(\int_{[0,1]} W_n(x_{k-1}, x_k)dx_k - 1 \right) \prod_{ij\in E\setminus\{k-1,k\}} W_n(x_i, x_j)dx_1 \ldots dx_{k-1}.
\end{aligned}$$

By Dominated Convergence Theorem and (2.4.1) we obtain

$$\lim_{n\to\infty} |t(T, W_n) - t(T', W_n)| = 0.$$

Moreover, by our assumption of the induction, $\lim_{n\to\infty} t(T', W_n) = 1$, therefore $\lim_{n\to\infty} t(T, W_n) = 1$. This completes the proof. $\square$

Now we can give a quick proof of the semicircle law for generalized Wigner matrices in the following theorem, which is a quick consequence of Lemma 2.4.1 and Theorem 2.3.2.

Theorem 2.4.2. *Let A_n be a generalized Wigner matrix with assumptions above. The limiting spectral distribution of $M_n := \frac{A_n}{\sqrt{n}}$ converges weakly almost surely to the semicircle law.*

Proof. Let W_n be the graphon representation of the variance profile for A_n. From Condition (2), we have

$$\lim_{n\to\infty}\int_{[0,1]} W_n(x,y)dy = 1$$

for $x \in [0,1]$ almost everywhere. Then by Lemma 2.4.1, $\lim_{n\to\infty} t(T,W_n) = 1$ for any finite tree T.

By part (1) in Theorem 2.3.2, the empirical spectral distribution of M_n converges almost surely to a probability measure μ such that for all $k \geq 0$.

$$\int x^{2k} d\mu = C_k, \quad \int x^{2k+1} d\mu = 0. \tag{2.4.2}$$

It's known that the semicircle law is uniquely determined by its moments, therefore the limiting spectral distribution for M_n is the semicircle law. □

Theorem 2.4.2 can be applied to study the spectrum of inhomogeneous random graphs with roughly equal expected degrees. This is a sparse random graph model where no limiting variance profile is assumed, so the theorems in [157, 19] do not apply here. Consider the inhomogeneous Erdős-Rényi model $\mathcal{G}(n,(p_{ij}))$ with adjacency matrix A_n, where edges exist independently with given probabilities p_{ij} such that $p_{ij} = p_{ji}$. Assume

$$\sum_{i=1}^{n} p_{ij} = (1+o(1))n\alpha \quad \text{for all } j \in [n] \tag{2.4.3}$$

with some $\alpha \to 0, \alpha = \omega\left(\frac{1}{n}\right)$, and

$$\max_{ij} p_{ij} \leq C\alpha \quad \text{for some constant } C \geq 1. \tag{2.4.4}$$

Corollary 2.4.3. Under the assumptions (2.4.3) and (2.4.4), the empirical spectral distribution of the scaled adjacency matrix $\frac{A_n}{\sqrt{n\alpha}}$ converges almost surely to the semicircle law.

Proof. Consider the matrix $M_n = \frac{A_n - \mathbb{E}A_n}{\sqrt{\alpha}}$. Then by (2.4.3) and (2.4.4), one can check that M_n satisfies the assumptions (1)-(4) above for the generalized Wigner matrices. By Theorem 2.4.2, the empirical spectral distribution of $\frac{A_n - \mathbb{E}A_n}{\sqrt{n\alpha}}$ converges to the semicircle law almost surely. By Lemma 2.3.9, we have almost surely

$$\begin{aligned} L^3\left(F^{\frac{A_n}{\sqrt{n\alpha}}}, F^{\frac{A_n-\mathbb{E}A_n}{\sqrt{n\alpha}}}\right) &\leq \frac{1}{n}\mathrm{tr}\left[\left(\frac{\mathbb{E}A_n}{\sqrt{n\alpha}}\right)^2\right] = \frac{1}{n^2\alpha}\sum_{i,j=1}^{n}(\mathbb{E}a_{ij})^2 \\ &= \frac{\sum_{i,j=1}^{n} p_{ij}^2}{n^2\alpha} \leq \frac{n^2C^2\alpha^2}{n^2\alpha} = C^2\alpha = o(1), \end{aligned} \tag{2.4.5}$$

where the last line of inequalities are from (2.4.4). Then $\frac{A_n}{\sqrt{n\alpha}}$ and $\frac{A_n - \mathbb{E}A_n}{\sqrt{n\alpha}}$ have the same limiting spectral distribution almost surely. This completes the proof. □

2.5 Sparse W–random graphs

Given a graphon $W : [0,1]^2 \to [0,1]$, following the definitions in [42], one can generate a sequence of sparse random graphs G_n in the following way. We choose a sparsity parameter ρ_n such that

$$\sup_n \rho_n < 1 \text{ with } \rho_n \to 0 \text{ and } n\rho_n \to \infty.$$

Let $x_1, \dots, x_n$ be i.i.d. chosen uniformly from $[0,1]$. For a graph G_n, i and j are connected with probability $\rho_n W(x_i, x_j)$ independently for all $i \neq j$. We define G_n to be a *sparse W-random graph*, and the sequence $\{G_n\}$ is denoted by $\mathcal{G}(n, W, \rho_n)$. Note that we use the same i.i.d. sequence $x_1, \dots, x_n$ when constructing G_n for different values of n without resampling the x_i's. We determine the limiting spectral distributions for the adjacency matrices of sparse W-random graphs in the following theorem. This is a novel application of our theorem that cannot be covered by any previous results, since W can be any bounded measurable function.

Theorem 2.5.1. *Let $G(n, W, \rho_n)$ be a sequence of sparse W-random graphs with adjacency matrices $\{A_n\}_{n\geq 1}$. The limiting spectral distribution of $\frac{A_n}{\sqrt{n\rho_n}}$ converges almost surely to a probability measure μ such that*

$$\int_{\mathbb{R}} x^{2k} d\mu = \sum_{j=1}^{C_k} t(T_j^{k+1}, W), \quad \int_{\mathbb{R}} x^{2k+1} d\mu = 0.$$

Moreover, its Stieltjes transform $s(z)$ satisfies the following equation:

$$s(z) = \int_0^1 a(z,x)dx, \quad a(z,x)^{-1} = z - \int_0^1 W(x,y)a(z,y)dy, \quad \forall x \in [0,1].$$

Proof. Let

$$B_n := \frac{A_n - \mathbb{E}[A_n | x_1, \ldots, x_n]}{\sqrt{\rho_n}} = (b_{ij})_{1\leq i,j\leq n}.$$

Note that B_n is now a function of $x_1, \ldots, x_n$. Since $n\rho_n \to \infty$ and $|b_{ij}| \leq \frac{2}{\sqrt{\rho_n}}$, , we have that for any constant $\eta > 0$.

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{1\leq, i,j \leq n} \mathbb{E}\left[|b_{ij}|^2 \mathbf{1}(|b_{ij}| \geq \eta\sqrt{n}) \mid x_1, \ldots, x_n\right] = 0,$$

then the Lindeberg's condition (2.3.1) holds for B_n. Let S_n be the variance profile matrix of B_n. Then we have $s_{ii} = 0, 1 \leq i \leq n$ and for all $i \neq j$,

$$s_{ij} = \frac{\rho_n W(x_i, x_j)(1 - \rho_n W(x_i, x_j))}{\rho_n} = W(x_i, x_j) + o(1).$$

Let W_n be the graphon representation of the matrix S_n and let $\tilde{W}_n$ be the graphon of a weighted complete graph on $[n]$ with edge weights $W(x_i, x_j)$ for each edge ij. It implies that

$$W_n(x,y) = \tilde{W}_n(x,y) + o(1), \quad \forall (x,y) \in [0,1]^2.$$

By Dominated Convergence Theorem, we get $\lim_{n\to\infty} \delta_\square(\tilde{W}_n, W_n) = 0$. From Theorem 4.5 (a) in [44], we have $\lim_{n\to\infty} \delta_\square(\tilde{W}_n, W) = 0$ almost surely, which implies $\lim_{n\to\infty} \delta_\square(W_n, W) = 0$ almost surely. Therefore from Theorem 2.3.2 (2), the limiting spectral distribution of $\frac{B_n}{\sqrt{n}}$ exists almost surely and its moments and Stieltjes transform are given by Theorem 2.3.2 and Theorem 2.3.4. Next we show $\frac{B_n}{\sqrt{n}}$ and $\frac{A_n}{\sqrt{n\rho_n}}$ have the same limiting spectral distribution.

By Lemma 2.3.9, we have almost surely

$$L^3(F^{\frac{A_n}{\sqrt{n\rho_n}}}, F^{\frac{B_n}{\sqrt{n}}}) \leq \frac{1}{n}\mathrm{tr}\left[\left(\frac{A_n}{\sqrt{n\rho_n}} - \frac{B_n}{\sqrt{n}}\right)^2\right] = \frac{1}{n^2\rho_n}\mathrm{tr}\left(\mathbb{E}[A_n|x_1,\ldots,x_n]\right)^2. \tag{2.5.1}$$

By the way we generate our W-random graphs, we have for all $i \neq j$,

$$\mathbb{E}[(A_n)_{ij} \mid x_1,\ldots,x_n] = \rho_n W(x_i,x_j).$$

Therefore the right hand side in (2.5.1) is almost surely bounded by

$$\frac{\rho_n}{n^2}\sum_{i\neq j} W^2(x_i,x_j) \leq \rho_n = o(1),$$

which implies $\lim_{n\to\infty} L^3(F^{\frac{A_n}{\sqrt{n\rho_n}}}, F^{\frac{B_n}{\sqrt{n}}}) = 0$ almost surely. This completes the proof. □

2.6 Random block matrices

Consider an $n \times n$ random Hermitian matrix A_n composed of d^2 many rectangular blocks as follows. We can write A_n as $A_n := \sum_{k,l=1}^d E_{kl} \otimes A_n^{(k,l)}$, where $\otimes$ denotes the Kronecker product of matrices, E_{kl} are the elementary $d \times d$ matrices having 1 at entry (k,l) and 0 otherwise. The blocks $A_n^{(k,l)}, 1 \leq k \leq l \leq d$ are of size $n_k \times n_l$ and consist of independent entries subject to symmetry. To summarize, we consider a *random block matrix* A_n with the following assumptions:

1. $\lim_{n\to\infty} \frac{n_k}{n} = \alpha_k \in [0,1], 1 \leq k \leq d.$

2. $\mathbb{E}a_{ij} = 0, 1 \le i, j \le n$, $\mathbb{E}|a_{ij}|^2 = s_{kl}$ if a_{ij} is in the (k,l)-th block. All entries are independent subject to symmetry.

3. $\sup_{kl} s_{kl} < C$ for some constant $C > 0$.

4. $\lim_{n\to\infty} \frac{1}{n^2} \sum_{ij} \mathbb{E}\left[(|a_{ij}|^2 \mathbf{1}(|a_{ij}| \ge \eta\sqrt{n})\right] = 0$, for any positive constant η.

For random block matrices with fixed d, the limiting spectral distributions are determined in [89, 73, 21] under various assumptions. However, explicit moment formulas were not known. With Theorem 2.3.2, we can compute the moments of the limiting spectral distribution. Let W_n be the graphon of the variance profile for A_n. Let $\beta_0 = 0, \beta_i = \sum_{j=1}^{i} \alpha_j, i \ge 1$. Then we can define the graphon W such that

$$W(x,y) = s_{kl}, \quad \text{if } (x,y) \in [\beta_{k-1}, \beta_k) \times [\beta_{l-1}, \beta_l). \tag{2.6.1}$$

Note that W is a step function defined on $[0,1]^2$. Below is a version of Theorem 2.3.2, written specifically to address this model.

Theorem 2.6.1. *Let A_n be a random block matrix satisfying the assumptions above. Let $M_n = \frac{A_n}{\sqrt{n}}$ and W be the graphon defined in* (2.6.1)*. Then the limiting spectral distribution of M_n converges almost surely to a probability measure μ such that*

$$\int_{\mathbb{R}} x^{2k} d\mu(x) = \sum_{j=1}^{C_k} t(T_j^{k+1}, W), \quad \int_{\mathbb{R}} x^{2k+1} d\mu(x) = 0, \tag{2.6.2}$$

and its Stieltjes transform $s(z)$ satisfies $s(z) = \sum_{k=1}^{d} \alpha_k a_k(z)$, where for all $1 \le k \le d$,

$$a_k(z)^{-1} = z - \sum_{i=1}^{d} \alpha_i s_{ik} a_i(z).$$

Proof. From the definition, we have $W_n(x,y) \to W(x,y)$ as $n \to \infty$ for $(x,y) \in [0,1]^2$ almost everywhere. Hence

$$\begin{aligned}\|W_n - W\|_\square &= \sup_{S,T\in[0,1]} \left|\int_{S\times T} W_n(x,y) - W(x,y)dxdy\right| \\ &\leq \int_{[0,1]^2} |W_n(x,y) - W(x,y)|dxdy.\end{aligned}$$

Since $|W_n(x,y)| \leq C$, by the Dominated Convergence Theorem, we have $\|W_n - W\|_\square \to 0$ as $n \to \infty$. (2.6.2) follow from Theorem 2.3.2. The existence and uniqueness of $a_k(z), 1 \leq k \leq d$ follows from Theorem 2.1 in [7]. $\square$

Now we consider the case where the number of blocks d depends on n such that $d \to \infty$ as $n \to \infty$. We partition the n vertices into d classes: $[n] = V_1 \cup V_2 \cup \cdots \cup V_d$. Let $m_0 = 0, m_i = \sum_{j=1}^{i} n_j$ and

$$V_i = \{m_{i-1}+1, m_{i-1}+2, \ldots, m_i\}$$

for $i = 1, \ldots, d$. We say the class V_i is *small* if $\frac{n_i}{n} \to \alpha_i = 0$, and V_i is *big* if $\frac{n_i}{n} \to \alpha_i > 0$.

It's not necessary that $\sum_{i=1}^{\infty} \alpha_i = 1$. For example, if $n_i \leq \log n$ for each i, we have $\frac{n_i}{n} \to 0$ for all $i = 1, 2, \ldots$, then $\sum_{i=1}^{\infty} \alpha_i = 0$. In such case, a limiting graphon might not be well defined for general variance profiles. However, if we make all variances for the off-diagonal blocks to be s_0 for some constant s_0, then the limiting graphon will be a constant function s_0 on $[0,1]^2$ since all diagonal blocks will vanish to a zero measure set in the limit. With these observations, we can extend our result to the case for $d \to \infty$ and $\sum_{i=1}^{\infty} \alpha_i \leq 1$ under more assumptions on the variance profile.

Theorem 2.6.2. *Let A_n be a random block matrix with $d \to \infty$ as $n \to \infty$ satisfying assumptions* (1)-(4)*, then the empirical spectral distribution of $\frac{A_n}{\sqrt{n}}$ converges almost surely to a probability measure μ if one of the extra conditions below holds.*

1. *$\sum_{i=1}^{\infty} \alpha_i = 1$ and $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$, or*

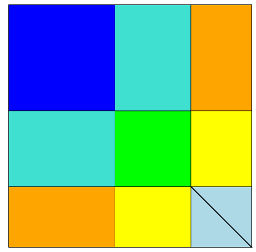

Figure 2.5: The limiting graphon with infinite many small classes

2. $\sum_{i=1}^{\infty} \alpha_i = \alpha < 1$, $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$*; also, for any two small classes* $V_k, V_l, k \neq l$, $s_{kl} = s_0$ *for some constant* s_0*. For any large class* V_k *and small class* V_l, $s_{kl} = s_{k0}$ *for some constant* s_{k0}.

We illustrate the limiting graphon for case (2) in Figure 2.5. Different colors represent different variances, and with our assumptions, all blocks of size $|V_k| \times |V_l|$ where V_k, V_l are small converge to a diagonal line inside the last big block.

Proof of Theorem 2.6.2. For case (1), assume $\sum_{i=1}^{\infty} \alpha_i = 1$. Define $\beta_0 = 0, \beta_i = \sum_{j=1}^{i} \alpha_j, i \geq 1$. Then we can define a graphon W as

$$W(x,y) = s_{ij}, \quad \forall (x,y) \in [\beta_{i-1}, \beta_i) \times [\beta_{j-1}, \beta_j)$$

if $\beta_{i-1} \neq \beta_i, \beta_{j-1} \neq \beta_j$. Then $W(x,y)$ is defined on $[0,1]^2$ almost everywhere. From our construction, $W_n(x,y) \to W(x,y)$ point-wise almost everywhere. By the Dominated Convergence Theorem, $\|W_n - W\|_\square \to 0$. For case (2), similarly, we define W in the following way,

$$W(x,y) = \begin{cases} s_{ij}, & \text{if } (x,y) \in [\beta_{i-1}, \beta_i) \times [\beta_{j-1}, \beta_j), \alpha_i, \alpha_j \neq 0, \\ s_0, & \text{if } (x,y) \in [\alpha, 1]^2, \\ s_{i0}, & \text{if } (x,y) \in [\beta_{i-1}, \beta_i) \times [\alpha, 1] \text{ or } [\alpha, 1] \times [\beta_{i-1}, \beta_i). \end{cases}$$

Then W is a graphon defined on $[0,1]^2$. Note that $\lim_{n\to\infty} W_n(x,y) = W(x,y)$ for all $(x,y) \in [0,1]^2$ outside the subset of the diagonal $\{(x,y) : x = y, x \in [\alpha, 1]\}$, which is a zero measure set on $[0,1]^2$.

So we have $\delta_\square(W_n, W) \to 0$. Then the result follows from Theorem 2.3.4. $\square$

2.7 Stochastic block models

The adjacency matrix A_n of a stochastic block model(SBM) with a growing number of classes is a random block matrix. A new issue here is $\mathbb{E}A_n \neq 0$, which does not fit our assumptions in Section 2.6. However some perturbation analysis of the empirical measures can be applied to address this issue. In this section, we consider the adjacency matrix A_n for both sparse and dense SBMs with the following assumptions:

1. $\frac{n_k}{n} \to \alpha_k \in [0,\infty), 1 \leq k \leq d$, where d depends on n.

2. Diagonal elements in A_n are 0. Entries in the block $V_i \times V_i$ are independent Bernoulli random variables with parameter p_{ii} depending on n up to symmetry. Entries in the block $V_k \times V_l, k \neq l$ are independent Bernoulli random variables with parameter p_{kl} depending on n.

3. Let $p = \sup_{ij} p_{ij}$. Assume $p = \omega(\frac{1}{n})$ and $\sup_n p < 1$.

4. Denote $\sigma^2 := p(1-p)$, and assume

$$\lim_{n\to\infty} \frac{p_{ij}(1-p_{ij})}{\sigma^2} = s_{ij} \in [0,1] \text{ for some constant } s_{ij}.$$

If $p \to 0$ (the sparse case), by the same argument in (2.4.5), $\frac{A_n - \mathbb{E}A_n}{\sigma\sqrt{n}}$ and $\frac{A_n}{\sigma\sqrt{n}}$ have the same limiting spectral distribution, we then have the following corollary from Theorem 2.6.2.

Corollary 2.7.1. Let A_n be the adjacency matrix of a sparse SBM with $p \to 0$, $d \to \infty$ as $n \to \infty$. The empirical spectral distribution of $\frac{A_n}{\sqrt{n}}$ converges almost surely to a probability measure μ if one of the extra conditions below holds.

1. $\sum_{i=1}^{\infty} \alpha_i = 1$ and $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$, or

2. $\sum_{i=1}^{\infty} \alpha_i = \alpha < 1$, $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$; also, for any two small classes $V_k, V_l, k \neq l$, $s_{kl} = s_0$ for some constant s_0. For any large class V_k and small class V_l, $s_{kl} = s_{k0}$ for some constant s_{k0}.

If $p \not\to 0$ (the dense case), to get the limiting spectral distribution of the non-centered matrix A_n, we need to consider the effect of $\mathbb{E}A_n$. If $\mathbb{E}A_n$ is of relatively low rank, we can still do a perturbation analysis from Lemma 2.3.8. The following theorem is a statement for the dense case.

Corollary 2.7.2. The empirical spectral distribution of the adjacent matrix $\frac{A_n}{\sqrt{n}\sigma}$ for a SBM with $p > c$ for a constant $c > 0$ converges almost surely if $d = o(n)$ and one of the following holds:

1. $\sum_{i=1}^{\infty} \alpha_i = 1$, $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$, or

2. $\sum_{i=1}^{\infty} \alpha_i = \alpha < 1$, $\alpha_1 \geq \alpha_2 \geq \cdots \geq 0$. For any two small classes $V_k, V_l, k \neq l$, $s_{kl} = s_0$ for some constant s_0. For any large class V_k and small class V_l, $s_{kl} = s_{k0}$ for some constant s_{k0}.

Proof. Let $\tilde{A}_n$ be a random block matrix such that $\tilde{a}_{ij} = a_{ij}$ for $i \neq j$ and $\{\tilde{a}_{ii}\}_{1 \leq i \leq n}$ be independent Bernoulli random variables with parameter p_{kk} if $i \in V_k$. Then $\operatorname{rank}(\mathbb{E}\tilde{A}_n) = d$.

Let $L\left(F^{\tilde{A}_n/\sigma\sqrt{n}}, F^{A_n/\sigma\sqrt{n}}\right)$ be the Lévy distance between the empirical spectral measures of $\frac{A_n}{\sigma\sqrt{n}}$ and $\frac{\tilde{A}_n}{\sigma\sqrt{n}}$, then by Lemma 2.3.9,

$$L^3\left(F^{\frac{\tilde{A}_n}{\sigma\sqrt{n}}}, F^{\frac{A_n}{\sigma\sqrt{n}}}\right) \leq \frac{1}{\sigma^2 n^2} \operatorname{tr}\left(\tilde{A}_n - A_n\right)^2 = \frac{1}{\sigma^2 n^2} \sum_{i=1}^{n} \tilde{a}_{ii}^2. \tag{2.7.1}$$

The right hand side of (2.7.1) is bounded by $\dfrac{1}{n\sigma^2} = o(1)$ almost surely. So we have almost surely

$$\lim_{n\to\infty} L^3\left(F^{\frac{\tilde{A}_n}{\sigma\sqrt{n}}}, F^{\frac{A_n}{\sigma\sqrt{n}}}\right) = 0. \tag{2.7.2}$$

Recall that the limiting distribution of $\frac{\tilde{A}_n - \mathbb{E}\tilde{A}_n}{\sigma\sqrt{n}}$ exists from Theorem 2.6.2 for random block matrices. By the Rank Inequality (Lemma 2.3.8), we have almost surely

$$\left\| F^{\frac{\tilde{A}_n - \mathbb{E}\tilde{A}_n}{\sigma\sqrt{n}}} - F^{\frac{\tilde{A}_n}{\sigma\sqrt{n}}} \right\| \leq \frac{\operatorname{rank}(\tilde{A}_n - \mathbb{E}\tilde{A}_n - \tilde{A}_n)}{n} = \frac{\operatorname{rank}(\mathbb{E}\tilde{A}_n)}{n} = \frac{d}{n} = o(1). \tag{2.7.3}$$

Then combining (2.7.2) and (2.7.3), almost surely $\frac{A_n}{\sigma\sqrt{n}}$ has the same limiting spectral distribution as $\frac{\tilde{A}_n - \mathbb{E}\tilde{A}_n}{\sigma\sqrt{n}}$. The conclusion then follows. □

Below, we give an example showing how to construct dense SBMs with a growing number of blocks which satisfies one of the assumptions in Corollary 2.7.2. Below is a lemma to justify that our two examples work.

Lemma 2.7.3. *Assume* $\sum_{i=1}^{\infty} \alpha_i = \alpha \leq 1$ *and* $1 \geq \alpha_1 \geq \alpha_2 \geq \cdots > 0$. *Let*

$$k(n) := \sup\left\{k : \alpha_k \geq \frac{1}{n}\right\},$$

then $\dfrac{k(n)}{n} = o(1)$.

Proof. If not, there exists a subsequence $\{n_l\}$ such that $\frac{k(n_l)}{n_l} \geq \varepsilon > 0$ for some ε. Then

$$\frac{1}{n_l} \leq \alpha_{k(n_l)} \quad \text{and} \quad \frac{k(n_l) - k(n_{l-1})}{n_l} \leq \sum_{i=k(n_{l-1})+1}^{k(n_l)} \alpha_i.$$

Hence

$$\begin{aligned} \sum_{l=1}^{\infty} \frac{k(n_l) - k(n_{l-1})}{n_l} &\leq \sum_{i=1}^{\infty} \alpha_i = \alpha, \\ \sum_{l=1}^{\infty} \frac{k(n_{l+1}) - k(n_l)}{k(n_{l+1})} &\leq \frac{\alpha}{\varepsilon} < \infty. \end{aligned} \tag{2.7.4}$$

This implies $\frac{k(n_{l+1})-k(n_l)}{k(n_{l+1})} \to 0$, so $\frac{k(n_{l+1})}{k(n_l)} \to 1$ as $n \to \infty$, therefore (2.7.4) implies

$$\sum_{l=1}^{\infty} \frac{k(n_{l+1})-k(n_l)}{k(n_l)} < \infty. \tag{2.7.5}$$

However,

$$\sum_{l=1}^{\infty} \frac{k(n_{l+1})-k(n_l)}{k(n_l)} \geq \int_{k(n_1)}^{\infty} \frac{1}{x} dx = \infty,$$

which is a contradiction to (2.7.5). Lemma 2.7.3 is then proved. □

Example 2.7.4. Let $\alpha_1 \geq \alpha_2 \geq \cdots > 0$ and $\sum_{i=1}^{\infty} \alpha_i = 1$. For each n, we generate the class V_i with size $n_i = \lfloor n\alpha_i \rfloor$ for $i = 1, 2, \ldots$ until $n_i = 0$. Then we generate the last class V_d with size $n_d = n - \sum_{i=1}^{d-1} n_i$. Note that for every fixed i, $\frac{n_i}{n} \to \alpha_i$. From Lemma 2.7.3, the number of blocks satisfies $d \leq k(n) + 1 = o(n)$. In particular, we have the following examples for the choice of α_i's:

1. $\alpha_i = \frac{C}{\gamma^i}$ for some constant $C, \gamma > 0$ with $\sum_{i=1}^{\infty} \alpha_i = 1$.

2. $\alpha_i = \frac{C}{i^{\beta}}$ for some $C > 0, \beta > 1$ with $\sum_{i=1}^{\infty} \alpha_i = 1$.

Example 2.7.5. Let $\alpha_1 \geq \alpha_2 \geq \cdots > 0$ and $\sum_{i=1}^{\infty} \alpha_i = \alpha < 1$. For each n, we can generate a class V_i with size $n_i = \lfloor n\alpha_i \rfloor$ for $i = 1, 2, \ldots$, until $n_i = 0$. Then generate $o(n)$ many small classes of size $o(n)$. By Lemma 2.7.3, $d = o(n)$.

2.8 Random Gram matrices

Let X be a $m \times n$ random matrix with independent, centered entries with unit variance, where $\frac{m}{n}$ converges to some positive constant as $n \to \infty$. It is known that the empirical spectral distribution converges to the Marčenko-Pastur law [140]. However, some applications in wireless communication require understanding the spectrum of $\frac{1}{n}XX^*$ where X has a variance profile [106, 69]. Such matrices are called *random Gram matrices*. The limiting spectral distribution of a random Gram matrix with non-centered diagonal entries and a variance profile was obtained in [105] under the assumptions that the $(4+\varepsilon)$-th moments of entries in X are bounded and the variance profile comes from a continuous function. The local law and singularities of the density of states of random Gram matrices were analyzed in [16, 13].

We use the symmetrization trick to connect the eigenvalues of $\frac{1}{n}XX^*$ to eigenvalues of a Hermitian matrix

$$H := \begin{bmatrix} 0 & X \\ X^* & 0 \end{bmatrix}.$$

As a corollary from our main theorem in Section 2.3, when $\mathbb{E}X = 0$, we obtain the moments and Stieltjes transforms of the limiting spectral distributions under weaker assumptions than [105]. In particular, we only need entries in X to have finite second moments, and the variance profile of H_n converges in terms of homomorphism densities.

Let X_n be a $m \times n$ complex random matrix whose entries are independent. Consider a *random Gram matrix* $M_n := \frac{1}{n}X_nX_n^*$ with a variance profile matrix $S_n = (s_{ij})_{1\leq i\leq m, 1\leq j\leq n}$ satisfies the following conditions:

1. $\mathbb{E}x_{ij} = 0, \mathbb{E}|x_{ij}|^2 = s_{ij}$, for all $1 \leq i \leq m, 1 \leq j \leq n$.

2. (Lindeberg's condition) for any constant $\eta > 0$,

$$\lim_{n\to\infty} \frac{1}{nm} \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbb{E}[|x_{ij}|^2 \mathbf{1}(|x_{ij}| \geq \eta\sqrt{n})] = 0.$$

3. $\sup_{ij} s_{ij} \leq C$ for some constant $C \geq 0$.

4. $\lim_{n\to\infty} \frac{m}{n} = y \in (0, \infty)$.

Let

$$H_n := \begin{bmatrix} 0 & X_n \\ X_n^* & 0 \end{bmatrix}. \tag{2.8.1}$$

We first find the relation between the trace of M_n and the trace of H_n in the following lemma.

Lemma 2.8.1. *For any integer $k \geq 1$, the following holds:*

$$\frac{1}{m}\mathrm{tr}M_n^k = \frac{(m+n)^k}{2mn^k}\mathrm{tr}\left(\frac{H_n}{\sqrt{n+m}}\right)^{2k}. \tag{2.8.2}$$

Proof. Nonzero eigenvalues of H come in pairs $\{-\sqrt{\lambda}, \sqrt{\lambda}\}$ where $\lambda > 0$ is a non-zero eigenvalue of $X_nX_n^*$. Therefore for $k \geq 1$,

$$\mathrm{tr}(H_n^{2k}) = 2\mathrm{tr}(X_nX_n^*)^k. \tag{2.8.3}$$

We then have for $k \geq 1$,

$$\frac{1}{m}\mathrm{tr}M_n^k = \frac{1}{m}\mathrm{tr}\left(\frac{1}{n}X_nX_n^*\right)^k = \frac{1}{2n^km}\cdot 2\mathrm{tr}(X_nX_n^*)^k = \frac{(m+n)^k}{2mn^k}\mathrm{tr}\left(\frac{H_n}{\sqrt{n+m}}\right)^{2k}. \tag{2.8.4}$$

$\square$

Since H_n is a $(n+m)\times(n+m)$ general Wigner-type matrix with a variance profile

$$\Sigma_n := \begin{bmatrix} 0 & S_n \\ S_n^T & 0 \end{bmatrix}, \tag{2.8.5}$$

we can decide the moments of the limiting spectral distribution of M_n from Theorem 2.3.2 and Lemma 2.8.1 in the following theorem.

Theorem 2.8.2. *Let M_n be a random Gram matrix with the assumptions above and W_n be the corresponding graphon of Σ_n. If for any finite tree T, $t(T,W_n)$ converges as $n \to \infty$, then the empirical spectral distribution of M_n converges almost surely to a probability measure μ such* *that for $k \geq 1$,*

$$\int x^k d\mu = \frac{(1+y)^{k+1}}{2y} \sum_{j=1}^{C_k} \lim_{n\to\infty} t(T_j^{k+1}, W_n).$$

Proof. From Lemma 2.8.1, for $k \geq 1$,

$$\frac{1}{m}\mathrm{tr} M_n^k = \frac{(m+n)^{k+1}}{2mn^k} \cdot \frac{1}{n+m}\mathrm{tr}\left(\frac{H_n}{\sqrt{n+m}}\right)^{2k}. \tag{2.8.6}$$

From Theorem 2.3.2, almost surely

$$\lim_{n\to\infty} \frac{1}{n+m}\mathrm{tr}\left(\frac{H_n}{\sqrt{n+m}}\right)^{2k} = \sum_{j=1}^{C_k} \lim_{n\to\infty} t(T_j^{k+1}, W_n).$$

Since $\lim_{n\to\infty} \frac{m}{n} = y > 0$, The result follows from (2.8.6). $\square$

Finally, we derive the Stieltjes transform of the limiting spectral distribution from Theorem 2.3.4.

Theorem 2.8.3. *Let M_n be a random Gram matrix with a variance profile S_n and W_n be the corresponding graphon of Σ_n defined in* (2.8.5). *If $\delta_\square(W_n, W) \to 0$ for some graphon W, then the empirical spectral distribution of $\frac{M_n}{\sqrt{n}}$ converges almost surely to a probability measure μ whose Stieltjes transform $s(z)$ is an analytic solution defined on $\mathbb{C}^+$ by the following equations:*

$$s(z) = \frac{1+y}{y}\int_0^{\frac{y}{1+y}} b(z,u)du, \tag{2.8.7}$$

$$b(z,u)^{-1} = z - \int_{\frac{y}{1+y}}^1 \frac{W(u,v)}{(1+y)^{-1} - \int_0^{\frac{y}{1+y}} W(u,t)b(z,t)dt} dv, \tag{2.8.8}$$

where $b(z,u)$ is an analytic function defined on $\mathbb{C}^+ \times \left[0, \frac{y}{1+y}\right]$.

Remark 2.8.4. Up to notational differences, (2.8.7), (2.8.8) are the centered case($\mathbb{E}M_n = 0$) of the equations in [105] (see Section 5.1 in [105]), where a non-centered form of the equations were also derived under the assumptions of $(4+\varepsilon)$-bounded moments and the continuity of the variance profile. Recently, (2.8.7), (2.8.8) were also studied in [16, 13], where the local law for the centered case was proved under stronger assumptions including bounded k-moments of each entry for each k and irreducibility condition on the variance profile. Our Theorem 2.8.2 and Theorem 2.8.3 give the weakest assumption so far for the existence of the limiting distribution and the quadratic vector equations only for the centered case.

Proof. Let $s(z)$ be the Stieltjes transform of the limiting spectral distribution of $\frac{M_n}{\sqrt{n}}$. Let

$$\gamma_k := \int x^k d\mu, \quad m_{2k} := \sum_{j=1}^{C_k} t(T_j^{k+1}, W), \quad \text{and} \quad m(z) := \sum_{k=0}^{\infty} \frac{m_{2k}}{z^{2k+1}}.$$

By Theorem 2.8.2, for $k \geq 1$,

$$\gamma_k = \frac{(1+y)^{k+1}}{2y} m_{2k}.$$

Note that $m_0 = \gamma_0 = 1$, we have for $|z|$ sufficiently large,

$$\begin{aligned} s(z) &= \sum_{k=0}^{\infty} \frac{\gamma_k}{z^{k+1}} = \frac{1}{z} + \sum_{k=1}^{\infty} \frac{m_{2k}}{z^{k+1}} \frac{1}{2y}(1+y)^{k+1} \\ &= \sum_{k=0}^{\infty} \frac{m_{2k}}{z^{k+1}} \frac{1}{2y}(1+y)^{k+1} + \frac{y-1}{2yz} = \frac{1}{2y}\sqrt{\frac{1+y}{z}}\, m\left(\sqrt{\frac{z}{1+y}}\right) + \frac{y-1}{2yz}. \end{aligned} \tag{2.8.9}$$

From Theorem 2.3.2 and (2.2.1), we know $m(z)$ is the Stieltjes transform of the limiting spectral distribution of $\frac{H_n}{\sqrt{n+m}}$. Moreover, from Theorem 2.3.4, we have

$$m(z) = \int_0^1 a(z,u)\,du, \tag{2.8.10}$$

$$a(z,u)^{-1} = z - \int_0^1 W(u,v)a(z,v)\,dv, \tag{2.8.11}$$

for some analytic function $a(z,u)$ defined on $\mathbb{C}^+ \times [0,1]$. It remains to translate the equations above to an equation for $s(z)$. Let

$$\begin{aligned} a_1(z,x) &:= a(z,x), \text{ for } x \in \left[0, \frac{y}{1+y}\right], \\ a_2(z,x) &:= a(z,x), \text{ for } x \in \left[\frac{y}{1+y}, 1\right]. \end{aligned}$$

Since $\frac{m}{n} \to y \in (0,\infty)$, and W_n is the corresponding graphon of Σ_n, its limit W will have a bipartite structure, i.e., $W(u,v) = 0$ for $(u,v) \in \left[0, \frac{y}{1+y}\right]^2 \cup \left[\frac{y}{1+y}, 1\right]^2$. Then we have the following equations from (2.8.11):

$$a_1(z,u)^{-1} = z - \int_{\frac{y}{1+y}}^{1} W(u,v) a_2(z,v) dv, \tag{2.8.12}$$

$$a_2(z,u)^{-1} = z - \int_{0}^{\frac{y}{1+y}} W(u,v) a_1(z,v) dv. \tag{2.8.13}$$

Combing (2.8.12) and (2.8.13), we have the following self-consistent equation for $a_1(z,u)$:

$$a_1(z,u)^{-1} = z - \int_{\frac{y}{1+y}}^{1} \frac{W(u,v)}{z - \int_0^{\frac{y}{1+y}} W(u,t) a_1(z,t) dt} dv. \tag{2.8.14}$$

Let $b(z,u) := \dfrac{a_1\left(\sqrt{\frac{z}{1+y}}, u\right)}{\sqrt{z(1+y)}}$. Then $b(z,u)$ is an analytic function defined on $\mathbb{C}^+ \times \left[0, \frac{y}{1+y}\right]$. From (2.8.14), we can substitute $a_1(z,u)$ with $b(z,u)$ and get

$$b(z,u)^{-1} = z - \int_{\frac{y}{1+y}}^{1} \frac{W(u,v)}{(1+y)^{-1} - \int_0^{\frac{y}{1+y}} W(u,t) b(z,t) dt} dv. \tag{2.8.15}$$

By multiplying with $a_1(z,u)$, $a_2(z,u)$ on both sides in (2.8.12) and (2.8.13) respectively, we have

$$1 = z a_1(z,u) - a_1(z,u) \int_{\frac{y}{1+y}}^{1} W(u,v) a_2(z,v) dv, \tag{2.8.16}$$

$$1 = z a_2(z,u) - a_2(z,u) \int_{0}^{\frac{y}{1+y}} W(u,v) a_1(z,v) dv. \tag{2.8.17}$$

From (2.8.16) and (2.8.17), by integration with respect to u, we have

$$\frac{y}{1+y} = z\int_0^{\frac{y}{1+y}} a_1(z,u)du - \int_0^{\frac{y}{1+y}}\int_{\frac{y}{1+y}}^1 W(u,v)a_1(z,u)a_2(z,v)dudv,$$
$$\frac{1}{1+y} = z\int_{\frac{y}{1+y}}^1 a_1(z,u)du - \int_{\frac{y}{1+y}}^1\int_0^{\frac{y}{1+y}} W(u,v)a_2(z,u)a_1(z,v)dudv.$$

Therefore we have

$$\int_0^{\frac{y}{1+y}} a_1(z,u)du - \int_{\frac{y}{1+y}}^1 a_2(z,u)du = \frac{y-1}{z(1+y)}. \tag{2.8.18}$$

From (2.8.10) and (2.8.18), we have the following relation between $m(z)$ and $a_1(z,u)$:

$$m(z) = \int_0^{\frac{y}{1+y}} a_1(z,u)du + \int_{\frac{y}{1+y}}^1 a_2(z,u)du = 2\int_0^{\frac{y}{1+y}} a_1(z,u)du - \frac{y-1}{z(1+y)}. \tag{2.8.19}$$

With (2.8.9) and (2.8.19), we obtain the following equation for $s(z)$:

$$s(z) = \frac{1+y}{y}\int_0^{\frac{y}{1+y}} b(z,u)du,$$

where $b(z,u)$ satisfies the equation (2.8.15). This completes the proof. $\square$

2.9 Acknowledgment

Chapter 2 is extracted from “Yizhe Zhu. A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Structures & Algorithms*, 56.1 (2020): 251-279”. The thesis author is the author of this paper.

Chapter 3

Global eigenvalue fluctuations of random bipartite biregular graphs

3.1 Introduction

3.1.1 Eigenvalue fluctuations of random matrices

The study of fluctuations from the limiting empirical spectral distributions (ESDs) for random matrices is a well-established topic of interest in random matrix theory, originated in [120, 123, 159], see also [18] and all references therein. More recently, it has been extended to sparse random matrices and random graph-related matrices in various regimes of sparsity and independence ([155, 156, 33, 76, 30]).

The ultimate goal in these studies is to see the equivalent of the one-dimensional Central Limit Theorem (CLT) emerge, when examining linear statistics of the spectra of random matrices and random graphs. More precisely, denote by $\lambda_1, \ldots, \lambda_n$ the eigenvalues of the random matrix, suitably scaled to put them with high probability on a compact set, and let f be a suitably smooth function. When the matrices in question are not extremely sparse, one can almost invariably prove that the linear statistic

$$\mathcal{L}(f) = \sum_{i=1}^{n} f(\lambda_i)$$

has the property that, when centered, it converges to a normal distribution whose variance depends on f:

$$\mathcal{L}(f) - \mathbb{E}(\mathcal{L}(f)) \ \rightarrow \ N(0, \sigma_f^2) \ .$$

Dense and not-too-sparse Wigner cases.

There is an interesting phenomenon taking place with respect to sparsity; the variance σ_f^2 is the same in the case of Gaussian Orthogonal Ensembles (GOEs) as in the case of the random regular graph under the permutation model with growing degrees [76]:

$$\sigma_f^2 \ = \ 2\sum_{k=1}^{\infty} k a_k^2 \ , \tag{3.1.1}$$

where a_k is the k-th coefficient of f in the Chebyshev polynomial basis expansion. Small variations of this expression occur also in dense Wigner variants and the uniform regular graph model, as follows. For real Wigner and generalized Wigner matrices in the dense case [118, 25, 19, 58, 162], σ_f also depends on the fourth moments of the off-diagonal entries and the variance of the diagonal entries, which yields corrections to the constants in front of a_1^2, a_2^2 (see Theorem 1.1 in [25] for an explicit expression). Similarly, in the uniform regular graph model [119], a correction must be introduced as there are no one- or two-cycles (as the graph is simple) and so the terms corresponding to $k = 1$ and $k = 2$ in the sum (3.1.1) are not present.

However, in the case of sparse Wigner matrices (corresponding to Erdős-Rényi graphs $G(n, p)$ with $p \to 0, np \to \infty$, [156]), the fluctuations are impacted by the fact that the number of nonzero entries in each row (i.e., the degree of each vertex) fluctuates, and the 4th moment of the scaled adjacency matrix entries grows. The variance σ_f^2 blows up, necessitating another multiplicative scaling of the linear statistic $\mathcal{L}(f) - \mathbb{E}(\mathcal{L}(f))$ by $\sqrt{p}$, and extracting only part of the expression (3.1.1) (see Theorem 1 in [156]).

Dense Wishart cases.

A similar phenomenon occurs in the Wishart case, i.e., for sample covariance matrices (corresponding to bipartite graphs); in the case of dense matrices with converging aspect ratio, the variance is given in different forms in [22, 24]. Although these expressions are not explicit in terms of a Chebyshev polynomial expansion, in [52, 128] it is shown that the covariance between two linear statistics is diagonalized by shifted Chebyshev polynomials. When the aspect ratio goes to ∞, [60] computes the variance which is consistent with the Wigner case in [25]. So far we are not aware of any CLT results for sparse bipartite Erdős-Rényi graphs, but a similar argument as in [156] should apply.

For dependent entries (biregular bipartite graphs), we obtain here the variance of the eigenvalue fluctuation in Theorem 3.4.7, and it matches the one in [60], except for the first coefficient.

Constant (expected or deterministic) degree.

When $p = \frac{c}{n}$, the explicit limiting spectral distribution for Erdős-Rényi graphs $G(n, p)$ is not known, although it is known that the measure μ_c exists for every c (given, e.g., by a Stieltjes transform equation as in [39]), and if $c > 1$ it consists of a continuous part and an atomic part [41]. Convergence of μ_c to the semicircular distribution is studied in [81, 121], where asymptotic expressions for the moments of μ_c with an $o(1/c)$ term are computed (as $c \to \infty$, μ_c converges to the semicircle law).

However, a CLT for Erdős-Rényi graphs $G(n, \frac{c}{n})$ still holds [155, 33] with a more complicated variance that does not follow the same expression as in (3.1.1), see Theorem 2.2 in [33].

By contrast, in the random d-regular graph case with d finite, the fluctuations are no longer

Gaussian. Instead, they are modeled by an infinitely divisible distribution, expressed as a sum of Poisson variables (see [76] for the permutation model and [119] for the uniform model). Notably, in the case when the matrix is not symmetric and corresponds to (directed) cycle structure of a random permutation, [30] showed that the global fluctuations can be computed, and whether or not the limiting distribution is Gaussian depends on how smooth the test function is.

For the bipartite Erdős-Rényi case, once again the limiting distribution is not known but results similar to [81] can be found in [150]. We are not aware of any CLT-like results for the fluctuations in this case.

We compute here the fluctuations for the uniformly random biregular bipartite with fixed degrees. Just like in the regular case [76], we see that the fluctuations are modeled by a sum of Poisson variables (Theorem 3.4.4).

Remark 3.1.1. In all cases where a formula for the variance has been obtained by expressing f in a polynomial basis which diagonalizes the covariance, the correct basis has been given by the orthogonal polynomials with respect to the limiting distribution, [161, 76].

Random biregular bipartite graphs

The adjacency matrix of a (d_1,d_2)-biregular bipartite graph $G(V_1 \cup V_2, E)$ with $V_1 = [n], V_2 = [m]$ can be written as

$$A = \begin{bmatrix} 0 & X \\ X^\top & 0 \end{bmatrix},$$

where $X \in \{0,1\}^{n \times m}$ is a matrix indexed by $V_1 \times V_2$ such that $X_{ij} = 1$ if and only if $(i,j) \in E$. All eigenvalues of A come in pairs as $\{-\lambda, \lambda\}$, where $|\lambda|$ is a singular value of X, along with extra $|n-m|$ zero eigenvalues. It's easy to see $\lambda_1(A) = -\lambda_{n+m}(A) = \sqrt{d_1 d_2}$.

In this chapter, instead of examining the spectrum of A, we will be looking at the spectrum of the matrix $XX^\top - d_1 I$. This serves two purposes: one, it allows for an immediate parallel to the sample covariance matrix (Wishart) case, and two, it allows us to deal with all regimes in a unitary fashion. The eigenvalues of $XX^\top - d_1 I$ are the shifted squares of the eigenvalues of A. Any result on global fluctuations for linear statistics of the spectrum of $XX^\top - d_1 I$ is automatically converted into an equivalent result for the spectrum of A. However, because any result of fluctuations must necessarily put most of the eigenvalues (with the exception of the deterministic outliers) on a compact interval, scaling must be involved. This works perfectly fine when the ratio d_1/d_2 is bounded, but it becomes tricky when it is not, and the matrix $XX^\top - d_1 I$ allows us to do the scaling in a more natural way, similarly to the sample covariance (Wishart) matrix with unbounded aspect ratio in [60].

To prove a result on eigenvalue fluctuations, we need two special ingredients: *eigenvalue confinement on a compact interval* and *asymptotic behavior of cycle counts*. For the former, we make use of the spectral gap shown in [48] for the fixed degree case and [179] for the growing degree case. Previous results of this kind were obtained for random regular graphs [94, 37] for fixed degree, and [49, 67, 168, 28] for growing degrees.

For the latter, we use Stein's method to approximate cycle counts as Poisson random variables by bounding the total variation distance (Theorem 3.2.10) and obtain a Poisson approximation of the number of cyclically non-backtracking walks (Corollary 3.2.15). Note that computing cycle counts is a fundamental problem in the study of random graphs, ever since the seminal papers of [142] and more general [143, 144].

To prove our results, we follow the recipe of [119] by using switching to construct exchangeable pairs of graphs that allow us to estimate cycle counts. The switching we use here is different from [119] and is suitable for biregular bipartite graphs. In the analysis of switchings, a new challenge is the imbalance between the parameters d_1, d_2 when the aspect ratio is unbounded. Our results on cycle counts hold for a large range of d_1, d_2, and are notably independent of the aspect ratio as long as the cycle length is small. It is also worth noting that the method of switching has been applied to other problems on random biregular bipartite graphs, for example, [54, 53, 153].

Finally, we also obtain an algebraic relation between linear eigenvalue statistics on modified Chebyshev polynomials and cyclically non-backtracking walks (Theorem 3.2.17). Then based on the spectral gap results in [48, 179] and approximation theory for Chebyshev polynomials [171], we extend the eigenvalue fluctuation results to a general class of analytic functions.

3.1.2 Main results

Our main contributions are represented by Theorems 3.4.4, 3.4.7, establishing the behavior of the global fluctuations for the linear statistics of eigenvalues of RBBGs in the fixed d_1, d_2, respectively, in the $d_1 \cdot d_2 \to \infty$ cases. Note that Theorem 3.4.7 describes the behavior of the fluctuations even in the case when the limiting ESD does not exist, since it merely requires d_1/d_2 to be bounded, rather than to converge to a number in $[1, \infty)$ (which would be the necessary condition for the ESD to converge). In addition, we show that the covariance between two linear statistics with different test functions is given by the coefficients in their Chebyshev expansions.

As part of the proofs for our main results, we also describe the asymptotic behavior of the cycle counts (Theorem 3.2.10). Based on the cycle counts estimates, we then use the locally tree-like structure of RBBGs to prove a global semicircle law in the case when the degree goes slowly and d_1/d_2 is unbounded. Finally, as an important application, we obtain equivalent results for uniformly distributed random regular hypergraphs, including cycle counts, global laws, spectral gaps, and eigenvalue fluctuations.

In Section 3.2 we prove our results on cycle counts in random biregular bipartite graphs. Section 3.3 collects relevant results for the spectral gap and eigenvalue confinement on a compact interval from the literature. Section 3.4 proves our main results, Theorems 3.4.4 and 3.4.7. Section 3.5 proves a global semicircle law for RBBGs when d_1/d_2 is unbounded. In Section 3.6, we uses the connections established in [79] to prove several results on uniformly distributed regular hypergraphs.

3.2 Cycle counts

3.2.1 Counting switchings

In this section, we estimate the number of switchings that create or delete a cycle in a biregular bipartite graph. The precise definitions of switchings for our purposes are given in Definition 3.2.4 and Definition 3.2.5. These estimates will be used in Section 3.2.2 to show that cycle counts converge in distribution to Poisson random variables.

Definition 3.2.1 (cycle)**.** Throughout the paper, when we say a *cycle*, we mean a *simple cycle*, i.e., all vertices in a cycle are distinct.

Let $K_{n,m}$ be the complete bipartite graph on $n+m$ vertices with $V_1=[n], V_2=[m]$. Let $H\subseteq K_{n,m}$ be a subgraph with v vertices. For any $i\in K_{n,m}$, let g_i, h_i denote the degree of i considered as a vertex in a biregular bipartite graph $G=(V_1,V_2,E)$ and the subgraph H, respectively. Let $h_{\max}$ be the largest value of h_i and $|H|$ be the number of edges of H. Denote by

$$[x]_a = x(x-1)\cdots(x-a+1)$$

the falling factorial. The following estimate is given in [143].

Proposition 3.2.2 (Theorem 3.5 in [143])**.** Assume $d_1\geq d_2$ and $nd_1\geq 2d_1(d_1+h_{\max}-2)+|H|+1$. Then

$$\mathbb{P}(H\subseteq G)\leq \frac{\prod_{i=1}^{v}[g_i]_{h_i}}{[nd_1-4d_1^2-1]_{|H|}}.$$

We first prove several estimates on random biregular bipartite graphs based on Proposition 3.2.2.

Lemma 3.2.3. *Let G be a random (d_1,d_2)-biregular bipartite graph with $d_2\leq d_1\leq n^{1/3}$.*

1. *Suppose H is a subgraph of the complete graph $K_{n,m}$ in which every vertex has degree at least 2. Let e be the number of edges in H. Suppose $e = o(n^{1/3})$. Then*

$$\mathbb{P}(H \subseteq G) \leq c_1 \left(\frac{(d_1-1)(d_2-1)}{nm} \right)^{e/2}. \tag{3.2.1}$$

2. *Let α be a cycle of length $2k$ in the complete bipartite graph $K_{n,m}$. Suppose $k \leq n^{1/10}$, then*

$$\mathbb{P}(\alpha \subseteq G) \leq c_1 \left(\frac{(d_1-1)(d_2-1)}{nm} \right)^{k}. \tag{3.2.2}$$

3. *Let β be another cycle of length $2j \leq 2n^{1/10}$ in the complete bipartite graph $K_{n,m}$. Suppose α, β share f edges. Then*

$$\mathbb{P}(\alpha \cup \beta \subseteq G) \leq c_1 \left(\frac{(d_1-1)(d_2-1)}{nm} \right)^{j+k-f/2}. \tag{3.2.3}$$

Proof. It suffices to prove (3.2.1). Then (3.2.2) and (3.2.3) follow as special cases. Since H has e edges, and H is bipartite, it satisfies

$$\sum_{i \in V_1} h_i = \sum_{i \in V_2} h_i = e.$$

Since $h_i \geq 2$ for all $i \in V(H)$, we know $[g_i]_{h_i} \leq (g_i(g_i-1))^{h_i/2}$. Therefore from Proposition 3.2.2,

$$\begin{aligned} \mathbb{P}(H \subseteq G) &\leq \frac{(d_1(d_1-1))^{e/2}(d_2(d_2-1))^{e/2}}{[nd_1 - 4d_1^2 - 1]_e} \\ &= \left(\frac{(d_1-1)(d_2-1)}{nm} \right)^{e/2} \frac{(nd_1)^e}{[nd_1 - 4d_1^2 - 1]_e}. \end{aligned}$$

Recall $d_1 \leq n^{1/3}, e = o(n^{1/3})$, and $(1+x)^r = 1 + O(rx)$ if $rx \to 0$. We have for some absolute constant $c_1 > 0$,

$$\frac{(nd_1)^e}{[nd_1 - 4d_1^2 - 1]_e} \le \left(\frac{nd_1}{nd_1 - 4d_1^2 - e}\right)^e = \left(1 + \frac{4d_1^2 + e}{nd_1 - 4d_1^2 - e}\right)^e \le c_1. \tag{3.2.4}$$

This proves (3.2.1). □

Let G be a (d_1, d_2)-biregular bipartite graph. Let C_j be the number of cycles of length $2j$ in G. We will always represent a cycle by a vertex sequence starting from a vertex in V_1. Suppose $\alpha = (x_1, y_1, \cdots, x_k, y_k)$ is a cycle of length $2k$ in G with $x_i \in V_1, y_i \in V_2$, $1 \le i \le k$, where y_k is connected to x_1 in the cycle α.

Let $e_i = u_i v_i, e_i' = u_i' v_i'$ be the edges with with $u_i, u_i' \in V_1, v_i, v_i' \in V_2$, $1 \le i \le k$ such that neither u_i, u_i' is adjacent to y_i for $1 \le i \le k$ and neither v_i, v_i' is adjacent to x_i. See the left part of Figure 3.1 for an example.

We now introduce our definitions of switching for biregular bipartite graphs.

Definition 3.2.4 (forward α-switching)**.** Consider the action of deleting all $4k$ edges $\tilde{e}_i, 1 \le i \le 2k$ and $e_i, e_i', 1 \le i \le k$, and replacing them by the edges $x_i v_i, x_i v_i', y_i u_i, y_i u_i'$ for $1 \le i \le k$. We obtain a new biregular bipartite graph G' with the cycle α deleted. We call this action induced by the 6 sequences $(x_i), (y_i), (u_i), (u_i'), (v_i), (v_i'), 1 \le i \le k$ a *forward* α*-switching*. See Figure 3.1 for an example. We will consider forward α-switchings only up to cyclic rotation and inversion of indices in $[k]$; that is, we identify the $2k$ different forward α-switchings obtained by applying the same cyclic rotation or inversion on $[k]$ to the 6 sequences $(x_i), (y_i), (u_i), (u_i'), (v_i), (v_i'), 1 \le i \le k$.

Definition 3.2.5 (backward α-switching)**.** Suppose G contains paths $v_i x_i v_i'$ and $u_i y_i u_i'$ for $1 \le i \le k$, where $x_i, u_i, u_i' \in V_1, y_i, v_i, v_i' \in V_2$. Consider deleting all $4k$ edges $v_i x_i, v_i' x_i, u_i y_i, u_i' y_i$ for $1 \le i \le k$, and replacing them with $u_i v_i, u_i' v_i', x_i y_i, y_i x_{i+1}$ for $1 \le i \le k$. We obtain a new graph G' with a cycle $\alpha = (x_1, y_1, \cdots, x_k, y_k)$. Such action is called a *backward* α*-switching* induced by the sequences $(x_i), (y_i), (u_i), (u_i'), (v_i), (v_i'), 1 \le i \le k$. We also identify the $2k$ different backward α-switchings obtained by applying the same cyclic rotation or inversion on the index set $[k]$.

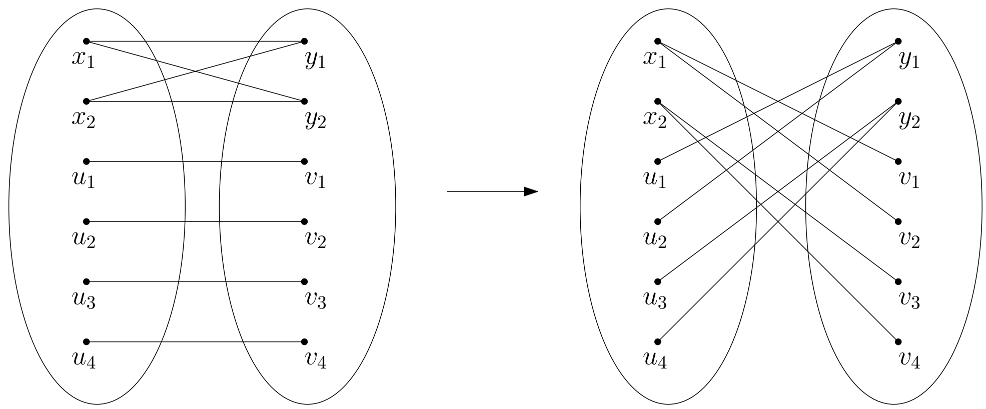

Figure 3.1: A forward α switching from the left to the right, where $\alpha = (x_1, y_1, x_2, y_2)$.

Definition 3.2.6 (short cycles)**.** Let r be an integer; we say that a cycle is *short* if its length is less than or equal to $2r$.

We call a α-switching *valid* if α is the only short cycle created or destroyed by the switching. For each forward α-switching from G to G', there is a corresponding backward α-switching from G' to G by simply reversing the operation (i.e. from the right to the left in Figure 3.1).

Let F_α be the number of all valid forward α-switchings from G to some G' and let B_α be the number of all valid backward α-switchings from some G' to G. In the following two lemmas, we estimate F_α and B_α for biregular bipartite graphs.

Lemma 3.2.7. *Let G be a deterministic (d_1, d_2)-biregular bipartite graph with $d_1 \geq d_2$ and cycle counts $C_k, 2 \leq k \leq r$. For any short cycle $\alpha \subseteq G$ of length $2k$, we have*

$$F_\alpha \leq [n]_k [m]_k d_1^k d_2^k. \tag{3.2.5}$$

If α does not share an edge with another short cycle, then for an absolute constant $c_1 > 0$, we have

$$F_\alpha \geq [n]_k [m]_k d_1^k d_2^k \left(1 - \frac{4k \sum_{j=2}^r jC_j + c_1 k (d_1 - 1)^r (d_2 - 1)^r}{n d_1}\right). \tag{3.2.6}$$

Proof. Consider a cycle denoted by $\alpha = (x_1, y_1, \cdots, x_k, y_k)$. Denote edges

$$\tilde{e}_i = x_i y_i, \quad \tilde{e}_{i+k} = y_i x_{i+1}, 1 \le i \le k, \tag{3.2.7}$$

where $x_{k+1} := x_1$. There are at most $[n]_k d_1^k [m]_k d_2^k$ many ways to choose edges $e_i = u_i v_i$ and $e_i' = u_i' v_i'$ for $1 \le i \le k$, which gives the upper bound (3.2.5). For the k edges $e_i, 1 \le i \le k$, we require distinct $u_i \in V_1, 1 \le i \le k$, and we have d_1 choices for each v_i, given the degree constraint on u_i. This gives $[n]_k d_1^k$ many choices altogether. For the remaining edges $e_i', 1 \le i \le k$ we require distinct $v_i' \in V_2, 1 \le i \le k$ and each for each u_i' we have d_2 choices, giving us a factor of $[m]_k d_2^k$. Therefore (3.2.5) holds.

For the rest of the proof, we always use the same way to count α-switchings by counting the choices of e_i, e_i'. We use the parameter d_1 to control the choices from $e_i, 1 \le i \le k$ and the parameter d_2 for the choices from $e_i', 1 \le i \le k$.

To prove the lower bound in (3.2.6), we choose a subset of configurations that are guaranteed to have a valid forward α-switching. Consider $e_i, e_i', 1 \le i \le k$ such that the following holds:

1. e_i and e_i' are not contained in any short cycle in G for $1 \le i \le k$.
2. The distance from any vertex in $\{e_i, e_i'\}$ to any vertex in $\tilde{e}_i$ is at least $2r$ for any $1 \le i \le k$.
3. The distance between any two different edges among the $2k$ edges $\{e_i, e_i', 1 \le i \le k\}$ is at least r.
4. For all $1 \le i \le k$, the distance between v_i and v_i' is at least $2r$, and the distance between u_i and u_i' is at least $2r$.

Recall the definition of $\tilde{e}_i$ in (3.2.7). By Condition (2), for all $1 \le i \le k$, u_i, u_i' are not adjacent to y_i, also v_i, v_i' are not adjacent to x_i, which satisfies the definition of a forward α-switching. Let G' be the graph obtained by applying the forward α-switching from G. We need to check that α is the only cycle deleted in G by this switching and no other short cycles are created in G'.

Since α shares no edges with other short cycles by our assumptions, deleting α will not destroy other short cycles. From Condition (1), deleting e_i, e_i' will not destroy any short cycles either.

Next we show no other short cycles are created in G'. Suppose there exists a new short cycle β in G' created by the switching. Then β contains paths in $G \cap G'$ separated by edges created in the forward switching in G' (β must contain at least such edge because it is created). Any such path in $G \cap G'$ must have length at least r, because

- if it starts and ends at vertices in α and has length less than r, then combining this path with a path in α gives a short cycle in G that intersects α, which is a contradiction to our assumption on α;

- if it starts in α and ends in $\{u_i, v_i, u_i', v_i'\}$ for some i and has length less than r, then combining this path with a path in α gives a path between $\tilde{e}_i$, e_i or between $\tilde{e}_i$, e_i' of length less than $2r$, which violates Condition (2);

- if it starts and ends in different edges among $\{e_i, e_i', 1 \leq i \leq k\}$, then it must have length at least r by Condition (3);

- if it starts at some vertex in e_i and ends at some vertex in e_i, then the path must start and end at different vertices in e_i, otherwise β is not a cycle in the sense of Definition 3.2.1. Then the path combined with e_i is a cycle. By Condition (1), it has length at least r, a contradiction. In the same way, it cannot start at some vertex in e_i' and end at some vertex in e_i'.

This implies β contains exactly one path in $G \cap G'$. If not, the two separated paths together with new edges in G' have length greater than $2r$, a contradiction to the condition that β is a short cycle. Given the path in $G \cap G'$, the remainder of β has two cases:

- a single edge that can be x_iv_i, x_iv_i', y_iu_i, or y_iu_i' for some $1 \le i \le k$, then by Condition (2), the path in $G \cap G'$ connecting the two vertices in the edge has length at least $2r$, which is a contradiction to the fact that β is a short cycle;
- a single path $v_ix_iv_i'$ or $u_iy_iu_i'$, which is impossible by Condition (4).

From the analysis above, no such β can exist, hence any α-switching satisfying Conditions (1)-(4) is valid.

Next we find the number of all switchings satisfying Conditions (1) to (4) to have a lower bound on F_α. We will do this by bounding from above the number of switchings out of the $[n]_k[m]_kd_1^kd_2^k$ many choices counted in (3.2.5) that fail one of the Conditions (1)-(4). We treat the 4 conditions in the following (a)-(d) parts.

(a) There are a total of at most $\sum_{i=2}^{r} 2jC_j$ edges in all short cycles of G. For some $1 \le i \le k$, if we choose one edge e_i from a short cycle and the other $(2k-1)$ edges arbitrarily, we obtain a forward α-switching that fails Condition (1). The number of all possible choices is at most

$$k\sum_{j=2}^{r} 2jC_j \cdot [n-1]_{k-1}[m]_kd_1^{k-1}d_2^k.$$

And if we choose e_i' from a short cycle and the other $(2k-1)$ edges arbitrarily, the number of all possible choices is at most

$$k\sum_{j=2}^{r} 2jC_j \cdot [n]_k[m-1]_{k-1}d_1^kd_2^{k-1}.$$

Altogether the number of choices is at most

$$\frac{4}{nd_1}[n]_k[m]_k(d_1d_2)^kk\sum_{j=2}^{r} jC_j. \tag{3.2.8}$$

(b) To fail Condition (2), we can obtain α-forward switchings by choosing $(2k-1)$ edges arbitrarily, and then choose one edge e_i or e'_i that is at most of distance $2r-1$ from $\tilde{e}_i$ for some $1 \le i \le k$. From the degree constraints, the number of edges of distance less than $2r$ from some edge is at most $O((d_1-1)^r(d_2-1)^r)$. Similar to Part (a), by considering whether e_i or e'_i is chosen for $1 \le i \le k$, the number of such switchings is at most

$$\begin{aligned}&\left([n-1]_{k-1}[m]_k d_1^{k-1} d_2^k + [n]_k[m-1]_k d_1^k d_2^{k-1}\right) \cdot k \cdot O((d_1-1)^r(d_2-1)^r)\\ =&\frac{1}{nd_1}[n]_k[m]_k(d_1d_2)^k kO((d_1-1)^r(d_2-1)^r). \end{aligned} \tag{3.2.9}$$

(c) For Condition (3), there are three cases to consider depending on whether the pair is (e_i, e_j), (e'_i, e'_j) or (e_i, e'_j).

Suppose the pair e_i, e_j violates Condition (3). We pick the pair of edges that are within distance $r-1$ and pick the remaining $(2k-2)$ edges arbitrarily. There are (nd_1) many ways to choose e_i. When e_i is fixed, there are at most $O((d_1-1)^{(r+1)/2}(d_2-1)^{(r+1)/2})$ choices for e_j. Hence the number of switchings that fail Condition (3) is at most

$$\begin{aligned}&(nd_1)\cdot[O((d_1-1)^{(r+1)/2}(d_2-1)^{(r+1)/2})]\cdot k(k-1)\cdot([n-2]_{k-2}[m]_k d_1^{k-2} d_2^k)\\ =&\frac{1}{nd_1}[n]_k[m]_k d_1^k d_2^k \cdot k^2 \cdot O\left((d_1-1)^{(r+1)/2}(d_2-1)^{(r+1)/2}\right). \end{aligned}$$

By the same argument, if the pair is (e'_i, e'_j), the number of switchings that fail Condition (3) is at most

$$\frac{1}{nd_1}[n]_k[m]_k d_1^k d_2^k k^2 \cdot O\left((d_1-1)^{(r+1)/2}(d_2-1)^{(r+1)/2}\right).$$

When the two edges of the pair violating Condition (3) are e_i, e'_j for some i, j, the number is at most

$$
\begin{aligned}
&(nd_1)\cdot[O((d_1-1)^{(r+1)/2}(d_2-1)^{(r+1)/2})]\cdot(2k^2)\cdot([n-1]_{k-1}[m-1]_{k-1}d_1^{k-1}d_2^{k-1})\\
=&\frac{1}{nd_1}[n]_k[m]_kd_1^kd_2^kk^2O\left((d_1-1)^{(r+1)/2}(d_2-1)^{(r+1)/2}\right).
\end{aligned}
$$

Combining the three cases in Part (c), the number of switchings that violate Condition (3) is at most

$$
\frac{1}{nd_1}[n]_k[m]_k(d_1d_2)^kk^2\cdot O\left((d_1-1)^{(r+1)/2}(d_2-1)^{(r+1)/2}\right). \tag{3.2.10}
$$

(d) Since the distance between a pair of vertices in V_1 or V_2 must be even, to violate Condition (4), we can choose a pair $u_i, u_i' \in V_1$ or $v_i, v_i' \in V_2$ that are within distance $2r-2$ first, then choose other edges arbitrarily. Similar to the cases above, the number of switchings that fail Condition (4) is at most

$$
\frac{1}{nd_1}[n]_k[m]_kd_1^kd_2^kkO((d_1-1)^r(d_2-1)^r). \tag{3.2.11}
$$

Combining the 4 Cases (a)-(d) above, from (3.2.8), (3.2.9),(3.2.10), and (3.2.11), we have at most

$$
\frac{4}{nd_1}[n]_k[m]_k(d_1d_2)^k\left(k\sum_{j=2}^{r}jC_j+O(k(d_1-1)^r(d_2-1)^r)\right)
$$

many switchings that fail one of the Conditions (1)-(4) among the $[n]_k[m]_k(d_1d_2)^k$ possible switchings. Then for an absolute constant $c_1>0$,

$$
F_\alpha \geq [n]_k[m]_kd_1^kd_2^k\left(1-\frac{4k\sum_{j=2}^r jC_j+c_1k(d_1-1)^r(d_2-1)^r}{nd_1}\right).
$$

Therefore (3.2.6) holds. $\square$

For the number of backward switchings, we obtain a similar upper bound, but the lower bound is only in expectation.

Lemma 3.2.8. *Let G be a random (d_1,d_2)-biregular bipartite graph and let α be a cycle of length $2k \leq 2r$ in the complete bipartite graph $K_{n,m}$. Let B_α be the number of valid backward switchings from G that create α. Then*

$$B_\alpha \leq (d_1(d_1-1))^k(d_2(d_2-1))^k, \tag{3.2.12}$$

and there is an absolute constant $c_2 > 0$ such that

$$\mathbb{E}B_\alpha \geq (d_1(d_1-1))^k(d_2(d_2-1))^k\left(1-\frac{c_2k(d_1-1)^r(d_2-1)^r}{nd_1}\right). \tag{3.2.13}$$

Proof. Given α, from the degree constraints, the number of choices for $u_i, u_i', v_i, v_i', 1 \leq i \leq k$ that yield a valid backward α switching is at most $(d_1(d_1-1))^k(d_2(d_2-1))^k$, which gives (3.2.12).

For the lower bound, we consider the quantity $B := \sum_\beta B_\beta$, where β is summing over all possible cycles of length $2k$ in the complete bipartite graph $K_{n,m}$. As in the proof of Lemma 3.2.7, we give conditions that guarantee a valid backward switching.

Assume $\beta = (x_1, y_1, \cdots, x_k, y_k)$. We first consider backward switchings that create β. Suppose the paths $v_ix_iv_i', u_iy_iu_i', 1 \leq i \leq k$ in G satisfy the following conditions:

1. The edges x_iv_i, x_iv_i', y_iu_i, and y_iu_i' are not contained in any short cycles.
2. For $1 \leq i \leq k$, the distance between any vertex in the path $v_ix_iv_i'$ and any vertex in the path $u_iy_iu_i'$ is at least $2r$.
3. For all $1 \leq i \leq k$ and $1 \leq j \leq k/2$, the distance between the paths $v_ix_iv_i'$ and $v_{i+j}x_{i+j}v_{i+j}'$ (the index $i+j$ is calculated modulo k) and the distance between $u_iy_iu_i'$ and $u_{i+j}y_{i+j}u_{i+j}'$

are at least $2r-2j+1$.

4. For $1 \le i \le k, 1 \le j \le k/2$, the distance between $v_i x_i v_i'$ and $u_{i+j} y_{i+j} u_{i+j}'$, and the distance between $u_i y_i u_i'$ and $v_{i+j} x_{i+j} v_{i+j}'$ are at least $2r-2j+2$.

We will show the four conditions above guarantee a valid backward β-switching.

By Condition (1), no short cycles are deleted. We denote $x \not\sim y$ if two vertices x, y are not connected in G. An immediate consequence of Condition (2) ensures that $x_i \not\sim y_i$ and $u_i \not\sim v_i$, $u_i' \not\sim v_i'$, and Condition (4) ensures that $y_i \not\sim x_{i+1}$. Therefore such switching can be applied.

Let G' be the graph obtained by applying the backward β-switching. We need to check that no short cycles other than β are created in G'.

Suppose a short cycle $\beta' \neq \beta$ is created. Then β' possibly consists of paths in $G \cap G'$, portions of β, and edges $u_i v_i, u_i' v_i'$ for some $1 \le i \le k$. Any such path in $G \cap G'$ must have length at least r because

- if it starts in one of the sets $\{x_i, v_i, v_i'\}$ or $\{y_i, u_i, u_i'\}$ for $1 \le i \le k$, and ends at a different set $\{x_j, v_j, v_j'\}$ or $\{y_j, u_j, u_j'\}$ for $1 \le j \le k$, then Conditions (2), (3) and (4) imply this;
- if it starts and ends in the same set $\{x_i, v_i, v_i'\}$ or $\{y_i, u_i, u_i'\}$, then it follows from Condition (1) that the path must have length at least r.

It follows that β' must contain exactly one such path, otherwise if two such paths are included in β', the length of β' is greater than $2r$, a contradiction to the fact that β' is a short cycle.

Besides this path in $G \cap G'$, the remainder of β' must either be an edge $u_i v_i$ or $u_i' v_i'$, or a portion of β. If the remainder is some $u_i v_i$, then the distance between u_i and v_i in G is at most $2r-1$, a contradiction to Condition (2). The same holds if the remainder is some $u_i' v_i'$.

If the remainder is a portion of β, then there exists two vertices in β connected by the path in $G \cap G'$ contained in β'. If the two vertices are x_i, x_{i+j} for some $1 \le i \le k, 1 \le j \le k/2$, then from Condition (3), the path in $G \cap G'$ contained in β' that connects the two vertices has length at least $2r-2j+1$. Since the path in β connecting x_i, x_{i+j} has length $2j$, this implies β' has length at least $(2r-2j+1)+2j = 2r+1$, a contradiction. In the same way, if the two vertices are y_i, y_{i+j} for some $1 \le i \le k, 1 \le j \le k/2$, we can find a contradiction for β' from Condition (3).

If the two vertices connected by the path are x_i, y_{i+j} with $1 \le i \le k, 1 \le j \le k/2$, then the path in β connecting the two vertices has length at least $2j-1$. Combining the path in $G \cap G'$ contained in β', from Condition (4), we conclude that β' has length at least $(2r-2j+2)+(2j-1) = 2r+1$, a contradiction. By the same argument, if the two vertices connected by the path are y_i, x_{i+j} for some $1 \le i \le k, 1 \le j \le k/2$, we can find a contradiction that β' is not a short cycle.

Therefore such β' does not exist, and all backward switchings satisfying Conditions (1)-(4) are valid.

There are $[n]_k[m]_k/(2k)$ choices for the $2k$-cycle β in the complete bipartite graph $K_{n,m}$, and at most $(d_1(d_1-1))^k(d_2(d_2-1))^k$ choices for $u_i, u_i', v_i, v_i', 1 \le i \le k$ given β. We now count how many possible backward switchings violate one of the four Conditions (1)-(4) to get a lower bound on B. We treat the Conditions (1)-(4) in 4 parts.

(a) Suppose Condition (1) is violated. We estimate the number of switchings by choosing one edge from the set of edges in short cycles and the other edges arbitrarily. Note that by our definition of switchings, we identify $2k$ different switchings by applying the cyclic rotation or inversion on $[k]$. Suppose we choose an edge $x_i v_i$ or $x_i v_i'$ from short cycles, similar to the analysis in Lemma (3.2.7), the number of switchings is at most

$$\begin{aligned}&\left(2k\sum_{j=2}^{r} 2jC_j \cdot (d_1-1)\right) \cdot \left(\frac{1}{2k}[n-1]_{k-1}(d_1(d_1-1))^{k-1} \cdot [m]_k(d_2(d_2-1))^k\right)\\ =&\frac{2}{nd_1}[n]_k[m]_k[d_1(d_1-1)d_2(d_2-1)]^{2k}\sum_{j=2}^{r} jC_j.\end{aligned}$$

Similarly, if we choose an edge y_iu_i or y_iu_i' from short cycles, the number of switchings is at most

$$\begin{aligned}&\left(2k\sum_{j=2}^{r}2jC_j\cdot(d_2-1)\right)\cdot\left(\frac{1}{2k}[m-1]_{k-1}(d_2(d_2-1))^{k-1}\cdot[n]_k(d_1(d_1-1))^k\right)\\=&\frac{2}{nd_1}[n]_k[m]_k[d_1(d_1-1)d_2(d_2-1)]^{2k}\sum_{j=2}^{r}jC_j.\end{aligned}$$

Combining two parts, the number of switchings that violate Condition (1) is at most

$$\frac{8k}{nd_1}[n]_k[m]_k[d_1(d_1-1)d_2(d_2-1)]^k\sum_{j=2}^{r}jC_j. \tag{3.2.14}$$

(b) Suppose for some $1\le i\le k$, two paths $v_ix_iv_i'$ and $u_iy_iu_i'$ are within distance $2r-1$. The number of switching is at most

$$\begin{aligned}&\frac{[n-1]_{k-1}[m-1]_{k-1}}{2k}[d_1(d_1-1)d_2(d_2-1)]^{k-1}\cdot(knd_1(d_1-1))\cdot O((d_1-1)^r(d_2-1)^{r+1})\\=&\frac{1}{nd_1}[n]_k[m]_k(d_1(d_1-1))^k(d_2(d_2-1))^kO((d_1-1)^r(d_2-1)^r).\end{aligned} \tag{3.2.15}$$

(c) Suppose for some $1\le i\le k, 1\le j\le k/2$, two paths $\{v_ix_iv_i', v_{i+j}x_{i+j}v_{i+j}'\}$ are within distance $2r-2j$. The number of switchings is at most

$$\begin{aligned}&\frac{[n-2]_{k-2}[m]_k}{2k}(d_1(d_1-1))^{k-2}(d_2(d_2-1))^k\cdot nd_1(d_1-1)\sum_{i=1}^{k}\sum_{j=1}^{\lfloor k/2\rfloor}O((d_1-1)^{r-j+2}(d_2-1)^{r-j})\\=&\frac{1}{nd_1}[n]_k[m]_k(d_1(d_1-1))^k(d_2(d_2-1))^kO((d_1-1)^r(d_2-1)^{r-1}).\end{aligned}$$

Suppose for some $1\le i\le k, 1\le j\le k/2$, two paths $\{u_iy_iu_i', u_{i+j}y_{i+j}u_{i+j}'\}$ are within distance $2r-2j$. Similarly, the number of switchings is bounded by

$$\frac{1}{nd_1}[n]_k[m]_k(d_1(d_1-1))^k(d_2(d_2-1))^kO((d_1-1)^r(d_2-1)^{r-1}).$$

Therefore the number of switchings that violate Condition (3) is at most

$$\frac{1}{nd_1}[n]_k[m]_k(d_1(d_1-1))^k(d_2(d_2-1))^k O((d_1-1)^r(d_2-1)^{r-1}). \tag{3.2.16}$$

(d) Suppose two paths $v_i x_i v'_i, u_{i+j} y_{i+j} u'_{i+j}$ for some $1 \le i \le k, 1 \le j \le k/2$ are within distance $2r-2j+1$. The number of choices is at most

$$\begin{aligned}
&\frac{[n]_k[m-1]_{k-1}}{2k}(d_1(d_1-1))^k(d_2(d_2-1))^{k-1}\sum_{i=1}^{k}\sum_{j=1}^{\lfloor k/2\rfloor} O((d_1-1)^{r-j+1}(d_2-1)^{r-j+2}\\
=&\frac{1}{nd_1}[n]_k[m]_k(d_1(d_1-1))^k(d_2(d_2-1))^k O\left((d_1-1)^r(d_2-1)^r\right).
\end{aligned}$$

Suppose two paths $u_i y_i u'_i, v_{i+j} x_{i+j} v'_{i+j}$ for some $1 \le i \le k, 1 \le j \le k/2$ are within distance $2r-2j+1$. By the same argument, the number of choices is at most

$$\frac{1}{nd_1}[n]_k[m]_k(d_1(d_1-1))^k(d_2(d_2-1))^k O\left((d_1-1)^r(d_2-1)^r\right).$$

Then the number of switchings that violate Condition (4) is at most

$$\frac{1}{nd_1}[n]_k[m]_k(d_1(d_1-1))^k(d_2(d_2-1))^k O\left((d_1-1)^r(d_2-1)^r\right). \tag{3.2.17}$$

From (3.2.14), (3.2.15), (3.2.16) and (3.2.17), the lower bound of B is given by

$$B \ge \frac{[n]_k[m]_k}{2k}(d_1(d_1-1))^k(d_2(d_2-1))^k\left(1-\frac{8k\sum_{j=2}^{r} jC_j + O(k(d_1-1)^r(d_2-1)^r)}{nd_1}\right). \tag{3.2.18}$$

By Lemma 3.2.3 (b),

$$\mathbb{E}C_k \le \frac{[n]_k[m]_k}{2k}\frac{c_1(d_1-1)^k(d_2-1)^k}{n^k m^k} \le \frac{c_1(d_1-1)^k(d_2-1)^k}{2k}.$$

Applying the inequality above to (3.2.18), we obtain

$$\mathbb{E}B \geq \frac{[n]_k[m]_k}{2k}(d_1(d_1-1))^k(d_2(d_2-1))^k\left(1-\frac{O(k(d_1-1)^r(d_2-1)^r)}{nd_1}\right).$$

By the exchangeability of the vertex labels in the uniformly distributed RBBG model, the law of B_β is the same for any $2k$-cycle β. Then

$$\mathbb{E}B_\alpha = \frac{2k}{[n]_k[m]_k}\mathbb{E}B \geq (d_1(d_1-1)d_2(d_2-1))^k\left(1-\frac{c_2k(d_1-1)^r(d_2-1)^r}{nd_1}\right),$$

for an absolute constant $c_2 > 0$. This completes the proof. $\square$

3.2.2 Poisson approximation of cycle counts

In this section, we prove the cycle counts in RBBGs are asymptotically distributed as Poisson random variables. The main tool we will use is the following total variation distance bound from [59].

Lemma 3.2.9 (Proposition 10 in [59]). *Let $W = (W_1, \dots, W_r)$ be a random vector taking values in $\mathbb{N}^r$, and let the coordinates of $Z = (Z_1, \dots, Z_r)$ be independent Poisson random variables with $\mathbb{E}Z_k = \mu_k$. Let $W' = (W_1', \dots, W_r')$ be defined on the same space as W, with (W, W') an exchangeable pair. For any choice of σ-algebra $\mathcal{F}$ with respect to which W is measurable and any choice of constants c_k, we have*

$$d_{\mathrm{TV}}(W,Z) \leq \sum_{k=1}^{r} \xi_k \left(\mathbb{E}|\mu_k - c_k\mathbb{P}(\Delta_k^+ \mid \mathcal{F})| + \mathbb{E}|W_k - c_k\mathbb{P}(\Delta_k^- \mid \mathcal{F})|\right), \tag{3.2.19}$$

where $\xi_k := \min\{1, 1.4\mu_k^{-1/2}\}$ *and*

$$\Delta_k^+ := \{W_k' = W_k + 1, W_j = W_j', k < j \le r\}, \tag{3.2.20}$$

$$\Delta_k^- := \{W_k' = W_k - 1, W_j = W_j', k < j \le r\}. \tag{3.2.21}$$

We apply Stein's method to obtain the following Poisson approximation in total variation distance.

Theorem 3.2.10. *Let G be a random (d_1, d_2)-biregular bipartite graph with cycle counts $(C_k, k \ge 2)$. Let $(Z_k, k \ge 2)$ be independent Poisson random variables with*

$$\mu_k := \mathbb{E}Z_k = \frac{(d_1-1)^k (d_2-1)^k}{2k}.$$

For any $n, m \ge 1$ and $r \ge 2, d_1 \ge 3$, there exists an absolute constant $c_6 > 0$ such that

$$d_{\mathrm{TV}}((C_2, \ldots, C_r), (Z_2, \ldots, Z_r)) \le \frac{c_6 \sqrt{r} (d_1-1)^{3r/2} (d_2-1)^{3r/2}}{nd_1}.$$

Proof. If $d_1 > n^{1/3}$ or $r > n^{1/10}$, then

$$\frac{c_6 \sqrt{r} (d_1-1)^{3r/2} (d_2-1)^{3r/2}}{nd_1} > 1$$

for sufficiently large choice of c_6 and the theorem holds trivially. Thus we assume $d_1 \le n^{1/3}$ and $r \le n^{1/10}$. We now construct an exchangeable pair of random biregular bipartite graphs by taking a step in a reversible Markov chain.

Define a graph $\mathcal{G}$ whose vertex set consists of all (d_1, d_2)-biregular bipartite graphs. If there is a valid forward or backward α-switching from a (d_1, d_2)-biregular bipartite graph G_0 to another graph G_1 with the length of α being $2k$, we make an undirected edge in $\mathcal{G}$ between

G_0, G_1 and place a weight of

$$\frac{1}{[n]_k[m]_k(d_1d_2)^k}$$

on each such edge. Define the degree of a vertex in $\mathcal{G}$ to be the sum of weights from all adjacent edges. Let d_0 be the largest degree in $\mathcal{G}$. To make $\mathcal{G}$ regular, we add a weighted loop to each vertex if necessary to increase the degree of all vertices to d_0.

Now consider the simple random walk on $\mathcal{G}$. This is a reversible Markov chain with respect to the uniform distribution on (d_1, d_2)-biregular bipartite graphs. Thus suppose G is a uniformly chosen random biregular bipartite graph, we can obtain another random biregular bipartite graph G' by taking an extra step in the random walk from G, and the pair (G, G') is exchangeable.

Let $\mathcal{J}_k$ be the collection of cycles of length $2k$ in $K_{n,m}$ with $k \leq r$. We have $|\mathcal{J}_k| = [n]_k[m]_k/2k$. Define $I_\alpha = \mathbf{1}\{\alpha \subseteq G\}$. Then $C_k = \sum_{\alpha \in \mathcal{J}_k} \mathbf{1}_\alpha$. Let I'_α, C'_k be defined on G' in the same way. Since G and G' are exchangeable, the vectors $(C_2, \ldots, C_r)$ and $(C'_2, \ldots, C'_r)$ are also exchangeable. We can then apply Lemma 3.2.9 to this exchangeable pair of vectors. Now define two events

$$\begin{aligned}
\Delta_k^+ :=& \{C'_k = C_k + 1, C_j = C'_j, k < j \leq r\}, \\
\Delta_k^- :=& \{C_k = C'_k + 1, C_j = C'_j, k < j \leq r\}.
\end{aligned}$$

By our construction of the exchangeable pair,

$$\begin{aligned}
\mathbb{P}(\Delta_k^+ \mid G) &= \sum_{\alpha \in \mathcal{J}_k} \frac{B_\alpha}{d_0[n]_k[m]_k(d_1d_2)^k}, \\
\mathbb{P}(\Delta_k^- \mid G) &= \sum_{\alpha \in \mathcal{J}_k} \frac{F_\alpha}{d_0[n]_k[m]_k(d_1d_2)^k}.
\end{aligned}$$

Applying Lemma 3.2.9 with all $c_k = d_0, 1 \le k \le r$, we have

$$\begin{aligned}
&d_{\mathrm{TV}}((C_3,\ldots,C_r),(Z_3,\ldots,Z_r)) \\
&\le \sum_{k=2}^{r} \xi_k \mathbb{E}\left|\mu_k - \sum_{\alpha\in\mathcal{I}_k} \frac{B_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right| + \sum_{k=2}^{r} \xi_k \mathbb{E}\left|C_k - \sum_{\alpha\in\mathcal{I}_k} \frac{F_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right| \\
&= \sum_{k=2}^{r} \xi_k \mathbb{E}\left|\sum_{\alpha\in\mathcal{I}_k} \frac{(d_1-1)^k(d_2-1)^k}{[n]_k[m]_k} - \frac{B_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right| + \sum_{k=2}^{r} \xi_k \mathbb{E}\left|\sum_{\alpha\in\mathcal{I}_k} I_\alpha - \frac{F_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right| \\
&\le \sum_{k=2}^{r} \xi_k \left(\sum_{\alpha\in\mathcal{I}_k} \mathbb{E}\left|\frac{(d_1-1)^k(d_2-1)^k}{[n]_k[m]_k} - \frac{B_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right| + \sum_{\alpha\in\mathcal{I}_k} \mathbb{E}\left|I_\alpha - \frac{F_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right|\right).
\end{aligned} \tag{3.2.22}$$

For the rest of the proof, we estimate the following two sums

$$\sum_{\alpha\in\mathcal{I}_k} \mathbb{E}\left|\frac{(d_1-1)^k(d_2-1)^k}{[n]_k[m]_k} - \frac{B_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right|, \tag{3.2.23}$$

$$\sum_{\alpha\in\mathcal{I}_k} \mathbb{E}\left|I_\alpha - \frac{F_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right| \tag{3.2.24}$$

from (3.2.22) in different ways.

(1) *The upper bound on* (3.2.23). From Lemma 3.2.8, for all $\alpha \in \mathcal{I}_k$,

$$\begin{aligned}
\mathbb{E}\left|\frac{(d_1-1)^k(d_2-1)^k}{[n]_k[m]_k} - \frac{B_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right| &= \frac{(d_1-1)^k(d_2-1)^k}{[n]_k[m]_k} - \frac{\mathbb{E}B_\alpha}{[n]_k[m]_k(d_1d_2)^k} \\
&\le \frac{c_2k(d_1-1)^{r+k}(d_2-1)^{r+k}}{nd_1[n]_k[m]_k},
\end{aligned}$$

where the first line is from (3.2.12) and the second line is from (3.2.13). Therefore (3.2.23) satisfies

$$\sum_{\alpha\in\mathcal{I}_k} \mathbb{E}\left|\frac{(d_1-1)^k(d_2-1)^k}{[n]_k[m]_k} - \frac{B_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right| \le \frac{c_2[(d_1-1)(d_2-1)]^{r+k}}{2nd_1}. \tag{3.2.25}$$

(2) *The upper bound on* (3.2.24). To bound the summation in (3.2.24), for a given short

cycle α, we consider a partition of $\mathcal{G}$ in the following way:

$$A_1^\alpha = \{G \text{ does not contain } \alpha\},$$

$$A_2^\alpha = \{G \text{ contains } \alpha, \text{ which does not share an edge with another short cycle in } G\},$$

$$A_3^\alpha = \{G \text{ contains } \alpha, \text{ which shares an edge with another short cycle in } G\}.$$

Conditioned on A_1^α, we have $I_\alpha = F_\alpha = 0$. Conditioned on A_2^α, both the upper and lower bounds in Lemma 3.2.7 can apply, which yield the following inequality:

$$\left| I_\alpha - \frac{F_\alpha}{[n]_k[m]_k(d_1d_2)^k} \right| \le \frac{4k\sum_{j=2}^r jC_j + c_1k(d_1-1)^r(d_2-1)^r}{nd_1}. \tag{3.2.26}$$

Conditioned on A_3^α, we have $I_\alpha = 1, F_\alpha = 0$.

With the partition of $\mathcal{G}$, the following inequality holds:

$$\begin{aligned} \mathbb{E}\left| I_\alpha - \frac{F_\alpha}{[n]_k[m]_k(d_1d_2)^k} \right| &= \mathbb{E}\left[\mathbf{1}_{A_2^\alpha} \left| I_\alpha - \frac{F_\alpha}{[n]_k[m]_k(d_1d_2)^k} \right| \right] + \mathbb{P}(A_3^\alpha) \\ &\le \frac{2k}{nd_1}\mathbb{E}\left[\mathbf{1}_{A_2^\alpha} \sum_{j=2}^r 2jC_j \right] + \frac{c_1k(d_1-1)^r(d_2-1)^r}{nd_1}\mathbb{P}(A_2^\alpha) + \mathbb{P}(A_3^\alpha). \end{aligned} \tag{3.2.27}$$

Let $\mathcal{J}_\alpha$ be the set of all short cycles in $K_{n,m}$ that share no edges with α. On the event A_2^α, the graph G contains no short cycles outside $\mathcal{J}_\alpha$ except for α. Define $|\beta|$ be the length of the cycle β. Then

$$\sum_{j=2}^r 2jC_j = 2k + \sum_{\beta \in \mathcal{J}_\alpha} |\beta| I_\beta.$$

Therefore the right hand side of (3.2.27) can be bounded by

$$\begin{aligned}
&\frac{4k^2}{nd_1}\mathbb{P}(A_2^\alpha)+\frac{2k}{nd_1}\sum_{\beta\in\mathcal{I}_\alpha}|\beta|\mathbb{E}I_\alpha I_\beta+\frac{c_1k(d_1-1)^r(d_2-1)^r}{nd_1}\mathbb{P}(A_2^\alpha)+\mathbb{P}(A_3^\alpha)\\
\le&\frac{4k^2}{nd_1}\mathbb{P}(\alpha\subseteq G)+\frac{c_1k(d_1-1)^r(d_2-1)^r}{nd_1}\mathbb{P}(\alpha\subseteq G)+\frac{2k}{nd_1}\sum_{\beta\in\mathcal{I}_\alpha}|\beta|\mathbb{E}I_\alpha I_\beta+\mathbb{P}(A_3^\alpha). \qquad (3.2.28)
\end{aligned}$$

By Lemma 3.2.3(1),

$$\begin{aligned}
\frac{4k^2}{nd_1}\mathbb{P}(\alpha\subseteq G)&=O\left(\frac{k^2[(d_1-1)(d_2-1)]^k}{nd_1(nm)^k}\right),\\
\frac{c_1k(d_1-1)^r(d_2-1)^r}{nd_1}\mathbb{P}(\alpha\subseteq G)&=O\left(\frac{k[(d_1-1)(d_2-1)]^{k+r}}{nd_1(nm)^k}\right).
\end{aligned}$$

Hence the first and the second term in (3.2.28) combine to yield a corresponding upper bound in (3.2.24) of

$$\sum_{\alpha\in\mathcal{I}_k}\left(\frac{4k^2}{nd_1}\mathbb{P}(\alpha\subseteq G)+\frac{c_1k(d_1-1)^r(d_2-1)^r}{nd_1}\mathbb{P}(\alpha\subseteq G)\right)=O\left(\frac{[(d_1-1)(d_2-1)]^{k+r}}{nd_1}\right). \qquad (3.2.29)$$

From Lemma 3.2.3 (3), we have for any $\beta\in\mathcal{I}_\alpha$ with $|\beta|=2j$,

$$\mathbb{E}I_\alpha I_\beta=\mathbb{P}(\alpha\cup\beta\in G)\le\frac{c_1[(d_1-1)(d_2-1)]^{j+k}}{(nm)^{j+k}}.$$

For $2\le j\le r$, there are at most $[n]_j[m]_j/(2j)$ cycles in $\mathcal{I}_\alpha$ of length $2j$. The third term in (3.2.28) then satisfies

$$\begin{aligned}
\frac{2k}{nd_1}\sum_{\beta\in\mathcal{I}_\alpha}|\beta|\mathbb{E}I_\alpha I_\beta&\le\frac{2k}{nd_1}\sum_{j=2}^{r}\frac{[n]_j[m]_j}{2j}\cdot 2j\cdot\frac{c_1[(d_1-1)(d_2-1)]^{j+k}}{(nm)^{j+k}}\\
&=O\left(\frac{k[(d_1-1)(d_2-1)]^{r+k}}{nd_1(nm)^k}\right).
\end{aligned}$$

Summing over all possible $\alpha \in \mathcal{J}_k$, we obtain a corresponding term in (3.2.24) of

$$\sum_{\alpha \in \mathcal{J}_k} \frac{2k}{nd_1} \sum_{\beta \in \mathcal{J}_\alpha} |\beta| \mathbb{E} I_\alpha I_\beta = O\left(\frac{[(d_1-1)(d_2-1)]^{r+k}}{nd_1} \right). \tag{3.2.30}$$

Now given (3.2.29) and (3.2.30), to control (3.2.24), it remains to estimate $\sum_{\alpha \in \mathcal{J}_k} \mathbb{P}(A_3^\alpha)$. Let $\mathcal{K}_\alpha$ be the set of all short cycles in $K_{n,m}$ that share an edge with α, not including α itself. By a union bound,

$$\sum_{\alpha \in \mathcal{J}_k} \mathbb{P}(A_3^\alpha) \leq \sum_{\alpha \in \mathcal{J}_k} \sum_{\beta \in \mathcal{K}_\alpha} \mathbb{P}(\alpha \cup \beta \subset G). \tag{3.2.31}$$

From (3.2.3) in Lemma 3.2.3, the upper bound for $\mathbb{P}(\alpha \cup \beta \subset G)$ depends on the lengths of α, β, and the number of edges that α, β share. To get an upper bound on (3.2.31), we will classify and count the number of pairs (α, β) based on the structure of $\alpha \cup \beta$.

Recall α has length $2k$. Suppose β has length $2j$. Let $H = (V(\alpha) \cap V(\beta), E(\alpha) \cap E(\beta))$ be the intersection of α and β. Suppose H has p components and f edges. Since H is the intersection of two different cycles, H must be a forest with $p+f$ vertices. So $\alpha \cup \beta$ has $2j+2k-p-f$ vertices and $2j+2k-f$ edges. Let a, b be the number of vertices in $\alpha \cup \beta$ that are from V_1 and V_2, respectively. Then

$$a + b = 2j + 2k - p - f. \tag{3.2.32}$$

Let v_1, v_2 be the number of vertices in V_1 and V_2 for H, respectively. Then we have $a = j + k - v_1, b = j + k - v_2$, and $|a - b| = |v_1 - v_2|$. Note that each component in H is a path. For each path, the difference between the number of vertices from V_1 and V_2 is at most 1. This implies

$$|a - b| = |v_1 - v_2| \leq p. \tag{3.2.33}$$

From the proof of Corollary 21 in [75], the number of all possible isomorphism types of $\alpha \cup \beta$ given $|\alpha|, |\beta| \leq 2r$ and $p, f \leq 2r$ is at most

$$\frac{(16r^3)^{p-1}}{((p-1)!)^2}.$$

For each isomorphism type, as a subgraph in $K_{n,m}$, the number of ways to label it is at most $[n]_a[m]_b + [n]_b[m]_a$, where the two terms come from assigning vertices in V_1, V_2 in two ways (pick an arbitrary starting vertex, decide whether it is from V_1 or V_2, then choose labels accordingly).

From (3.2.32), (3.2.33), and the assumption that $n \leq m$, we have that when f is even,

$$[n]_a[m]_b + [n]_b[m]_a \leq 2n^{j+k-p-f/2} m^{j+k-f/2} = 2n^{-p}(nm)^{j+k-f/2}.$$

And when f is odd,

$$[n]_a[m]_b + [n]_b[m]_a \leq 2n^{-p+1}(nm)^{j+k-f/2-1/2}.$$

By (3.2.3) in Lemma 3.2.3, the probability of any realization of an the isomorphism type as a subgraph in G is bounded by

$$\frac{c_1[(d_1-1)(d_2-1)]^{j+k-f/2}}{(nm)^{j+k-f/2}}.$$

With all the estimates above, the right hand side of (3.2.31) is now bounded by

$$
\begin{aligned}
&\sum_{j=2}^{r}\sum_{1\le p,f\le 2r}\frac{(16r^3)^{p-1}}{((p-1)!)^2}\cdot([n]_a[m]_b+[n]_b[m]_a)\cdot\frac{c_1[(d_1-1)(d_2-1)]^{j+k-f/2}}{(nm)^{j+k-f/2}}\\
\le&\sum_{j=2}^{r}\sum_{1\le p,f\le 2r}\frac{(16r^3)^{p-1}}{((p-1)!)^2}\cdot(2n^{-p}(nm)^{j+k-f/2})\cdot\frac{c_1[(d_1-1)(d_2-1)]^{j+k-f/2}}{(nm)^{j+k-f/2}}\mathbf{1}\{f\text{ is even}\}\\
&+\sum_{j=2}^{r}\sum_{1\le p,f\le 2r}\frac{(16r^3)^{p-1}}{((p-1)!)^2}\cdot(2n^{-p+1}(nm)^{j+k-f/2-1/2})\\
&\cdot\frac{c_1[(d_1-1)(d_2-1)]^{j+k-f/2}}{(nm)^{j+k-f/2}}\mathbf{1}\{f\text{ is odd}\}\\
=&O\left(\frac{[(d_1-1)(d_2-1)]^{k+r-1}}{n}\right)+O\left(\frac{[(d_1-1)(d_2-1)]^{r+k-1/2}}{(nm)^{1/2}}\right)\\
=&O\left(\frac{[(d_1-1)(d_2-1)]^{k+r}}{nd_1}\right).
\end{aligned}
\tag{3.2.34}
$$

Combining all estimates from (3.2.29), (3.2.30) and (3.2.34), we finally obtain

$$
\sum_{\alpha\in\mathcal{I}_k}\mathbb{E}\left|I_\alpha-\frac{F_\alpha}{[n]_k[m]_k(d_1d_2)^k}\right|=O\left(\frac{[(d_1-1)(d_2-1)]^{r+k}}{nd_1}\right). \tag{3.2.35}
$$

This provides an upper bound for (3.2.24).

(3) *The upper bound on* (3.2.22). Now the upper bounds on (3.2.23) and (3.2.24) have been provided in (3.2.25) and (3.2.35), respectively. We are ready to estimate (3.2.22). Recall

$$
\xi_k=\min\{1,1.4\mu_k^{-1/2}\}=\frac{2.8\sqrt{k}}{[(d_1-1)(d_2-1)]^{k/2}}.
$$

Then from (3.2.25) and (3.2.35), there is an absolute constant $c_7>0$ such that (3.2.22) is bounded by

$$
\sum_{k=2}^{r}\frac{c_7\sqrt{k}[(d_1-1)(d_2-1)]^{r+k/2}}{nd_1}=O\left(\frac{\sqrt{r}[(d_1-1)(d_2-1)]^{3r/2}}{nd_1}\right). \tag{3.2.36}
$$

This completes the proof. $\square$

3.2.3 Cyclically non-backtracking walks and the Chebyshev polynomials

In this section, we study non-backtracking walks in biregular bipartite graphs and relate them to the Chebyshev polynomials. The relation will be used in Section 3.4 to study eigenvalue fluctuations for random biregular bipartite graphs.

Definition 3.2.11 (non-backtracking walk)**.** We define a *non-backtracking walk* of length $2k$ in a biregular bipartite graph to be a walk $(u_1, v_1, \ldots, u_k, v_k, u_{k+1})$ such that $u_i \in V_1, v_i \in V_2$, $u_{i+1} \neq u_i$, for all $1 \leq i \leq k$ and $v_{i+1} \neq v_i$ for all $1 \leq i \leq k-1$. Note that in our definition, all such walks start and end at some vertices from V_1.

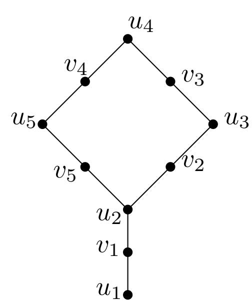

Figure 3.2: In this example, $(u_2, v_2, u_3, v_3, u_4, v_4, u_5, v_5, u_2)$ is a cyclically non-backtracking walk. $(u_1, v_1, u_2, v_2, u_3, v_3, u_4, v_4, u_5, v_5, u_2, v_1, u_1)$ is a closed non-backtracking walk, but it is not cyclically non-backtracking.

Definition 3.2.12 (cyclically non-backtracking walk)**.** A walk of length $2k$ denoted by

$$(u_1, v_1, \ldots, u_k, v_k, u_{k+1})$$

is closed if $u_{k+1} = u_1$. *A cyclically non-backtracking walk* is a closed non-backtracking walk such that its last two steps are not the reverse of its first two steps. Namely, $(u_1, v_1, u_2) \neq (u_{k+1}, v_k, u_k)$. Figure 3.2 gives an example of a closed non-backtracking walk that is not cyclic non-backtracking.

Let G_n be a random (d_1,d_2)-biregular bipartite graph and $C_k^{(n)}$ be the number of cycles of length $2k$ in G_n. Denote $\mathrm{NBW}_k^{(n)}$ to be the number of non-backtracking walk of length $2k$, and $\mathrm{CNBW}_k^{(n)}$ to be the number of cyclically non-backtracking walks of length $2k$ in G_n. Let $(C_k^{(\infty)}, k \geq 2)$ be independent Poisson random variables with mean

$$\mu_k = \frac{[(d_1-1)(d_2-1)]^k}{2k}.$$

We also define $C_1^{(\infty)} = C_1^{(n)} = 0$. For $k \geq 1$, denote

$$\mathrm{CNBW}_k^{(\infty)} = \sum_{j|k} 2jC_j^{(\infty)}. \tag{3.2.37}$$

For any cycle of length $2j$ in G_n with $j \mid k$, we can obtain $2j$ cyclically non-backtracking walks by choosing a starting point from V_1, fixing a direction and then walking around the cycle of length $2k$ repeatedly. The next lemma shows that $\mathrm{CNBW}_k^{(n)}$ can be approximated by the count of those repeated walks around cycles.

Lemma 3.2.13. *Let G_n be a random (d_1,d_2)-biregular bipartite graph. Suppose $d_1 \leq n^{1/3}, k \leq n^{1/10}$, define*

$$B_k^{(n)} = \mathrm{CNBW}_k^{(n)} - \sum_{j|k} 2jC_j^{(n)} \tag{3.2.38}$$

to be the number of cyclically non-backtracking walks of length $2k$ in G_n that are not repeated walks around cycles. Then

$$\mathbb{E}B_k^{(n)} \leq \frac{c_7 k^7 [(d_1-1)(d_2-1)]^k}{n}.$$

We call a cyclically non-backtracking walk *bad* if it's not a repeated walk on a cycle. Then from (3.2.38), $B_k^{(n)}$ counts the number of bad cyclically non-backtracking walks of length

$2k$.

Let $(w_0, w_1, \ldots, w_{2k})$ with $w_{2k} = w_0 \in V_1$ be a bad cyclically non-backtracking walk in $K_{n,m}$ of length $2k$. For any $1 \leq i \leq 2k$, we say that the i-th step of the walk is

- *free* if w_i did not previously occur in the walk;
- *a coincidence* if w_i previously occurred in the walk, but the edge $w_{i-1}w_i$ didn't;
- *forced* if the edge $w_{i-1}w_i$ previously occurred in the walk.

Let $\chi+1$ be the number of coincidences and f be the number of forced steps in the walk. Let χ_1+1 and χ_2 be the number of coincidence steps ending at a vertex from V_1 and V_2, respectively. Let f_1, f_2 be the number of forced steps ending at a vertex from V_1 and V_2, respectively. Denote v, e the number of distinct vertices and edges in the cyclically non-backtracking walk, respectively. We now have the following relations:

$$
\begin{aligned}
\chi+1 &= \chi_1+\chi_2+1, \\
f &= f_1+f_2, \\
v &= (2k+1)-(\chi+1)-f = 2k-\chi-f, \\
e &= 2k-f.
\end{aligned}
$$

For any repeated walk on a cycle, the the number of coincidences is 1 and $\chi = 0$. Therefore if the walk is bad, we must have $\chi \geq 1$.

The following lemma bounds the number of cyclically non-backtracking walks with given parameters χ_1, χ_2, f_1, and f_2.

Lemma 3.2.14. *Consider cyclically non-backtracking walks of length $2k$ on $K_{n,m}$ such that in the subgraph spanned by this walk, all vertices from V_1 have degrees at most d_1 and vertices from V_2 have degrees at most d_2. Then the number of such walks with given χ_1, χ_2, f_1, f_2 satisfying $\chi \geq 1$*

is at most

$$(2k)^{3(\chi_1+\chi_2)+2}(d_1-1)^{f_2}(d_2-1)^{f_1}n^{k-\chi_1-f_1}m^{k-\chi_2-f_2}.$$

Moreover, we must have $|f_1-f_2|\leq\chi+1$.

Proof. We count the number of such cyclically non-backtracking walks by choosing the coincidences, forced steps, and free steps separately. Given that there are $\chi+1$ coincidences, there are $\binom{2k}{\chi+1}$ many possible subsets of indices in $\{1,\dots,2k\}$ where coincidences can happen. The vertices at a coincidence has already occurred in the walk, so there are at most $2k$ choices for each of them, giving us a total of $\binom{2k}{\chi+1}(2k)^{\chi+1}\leq(2k)^{2\chi+2}$ many choices.

For forced steps, they can only occur after a coincidence or another forced step. After each coincidence, imagine assigning some number of steps to be forced. The number of ways to do this is at most the number of weak compositions of f elements into $\chi+1$ parts, which is $\binom{f+\chi}{\chi}\leq(2k)^{\chi}$. For each forced step ending at a vertex from V_1, the walk can only move along an edge that has already been traversed, so there are at most (d_2-1) possible choices at every step due to the non-backtracking property. Similarly, for each forced step ending at a vertex from V_2 there are at most d_1-1 possible choices. Altogether this gives us at most $(2k)^{\chi}(d_1-1)^{f_2}(d_2-1)^{f_1}$ choices for all forced steps.

There are $k-\chi-1-f_1$ many free steps ending at a vertex from V_1, we have at most n choices for the next vertex, and we have an additional n choices for $w_0\in V_1$, which gives a total of at most $n^{k-\chi_1-f_1}$ many choices. Similarly, the number of free steps ending at a vertex from V_2 is at most $m^{k-\chi_2-f_2}$. Multiplying together every parts from coincidences, forced steps and free steps gives us at most

$$(2k)^{3\chi+2}(d_1-1)^{f_2}(d_2-1)^{f_1}n^{k-\chi_1-f_1}m^{k-\chi_2-f_2}$$

many such cyclically non-backtracking walks.

Next we bound $|f_1-f_2|$. Recall forced steps can only occur after a coincidence or another forced step. Then there are at most $\chi+1$ many consecutive forced steps starting from a certain coincidence step. In each consecutive forced steps, the number of vertices from V_1 and V_2 differ by at most 1, since the subgraph spanned by any consecutive forced steps is a path. Hence we have $|f_1-f_2|\le\chi+1$. $\square$

Equipped with Lemma 3.2.14, we continue to prove Lemma 3.2.13.

Proof of Lemma 3.2.13. By Part (a) in Lemma 3.2.3, the probability that a given bad walk appears in G_n is at most

$$c_1\left(\frac{(d_1-1)(d_2-1)}{nm}\right)^{k-f/2}.$$

From the upper bound on the number of such walks in Lemma 3.2.14, summing over all possibilities of χ_1,χ_2,f_1,f_2, we have

$$\begin{aligned}
&\mathbb{E}B_k^{(n)}\le \\
&\sum_{\substack{\chi_1;\chi_2:\\ \chi_1+\chi_2\ge 1}}\sum_{\substack{0\le f_1,f_2\le k-1\\ |f_1-f_2|\le\chi+1}}(2k)^{3\chi+2}(d_1-1)^{f_2}(d_2-1)^{f_1}n^{k-\chi_1-f_1}m^{k-\chi_2-f_2}\cdot c_1\left(\frac{(d_1-1)(d_2-1)}{nm}\right)^{k-\frac{f}{2}}\\
=&c_1[(d_1-1)(d_2-1)]^k\sum_{\chi_1+\chi_2\ge 1}n^{-\chi_1}m^{-\chi_2}(2k)^{3(\chi_1+\chi_2)+2}\sum_{\substack{0\le f_1,f_2\le k-1\\ |f_1-f_2|\le\chi+1}}\left(\frac{(d_2-1)m}{(d_1-1)n}\right)^{(f_1-f_2)/2}.
\end{aligned}$$

Since $(d_2-1)d_1\le(d_1-1)d_2$, the following inequality holds:

$$\begin{aligned}
\sum_{\substack{0\le f_1,f_2\le k-1\\ |f_1-f_2|\le\chi+1}}\left(\frac{(d_2-1)m}{(d_1-1)n}\right)^{(f_1-f_2)/2}=&\sum_{\substack{0\le f_1,f_2\le k-1\\ |f_1-f_2|\le\chi+1}}\left(\frac{(d_2-1)d_1}{(d_1-1)d_2}\right)^{(f_1-f_2)/2}\\
\le&k^2\left(\frac{(d_1-1)d_2}{(d_2-1)d_1}\right)^{(\chi+1)/2}.
\end{aligned}\tag{3.2.39}$$

Since $d_1 \le n^{1/3}, k \le n^{1/10}$, (3.2.39) implies

$$\begin{aligned}
\mathbb{E}B_k^{(n)} &\le c_1 k^2 [(d_1-1)(d_2-1)]^k \sum_{\chi_1+\chi_2\ge 1} n^{-\chi_1} m^{-\chi_2} (2k)^{3(\chi_1+\chi_2)+2} \left(\frac{(d_1-1)d_2}{(d_2-1)d_1}\right)^{(\chi+1)/2} \\
&= k^2[(d_1-1)(d_2-1)]^k O\left(\frac{(2k)^5(d_1-1)d_2}{n(d_2-1)d_1}\right) \\
&= O\left(\frac{k^7(d_1-1)^k(d_2-1)^k}{n}\right).
\end{aligned}$$

This completes the proof of Lemma 3.2.13. □

Recall the definition of $\mathrm{CNBW}_k^{(\infty)}$ from (3.2.37). The following corollary holds.

Corollary 3.2.15. Suppose $d_1 \le n^{1/3}$ and $r \le n^{1/10}$. There exists a constant $c_8 > 0$ such that

$$d_{\mathrm{TV}}\Big((\mathrm{CNBW}_k^{(n)}, 2 \le k \le r), (\mathrm{CNBW}_k^{(\infty)}, 2 \le k \le r)\Big) \le \frac{c_8\sqrt{r}[(d_1-1)(d_2-1)]^{3r/2}}{nd_1}. \quad (3.2.40)$$

Proof. By the definition of total variation distance, for any measurable map f and random variable X, Y, we have

$$d_{\mathrm{TV}}(f(X), f(Y)) \le d_{\mathrm{TV}}(X,Y). \quad (3.2.41)$$

It follows from Theorem 3.2.10 that

$$d_{\mathrm{TV}}\left(\left(\sum_{j|k} 2jC_j^{(n)}, 2 \le k \le r\right), \Big(\mathrm{CNBW}_k^{(\infty)}, 2 \le k \le r\Big)\right) \le \frac{c_6\sqrt{r}[(d_1-1)(d_2-1)]^{3r/2}}{nd_1}. \quad (3.2.42)$$

By Markov's inequality and Lemma 3.2.13,

$$\mathbb{P}(B_k^{(n)} \ge 1) \le \frac{c_7 k^7[(d_1-1)(d_2-1)]^k}{n}. \quad (3.2.43)$$

Summing these probabilities for $k = 2, \dots, r$ implies

$$\left(\sum_{j|k} 2jC_j^{(n)}, 2 \le k \le r \right) = (\mathrm{CNBW}_k^{(n)}, 2 \le k \le r) \tag{3.2.44}$$

with probability $1 - O\left(\frac{r^7[(d_1-1)(d_2-1)]^r}{n}\right)$. Therefore by the coupling inequality,

$$d_{\mathrm{TV}}\left(\left(\sum_{j|k} 2jC_j^{(n)}, 2 \le k \le r \right), (\mathrm{CNBW}_k^{(n)}, 2 \le k \le r) \right) = O\left(\frac{r^7[(d_1-1)(d_2-1)]^r}{n} \right). \tag{3.2.45}$$

From (3.2.42) and (3.2.45),

$$\begin{aligned} & d_{\mathrm{TV}}\left((\mathrm{CNBW}_k^{(n)}, 2 \le k \le r), (\mathrm{CNBW}_k^{(\infty)}, 2 \le k \le r) \right) \\ \le & \frac{c_6\sqrt{r}[(d_1-1)(d_2-1)]^{3r/2}}{nd_1} + O\left(\frac{r^7[(d_1-1)(d_2-1)]^r}{n} \right) = O\left(\frac{\sqrt{r}[(d_1-1)(d_2-1)]^{3r/2}}{nd_1} \right). \end{aligned}$$

$\square$

Let $\lambda_1 \ge \cdots \ge \lambda_n$ be the eigenvalues of $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$. For the rest of this section, we connect the spectrum of $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$ with Chebyshev polynomials and cyclically non-backtracking walks. Define

$$\Gamma_0(x) = 1, \quad \Gamma_{2k}(x) = 2T_{2k}\left(\frac{x}{2}\right) + \frac{d_1-2}{(d_1-1)^k}, \tag{3.2.46}$$

$$\Gamma_{2k+1}(x) = 2T_{2k+1}\left(\frac{x}{2}\right). \tag{3.2.47}$$

Here $\{T_k(x)\}$ are the Chebyshev polynomials of the first kind on $[-1,1]$ which satisfy

$$\begin{aligned} T_0(x) = 1, \quad T_1(x) = x, \\ T_{k+1}(x) = 2xT_k(x) - T_{k-1}(x). \end{aligned} \tag{3.2.48}$$

Let $\{U_k(x)\}$ be the Chebyshev polynomials of the second kind on $[-1,1]$ such that

$$\begin{aligned} U_{-1}(x) = 0, \quad U_0(x) = 1, \\ U_{k+1}(x) = 2xU_k(x) - U_{k-1}(x). \end{aligned}$$

Define

$$p_k(x) = U_k\left(\frac{x}{2}\right) - \frac{1}{d_1 - 1} U_{k-2}\left(\frac{x}{2}\right). \tag{3.2.49}$$

We begin with representing closed non-backtracking walks with $p_k(x)$. The following lemma gives a deterministic identity. Recall in our Definition 3.2.11, all closed non-backtracking walks start and end at vertices in V_1.

Lemma 3.2.16. *Let* $\mathrm{NBW}_k^{(n)}$ *be the number of closed non-backtracking walks of length* $2k$ *in a* (d_1, d_2)*-biregular bipartite graph* G*. Let* $\lambda_1 \geq \cdots \geq \lambda_n$ *be the eigenvalues of* $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$*. We have*

$$\sum_{i=1}^{n} p_k(\lambda_i) = (d_1 - 1)^{-k/2}(d_2 - 1)^{-k/2}\mathrm{NBW}_k^{(n)}. \tag{3.2.50}$$

Proof. Let $A^{(k)}$ be the $n \times n$ matrix such that $A_{ij}^{(k)}$ is the number of non-backtracking walks of length $2k$ from i to j, where $i, j \in V_1$. We have the following relations:

$$A^{(1)} = XX^\top - d_1 I, \quad A^{(2)} = (A^{(1)})^2 - d_1(d_2-1)I,$$
$$A^{(k+1)} = A^{(1)}A^{(k)} - (d_1-1)(d_2-1)A^{(k-1)}, \quad \forall k \geq 2. \tag{3.2.51}$$

The expressions of $A^{(1)}$ and $A^{(2)}$ follow from the definition of non-backtracking walks. Since a non-backtracking walk of length $2k+2$ can be decomposed as a non-backtracking walk of length $2k$ and a non-backtracking walk of length 2 which avoid backtracking at the $2k$-th step, the expression (3.2.51) holds. We now claim that for $k \geq 1$,

$$p_k\left(\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}\right) = [(d_1-1)(d_2-1)]^{-k/2}A^{(k)}, \tag{3.2.52}$$

and prove it by induction. Note that from (3.2.49),

$$p_1(x) = x, \quad p_2(x) = x^2 - 1 - \frac{1}{d_1-1}.$$

It is easy to check (3.2.52) holds for $k = 1, 2$. Since $p_k(x)$ is a linear combination of $U_k(x/2)$ and $U_{k-2}(x/2)$, it satisfies the recursive relation for $U_k(x/2)$, which is

$$p_k(x) = xp_k(x) - p_{k-1}(x).$$

Assume (3.2.52) holds for $k \leq s$. Let $M = XX^\top - d_1 I$. Then

$$\begin{aligned}
&p_{s+1}\left(\frac{M}{\sqrt{(d_1-1)(d_2-1)}}\right) \\
=&M[(d_1-1)(d_2-1)]^{-(s+1)/2}A^{(s)} - [(d_1-1)(d_2-1)]^{-(s-1)/2}A^{(s-1)} \\
=&[(d_1-1)(d_2-1)]^{-(s+1)/2}\left(MA^{(s)} - (d_1-1)(d_2-1)A^{(s-1)}\right) \\
=&[(d_1-1)(d_2-1)]^{-(s+1)/2}A^{(s+1)},
\end{aligned}$$

where the last equality is from (3.2.51). Therefore (3.2.52) holds. Taking trace on both sides in (3.2.52), we obtain (3.2.50). □

The next theorem is an algebraic relation between Γ_k and the number of cyclic non-backtracking walks. Together with Lemma 3.2.15, it implies the polynomials $\Gamma_k(x)$ of the eigenvalues for RBBGs converges in distribution to a sum of Poisson random variables.

Theorem 3.2.17. *Let G be a (d_1, d_2)-biregular bipartite graph and $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$. Then for any $k \geq 1$, we have*

$$\sum_{i=1}^{n} \Gamma_k(\lambda_i) = (d_1-1)^{-k/2}(d_2-1)^{-k/2}\mathrm{CNBW}_k^{(n)}. \tag{3.2.53}$$

Proof. We first relate the number of cyclically non-backtracking closed walks $\mathrm{CNBW}_k^{(n)}$ to the number of closed non-backtracking walks $\mathrm{NBW}_k^{(n)}$.

A closed non-backtracking walk of length $2k$ is either cyclically non-backtracking or it can be obtained from a closed non-backtracking walk of length $2(k-2)$ by adding a new walk of length 2 (which we call a *tail*) to the beginning of the walk and its reverse to the end (see Figure 3.2 for an example). For any cyclically non-backtracking walk of length $2(k-2)$, we can add a tail in $(d_1-2)(d_2-1)$ many ways. For any closed non-backtracking walk of length $2(k-2)$ that is not cyclically non-backtracking, we can add a tail in $(d_1-1)(d_2-1)$ many ways. Therefore for $k \geq 3$, we have the following equation

$$\begin{aligned}\mathrm{NBW}_k^{(n)} &= \mathrm{CNBW}_k^{(n)} + (d_1-2)(d_2-1)\mathrm{CNBW}_{k-2}^{(n)} + (d_1-1)(d_2-1)(\mathrm{NBW}_{k-2}^{(n)} - \mathrm{CNBW}_{k-2}^{(n)}) \\ &= \mathrm{CNBW}_k^{(n)} + (d_1-1)(d_2-1)\mathrm{NBW}_{k-2}^{(n)} - (d_2-1)\mathrm{CNBW}_{k-2}^{(n)},\end{aligned}$$

which can be written as

$$\mathrm{CNBW}_k^{(n)} - (d_2-1)\mathrm{CNBW}_{k-2}^{(n)} = \mathrm{NBW}_k^{(n)} - (d_1-1)(d_2-1)\mathrm{NBW}_{k-2}^{(n)}. \tag{3.2.54}$$

Note that $\mathrm{CNBW}_k^{(n)} = \mathrm{NBW}_k^{(n)}$ for $k = 1, 2$. Applying (3.2.54) recursively, we have when k is even,

$$\begin{aligned}&\mathrm{CNBW}_k^{(n)} \\ =&\mathrm{NBW}_k^{(n)} - (d_1-2)[(d_2-1)\mathrm{NBW}_{k-2}^{(n)} + (d_2-1)^2\mathrm{NBW}_{k-4}^{(n)} + \cdots + (d_2-1)^{\frac{k-2}{2}}\mathrm{NBW}_2^{(n)})].\end{aligned} \tag{3.2.55}$$

And when k is odd,

$$\begin{aligned}&\mathrm{CNBW}_k^{(n)} \\ =&\mathrm{NBW}_k^{(n)} - (d_1-2)[(d_2-1)\mathrm{NBW}_{k-2}^{(n)} + (d_2-1)^2\mathrm{NBW}_{k-4}^{(n)} + \cdots + (d_2-1)^{\frac{k-3}{2}}\mathrm{NBW}_3^{(n)}].\end{aligned} \tag{3.2.56}$$

Denote

$$\overline{\mathrm{NBW}}_k^{(n)} :=(d_2-1)^{-k/2}\mathrm{NBW}_k^{(n)}, \quad \overline{\mathrm{CNBW}}_k^{(n)} := (d_2-1)^{-k/2}\mathrm{CNBW}_k^{(n)}.$$

We can simplify the above equations (3.2.55) and (3.2.56) as

$$\overline{\mathrm{CNBW}}_k^{(n)} = \overline{\mathrm{NBW}}_k^{(n)} - (d_1-2)\left(\overline{\mathrm{NBW}}_{k-2}^{(n)} + \overline{\mathrm{NBW}}_{k-4}^{(n)} + \cdots + \overline{\mathrm{NBW}}_a^{(n)}\right), \tag{3.2.57}$$

where $a = 2$ if k is even and $a = 1$ if k is odd. Also (3.2.50) can be written as

$$\sum_{i=1}^{n} p_k(\lambda_i) = (d_1 - 1)^{-k/2}\overline{\mathrm{NBW}}_k^{(n)}. \tag{3.2.58}$$

From the proof of Proposition 32 in [76], we have the following relation between $\Gamma_k(x)$ and $p_k(x)$ for $k \geq 1$:

$$\Gamma_{2k}(x) = p_{2k}(x) - (d_1 - 2)\left(\frac{p_{2k-2}(x)}{d_1 - 1} + \frac{p_{2k-4}(x)}{(d_1 - 1)^2} + \cdots + \frac{p_2(x)}{(d_1 - 1)^{k-1}}\right), \tag{3.2.59}$$

$$\Gamma_{2k-1}(x) = p_{2k-1}(x) - (d_1 - 2)\left(\frac{p_{2k-3}(x)}{d_1 - 1} + \frac{p_{2k-5}(x)}{(d_1 - 1)^2} + \cdots + \frac{p_1(x)}{(d_1 - 1)^{k-1}}\right). \tag{3.2.60}$$

Then from (3.2.58) and (3.2.57),

$$\begin{aligned}
&(d_1 - 1)^{-k}\overline{\mathrm{CNBW}}_{2k}^{(n)} \\
=&\sum_{i=1}^{n}\left(p_{2k}(\lambda_i) - (d_1 - 2)\left(\frac{p_{2k-2}(\lambda_i)}{d_1 - 1} + \cdots + \frac{p_2(\lambda_i)}{(d_1 - 1)^{k-1}}\right)\right) = \sum_{i=1}^{n}\Gamma_{2k}(\lambda_i),
\end{aligned}$$

where the last equality is from (3.2.59). Similarly, from (3.2.60),

$$(d_1 - 1)^{(2k-1)/2}\overline{\mathrm{CNBW}}_{2k-1}^{(n)} = \sum_{i=1}^{n}\Gamma_{2k-1}(\lambda_i).$$

Therefore for all $k \geq 1$,

$$\sum_{i=1}^{n}\Gamma_k(\lambda_i) = (d_1 - 1)^{-k/2}\overline{\mathrm{CNBW}}_k^{(n)} = [(d_1 - 1)(d_2 - 1)]^{-k/2}\mathrm{CNBW}_k^{(n)}.$$

This completes the proof of Theorem 3.2.17. $\square$

3.3 Spectral gap

In this section, we provide some estimates on the second largest eigenvalue of the random biregular bipartite graphs that will be used to study eigenvalue fluctuations in Section 3.4. Note that the largest eigenvalue of $XX^\top - d_1 I$ is $\lambda_1 = d_1(d_2-1)$. In the next theorem, we provide upper bounds on $|\lambda|$ for all eigenvalues $\lambda \neq \lambda_1$.

Theorem 3.3.1. *Let G be a (d_1, d_2)-random biregular bipartite graph with $d_1 \geq d_2$. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $XX^\top - d_1 I$.*

1. *For fixed d_1, d_2, there exists a sequence $\varepsilon_n \to 0$ such that for any eigenvalue $\lambda \neq \lambda_1$,*

$$\mathbb{P}(|\lambda - (d_2-2)| \geq 2\sqrt{(d_1-1)(d_2-1)} + \varepsilon_n) \to 0 \tag{3.3.1}$$

   as $n \to \infty$.

2. *Suppose $d_2 \leq \frac{1}{2} n^{2/3}$, $d_1 \geq d_2 \geq c d_1$ for some constant $c \in (0,1)$. Then for some constant $\alpha_1 > 0$ depending on c and any eigenvalue $\lambda \neq \lambda_1$,*

$$\mathbb{P}\left(|\lambda| \geq \alpha_1 \sqrt{(d_1-1)(d_2-1)}\right) \leq \frac{1}{n^2}. \tag{3.3.2}$$

3. *Suppose $d_2 \leq C_1$, $d_1 \leq n^2$, there exists a constant α_2 depending on C_1 such that for any eigenvalue $\lambda \neq \lambda_1$,*

$$\mathbb{P}\left(|\lambda| \geq \alpha_2 \sqrt{(d_1-1)(d_2-1)}\right) \leq \frac{1}{n^2}. \tag{3.3.3}$$

Remark 3.3.2. The probability estimates in (3.3.2) and (3.3.3) can be improved, see [179]. In order to prove the main theorems in Section 3.4, we only include a weaker version for simplicity.

proof of Theorem 3.3.1. Theorem 4 in [48] states that for a random biregular bipartite graph with

$d_1 \geq d_2$, the eigenvalues of the adjacency matrix A satisfy the following estimates with high probability:

1. the second eigenvalue of A satisfies $\lambda_2(A) \leq \sqrt{d_1 - 1} + \sqrt{d_2 - 1} + o(1)$,
2. the smallest positive eigenvalue of A satisfies $\lambda^+_{\min}(A) \geq \sqrt{d_1 - 1} - \sqrt{d_2 - 1} - o(1)$.

Since eigenvalues of $XX^\top$ are the squares of the eigenvalues for A, we have with high probability,

$$\lambda_2(XX^\top) - d_1 - (d_2 - 2) \leq 2\sqrt{(d_1 - 1)(d_2 - 1)} + o(1),$$
$$\lambda_n(XX^\top) - d_1 - (d_2 - 2) \geq -2\sqrt{(d_1 - 1)(d_2 - 1)} - o(1),$$

therefore (3.3.1) holds.

Theorem 1.1 in [179] states that if $d_2 \leq \frac{1}{2}n^{2/3}$ and $d_1 \geq d_2$, there exists a constant $\alpha > 0$ such that $\lambda_2(A) \leq \alpha\sqrt{d_1}$ with probability at least $1 - m^{-2}$. This implies for any eigenvalue λ of $XX^\top - d_1 I$ with $\lambda \neq d_1(d_2 - 1)$, we have

$$\mathbb{P}\left(-d_1 \leq \lambda \leq \alpha^2 d_1 - d_1\right) \geq 1 - m^{-2} \geq 1 - n^{-2}.$$

Since $d_1 \geq d_2 \geq cd_1$, we can find a constant $\alpha_1 > 0$ depending on α and c such that

$$\mathbb{P}\left(|\lambda| \leq \alpha_1\sqrt{(d_1 - 1)(d_2 - 1)}\right) \geq 1 - n^{-2}.$$

Therefore (3.3.2) holds. Theorem 1.5 in [179] states that if $d_2 \leq C_1, d_1 \leq n^2$, there exists a constant α_2 depending on C_1 such that

$$\mathbb{P}\left(\max_{2 \leq i \leq m+n-1} |\lambda_i^2(A) - d_1| \geq \alpha_2\sqrt{(d_1 - 1)(d_2 - 1)}\right) \leq n^{-2}.$$

Then (3.3.3) follows from the algebraic relation between the spectra of A and $XX^\top - d_1 I$. $\square$

3.4 Eigenvalue fluctuations

Lemma 3.2.17 and Corollary 3.2.15 imply the limiting laws for $\sum_{i=1}^{n}\Gamma_k(\lambda_i)$ are given by a sum of Poisson random variables. In this section we extend the results to a more general class of function f and study the behavior of $\sum_{i=1}^{n} f(\lambda_i)$ for RBBGs with fixed and growing degrees.

The following set-up for weak convergence will be used in Section 3.4.2 to prove Theorem 3.4.7. We will closely follow the definitions and notations used in [76]. See Section 2 in [76] for more details.

Denote $\mathbb{N} := \{1,2,\dots\}$. Let $\vec{w} = (w_m)_{m\in\mathbb{N}}$ be a sequence of positive weights. Let $L^2(\vec{w})$ be the space of sequences $(x_m)_{m\in\mathbb{N}}$ that are square-integrable with respect to $\vec{w}$, i.e., $\sum_{m=1}^{\infty} x_m^2 w_m < \infty$. We define a complete separable metric space $\mathcal{X} = (L^2(\vec{w}), \|\cdot\|)$, where for any sequence $(x_m)_{m\in\mathbb{N}}$,

$$\|x\| = \left(\sum_{m=1}^{\infty} x_m^2 w_m\right)^{1/2}.$$

Denote the space of probability measures on the Borel σ-algebra of $\mathcal{X}$ by $\mathcal{P}(\mathcal{X})$. We use the Prokhorov metric for weak convergence as the metric on $\mathcal{P}(\mathcal{X})$. The following results are proved in Section 2 of [76].

Proposition 3.4.1 (Lemma 2-4 in [76])**.** The following holds for the complete separable metric space $\mathcal{X}$.

1. Let $(a_m)_{m\in\mathbb{N}} \in L^2(\vec{w})$ be such that $a_m \geq 0$ for every m. Then the set

$$\{(b_m)_{m\in\mathbb{N}} \in L^2(\vec{w}) : 0 \leq |b_m| \leq a_m, \forall m \in \mathbb{N}\}$$

   is compact in $(L^2(\vec{w}), \|\cdot\|)$.

2. Suppose $\{X_n\}$ and X are random sequences taking values in $L^2(\vec{w})$ such that X_n converges in distribution to X. Then for any $b \in L^2(\vec{w})$, the random variables $\langle b, X_n\rangle$ converges in

distribution to $\langle b, X\rangle$.

3. Let $x \in \mathcal{X}$ and P, Q be two probability measures in $\mathcal{P}(\mathcal{X})$. Suppose for any finite collection of indices $(i_1, \dots, i_k)$, the law of random vector $(x_{i_1}, \dots, x_{i_k})$ is the same under both P and Q. Then $P = Q$ on the entire Borel σ-algebra of $\mathcal{X}$.

We also need the following results from the approximation theory.

Definition 3.4.2 (Bernstein ellipse)**.** For $\rho > 1$, let E_ρ be the image of the map $z \mapsto (z + z^{-1})/2$ of the open disc of radius ρ in the complex plain centered at the origin. We can E_ρ the *Bernstein ellipse* of radius ρ. The ellipse has foci at ± 1 and the sum of the major semi-axis and minor semi-axis is exactly ρ.

Proposition 3.4.3 ([171], Theorem 8.1)**.** Suppose $f : [-1, 1] \to \mathbb{R}$ can be analytically extended to E_ρ and is bounded by M on E_ρ. Then f has a unique expansion on $[-1, 1]$ as

$$f(x) = \sum_{k=0}^{\infty} a_k T_k(x),$$

where $T_k(x)$ is the Chebyshev polynomial of the first kind defined in (3.2.48), and the coefficients of this expansion satisfy

$$|a_0| \le M, \quad |a_k| \le \frac{2M}{\rho^k}.$$

Define $f_k(x) = \sum_{i=0}^{k} a_k T_k(x)$. Applying the bound $|T_k(x)| \le 1$ when $x \in [-1, 1]$ and Proposition 3.4.3, we obtain for all $x \in [-1, 1]$,

$$|f(x) - f_k(x)| \le \frac{2M}{\rho^k(\rho - 1)}. \tag{3.4.1}$$

3.4.1 Poisson fluctuations with fixed degrees

Now fix d_1 and d_2 as constants. We are ready to extend our results in Section 3.2.3 to a more general class of functions as follows. Note that the following theorem is given for a sequence of RBBGs with growing n. For the ease of notations, we drop the dependence on n when writing the matrix X and eigenvalues $\lambda_1, \dots, \lambda_n$.

Theorem 3.4.4. *For fixed $d_1 \geq d_2 \geq 2$ and $(d_1, d_2) \neq (2,2)$, let G_n be a sequence of random (d_1, d_2)-biregular bipartite graph. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$. Suppose f is a function such that $f(2z)$ is analytic on E_ρ, where $\rho = [(d_1-1)(d_2-1)]^\alpha$ for some $\alpha > \frac{7}{2}$. Then $f(x)$ can be expanded on $[-2,2]$ as*

$$f(x) = \sum_{k=0}^{\infty} a_k \Gamma_k(x), \tag{3.4.2}$$

and the random variable

$$Y_f^{(n)} := \sum_{i=1}^{n} f(\lambda_i) - n a_0 \tag{3.4.3}$$

converges in distribution as $n \to \infty$ to the infinitely divisible random variable

$$Y_f := \sum_{k=2}^{\infty} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(\infty)}, \tag{3.4.4}$$

where $\mathrm{CNBW}_k^{(\infty)}$ *is defined in* (3.2.37).

Proof. Define

$$f_k(x) := \sum_{i=0}^{k} a_i \Gamma_i(x).$$

We first show that $f_k(x)$ is a good approximation of $f(x)$. Applying Proposition 3.4.3 to $f(2x)$ gives an expansion (3.4.2) with

$$|a_k| \leq C[(d_1-1)(d_2-1)]^{-\alpha k} \tag{3.4.5}$$

for some constant C that depends only on d_1, d_2 and the constant M given in Proposition 3.4.3.

By the proprieties of Chebyshev polynomials, on any interval $[-K,K]$, we have

$$\max_{|x|\leq K}|T_k(x)| = \frac{(K-\sqrt{K^2-1})^k+(K+\sqrt{K^2-1})^k}{2}. \tag{3.4.6}$$

From (3.2.46) and (3.2.47), we have $\Gamma_1(x)=x$, and for any $k\geq 2$,

$$|\Gamma_k(x)| \leq 2\left|T_k\left(\frac{x}{2}\right)\right| + \frac{d_1-2}{(d_1-1)^{k/2}} \leq 2\left|T_k\left(\frac{x}{2}\right)\right|+1. \tag{3.4.7}$$

From (3.4.6),

$$\max_{|x|\leq 3} 2\left|T_k\left(\frac{x}{2}\right)\right| = \left(\frac{3}{2}-\frac{\sqrt{5}}{2}\right)^k + \left(\frac{3}{2}+\frac{\sqrt{5}}{2}\right)^k,$$

Then for all $k\geq 2$, with (3.4.7) we obtain

$$\sup_{|x|\leq 3}|\Gamma_k(x)| \leq \left(\frac{3}{2}+\frac{\sqrt{5}}{2}\right)^k + 2 \leq 3^{k+1}, \tag{3.4.8}$$

and the same bound holds when $k=1$. From (3.4.5) and (3.4.8), for all $x\in[-3,3]$,

$$\sum_{k=0}^{\infty}|a_k\Gamma_k(x)| \leq 3C\sum_{k=0}^{\infty}[3((d_1-1)(d_2-1))^{-\alpha}]^k < \infty,$$

where the last inequality comes from the fact that $(d_1-1)(d_2-1)\geq 2$ and $\alpha > \frac{7}{2}$. Hence the series $\sum_{k=0}^{\infty} a_k\Gamma_k(x)$ is absolutely convergent on $[-3,3]$, which implies the expansion of f in (3.4.2) is valid on $[-3,3]$. Then we have for a constant $C_1>0$ depending on C,

$$\sup_{|x|\leq 3}|f(x)-f_k(x)| \leq \sup_{|x|\leq 3}\sum_{i=k+1}^{\infty}|a_i\Gamma_i(x)| \leq C_1[3(d_1-1)(d_2-1)^{-\alpha}]^{k+1}. \tag{3.4.9}$$

Denote

$$K_1 := \lambda_1 = \frac{d_1\sqrt{(d_2-1)}}{\sqrt{d_1-1}}.$$

For sufficiently large k, from (3.4.6) and (3.4.7),

$$\sup_{|x|\le K_1} |\Gamma_k(x)| \le (2K_1)^k.$$

And from (3.4.5) and the assumption $\alpha > 7/2$,

$$\sum_{k=0}^{\infty} |a_k\Gamma_k(x)| \le C\sum_{k=0}^{\infty}[2K_1((d_1-1)(d_2-1))^{-\alpha}]^k < \infty.$$

It implies the series $\sum_{k=0}^{\infty} a_k\Gamma_k(x)$ is also absolutely convergent on $[-K_1, K_1]$, and the expansion of f in (3.4.2) is valid on $[-K_1, K_1]$.

Since $2K_1 \le 4[(d_1-1)(d_2-1)]^{1/2}$ and $(d_1-1)(d_2-1) \ge 2$, for a constant $C_2 > 0$,

$$\begin{aligned}\sup_{|x|\le K_1} |f(x) - f_k(x)| &\le C_2[2K_1((d_1-1)(d_2-1))^{-\alpha}]^{k+1} \\ &\le C_2\left[4((d_1-1)(d_2-1))^{-\alpha+\frac{1}{2}}\right]^{k+1} \le C_2\left[((d_1-1)(d_2-1))^{-\alpha+\frac{5}{2}}\right]^{k+1}.\end{aligned} \tag{3.4.10}$$

Therefore f_k converges to f uniformly on $[-K_1, K_1]$, and the interval $[-K_1, K_1]$ deterministically contains all eigenvalues of $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$.

By the definition of $\mathrm{CNBW}_k^{(\infty)}$ in (3.2.37), Equation (3.4.4) can be written as

$$Y_f := \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} \frac{a_{ij}}{[(d_1-1)(d_2-1)]^{ij/2}} 2jC_j^{(\infty)},$$

where Y_f is a sum of independent random variables, and $\mathbb{E}|Y_f|^2 < \infty$ by (3.4.5).

Denote $\alpha' := \alpha - 2 > \frac{3}{2}$. Choose β satisfying $\frac{1}{\alpha'} < \beta < \frac{2}{3}$ and define

$$\begin{aligned} r_n &= \left\lfloor \frac{\beta \log n}{\log[(d_1-1)(d_2-1)]} \right\rfloor, \\ X_f^{(n)} &= \sum_{k=1}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(n)}, \\ \tilde{Y}_f^{(n)} &= \sum_{k=1}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(\infty)}. \end{aligned}$$

Note that $\mathrm{CNBW}_1^{(n)} = 0$, from (3.2.53),

$$X_f^{(n)} = \sum_{k=2}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(n)} = \sum_{i=1}^{n} f_{r_n}(\lambda_i) - n a_0. \tag{3.4.11}$$

By Corollary 3.2.15,

$$d_{\mathrm{TV}}\left((\mathrm{CNBW}_k^{(n)}, 2 \le k \le r_n), (\mathrm{CNBW}_k^{(\infty)}, 2 \le k \le r_n) \right) \le \frac{c_8 \sqrt{r_n}[(d_1-1)(d_2-1)]^{3r_n/2}}{n d_1} = o(1).$$

Since $X_f^{(n)}$ and $\tilde{Y}_f^{(n)}$ are measurable functions of

$$(\mathrm{CNBW}_k^{(n)}, 2 \le k \le r_n) \quad \text{and} \quad (\mathrm{CNBW}_k^{(\infty)}, 2 \le k \le r_n),$$

respectively, we have

$$d_{\mathrm{TV}}\left(X_f^{(n)}, \tilde{Y}_f^{(n)} \right) \le d_{\mathrm{TV}}\left((\mathrm{CNBW}_k^{(n)}, 2 \le k \le r_n), (\mathrm{CNBW}_k^{(\infty)}, 2 \le k \le r_n) \right) = o(1).$$

Note that $\tilde{Y}_f^{(n)}$ converges almost surely to Y_f by (3.4.5), so $X_f^{(n)}$ converges in distribution to Y_f.

By Slutsky's theorem, to show $Y_f^{(n)}$ defined in (3.4.3) converges in distribution to Y_f, it remains to show that $Y_f^{(n)} - X_f^{(n)}$ converges to zero in probability. The largest eigenvalue of

$\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$ is K_1, so from (3.4.10) we have

$$\lim_{k\to\infty} f_k(\lambda_1) = f(\lambda_1).$$

Then for any $\delta > 0$ and sufficiently large n,

$$|f(\lambda_1) - f_{r_n}(\lambda_1)| \le \delta/2. \tag{3.4.12}$$

From (3.4.3), (3.4.11) and (3.4.12), we have for sufficiently large n,

$$\left|Y_f^{(n)} - X_f^{(n)}\right| \le \sum_{i=1}^{n} |f(\lambda_i) - f_{r_n}(\lambda_i)| \le \frac{\delta}{2} + \sum_{i=2}^{n} |f(\lambda_i) - f_{r_n}(\lambda_i)|. \tag{3.4.13}$$

Suppose that all the non-trivial eigenvalues $\lambda \neq \lambda_1$ are contained in $[-3,3]$, from (3.4.9),

$$\begin{aligned}\sum_{i=2}^{n} |f(\lambda_i) - f_{r_n}(\lambda_i)| &\le C_1(n-1)[3(d_1-1)(d_2-1)^{-\alpha}]^{r_n+1} \\ &\le C_1 n[(d_1-1)(d_2-1)^{-\alpha+2}]^{r_n} \le C_1 n^{1-\alpha'\beta} = o(1),\end{aligned}$$

which combining (3.4.13) implies for sufficiently large n,

$$\left|Y_f^{(n)} - X_f^{(n)}\right| \le \delta.$$

Recall (3.3.1) and the assumption $d_1 \ge d_2$. With high probability, for a sequence $\varepsilon_n \to 0$, we have the nontrivial eigenvalues of $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$ is contained in

$$\left[-2-\varepsilon_n + \frac{d_2-2}{\sqrt{(d_1-1)(d_2-1)}}, 2+\varepsilon_n + \frac{d_2-2}{\sqrt{(d_1-1)(d_2-1)}}\right] \subseteq [-3,3]$$

for sufficiently large n. Therefore

$$\mathbb{P}\left(\left|Y_f^{(n)} - X_f^{(n)}\right| \geq \delta\right) \leq \mathbb{P}\left(\max_{2\leq i\leq n} |\lambda_i| \geq 3\right) = o(1).$$

This finishes the proof. $\square$

As a corollary of Theorem 3.4.4, we obtain eigenvalue fluctuations for the adjacency matrices of RBBGs as follows.

Corollary 3.4.5. For fixed $d_1 \geq d_2 \geq 2$ and $(d_1,d_2) \neq (2,2)$, let G_n be a sequence of random (d_1,d_2)-biregular bipartite graph. Let $\lambda_1 \geq \cdots \geq \lambda_{n+m}$ be the eigenvalues of its adjacency matrix A. Suppose f satisfies the same conditions as in Theorem 3.4.4. Then the random variable

$$Y_f^{(n)} := \frac{1}{2}\left[\sum_{i=1}^{n+m} f\left(\frac{\lambda_i^2 - d_1}{\sqrt{(d_1-1)(d_2-1)}}\right) - (m-n) f\left(\frac{-d_1}{\sqrt{(d_1-1)(d_2-1)}}\right)\right] - na_0$$

converges in distribution as $n \to \infty$ to the infinitely divisible random variable

$$Y_f := \sum_{k=2}^{\infty} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(\infty)},$$

where $\mathrm{CNBW}_k^{(\infty)}$ is defined in (3.2.37).

Proof. Recall that all eigenvalues of A consist of two parts. There are $2n$ eigenvalues in pair as $\{-\lambda,\lambda\}$ where λ is a singular value of X. In addition, there are $(m-n)$ extra zero eigenvalues. the result then follows from the algebraic relation between eigenvalues of A and eigenvalues of $XX^\top - d_1 I$. $\square$

3.4.2 Gaussian fluctuations with growing degrees

In this section, we consider the eigenvalue fluctuations of RBBGs when $d_1 \cdot d_2 \to \infty$.

We first prove the following weak convergence result for a normalized and centered version of $\text{CNBW}_k^{(\infty)}$.

Lemma 3.4.6. *Suppose that $d_1 \cdot d_2 \to \infty$, $r_n \to \infty$ as $n \to \infty$. For $k \geq 2$, define*

$$N_k^{(n)} := \frac{1}{[d_1-1)(d_2-1)]^{k/2}} \left(\text{CNBW}_k^{(\infty)} - \mathbb{E}\text{CNBW}_k^{(\infty)} \right) \mathbf{1}_{\{k \leq r_n\}}. \tag{3.4.14}$$

Let $\{Z_k\}_{k\geq 2}$ be independent Gaussian random variables with $\mathbb{E}Z_k = 0$ and $\mathbb{E}Z_k^2 = 2k$. Define the weight $w_k = b_k/(k^2 \log(k+1))$, where $(b_k)_{k\in\mathbb{N}}$ is any fixed positive summable sequence.

Let P_n be the law of the sequence $(N_k^{(n)})_{k\geq 2}$. Then as an element in $\mathcal{P}(\mathcal{X})$, P_n converges weakly to the law of the random vector $(Z_k)_{k\geq 2}$.

Proof. We first prove the following *Claim (1)*: for any fixed r, $(N_k^{(n)})_{2\leq k\leq r}$ converges in distribution to $(Z_k)_{2\leq k\leq r}$.

For any fixed k, when n is sufficiently large, we can write (3.4.14) as

$$\begin{aligned} N_k^{(n)} =& \frac{1}{[d_1-1)(d_2-1)]^{k/2}} \left(2kC_k^{(\infty)} - [(d_1-1)(d_2-1)]^k \right) \\ &+ \frac{1}{[d_1-1)(d_2-1)]^{k/2}} \sum_{j|k, j<k} \left(2jC_j^{(\infty)} - [(d_1-1)(d_2-1)]^j \right). \end{aligned} \tag{3.4.15}$$

Recall $C_k^{(\infty)}$ is a Poisson random variable with mean $\frac{(d_1-1)^k(d_2-1)^k}{2k}$. The first term in (3.4.15) converges in distribution to a centered Gaussian random variable Z_k with variance $2k$ as $n \to \infty$ from the Gaussian approximation of Poisson distribution.

To show the convergence of $N_k^{(n)}$ for a fixed k, it remains to show the second term in (3.4.15) converges to zero in probability. Note that the second term in (3.4.15) has mean zero and its variance is given by

$$\mathrm{Var}\left[\frac{1}{[d_1-1)(d_2-1)]^{k/2}}\sum_{j|k,j<k}\left(2jC_j^{(\infty)}-[(d_1-1)(d_2-1)]^j\right)\right]$$
$$=\sum_{j|k,j<k}2j[(d_1-1)(d_2-1)]^{j-k},$$

which goes to 0 as $n\to\infty$. Then by Chebyshev's inequality, this term converges to 0 in probability. Therefore Claim (1) holds.

We further define $N_1^{(n)}=0, Z_1=0$, and consider the weak convergence of $(N_k^{(n)})_{k\in\mathbb{N}}$ as an element in $L^2(\vec{w})$. Since

$$\mathbb{E}\sum_{k=1}^{\infty}(Z_k)^2 w_k=\sum_{k=2}^{\infty}\frac{b_k}{k\log(k+1)}<\infty,$$

$(Z_k)_{k\in\mathbb{N}}\in L^2(\vec{w})$ almost surely. From Claim (1), every sub-sequential limit of P_n has the same finite dimensional distributions as $(Z_k)_{k\in\mathbb{N}}$. From Proposition 3.4.1 (3), every sub-sequential weak limit of P_n in $\mathcal{P}(\mathcal{X})$ is equal to the law of $(Z_k)_{k\in\mathbb{N}}$.

By Prokhorov's Theorem (see for example [158, Chapter 14, Theorem 1.5]), if $\{P_n\}_{n\in\mathbb{N}}$ is tight, and every weakly convergent sub-sequence has the same limit μ in $\mathcal{P}(\mathcal{X})$, then the sequence $\{P_n\}_{n\in\mathbb{N}}$ converges weakly to μ. Since we have already shown every sub-sequential weak limit of P_n is the law of $(Z_k)_{k\in\mathbb{N}}$ in $\mathcal{P}(\mathcal{X})$, to finish the proof, it remains to show $\{P_n\}_{n\in\mathbb{N}}$ is tight.

From the description of compact sets in $L^2(\vec{w})$ given in Proposition 3.4.1 (1), it suffices to show for any $\varepsilon>0$, there exists an element $(a_k)_{k\in\mathbb{N}}\in L^2(\vec{w})$ with $a_k>0, \forall k\in\mathbb{N}$, such that

$$\sup_n \mathbb{P}\left[\bigcup_{k\in\mathbb{N}}\left\{|N_k^{(n)}|>a_k\right\}\right]=\sup_n \mathbb{P}\left[\bigcup_{k=1}^{r_n}\left\{|N_k^{(n)}|>a_k\right\}\right]<\varepsilon, \tag{3.4.16}$$

where $\bigcup_{k\in\mathbb{N}}\left\{|N_k^{(n)}|>a_k\right\}$ is the complement of a compact set in $L^2(\vec{w})$.

For any fixed $\varepsilon>0$, choose $a_k=\alpha k\sqrt{\log(k+1)}$ for a constant $\alpha^2>32$ depending on

ε, then $(a_k)_{k\in\mathbb{N}} \in L^2(\vec{w})$ and $a_k > 0, \forall k \in \mathbb{N}$. According to the definition of $N_k^{(n)}$ in (3.4.14), the above Condition (3.4.16) is equivalent to

$$\sup_n \mathbb{P}\left[\bigcup_{k=1}^{r_n}\left\{|\mathrm{CNBW}_k^{(\infty)} - \mathbb{E}\mathrm{CNBW}_k^{(\infty)}| > a_k[(d_1-1)(d_2-1)]^{k/2}\right\}\right] < \varepsilon. \tag{3.4.17}$$

From the proof of Theorem 22 in [76], $\mathrm{CNBW}_k^{(\infty)}$, as a sum of independent Poisson random variables, satisfies the following concentration inequality: for any $t > 0$,

$$\mathbb{P}\left(|\mathrm{CNBW}_k^{(\infty)} - \mathbb{E}\mathrm{CNBW}_k^{(\infty)}| > t\right) \leq 2\exp\left(-\frac{t}{8k}\log\left(1+\frac{t}{2k[(d_1-1)(d_2-1)]^k}\right)\right). \tag{3.4.18}$$

Since $\log(1+x) \geq x/2$ for $x \in [0,1]$, we have from (3.4.18), for sufficiently large n and all $k \leq r_n$,

$$\begin{aligned}
&\mathbb{P}\left(|\mathrm{CNBW}_k^{(\infty)} - \mathbb{E}\mathrm{CNBW}_k^{(\infty)}| > a_k[(d_1-1)(d_2-1)]^{k/2}\right) \\
\leq& 2\exp\left(-\frac{a_k[(d_1-1)(d_2-1)]^{k/2}}{8k}\log\left(1+\frac{a_k}{2k[(d_1-1)(d_2-1)]^{k/2}}\right)\right) \\
\leq& 2\exp\left(-\frac{a_k^2}{32k^2}\right) = 2(k+1)^{-\alpha^2/32}.
\end{aligned}$$

With the assumption $\alpha^2 > 32$, we can make $\sum_{k=1}^{\infty} 2(k+1)^{-\alpha^2/32} < \varepsilon$ by choosing a sufficiently large constant α depending on ε, which guarantees (3.4.16). Hence $\{P_n\}_{n\in\mathbb{N}}$ is tight. This completes the proof. $\square$

We now continue to study the eigenvalue fluctuation for $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$ when $d_1 \cdot d_2 \to \infty$. Before stating the main result, we make several assumptions on the test function f. Define

$$\Phi_0(x) = 1, \quad \Phi_k(x) = 2T_k(x/2), \quad \forall k \geq 1.$$

Assume f is an entire function on $\mathbb{C}$. Let $K_1 = \max\{\alpha_1, \alpha_2\}$, where α_1 and α_2 are the constants in (3.3.2), (3.3.3), respectively. Then from Proposition 3.4.3, f has the expansion

$$f(x) = \sum_{i=0}^{\infty} a_i \Phi_i(x) \tag{3.4.19}$$

on $[-K_1, K_1]$. Denote

$$f_k(x) := \sum_{i=0}^{k} a_i \Phi_i(x).$$

Suppose the following conditions hold for f:

1. For some $\alpha > 3/2$ and $M > 0$,

$$\sup_{|x| \le K_1} |f(x) - f_k(x)| \le M \exp(-\alpha k h(k)), \tag{3.4.20}$$

   where h is a function such that $h(r_n) \ge \log[(d_1-1)(d_2-1)]$ for a sequence

$$r_n = \left\lfloor \frac{\beta \log n}{\log[(d_1-1)(d_2-1)]} \right\rfloor \tag{3.4.21}$$

   with a constant $\beta < 1/\alpha$.

2.

$$\lim_{n\to\infty} \left| f_{r_n}\left(\frac{d_1(d_2-1)}{\sqrt{(d_1-1)(d_2-1)}}\right) - f\left(\frac{d_1(d_2-1)}{\sqrt{(d_1-1)(d_2-1)}}\right) \right| = 0. \tag{3.4.22}$$

Let $\mu_k(d_1, d_2) := \mathbb{E}\mathrm{CNBW}_k^{(\infty)}$. We define the following sequence:

$$m_f^{(n)} := na_0 + \sum_{k=1}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \left(\mu_k(d_1, d_2) - n(d_1-2)\cdot(d_2-1)^{k/2}\mathbf{1}\{k \text{ is even}\}\right). \tag{3.4.23}$$

Now we are ready to state our results for eigenvalue fluctuations when $d_1 \cdot d_2 \to \infty$. Here d_1, d_2, $(\lambda_i)_{1\le i\le n}$ and the matrix X are quantities depending on n, but for simplicity of notations, we drop the dependence on n.

Theorem 3.4.7. *Let G_n be a sequence of random (d_1,d_2)-biregular bipartite graphs with*

$$d_1 d_2 \to \infty, \quad d_1 d_2 = n^{o(1)}.$$

Let $\lambda_1 \ge \cdots \ge \lambda_n$ be the eigenvalues of $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$. Suppose one of the following two assumptions holds:

1. *There exists a constant $c \ge 1$ such that $1 \le \frac{d_1}{d_2} \le c$.*

2. *There exists a constant c_1 such that $d_2 \le c_1$ for all n.*

Let f be an entire function on $\mathbb{C}$ satisfying (3.4.20) *and* (3.4.22). *Then as $n \to \infty$, the random variable*

$$Y_f^{(n)} = \sum_{i=1}^{n} f(\lambda_i) - m_f^{(n)} \tag{3.4.24}$$

converges in distribution to a centered Gaussian random variable with variance $\sigma_f = 2\sum_{k=2}^{\infty} k a_k^2$.

Moreover, for any fixed t, consider the entire functions $g_1, \dots, g_t$ satisfying (3.4.20) *and* (3.4.22). *The corresponding random vector $(Y_{g_1}^{(n)}, \dots, Y_{g_t}^{(n)})$ converges in distribution to a centered Gaussian random vector $(Z_{g_1}, \dots, Z_{g_t})$ with covariance*

$$\mathrm{Cov}(Z_{g_i}, Z_{g_j}) = 2\sum_{k=2}^{\infty} k a_k(g_i) a_k(g_j) \tag{3.4.25}$$

for $1 \le i, j \le t$, where $a_k(g_i), a_k(g_j)$ are the k-th coefficients in the expansion (3.4.19) *for g_i, g_j, respectively.*

Proof. We first prove the CLT for a single test function f. Define

$$X_f^{(n)} := \sum_{k=2}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(n)} - \mathbb{E} \sum_{k=2}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(\infty)},$$
$$\tilde{X}_f^{(n)} := \sum_{k=2}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(\infty)} - \mathbb{E} \sum_{k=2}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(\infty)}.$$

Recall the definition of $m_f(n)$ in (3.4.23). From (3.2.46), (3.2.47), and (3.2.53), $X_f^{(n)}$ can be written as

$$\begin{aligned}
X_f^{(n)} &= \sum_{k=2}^{r_n} \sum_{i=1}^{n} a_k \Gamma_k(\lambda_i) - \sum_{k=2}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mu_k(d_1,d_2) \\
&= \sum_{k=2}^{r_n} \sum_{i=1}^{n} \left(2a_k T_k(\lambda_i/2) + \frac{a_k(d_1-2)}{(d_1-1)^{k/2}} \mathbf{1}_{\{k \text{ is even}\}} \right) - \sum_{k=2}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mu_k(d_1,d_2) \\
&= \sum_{i=1}^{n} f_{r_n}(\lambda_i) - na_0 + \sum_{k=2}^{r_n} \frac{na_k(d_1-2)}{(d_1-1)^{k/2}} \mathbf{1}_{\{k \text{ is even}\}} - \sum_{k=2}^{r_n} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mu_k(d_1,d_2) \\
&= \sum_{i=1}^{n} f_{r_n}(\lambda_i) - m_f^{(n)},
\end{aligned}$$

where in the third line we use the fact given in (3.2.53) that

$$\sum_{i=1}^{n} 2a_1 T_1(\lambda_i/2) = \sum_{i=1}^{n} a_1 \Gamma_1(\lambda_i) = a_1[(d_1-1)(d_2-1)]^{-1/2} \mathrm{CNBW}_1^{(n)} = 0.$$

From the definition of $N_k^{(n)}$ in (3.4.14),

$$\tilde{X}_f^{(n)} = \sum_{k=2}^{r_n} a_k N_k^{(n)}.$$

By Lemma 3.4.6 and Proposition 3.4.1 (2), $\tilde{X}_f^{(n)}$ converges in distribution to a centered Gaussian random variable with variance $\sigma_f = \sum_{k=2}^{\infty} 2ka_k^2$.

From Corollary 3.2.15, the total variation distance between $X_f^{(n)}$ and $\tilde{X}_f^{(n)}$ satisfies

$$\begin{aligned} d_{\mathrm{TV}}(X_f^{(n)}, \tilde{X}_f^{(n)}) &\leq d_{\mathrm{TV}}\left((\mathrm{CNBW}_k^{(n)}, 2 \leq k \leq r_n), (\mathrm{CNBW}_k^{(\infty)}, 2 \leq k \leq r_n)\right) \\ &\leq \frac{c_8\sqrt{r_n}[(d_1-1)(d_2-1)]^{3r_n/2}}{nd_1}, \end{aligned}$$

which converges to 0 as $n \to \infty$ from the assumption (3.4.21). Therefore $X_f^{(n)}$ and $\tilde{X}_f^{(n)}$ converge to the same limit.

It remains to show $Y_f^{(n)}$ and $X_f^{(n)}$ converge in distribution to the same limit. We have $f_{r_n}(\lambda_1) \to f(\lambda_1)$ as $n \to \infty$ from (3.4.22). Then for any $\delta > 0$, $|f(\lambda_1) - f_{r_n}(\lambda_1)| \leq \delta/2$ for sufficiently large n.

Suppose that all the non-trivial eigenvalues are contained in $[-K_1, K_1]$. From Condition (3.4.20), we have for sufficiently large n,

$$\left|Y_f^{(n)} - X_f^{(n)}\right| \leq \sum_{i=1}^{n} |f(\lambda_i) - f_{r_n}(\lambda_i)| \leq \frac{\delta}{2} + (n-1)M\exp(-\alpha r_n h(r_n)) \leq \frac{\delta}{2} + Mn^{1-\alpha\beta} < \delta.$$

Therefore

$$\mathbb{P}\left(\left|Y_f^{(n)} - X_f^{(n)}\right| \geq \delta\right) \leq \mathbb{P}\left(\max_{2 \leq i \leq n} |\lambda_i| \geq K_1\right) = o(1), \tag{3.4.26}$$

where the last inequality is from part (2) and (3) in Theorem 3.3.1. Hence $Y_f^{(n)}$ and $X_f^{(n)}$ converges in distribution to the same limit. This proves the CLT for (3.4.24).

We now extend the results to a random vector $(Y_{g_1}, \ldots, Y_{g_t})$. By Lemma 3.4.6 and part (2) in Proposition 3.4.1, the random vector $(\tilde{X}_{g_1}^{(n)}, \ldots, \tilde{X}_{g_t}^{(n)})$ converges in distribution to the Gaussian random vector $(Z_{g_1}, \ldots, Z_{g_t})$ with covariance given in (3.4.25). Note that each entry in the vector $(X_{g_1}^{(n)}, \ldots, X_{g_t}^{(n)})$ is a measurable function of $(\mathrm{CNBW}_k^{(n)})_{2 \leq k \leq r_n}$, and we can find a measurable map

ψ such that

$$\psi((\mathrm{CNBW}_k^{(n)})_{2\leq k\leq r_n}) = (X_{g_1}^{(n)},\dots,X_{g_t}^{(n)}), \quad \psi((\mathrm{CNBW}_k^{(\infty)})_{2\leq k\leq r_n}) = (\tilde{X}_{g_1}^{(n)},\dots,\tilde{X}_{g_t}^{(n)}).$$

Since any measurable map reduces total variation distance between two random variables, we obtain from (3.2.41),

$$\begin{aligned}
&d_{\mathrm{TV}}\left((X_{g_1}^{(n)},\dots,X_{g_t}^{(n)}),(\tilde{X}_{g_1}^{(n)},\dots,\tilde{X}_{g_t}^{(n)})\right)\\
\leq& d_{\mathrm{TV}}\left((\mathrm{CNBW}_k^{(n)}, 2\leq k\leq r_n),(\mathrm{CNBW}_k^{(\infty)}, 2\leq k\leq r_n)\right)\\
\leq& \frac{c_8\sqrt{r_n}[(d_1-1)(d_2-1)]^{3r_n/2}}{nd_1} = o(1).
\end{aligned}$$

Therefore $\left(X_{g_1}^{(n)},\dots,X_{g_t}^{(n)}\right)$ converges in distribution to $(Z_{g_1},\dots,Z_{g_t})$. Finally, according to (3.4.26), $\left(X_{g_1}^{(n)},\dots,X_{g_t}^{(n)}\right)$ and $\left(Y_{g_1}^{(n)},\dots,Y_{g_t}^{(n)}\right)$ converge in distribution to the same limit. This finishes the proof. $\square$

Remark 3.4.8. In [60], the authors proved a CLT for linear spectral statistics for normalized sample covariance matrices $A = \frac{1}{\sqrt{np}}(XX^\top - pI)$, where $p/n \to \infty$ and $X = (X_{ij})_{n\times p}$ has i.i.d. entries with mean 0 variance 1. It is shown in Theorem 1 of [60] that the fluctuations of linear statistics for two analytic functions g_1, g_2 converge in distribution to a centered Gaussian vector with covariance given by $(\nu_4 - 3)a_1(g_1)a_1(g_2) + 2\sum_{k=1}^{\infty} k a_k(g_1)a_k(g_2)$, where $\nu_4 = \mathbb{E}X_{11}^4$. The covariance given in (3.4.25) is the same, except for the fact that the coefficient in front of $a_1(f_1)a_1(f_2)$ is 0. This can be explained by the fact that the number of 2-cycles is 0 in RBBGs, whereas in the model used in [60] it is not. The same phenomenon was also observed in uniform random regular graphs [119], where the limiting variance is the same as the eigenvalue fluctuations for the GOE except for the first two terms, see Remark 22 in [119].

3.5 Global semicircle law

Consider a random (n,m,d_1,d_2)-biregular bipartite graph with $d_1 \geq d_2$. We assume d_1, d_2 satisfy the following:

$$\lim_{n\to\infty} d_1 = \infty, \tag{3.5.1}$$

$$d_1 = o(n^{\varepsilon}), \quad \forall \varepsilon > 0, \tag{3.5.2}$$

$$\frac{d_1}{d_2} \to \infty. \tag{3.5.3}$$

Here d_2 can be fixed or a parameter depending on n. In this section, we prove a semicircle law for the matrix $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$ under the assumptions (3.5.1)-(3.5.3).

For RBBGs in this regime, we have the locally tree-like structure in the following sense. Let R be fixed and τ_1 be the set of vertices in V_1 without any cycles in the R-neighborhood. The following lemma holds.

Lemma 3.5.1. *Then under Condition* (3.5.2)*,*

$$\mathbb{P}\left(\frac{n - |\tau_1|}{n} > n^{-1/4}\right) = o(n^{-5/4}).$$

To prove Lemma 3.5.1, the following estimates on the expectation and variance of the cycle counts of RBBGs given in [75] are needed.

Lemma 3.5.2 (Proposition 4 in [75])**.** *Let C_k be the number of cycles of length $2k$ in a random (d_1,d_2)-biregular bipartite graph. Denote $\mu_k = \frac{[(d_1-1)(d_2-1)]^k}{2k}$. If $d_1 = o(n), k = O(\log n)$ and $kd_1 = o(n)$, then*

$$\mathbb{E}C_k = \mu_k\left(1 + O\left(\frac{k(k+d_1)}{n}\right)\right), \tag{3.5.4}$$

$$\mathrm{Var}[C_k] = \mu_k\left(1 + O\left(\frac{d_1^{2k}(k(d_1/d_2)^{2k-1} + (d_1/d_2)^{-k}d_2)}{n}\right)\right). \tag{3.5.5}$$

Proof of Lemma 3.5.1. From Lemma 3.5.2, for each fixed k, under Condition (3.5.2),

$$\mathbb{E}C_k = (1+o(1))\mu_k, \quad \operatorname{Var}[C_k] = (1+o(1))\mu_k. \tag{3.5.6}$$

If a vertex $v_1 \in V_1$ is not in τ_1, then for some s with $2 \le s \le R$, there exists a $2s$-cycle within $(R-s)$-neighborhood of v_1. Hence the size of all $(R-s)$-neighborhoods of $2s$-cycles from V_1 gives an upper bound on $(n-|\tau_1|)$.

For any $2s$-cycle, the size of its $(R-s)$-neighborhood from V_1 is bounded by

$$c_1 s[(d_1-1)(d_2-1)]^{(R-s)/2+1}$$

with an absolute constant c_1. Define

$$N_R := c_1 \sum_{s=2}^{R} s[(d_1-1)(d_2-1)]^{(R-s)/2+1} C_s.$$

We then have $n-\tau_1 \le N_R$. From (3.5.6), $\mathbb{E}N_R = O([(d_1-1)(d_2-1)]^{R+1})$.

Recall R is fixed. By Cauchy inequality,

$$\operatorname{Var}[N_R] \le c_1^2 R \sum_{s=2}^{R} s^2[(d_1-1)(d_2-1)]^{R-s+2}\operatorname{Var}[C_s] = O([(d_1-1)(d_2-1)]^{R+2}).$$

Then from Markov's inequality, together with our assumptions (3.5.1)-(3.5.3),

$$\begin{aligned}\mathbb{P}\left(\frac{n-|\tau_1|}{n} > n^{-1/4}\right) &= \mathbb{P}(n-|\tau_1| > n^{3/4}) \le \mathbb{P}(N_R \ge n^{3/4}) \\ &\le \frac{\mathbb{E}[N_R^2]}{n^{3/2}} = O([(d_1-1)(d_2-1)]^{2R+2} n^{-3/2}) = o(n^{-5/4}).\end{aligned}$$

□

We now state our main result in this section. The proof is based on the moment method and the tree approximation of local neighborhoods, which were previously applied to random regular graphs in [77].

Theorem 3.5.3. *Let G_n be a sequence of random (d_1, d_2)-biregular bipartite graph. Under assumptions* (3.5.1)-(3.5.3)*, the empirical spectral distribution of $\frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$ converges weakly to the semicircle law almost surely.*

Remark 3.5.4. Recall in [75], when the ratio $d_1/d_2 \geq 1$ converges to a positive constant, the ESD of $\frac{XX^\top}{d_1}$ converges to Marčenko-Pastur law. With different scaling parameters, we obtain a different semicircle law when $d_1/d_2 \to \infty$. This can be seen as an analog of the semicircle law for sample covariance matrices proved in [26] when the aspect ratio is unbounded.

Proof of Theorem 3.5.3. Note that for all $i \in V_1$, by the degree constraint,

$$(XX^\top)_{ii} = \sum_j X_{ij} X_{ji} = \sum_j X_{ij} = \deg(i) = d_1. \tag{3.5.7}$$

Denote $M = \frac{XX^\top - d_1 I}{\sqrt{(d_1-1)(d_2-1)}}$. We start with the trace expansion of M.

$$\begin{aligned} \frac{1}{n}\mathrm{tr} M^k &= \frac{1}{n((d_1-1)(d_2-1))^{k/2}} \mathrm{tr}(XX^\top - d_1 I)^k \\ &= \frac{1}{n((d_1-1)(d_2-1))^{k/2}} \sum_{\substack{i_1,\dots,i_k \in [n] \\ i_1 \neq i_2, \dots, i_k \neq i_1 \\ j_1,\dots,j_k \in [m]}} X_{i_1 j_1} X_{i_2 j_1} \cdots X_{i_k j_k} X_{i_1 j_k}. \end{aligned} \tag{3.5.8}$$

From (3.5.7), the diagonal entries of $XX^T - d_1 I$ are 0, therefore we have the constraint that $i_1 \neq i_2, \dots, i_k \neq i_1$ in (3.5.8).

Let $A_k^{r,c}(v,v)$ be the number of all closed walks of length $2k$ in G starting from $v \in V_1$ that use r distinct vertices from V_1, c distinct vertices from V_2, with the restriction that $i_1 \neq i_2, \dots, i_k \neq i_1$. We have $r \leq k+1$ and $c \leq k$, since there are at most $k+1$ vertices in V_1 and k vertices in V_2 that are visited in one closed walk of length $2k$. From (3.5.8), the k-th moment of the empirical spectral distribution μ_n satisfies

$$\int x^k d\mu_n(x) = \frac{1}{n}\mathrm{tr}M^k = \frac{1}{n((d_1-1)(d_2-1))^{k/2}} \sum_{v\in V_1} \sum_{r=1}^{k+1} \sum_{c=1}^{k} A_k^{r,c}(v,v). \tag{3.5.9}$$

Since $d_1 \geq d_2$, for any fixed $v \in V_1$, we have

$$\sum_{r\leq k+1, c\leq k} A_k^{r,c}(v,v) \leq d_1^{2k}.$$

For the ease of notations, in the following equations we often omit the range of r,c in the summation.

We may decompose the sum in (3.5.9) into two parts depending on whether $v \in \tau_1$ or not. For any $v \in \tau_1$, we write $A_k^{r,c} =: A_k^{r,c}(v,v)$ since all neighborhood of $v \in \tau_1$ of radius k looks the same and the number of such closed walks is independent of v. Now we have the following upper bound on (3.5.9):

$$\begin{aligned}
\int x^k d\mu_n(x) &\leq \frac{1}{n((d_1-1)(d_2-1))^{k/2}} \sum_{v\in\tau_1} \sum_{r,c} A_k^{r,c}(v,v) + \frac{(n-|\tau_1|)d_1^{2k}}{n((d_1-1)(d_2-1))^{k/2}} \\
&= \frac{|\tau_1|}{n((d_1-1)(d_2-1))^{k/2}} \sum_{r,c} A_k^{r,c} + \frac{(n-|\tau_1|)d_1^{2k}}{n((d_1-1)(d_2-1))^{k/2}} \\
&\leq \frac{1}{((d_1-1)(d_2-1))^{k/2}} \sum_{r,c} A_k^{r,c} + \frac{(n-|\tau_1|)d_1^{2k}}{n((d_1-1)(d_2-1))^{k/2}}.
\end{aligned}$$

Similarly, a lower bound holds by only counting closed walks starting with vertices in τ_1:

$$\int x^k d\mu_n(x) \geq \frac{1}{n((d_1-1)(d_2-1))^{k/2}} \sum_{v\in\tau_1} \sum_{r} \sum_{c} A_k^{r,c}(v,v) = \frac{|\tau_1|}{n((d_1-1)(d_2-1))^{k/2}} \sum_{r} \sum_{c} A_k^{r,c}.$$

From Lemma 3.5.1 and assumption (3.5.2), with probability at least $1-o(n^{-5/4})$, for any fixed $k \geq 0$,

$$\frac{(n-|\tau_1|)}{n((d_1-1)(d_2-1))^{k/2}} d_1^{2k} = o(1), \quad \text{and} \quad \frac{|\tau_1|}{n} = 1 - o(n^{-1/4}).$$

To show the almost sure convergence of the empirical measure to semicircle law, by the upper and lower bounds above, it suffices to show

$$\lim_{n\to\infty} \frac{1}{((d_1-1)(d_2-1))^{k/2}} \sum_{r,c} A_k^{r,c} = \begin{cases} 0 & \text{if } k \text{ is odd,} \\ C_{k/2} & \text{if } k \text{ is even,} \end{cases} \tag{3.5.10}$$

where $C_k := \frac{1}{k+1}\binom{2k}{k}$ is the k-th Catalan number.

Recall $A_k^{r,c}$ counts the closed walks of length $2k$ on a rooted (d_1,d_2)-biregular tree starting from the a root with degree d_1, ending at the same root. Now we consider the quantity

$$\frac{1}{((d_1-1)(d_2-1))^{k/2}} \sum_{r,c} A_k^{r,c}$$

more carefully. We first consider possible ranges of r and c in the expression above.

The walk $(i_1, j_1, i_2, j_2, \dots, i_k, j_k, i_1)$ in the summation satisfies $i_1 \neq i_2, \cdots, i_{k-1} \neq i_k, i_k \neq i_1$. This implies when a walk goes from i_t to j_t for some t, it cannot backtrack immediately to i_t. Namely, any such walk is not allowed to backtrack at even depths (here we define the depth of the root in a tree is 1). To have a closed walk of length $2k$ on a tree, each edge is repeated at least twice, so the number of distinct edges is at most k, therefore number of distinct vertices satisfies

$$r + c \leq k + 1. \tag{3.5.11}$$

For fixed r and c, the number of such unlabeled rooted tree with $r+c-1$ distinct edges is C_{r+c-1}. Let I be the set of vertices in the odd depths of the biregular tree and J be the set of vertices in the even depths. Since the first vertex of the walk is fixed (we always start from the fixed root), for any closed walk, there are at most d_1^c many ways to choose distinct vertices from J and d_2^{r-1} many ways to choose distinct vertices from I. Therefore we have

$$A_k^{r,c} \le d_1^c d_2^{r-1} C_{r+c-1} \le d_1^c d_2^{r-1} C_k, \tag{3.5.12}$$

where the last inequality is from (3.5.11). We also know that $r-1 \ge c$, because whenever a new vertex in J is reached by the walk, the walk cannot backtrack, so it must reach a new vertex in I. Therefore we have

$$c \le r-1 \quad \text{and} \quad r+c \le k+1,$$

which implies the following conditions on c and r:

$$c \le k/2 \quad \text{and} \quad r-1 \le k-c. \tag{3.5.13}$$

From (3.5.12), for any (r,c) satisfying (3.5.13), the following holds:

$$\frac{A_k^{r,c}}{((d_1-1)(d_2-1))^{k/2}} \le \frac{d_1^c}{(d_1-1)^{k/2}} \frac{d_2^{r-1}}{(d_2-1)^{k/2}} C_k \le \frac{d_1^c}{(d_1-1)^{k/2}} \frac{d_2^{k-c}}{(d_2-1)^{k/2}} C_k. \tag{3.5.14}$$

Now we discuss two cases depending on the parity of k. When k is odd, from (3.5.13), $c \le \frac{k-1}{2}$. Since $d_1/d_2 \to \infty$, we obtain

$$\frac{1}{((d_1-1)(d_2-1))^{k/2}} \sum_{r,c} A_k^{r,c} \le \left(\frac{d_1}{d_2}\right)^c \frac{d_2^k C_k}{[(d_1-1)(d_2-1)]^{k/2}} \le \left(\frac{d_1}{d_2}\right)^{c-k/2} 2^k C_k = o(1). \tag{3.5.15}$$

When k is even, to have a non-vanishing term in the limit for $A_k^{r,c}$, we must have $c = k/2$

and $r = k/2+1$. Then we have

$$\frac{1}{((d_1-1)(d_2-1))^{k/2}} \sum_{r,c} A_k^{r,c} = \frac{1}{((d_1-1)(d_2-1))^{k/2}} A_k^{k/2+1,k/2} + o(1). \tag{3.5.16}$$

We continue our proof with a more refined estimate on $A_k^{k/2+1,k/2}$. Since every edge is repeated exactly twice in the closed walk, it's a depth-first search on the biregular tree.

If the root is at level 1, and subsequent vertices are at a level $i+1$ where i is the distance from the root, then all leaves must be at odd levels, since we can never backtrack at an even level. This implies that every vertex at an even level has at least one child, which means $r \geq c+1$, with equality if and only if every vertex at an even level has *exactly* one child. Thus, one can see the tree as a subdivision of a smaller tree, where a vertex has been introduced on each edge (the "new" vertices being the vertices on an even level in the bigger tree). This is clearly a bijection between the kind of planar rooted tree on $k+1$ vertices we are trying to count, and the set of all planar rooted trees on $k/2+1$ vertices. There are precisely $C_{k/2}$ of the latter. See Figure 3.3 for an example of a valid closed walk and an illustration of the aforementioned bijection.

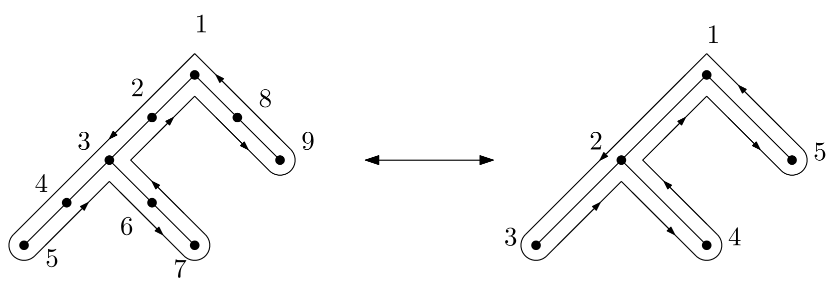

Figure 3.3: On the left we have a closed walk $(1,2,3,4,5,4,3,6,7,6,3,2,1,8,9,8,1)$ on a rooted planar tree which only backtracks at odd depths and the tree has no new branches at any even depth along the walk. Its correspondent under the bijection is the closed walk $(1,2,3,2,4,2,1,5,1)$ on the smaller rooted planar tree induced by the depth-first search on the right.

Moreover, given a fixed root with a vertex label, the number of all possible ways to label the tree with vertices in a biregular bipartite graph is between $d_1^{k/2}(d_2-1)^{k/2}$ and $(d_1 -$

$k/2)^{k/2}(d_2-1)^{k/2}$, so the following inequality for $A_k^{k/2+1,k/2}$ holds:

$$(d_1-k/2-1)^{k/2}(d_2-1)^{k/2}C_{k/2} \leq A_k^{k/2+1,k/2} \leq d_1^{k/2}(d_2-1)^{k/2}C_{k/2}. \tag{3.5.17}$$

From (3.5.16) and (3.5.17), we obtain for even k,

$$\lim_{n\to\infty} \frac{1}{((d_1-1)(d_2-1))^{k/2}} \sum_{r,c} A_k^{r,c} = C_{k/2}. \tag{3.5.18}$$

With (3.5.15) and (3.5.18), the asymptotic behavior of moments given in (3.5.10) holds. This completes the proof of Theorem 3.5.3. □

3.6 Random regular hypergraphs

We first describe a bijection between a subset of biregular bipartite graphs and the set of regular hypergraphs studied in [79]. We will use the map given in Definition 3.6.1 to apply some of our results for RBBGs to random regular hypergraphs, see [79] for more details.

Definition 3.6.1 (incidence matrix and associated bipartite graph)**.** A vertex i is *incident* to a hyperedge e if and only v is an element of e. We can define the *incidence matrix* X of a hypergraph $H=(V,E)$ to be a $|V|\times|E|$ matrix indexed by elements in V and E such that $X_{i,e}=1$ if $i\in e$ and 0 otherwise. Moreover, if we regard X as the adjacency matrix of a graph, it defines a bipartite graph G with two vertex sets V and E. We call G the *bipartite graph associated to* H, given by a map Φ (so $\Phi(H)=G$). See Figure 3.4 for an example.

Definition 3.6.2 (adjacency matrix)**.** For a hypergraph H with n vertices, we associate a $n\times n$ symmetric matrix A called the *adjacency matrix* of H. For $i\neq j$, we define A_{ij} as the number of hyperedges containing both i and j; we define $A_{ii}=0$ for all $1\leq i\leq n$. When the hypergraph is 2-uniform (i.e., it is a graph), this is the usual definition for the adjacency matrix of a graph.

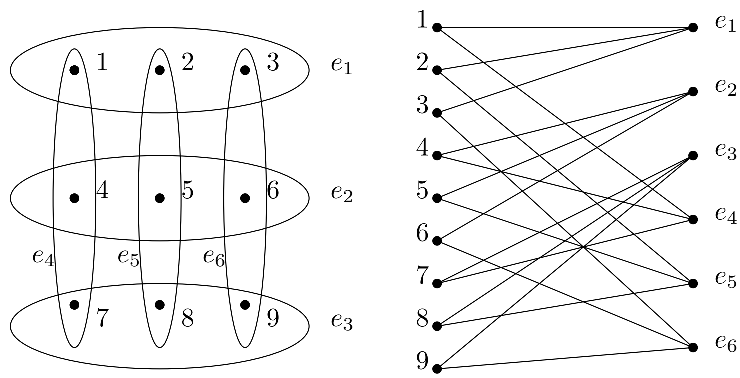

Figure 3.4: a $(2,3)$-regular hypergraph and its associated biregular bipartite graph

The following lemma connects the adjacency matrix of a regular hypergraph with its associated biregular bipartite graph. It formally appears in [132, 79], and it is also informally mentioned in [91].

Lemma 3.6.3 (Lemma 4.5 in [79])**.** *Let H be a (d_1,d_2)-regular hypergraph, and let G be the corresponding (d_1,d_2)-biregular bipartite graph. Let A_H be the adjacency matrix of H and A_G be the adjacency matrix of G given by*

$$A_G = \begin{pmatrix} 0 & X \\ X^\top & 0 \end{pmatrix}. \tag{3.6.1}$$

Then $A_H = XX^\top - d_1 I$.

Definition 3.6.4 (walks and cycles)**.** A *walk* of length l on a hypergraph H is a vertex-hyperedge sequence $(i_0, e_1, i_1, \cdots, e_l, i_l)$ such that $i_{j-1} \neq i_j$ and $\{i_{j-1}, i_j\} \subset e_j$ for all $1 \leq j \leq l$. A walk is closed if $i_0 = i_l$. A cycle of length l in a hypergraph H is a closed walk $(v_0, e_1, \ldots, v_{l-1}, e_l, v_{l+1})$ such that all edges are distinct and all vertices are distinct subject to $v_{l+1} = v_0$. In the associated bipartite graph G, a cycle of length $2l$ corresponds to a cycle of length l in H.

Let $\mathcal{G}(n,m,d_1,d_2)$ be the set of all simple biregular bipartite random graphs with vertex set $V = V_1 \cup V_2$ such that $|V_1| = n, |V_2| = m$, and every vertex in V_i has degree d_i for $i = 1,2$.

Without loss of generality, we assume $d_1 \geq d_2$. Let $\mathcal{H}(n,d_1,d_2)$ be the set of all simple (without multiple hyperedges) (d_1,d_2)-regular hypergraphs with labeled vertex set $[n]$ and $\frac{nd_1}{d_2}$ many labeled hyperedges denoted by $\{e_1,\dots,e_{nd_1/d_2}\}$.

Remark 3.6.5. We can also consider all (d_1,d_2)-regular hypergraphs with labeled vertices and unlabeled hyperedges. Since all hyperedges are distinct, any such regular hypergraph with unlabeled hyperedges corresponds to $(nd_1/d_2)!$ regular hypergraphs with labeled hyperedges.

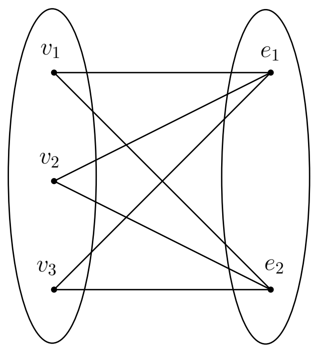

Figure 3.5: a subgraph in a biregular bipartite graph which gives multiple hyperedges e_1 and e_2 in the corresponding regular hypergraph

It is well known (see for example [91]) that the map Φ defined can be extended to a bijection $\tilde{\Phi}$ between labeled regular multi-hypergraphs and biregular bipartite graphs. See Figure 3.4 as an example of the bijection. For a given biregular bipartite graph, if there are two vertices in V_2 that have the same set of neighbors in V_1, the corresponding regular hypergraph will have multiple hyperedges, see Figure 3.5. Let $\mathcal{G}'(n,m,d_1,d_2)$ be a subset of $\mathcal{G}(n,m,d_1,d_2)$ such that for any $G \in \mathcal{G}'(n,m,d_1,d_2)$, any two vertices in V_2 have different neighborhoods in V_1. The following lemma holds.

Lemma 3.6.6 (Lemma 4.2 in [79]). *Φ is the restriction of the bijection $\tilde{\Phi}$ to $\mathcal{H}(n,d_1,d_2)$ and its image is $\mathcal{G}'(n,m,d_1,d_2)$. Hence $|\mathcal{H}(n,d_1,d_2)| = |\mathcal{G}'(n,m,d_1,d_2)|$.*

From Lemma 3.6.6, the uniform distribution on $\mathcal{G}'(n,m,d_1,d_2)$ for biregular bipartite graphs induces the uniform distribution on $\mathcal{H}(n,d_1,d_2)$ for regular hypergraphs. With this observation, we are able to translate some of the results for spectra of random biregular bipartite graphs into results for spectra of random regular hypergraphs. A similar approach was applied in [35] to enumerate uniform hypergraphs with given degrees.

Lemma 3.6.7 (Lemma 4.8 in [79])**.** *Let G be a random biregular bipartite graph sampled uniformly from $\mathcal{G}(n,m,d_1,d_2)$ such that $3 \leq d_2 \leq d_1 \leq \frac{n}{32}$. Let $\mathcal{G}'(n,m,d_1,d_2)$ be the set of biregular bipartite graphs that corresponds to simple regular hypergraphs. Then*

$$\mathbb{P}\left(G \in \mathcal{G}'(n,m,d_1,d_2)\right) \geq 1 - \left(\frac{nd_1}{d_2}\right)^2 \left(\frac{4ed_2}{n}\right)^{d_2}. \tag{3.6.2}$$

In particular,

$$\mathbb{P}\left(G \in \mathcal{G}'(n,m,d_1,d_2)\right) = 1 - O\left(\frac{d_1^2}{nd_2^2}\right). \tag{3.6.3}$$

Lemma 3.6.7 implies the following total variation bound.

Lemma 3.6.8 (total variation bound)**.** *Let μ_n be the probability measure of the random (d_1,d_2)-regular hypergraph with n vertices induced on the set of all (n,m,d_1,d_2)-biregular bipartite graphs, and let μ_n' be the uniform measure on the set of all (n,m,d_1,d_2)-biregular bipartite graphs. We have*

$$d_{\mathrm{TV}}(\mu_n,\mu_n') \leq \left(\frac{nd_1}{d_2}\right)^2 \left(\frac{4ed_2}{n}\right)^{d_2}. \tag{3.6.4}$$

Proof. Since $\mathcal{G}'(n,m,d_1,d_2)$ is the set of all biregular bipartite graphs that are bijective to regular hypergraphs. We have $\mu_n(\mathcal{G}'(n,m,d_1,d_2)) = 1$ and $\mu_n'(\mathcal{G}'(n,m,d_1,d_2)) = \frac{|G'(n,m,d_1,d_2)|}{|\mathcal{G}(n,m,d_1,d_2)|}$. Let $\mathcal{F}$ be the power set of $\mathcal{G}(n,m,d_1,d_2)$. Taking into account the fact that both μ_n and μ_n' are uniform measures, we obtain that

$$
\begin{aligned}
d_{\mathrm{TV}}(\mu_n,\mu_n') &= \sup_{A\in\mathcal{F}} |\mu_n(A)-\mu_n'(A)| = |1-\mu_n'(G'(n,m,d_1,d_2))| \\
&= \mathbb{P}\left(G \notin \mathcal{G}'(n,m,d_1,d_2)\right) \leq \left(\frac{nd_1}{d_2}\right)^2 \left(\frac{4ed_2}{n}\right)^{d_2},
\end{aligned}
$$

where the last inequality is from Lemma 3.6.7. □

Equipped with Lemma 3.6.8, we obtain several corollaries for random regular hypergraphs in the following subsections.

Cycle counts

Recall the definition of cycles in a hypergraph given in Definition 3.6.4. Let C_k be the number of cycles of length k in a (d_1,d_2)-regular hypergraph. The following result holds.

Corollary 3.6.9. Let H be a (d_1,d_2)-random regular hypergraph with cycle counts $(C_k, k\geq 2)$. Let $(Z_k, k\geq 2)$ be independent Poisson random variables with $\mathbb{E}Z_k = \dfrac{(d_1-1)^k(d_2-1)^k}{2k}$. For any $n,m\geq 1$, $r\geq 3$, and $3\leq d_2\leq d_1\leq \frac{n}{32}$,

$$
d_{\mathrm{TV}}((C_2,\dots,C_r),(Z_2,\dots,Z_r)) \leq \frac{c_6\sqrt{r}(d_1-1)^{3r/2}(d_2-1)^{3r/2}}{nd_1} + \left(\frac{nd_1}{d_2}\right)^2\left(\frac{4ed_2}{n}\right)^{d_2}.
$$

Proof. Let $\tilde{C}_k$ be the number of cycles with length $2k$ in a uniform random (d_1,d_2)-biregular bipartite graph. From Lemma 3.6.8,

$$
d_{\mathrm{TV}}((C_2,\dots,C_r),(\tilde{C}_2,\dots,\tilde{C}_r)) \leq d_{\mathrm{TV}}(\mu_n,\mu_n') \leq \left(\frac{nd_1}{d_2}\right)^2\left(\frac{4ed_2}{n}\right)^{d_2}. \tag{3.6.5}
$$

Then the conclusion follows from Theorem 3.2.10 and the triangle inequality. □

Global laws

The limiting spectral distributions for the adjacency matrix of a random regular hypergraph can be summarized in the following corollary.

Corollary 3.6.10. Let H be a random (d_1,d_2)-regular hypergraph.

1. If d_1,d_2 are fixed, the empirical spectral distribution of $\frac{A-(d_2-2)}{\sqrt{(d_1-1)(d_2-1)}}$ converges in probability to a measure μ with density function given by

$$f(x) := \frac{1+\frac{d_2-1}{q}}{(1+\frac{1}{q}-\frac{x}{\sqrt{q}})(1+\frac{(d_2-1)^2}{q}+\frac{(d_2-1)x}{\sqrt{q}})}\frac{1}{\pi}\sqrt{1-\frac{x^2}{4}}dx, \tag{3.6.6}$$

where $q=(d_1-1)(d_2-1)$.

2. For $d_1,d_2\to\infty$ with $\frac{d_1}{d_2}\to\alpha\geq 1$ and $d_1\leq\frac{n}{32}$, the empirical spectral distribution of $\frac{A-(d_2-2)}{\sqrt{(d_1-1)(d_2-1)}}$ converges in probability to a measure supported on $[-2,2]$ with a density function given by

$$g(x) = \frac{\alpha}{1+\alpha+\sqrt{\alpha}x}\frac{1}{\pi}\sqrt{1-\frac{x^2}{4}}. \tag{3.6.7}$$

3. If $d_1\to\infty, d_1=o(n^{\varepsilon})$ for any $\varepsilon>0$ and $\frac{d_1}{d_2}\to\infty$, the ESD of $\frac{A}{\sqrt{(d_1-1)(d_2-1)}}$ converges to the semicircle law in probability.

Proof of Corollary 3.6.10. Claim (1) is proved in Theorem 6.4 of [79] based on a result for deterministic regular hypergraphs in Theorem 5 of [91].

Claim (2) is a combination of several results. When $d_1=o(n^{1/2})$, it is proved in Theorem 6.6 of [79] based on the global law for random biregular bipartite graphs in [75] and [169]. When $d_1=\omega(\log^4 n)$, the optimal local law for RBBGs was recently proved in [175], which also implies the global law for RBBGs. When $d_1 \leq \frac{n}{32}$ and $\frac{d_1}{d_2} \to \alpha$, from Lemma 3.6.7,

$$\mathbb{P}\left(G \in \mathcal{G}'(n,m,d_1,d_2)\right) \to 1.$$

Therefore by the same proof of Theorem 6.6 in [79], the ESD of $\frac{A-(d_2-2)}{\sqrt{(d_1-1)(d_2-1)}}$ for random regular hypergraphs converges in probability.

Under the assumptions $d_1 \to \infty, d_1 = o(n^{\varepsilon})$ for any $\varepsilon > 0$ and $\frac{d_1}{d_2} \to \infty$, from Lemma 3.6.7, we have again $\mathbb{P}(G \in \mathcal{G}'(n,m,d_1,d_2)) \to 1$. Then Claim (3) follows from Theorem 3.5.3 and Lemma 3.6.8. $\square$

Remark 3.6.11. The ESD in Corollary 3.6.10 (2) is a shifted and scaled Marčenko-Pastur law. Taking $\alpha \to \infty$, $g(x)$ converges to the density function of the semicircle law. The transition from Marčenko-Pastur law to the semicircle law was also proved for sample covariance matrices in [26] when the aspect ratio goes to infinity.

Remark 3.6.12. A semicircle law for the adjacency matrix of d_2-uniform Erdős-Rényi random hypergraphs with growing expected degrees was proved in Theorem 5 of [137] when d_2 is a constant. Part (3) of Corollary 3.6.10 proves a corresponding semicircle law for random d_2-uniform d_1-regular hypergraphs where d_2 can be a parameter depending on n.

Spectral gaps

The spectral gap for random regular hypergraphs with fixed d_1, d_2 was studied in [79]. Here we include the results for the case when d_1, d_2 are growing with n.

Corollary 3.6.13. Let H be a random (d_1,d_2)-regular hypergraph with $d_1 \geq d_2$. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of A. Let $\lambda = \max_{2 \leq i \leq n} |\lambda_i|$.

1. Suppose $d_1 \geq d_2 \geq 3$ is fixed. There exists a sequence $\varepsilon_n \to 0$ such that

$$\mathbb{P}(|\lambda - (d_2 - 2)| \geq 2\sqrt{(d_1-1)(d_2-1)} + \varepsilon_n) \to 0$$

as $n \to \infty$.

2. Suppose $3 \leq d_2 \leq \frac{1}{2}n^{2/3}$, $d_1 \geq d_2 \geq cd_1$ for some $c \in (0,1)$. Then for some constant $K > 0$ depending on c, for all $n \geq 1$,

$$\mathbb{P}\left(\lambda \geq K\sqrt{(d_1-1)(d_2-1)}\right) = O\left(\frac{1}{n}\right).$$

3. Suppose $3 \leq d_2 \leq C_1$ for a constant C_1, and $d_1 = o(n^{1/2})$. There exists a constant C depending on C_1 such that

$$\mathbb{P}\left(\lambda \geq C\sqrt{(d_1-1)(d_2-1)}\right) = O\left(\frac{d_1^2}{n^2}\right).$$

Proof. Claim (1) is proved in Theorem 4.3 in [79]. Claim (2) and (3) follow from part (2) and (3) in Theorem 3.3.1 with Lemma 3.6.7. □

Remark 3.6.14. Results in [179] that Claim (2) and (3) are based on have stronger probability estimates. However, Lemma 3.6.7 we used here yields a weaker failure probability.

Eigenvalue fluctuations

The following eigenvalue fluctuation results for random regular hypergraphs can be derived from Lemma 3.6.3, Lemma 3.6.8, and the eigenvalue fluctuations results for random biregular bipartite graphs in Section 3.4.

Corollary 3.6.15. For fixed $d_1 \geq d_2 \geq 3$, let H be a random (d_1, d_2)-regular hypergraph with adjacency matrix A. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\frac{A}{\sqrt{(d_1-1)(d_2-1)}}$. Suppose f is a function satisfying the same conditions in Theorem 3.4.4. Then $Y_f^{(n)} := \sum_{i=1}^n f(\lambda_i) - na_0$ converges in distribution as $n \to \infty$ to the infinitely divisible random variable

$$Y_f := \sum_{k=2}^{\infty} \frac{a_k}{[(d_1-1)(d_2-1)]^{k/2}} \mathrm{CNBW}_k^{(\infty)},$$

where $\mathrm{CNBW}_k^{(\infty)}$ is defined in (3.2.37).

Proof. Let $\tilde{Y}_f^{(n)}$ be the corresponding random variable of $Y_f^{(n)}$ for the uniform random biregular bipartite graphs considered in Theorem 3.4.4. From the total variation distance bound in Lemma 3.6.8, we have

$$d_{\mathrm{TV}}(Y_f^{(n)}, \tilde{Y}_f^{(n)}) \leq d_{\mathrm{TV}}(\mu_n, \mu_n') \leq \left(\frac{nd_1}{d_2}\right)^2 \left(\frac{4ed_2}{n}\right)^{d_2} = o(1).$$

Therefore $\tilde{Y}_f^{(n)}$ and $Y_f^{(n)}$ converge in distribution to the same law. $\square$

Corollary 3.6.16. Let H be a random (d_1, d_2)-regular hypergraph with $d_1 d_2 \to \infty$ as $n \to \infty$ and $d_1 d_2 = n^{o(1)}$. Let $\lambda_1 \geq \cdots \geq \lambda_n$ be the eigenvalues of $\frac{A}{\sqrt{(d_1-1)(d_2-1)}}$. Let f be a function satisfying (3.4.20) and (3.4.22). Suppose one of the two assumptions holds:

1. there exists a constant $c \geq 1$ such that $1 \leq \frac{d_1}{d_2} \leq c$,

2. $3 \leq d_2 \leq c_1$ for a constant $c_1 \geq 3$.

Then the random variable

$$Y_f^{(n)} = \sum_{i=1}^n f(\lambda_i) - m_f^{(n)}$$

converges in law to a Gaussian random variable with mean zero and variance $\sigma_f = \sum_{k=2}^{\infty} 2ka_k^2$. Moreover, for any fixed t, consider the entire functions $g_1, \dots, g_t$ satisfying (3.4.20) and (3.4.22). The corresponding random vector $(Y_{g_1}^{(n)}, \dots, Y_{g_t}^{(n)})$ converges in distribution to a centered Gaussian random vector $(Z_{g_1},\dots,Z_{g_t})$ with covariance

$$\mathrm{Cov}(Z_{g_i},Z_{g_j}) = 2\sum_{k=2}^{\infty} k a_k(g_i) a_k(g_j)$$

for $1 \le i,j \le t$, where $a_k(g_i), a_k(g_j)$ are the k-th coefficients in the expansion (3.4.19) for g_i, g_j, respectively.

Proof. Recall Lemma 3.6.8 and our assumption $d_1 d_2 = n^{o(1)}$. Under Case (1), we have $d_2 \to \infty$ and

$$d_{\mathrm{TV}}(\mu_n,\mu_n') \le \left(\frac{nd_1}{d_2}\right)^2 \left(\frac{4ed_2}{n}\right)^{d_2} = O(n^2)(n^{(-1+o(1))d_2}) = o(1).$$

Under Case (2), we have

$$d_{\mathrm{TV}}(\mu_n,\mu_n') \le \left(\frac{nd_1}{d_2}\right)^2 \left(\frac{4ed_2}{n}\right)^{d_2} = O(n^2 d_1^2) \left(\frac{4ec_1}{n}\right)^3 = o(1).$$

Then with Lemma 3.6.8, in both cases $Y_f^{(n)}$ converges in distribution to the same limiting random variable defined in Theorem 3.4.7. The proof of the covariance part follows in the same way. □

3.7 Acknowledgment

Chapter 3 is extracted from the preprint "Ioana Dumitriu and Yizhe Zhu. Global eigenvalue fluctuations of random biregular bipartite graphs. *arXiv:2008.11760* (2020)". The thesis author is the co-author of this paper.

Chapter 4

Community detection in the sparse hypergraph block model

4.1 Introduction

Clustering is an important topic in network analysis, machine learning, and computer vision [111]. Many clustering algorithms are based on graphs, which represent pairwise relationships among data. Hypergraphs can be used to represent higher-order relationships among objects, including co-authorship and citation networks, and they have been shown empirically to have advantages over graphs [176]. Recently hypergraphs have been used as the data model in machine learning, including recommender system, image retrieval and bioinformatics [134, 6]. The stochastic block model (SBM) is a generative model for random graphs with community structures, which serves as a useful benchmark for clustering algorithms on graph data. It is natural to have an analogous model for random hypergraphs to model higher-order relations.

In this chapter, we consider a higher-order SBM called the hypergraph stochastic block model (HSBM). Before describing HSBMs, we first recall clustering on graph SBMs.

The Stochastic block model for graphs

In this section, we summarize the state-of-the-art results for graph SBM with two blocks of roughly equal size. Let Σ_n be the set of all pairs (G,σ), where $G = ([n],E)$ is a graph with vertex set $[n]$ and edge set E, $\sigma = (\sigma_1,\dots,\sigma_n) \in \{+1,-1\}^n$ are spins on $[n]$, i.e., each vertex $i \in [n]$ is assigned with a spin $\sigma_i \in \{-1,+1\}$. From this finite set Σ_n, one can generate a random element (G,σ) in two steps.

1. First generate i.i.d random variables $\sigma_i \in \{-1,+1\}$ equally likely for all $i \in [n]$.
2. Then given $\sigma = (\sigma_1,\dots,\sigma_n)$, we generate a random graph G where each edge $\{i, j\}$ is included independently with probability p if $\sigma_i = \sigma_j$ and with probability q if $\sigma_i \neq \sigma_j$.

The law of this pair (G,σ) will be denoted by $\mathcal{G}(n, p, q)$. In particular, we are interested in the model $\mathcal{G}(n, p_n, q_n)$ where p_n, q_n are parameters depending on n. We use the shorthand notation $\mathbb{P}_{\mathcal{G}_n}$ to emphasize that the integration is taken under the law $\mathcal{G}(n, p_n, q_n)$.

Imagine $C_1 = \{i : \sigma_i = +1\}$ and $C_2 = \{i : \sigma_i = -1\}$ as two communities in the graph G. Observing only G from a sample (G,σ) from the distribution $\mathcal{G}(n, p_n, q_n)$, the goal of community detection is to estimate the unknown vector σ up to a sign flip. Namely, we construct label estimators $\hat{\sigma}_i \in \{\pm 1\}$ for each i and consider the empirical overlap between $\hat{\sigma}$ and unknown σ defined by

$$ov_n(\hat{\sigma},\sigma) := \frac{1}{n}\sum_{i\in[n]} \sigma_i\hat{\sigma}_i. \tag{4.1.1}$$

We may ask the following questions about the estimation as n tends to infinity:

1. Exact recovery (strong consistency):

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{G}_n}(\{ov_n(\hat{\sigma},\sigma) = 1\} \cup \{ov_n(\hat{\sigma},\sigma) = -1\}) = 1.$$

2. Almost exact recovery (weak consistency): for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{G}_n}\left(\{|ov_n(\hat{\sigma},\sigma) - 1| > \varepsilon\} \cap \{|ov_n(\hat{\sigma},\sigma) + 1| > \varepsilon\}\right) = 0.$$

3. Detection: Find a partition which is correlated with the true partition. More precisely, there exists a constant $r > 0$ such that it satisfies the following: for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{G}_n}(\{|ov(\hat{\sigma},\sigma) - r| > \varepsilon\} \cap \{|ov(\hat{\sigma},\sigma) + r| > \varepsilon\}) = 0. \tag{4.1.2}$$

There are many works on these questions using different tools, we list some of them. A conjecture of [70] based on non-rigorous ideas from statistical physics predicts a threshold of detection in the SBM, which is called the Kesten-Stigum threshold. In particular, if $p_n = \frac{a}{n}$ and $q_n = \frac{b}{n}$ where a, b are positive constants independent of n, then the detection is possible if and only if $(a-b)^2 > 2(a+b)$. This conjecture was confirmed in [146, 148, 141, 40] where [148, 141, 40] provided efficient algorithms to achieve the threshold. Very recently, two alternative spectral algorithms were proposed based on distance matrices [163] and a graph powering method in [3], and they both achieved the detection threshold.

Suppose $p_n = \frac{a\log n}{n}, q_n = \frac{b\log n}{n}$ where a, b are constant independent of n. Then the exact recovery is possible if and only if $(\sqrt{a} - \sqrt{b})^2 > 2$, which was solved in [2, 107] with efficient algorithms achieving the threshold. Besides the phase transition behavior, various algorithms were proposed and analyzed in different regimes and more general settings beyond the 2-block SBM [47, 61, 104, 5, 125, 147, 65, 164, 38, 68], including spectral methods, semidefinite programming, belief-propagation, and approximate message-passing algorithms. We recommend [1] for further details.

Hypergraph stochastic block models

The hypergraph stochastic block model (HSBM) is a generalization of the SBM for graphs, which was first studied in [98], where the authors consider hypergraphs generated by the stochastic block models that are dense and uniform. A faithful representation of a hypergraph is its adjacency tensor (see Definition 4.2.1). However, most of the computations involving tensors are NP-hard [112]. Instead, they considered spectral algorithms for exact recovery using hypergraph Laplacians. Subsequently, they extended their results to sparse, non-uniform hypergraphs [99, 100, 101]. For exact recovery, it was shown that the phase transition occurs in the regime of logarithmic average degrees in [133, 62, 61] and the exact threshold was given in [124], by a generalization of the techniques in [2]. Almost exact recovery for HSBMs was studied in [61, 62, 101].

For detection of the HSBM with two blocks, the authors of [20] proposed a conjecture that the phase transition occurs in the regime of constant average degree, based on the performance of the belief-propagation algorithm. Also, they conjectured a spectral algorithm based on non-backtracking operators on hypergraphs could reach the threshold. In [92], the authors showed an algorithm for detection when the average degree is bigger than some constant by reducing it to a bipartite stochastic block model. They also mentioned a barrier to further improvement. We confirm the positive part of the conjecture in [20] for the case of two blocks: above the threshold, there is a spectral algorithm which asymptotically almost surely constructs a partition of the hypergraph correlated with the true partition.

Now we specify our d-uniform hypergraph stochastic block model with two clusters. Analogous to $\mathcal{G}(n,p_n,q_n)$, we define $\mathcal{H}(n,d,p_n,q_n)$ for d-uniform hypergraphs. Let Σ_n be the set of all pair (H,σ), where $H=([n],E)$ is a d-uniform hypergraph with vertex set $[n]$ and hyperedge set E, $\sigma=(\sigma_1,\dots,\sigma_n)\in\{+1,-1\}^n$ are the spins on $[n]$. From this finite set Σ_n, one can generate a random element (H,σ) in two steps.

1. First generate i.i.d random variables $\sigma_i\in\{-1,+1\}$ equally likely for all $i\in[n]$.

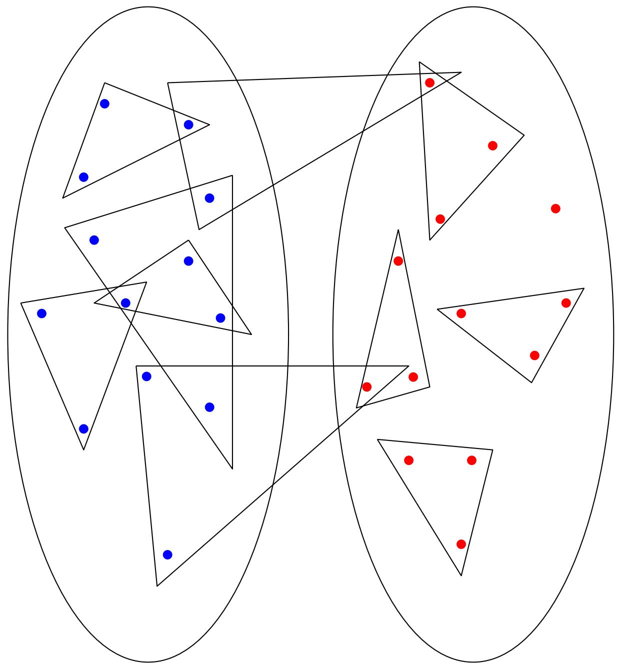

Figure 4.1: An HSBM with $d = 3$. Vertices in blue and red have spin $+$ and $-$, respectively.

2. Then given $\sigma = (\sigma_1, \ldots, \sigma_n)$, we generate a random hypergraph H where each hyperedge $\{i_1, \ldots i_d\}$ is included independently with probability p_n if $\sigma_{i_1} = \cdots = \sigma_{i_d}$ and with probability q_n if the spins $\sigma_{i_1}, \ldots \sigma_{i_d}$ are not the same.

The law of this pair (H, σ) will be denoted by $\mathcal{H}(n, d, p_n, q_n)$. We use the shorthand notation $\mathbb{P}_{\mathcal{H}_n}$ and $\mathbb{E}_{\mathcal{H}_n}$ to emphasize that integration is taken under the law $\mathcal{H}(n, d, p_n, q_n)$. Often we drop the index n from our notation, but it will be clear from $\mathbb{P}_{\mathcal{H}_n}$.

We consider the detection problem of the HSBM in the constant expected degree regime. Let

$$p_n := \frac{a}{\binom{n}{d-1}}, \quad q_n := \frac{b}{\binom{n}{d-1}}$$

for some constants $a \geq b > 0$ and a constant integer $d \geq 3$. Let

$$\alpha := (d-1)\frac{a + (2^{d-1} - 1)b}{2^{d-1}}, \quad \beta := (d-1)\frac{a-b}{2^{d-1}}. \tag{4.1.3}$$

Here α is a constant which measures the expected degree of any vertex, and β measures the discrepancy between the number of neighbors with $+$ sign and $-$ sign of any vertex. For $d=2$, α,β are the same parameters for the graph case in [141]. Now we are able to state our main result which is an extension of the result of for graph SBMs in [141]. Note that with the definition of α,β, we have $\alpha>\beta$. The condition $\beta^2>\alpha$ in the statement of Theorem (2.3.2) below implies $\alpha,\beta>1$, which will be assumed for the rest of the chapter.

Theorem 4.1.1. *Assume $\beta^2>\alpha$. Let (H,σ) be a random labeled hypergraph sampled from $\mathcal{H}(n,d,p_n,q_n)$ and $B^{(l)}$ be its l-th self-avoiding matrix (see Definition 4.2.5 below). Set $l=c\log(n)$ for a constant c such that $c\log(\alpha)<1/8$. Let x be a l_2-normalized eigenvector corresponding to the second largest eigenvalue of $B^{(l)}$. There exists a constant t such that, if we define the label estimator $\hat{\sigma}_i$ as*

$$\hat{\sigma}_i=\begin{cases}+1 & \textit{if } x_i\geq t/\sqrt{n},\\ -1 & \textit{otherwise},\end{cases}$$

then detection is possible. More precisely, there exists a constant $r>0$ such that the empirical overlap between $\hat{\sigma}$ and σ defined similar to (4.1.1) satisfies the following: for any $\varepsilon>0$,

$$\lim_{n\to\infty}\mathbb{P}_{\mathcal{H}_n}\left(\{|ov_n(\hat{\sigma},\sigma)-r|>\varepsilon\}\bigcap\{|ov_n(\hat{\sigma},\sigma)+r|>\varepsilon\}\right)=0.$$

Remark 4.1.2. If we take $d=2$, the condition $\beta^2>\alpha$ is the threshold for detection in graph SBMs proved in [141, 146, 148]. When $d\geq 3$, the conjectured detection threshold for HSBMs is given in Equation (48) of [20]. With our notations, in the 2-block case, Equation (48) in [20] can be written as $\frac{\alpha-\beta}{\alpha+\beta}=\frac{\sqrt{\alpha}-1}{\sqrt{\alpha}+1}$, which says $\beta^2=\alpha$ is the conjectured detection threshold for HSBMs. This is an analog of the Kesten-Stigum threshold proved in the graph case [70, 146, 148, 141, 40]. Our Theorem 2.3.2 proves the positive part of the conjecture.

Our algorithm can be summarized in two steps. The first step is a dimension reduction:

$B^{(l)}$ has n^2 many entries from the original adjacency tensor T (see Definition 4.2.1) of n^d many entries. Since the l-neighborhood of any vertex contains at most one cycle with high probability (see Lemma 4.4.4), by breadth-first search, the matrix $B^{(l)}$ can be constructed in polynomial time. The second step is a simple spectral clustering according to leading eigenvectors as the common clustering algorithm in the graph case.

Unlike graph SBMs, in the HSBMs, the random hypergraph H we observe is essentially a random tensor. Getting the spectral information of a tensor is NP-hard [112] in general, making the corresponding problems in HSBMs very different from graph SBMs. It is not immediately clear which operator to associate to H that encodes the community structure in the bounded expected degree regime. The novelty of our method is a way to project the random tensor into matrix forms (the self-avoiding matrix $B^{(l)}$ and the adjacency matrix A) that give us the community structure from their leading eigenvectors. In practice, the hypergraphs we observed are usually not d-uniform, which can not be represented as a tensor. However, we can still construct the matrix $B^{(l)}$ since the definition of self-avoiding walks does not depend on the uniformity assumption. In this chapter, we focus on the d-uniform case to simplify the presentation, but our proof techniques can be applied to the non-uniform case.

The analysis of HSBMs is harder than the original graph SBMs due to the extra dependency in the hypergraph structure and the lack of linear algebra tools for tensors. To overcome these difficulties, new techniques are developed in this chapter to establish the desired results.

There are multiple ways to define self-avoiding walks on hypergraphs, and our definition (see Definition 4.2.3) is the only one that works for us when applying the moment method. We develop a moment method suitable for sparse random hypergraphs in Section 4.7 that controls the spectral norms by counting concatenations of self-avoiding walks on hypergraphs. The combinatorial counting argument in the proof of Lemma 4.7.1 is more involved as we need to consider labeled vertices and labeled hyperedges. The moment method for hypergraphs developed here could be of independent interest for other random hypergraph problems.

The growth control of the size of the local neighborhood (Section 4.4) for HSBMs turns out to be more challenging compared to graph SBMs in [141] due to the dependency between the number of vertices with spin $+$ and $-$, and overlaps between different hyperedges. We use a new second-moment estimate to obtain a matching lower bound and upper bound for the size of the neighborhoods in the proof of Theorem 4.8.4. The issues mentioned above do not appear in the sparse random graph case.

To analyze the local structure of HSBMs, we prove a new coupling result between a typical neighborhood of a vertex in the sparse random hypergraph H and a multi-type Galton-Watson hypertree described in Section 4.5, which is a stronger version of local weak convergence of sparse random hypergraphs (local weak convergence for hypergraphs was recently introduced in [71]). Compared to the classical 2-type Galton-Watson tree in the graph case, the vertex $\pm$ labels in a hyperedge is not assigned independently. We carefully designed the probability of different types of hyperedges that appear in the hypertree to match the local structure of the HSBM. Combining all the new ingredients, we obtain the weak Ramanujan property of $B^{(l)}$ for sparse HSBMs in Theorem 4.6.1 as a generalization of the results in [141]. We conclude the proof of our Theorem 4.1.1 in Section 4.6.

Our Theorem 4.1.1 deals with the positive part of the phase transition conjecture in [20]. To have a complete characterization of the phase transition, one needs to show an impossibility result when $\beta^2 < \alpha$. Namely, below this threshold, no algorithms (even with exponential running time) will solve the detection problem with high probability. For graph SBMs, the impossibility result was proved in [146] based on a reduction to the broadcasting problem on Galton-Watson trees analyzed in [88]. To answer the corresponding problem in the HSBMs, one needs to establish a similar information-theoretical lower bound for the broadcasting problem on hypertrees and relate the problem to the detection problem on HSBMs. To the best of our knowledge, even for the very first step, the broadcasting problem on hypertrees has not been studied yet. The multi-type Galton-Watson hypertrees described in Section 4.5 can be used as a model to study this type of problem on hypergraphs. We leave it as a future direction.

4.2 Preliminaries

Definition 4.2.1 (adjacency tensor)**.** Let $H = (V,E)$ be a d-uniform hypergraph with $V = [n]$. We define T to be the adjacency tensor of H such that for any set of vertices $\{i_1, i_2, \dots, i_d\}$,

$$T_{i_1,\dots,i_d} = \begin{cases} 1 & \text{if } \{i_1,\dots,i_d\} \in E, \\ 0 & \text{otherwise.} \end{cases}$$

We set $T_{\sigma(i_1),\sigma(i_2),\dots,\sigma(i_d)} = T_{i_1,\dots,i_d}$ for any permutation σ. We may write T_e in place of $T_{i_1,\dots,i_d}$ where $e = \{i_1,\dots,i_d\}$.

Definition 4.2.2 (adjacency matrix)**.** The adjacency matrix A of a d-uniform hypergraph $H = (V,E)$ with vertex set $[n]$ is a $n \times n$ symmetric matrix such that for any $i \neq j$, A_{ij} is the number of hyperedges in E which contains i, j and $A_{ii} = 0$ for $i \in [n]$. Equivalently, we have

$$A_{ij} = \begin{cases} \sum_{e:\{i,j\}\in e} T_e & \text{if } i \neq j, \\ 0 & \text{if } i = j. \end{cases}$$

Definition 4.2.3 (walk)**.** A *walk* of length l on a hypergraph H is a sequence $(i_0, e_1, i_1, \cdots, e_l, i_l)$ such that $i_{j-1} \neq i_j$ and $\{i_{j-1}, i_j\} \subset e_j$ for all $1 \leq j \leq l$. A walk is closed if $i_0 = i_l$ and we call it a *circuit*. A *self-avoiding walk* of length l is a walk $(i_0, e_1, i_1, \cdots, e_l, i_l)$ such that

1. $|\{i_0, i_1, \dots, i_l\}| = l + 1$.
2. Any consecutive hyperedges e_{j-1}, e_j satisfy $e_{j-1} \cap e_j = \{i_{j-1}\}$ for $2 \leq j \leq l$.
3. Any two hyperedges e_j, e_k with $1 \leq j < k \leq l, k \neq j+1$ satisfy $e_j \cap e_k = \emptyset$.

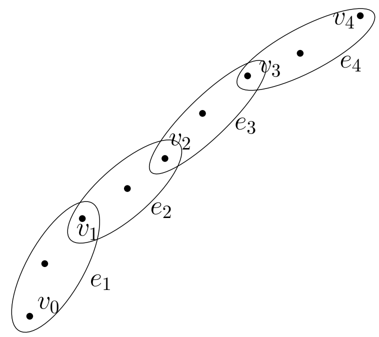

Figure 4.2: a self-avoiding walk of length 4 denoted by $(v_0, e_1, v_1, e_2, v_2, e_3, v_3, e_4, v_4)$

See Figure 4.2 for an example of a self-avoiding walk in a 3-uniform hypergraph. Recall that a self-avoiding walk of length l on a graph is a walk $(i_0, \ldots, i_l)$ without repeated vertices. Our definition is a generalization of the self-avoiding walk to hypergraphs.

Definition 4.2.4 (cycle and hypertree)**.** A *cycle* of length l with $l \geq 2$ in a hypergraph H is a walk $(i_0, e_1, \ldots, i_{l-1}, e_l, i_0)$ such that $i_0, \ldots i_{l-1}$ are distinct vertices and $e_1 \ldots e_l$ are distinct hyperedges. A *hypertree* is a hypergraph which contains no cycles.

Let $\binom{[n]}{d}$ be the collection of all subsets of $[n]$ with size d. For any subset $e \in \binom{[n]}{d}$ and $i \neq j \in [n]$, we define

$$A_{ij}^e = \begin{cases} 1 & \text{if } \{i, j\} \in e \text{ and } e \in E, \\ 0 & \text{otherwise,} \end{cases}$$

and we define $A_{ii}^e = 0$ for all $i \in [n]$. With our notation above, $A_{ij} = \sum_{e \in \binom{[n]}{d}} A_{ij}^e$. We have the following expansion of the trace of A^k for any integer $k \geq 0$:

$$\operatorname{tr} A^k = \sum_{i_0, i_2, \ldots, i_{k-1} \in [n]} A_{i_0 i_1} A_{i_2 i_3} \cdots A_{i_{k-1} i_0} = \sum_{\substack{i_0, i_1, \ldots, i_{k-1} \in [n] \\ e_1, \ldots, e_k \in \binom{[n]}{d}}} A_{i_0 i_1}^{e_1} \cdots A_{i_{k-2} i_{k-1}}^{e_{k-1}} A_{i_{k-1} i_0}^{e_k}.$$

Therefore, $\operatorname{tr} A^k$ counts the number of circuits $(i_0, e_1, i_1, \ldots, i_{k-1}, e_k, i_0)$ in the hypergraph H of length k. This connection was used in [137] to study the spectra of the Laplacian of random hypergraphs. From our definition of self-avoiding walks on hypergraphs, we associate a self-avoiding adjacency matrix to the hypergraph.

Definition 4.2.5 (self-avoiding matrix)**.** Let $H=(V,E)$ be a hypergraph with $V=[n]$. For any $l\geq 1$, a l-th *self-avoiding matrix* $B^{(l)}$ is a $n\times n$ matrix where for $i\neq j\in[n]$, $B^{(l)}_{ij}$ counts the number of self avoiding walks of length l from i to j and $B^{(l)}_{ii}=0$ for $i\in[n]$.

$B^{(l)}$ is a symmetric matrix since a time-reversing self avoiding walk from i to j is a self avoiding walk from j to i. Let SAW_{ij} be the set of all self-avoiding walks of length l connecting i and j in the complete d-uniform hypergraph on vertex set $[n]$. We denote a walk of length l by $w=(i_0,e_{i_1},\dots,i_{l-1},e_{i_l},i_l)$. Then for any $i,j\in[n]$,

$$B^{(l)}_{ij}=\sum_{w\in\mathrm{SAW}_{ij}}\prod_{t=1}^{l}A^{e_{i_t}}_{i_{t-1}i_t}. \tag{4.2.1}$$

4.3 Matrix expansion and spectral norm bounds

Consider a random labeled d-uniform hypergraph H sampled from $\mathcal{H}(n,d,p_n,q_n)$ with adjacency matrix A and self-avoiding matrix $B^{(l)}$. Let $\overline{A}:=\mathbb{E}_{\mathcal{H}_n}[A\mid\sigma]$. Let

$$\rho(A):=\sup_{x:\|x\|_2=1}\|Ax\|_2$$

be the spectral norm of a matrix A Recall (4.2.1), define

$$\Delta^{(l)}_{ij}:=\sum_{w\in\mathrm{SAW}_{ij}}\prod_{t=1}^{l}(A^{e_{i_t}}_{i_{t-1}i_t}-\overline{A}^{e_{i_t}}_{i_{t-1}i_t}), \tag{4.3.1}$$

where $\overline{A}^{e_{i_t}}_{i_{t-1}i_t}=\mathbb{E}_{\mathcal{H}_n}[A^{e_{i_t}}_{i_{t-1}i_t}\mid\sigma]$. $\Delta^{(l)}$ can be regarded as a centered version of $B^{(l)}$. We will apply the classical moment method to estimate the spectral norm of $\Delta^{(l)}$, since this method works well for centered random variables. Then we can relate the spectrum of $\Delta^{(l)}$ to the spectrum of $B^{(l)}$

through a matrix expansion formula which connects $\overline{A}$, $B^{(l)}$ and $\Delta^{(l)}$ in the following theorem. Recall the definition of α in (4.1.3).

Theorem 4.3.1. *Let H be a random hypergraph sampled from $\mathcal{H}(n,d,p_n,q_n)$ and $B^{(l)}$ be its l-th self avoiding matrix. Then the following holds.*

1. *There exist some matrices $\{\Gamma^{(l,m)}\}_{m=1}^{l}$ such that for any $l \geq 1$, $B^{(l)}$ satisfies the identity*

$$B^{(l)} = \Delta^{(l)} + \sum_{m=1}^{l} (\Delta^{(l-m)}\overline{A}B^{(m-1)}) - \sum_{m=1}^{l} \Gamma^{(l,m)}. \tag{4.3.2}$$

2. *For any sequence $l_n = O(\log n)$ and any fixed $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\rho(\Delta^{(l_n)}) \leq n^{\varepsilon}\alpha^{l_n/2}\right) = 1, \tag{4.3.3}$$

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\bigcap_{m=1}^{l_n}\left\{\rho(\Gamma^{(l_n,m)}) \leq n^{\varepsilon-1}\alpha^{(l_n+m)/2}\right\}\right) = 1. \tag{4.3.4}$$

Theorem 4.3.1 is one of the main ingredients to show $B^{(l)}$ has a spectral gap. Together with the local analysis in Section 4.4, we will show in Theorem 4.6.1 that the bulk eigenvalues of $B^{(l)}$ are separated from the first and second eigenvalues. The proof of Theorem 4.3.1 is deferred to Section 4.7. The matrices $\{\Gamma^{(l,m)}\}_{m=1}^{l}$ in Theorem 4.3.1 record concatenations of self-avoiding walks with different weights, which will be carefully analyzed in Lemma 4.7.2 of Section 4.7.

4.4 Local analysis

In this section, we study the structure of the local neighborhoods in the HSBM. Namely, what the neighborhood of a typical vertex in the random hypergraph looks like.

Definition 4.4.1. In a hypergraph H, we define the *distance* $d(i,j)$ between two vertices i,j to be the minimal length of walks between i and j. Define the *t-neighborhood* $V_t(i)$ of a fixed vertex i

to be the set of vertices which have distance t from i. Define $V_{\leq t}(i) := \bigcup_{k\leq t} V_k(i)$ to be the set all of vertices which have distance at most t from i and $V_{>t} = [n] \setminus V_{\leq t}$. Let $V_t^{\pm}(i)$ be the vertices in $V_t(i)$ with spin $\pm$ and define it similarly for $V_{\leq t}^{\pm}(i)$.

For $i \in [n]$, define

$$S_t(i) := |V_t(i)|, \quad D_t(i) := \sum_{j:d(i,j)=t} \sigma_j.$$

Let $\mathbf{1} = (1\ldots,1)\in\mathbb{R}^n$ and recall $\sigma \in \{-1,1\}^n$. We will show that when $l = c\log n$ with $c\log\alpha < 1/8$, $S_l(i), D_l(i)$ are close to the corresponding quantities $(B^{(l)}\mathbf{1})_i, (B^{(l)}\sigma)_i$ (see Lemma 4.11.1). In particular, the vector $(D_l(i))_{1\leq i\leq n}$ is asymptotically aligned with the second eigenvector of $B^{(l)}$, from which we get the information on the partitions. We give the following growth estimates of $S_t(i)$ and $D_t(i)$. The proof of Theorem 4.4.2 is given in Section 4.8.

Theorem 4.4.2. *Assume* $\beta^2 > \alpha > 1$ *and* $l = c\log n$*, for a constant* c *such that* $c\log\alpha < 1/4$*. There exists constants* $C, \gamma > 0$ *such that for sufficiently large n, with probability at least* $1 - O(n^{-\gamma})$ *the following holds for all* $i \in [n]$ *and* $1 \leq t \leq l$*:*

$$S_t(i) \leq C\log(n)\alpha^t, \tag{4.4.1}$$

$$|D_t(i)| \leq C\log(n)\beta^t, \tag{4.4.2}$$

$$S_t(i) = \alpha^{t-l}S_l(i) + O(\log(n)\alpha^{t/2}), \tag{4.4.3}$$

$$D_t(i) = \beta^{t-l}D_l(i) + O(\log(n)\alpha^{t/2}). \tag{4.4.4}$$

The approximate independence of neighborhoods of distinct vertices is given in the following lemma. It will be used later to analyze the martingales constructed on the Galton-Watson hypertree defined in Section 4.5. The proof of Lemma 4.4.3 is given in Appendix 4.12.

Lemma 4.4.3. *For any two fixed vertices $i \neq j$, let $l = c\log(n)$ with constant $c\log(\alpha) < 1/4$. Then the total variation distance between the joint law $\mathcal{L}((U_k^{\pm}(i))_{k\leq l}, (U_k^{\pm}(j))_{k\leq l})$ and the law with the same marginals and independence between them, denoted by $\mathcal{L}((U_k^{\pm}(i))_{k\leq l} \otimes (U_k^{\pm}(j))_{k\leq l})$, is $O(n^{-\gamma})$ for some $\gamma > 0$.*

Now we consider number of cycles in $V_{\leq l}(i)$ of any vertex $i \in [n]$. We say H is *l-tangle-free* if for any $i \in [n]$, there is no more than one cycle in $V_{\leq l}(i)$.

Lemma 4.4.4. *Assume $l = c\log n$ with $c\log(\alpha) < 1/4$. Let $(H, \sigma) \sim \mathcal{H}(n, d, p_n, q_n)$. Then*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(|\{i \in [n] : V_{\leq l}(i) \text{ contains at least one cycle}\}| \leq \log^4(n)\alpha^{2l}\right) = 1,$$
$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}(H \text{ is } l\text{-tangle-free}) = 1.$$

The proof of Lemma 4.4.4 is given in Appendix 4.12. In the next lemma, we translate the local analysis of the neighborhoods to the control of vectors $B^{(m)}\mathbf{1}, B^{(m)}\sigma$. The proof is similar to the proof of Lemma 4.3 in [141], and we include it in Appendix 4.12. For any event A_n, we say A_n happens *asymptotically almost surely* if $\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}(A_n) = 1$.

Lemma 4.4.5. *Let $\mathcal{B}$ be the set of vertices i whose $l-$neighborhood contains a cycle. For $l = c\log n$ with $c\log(\alpha) < 1/4$, asymptotically almost surely the following holds:*

1. *for all $m \leq l$ and all $i \notin \mathcal{B}$ the following holds*

$$(B^{(m-1)}\mathbf{1})_i = \alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{(m-1)/2}\log n), \tag{4.4.5}$$
$$(B^{(m-1)}\sigma)_i = \beta^{m-1-l}(B^{(l)}\sigma)_i + O(\alpha^{(m-1)/2}\log n). \tag{4.4.6}$$

2. *For all $i \in \mathcal{B}$:*

$$|(B^{(m)}\sigma)_i| \leq |(B^{(m)}\mathbf{1})_i| \leq 2\sum_{t=0}^{m} S_t(i) = O(\alpha^m \log n). \tag{4.4.7}$$

Combining Theorem 4.3.1, Theorem 4.4.2, and Lemma 4.4.5, we are able to prove the following theorem.

Theorem 4.4.6. *Assume* $\beta^2 > \alpha > 1$ *and* $l = c \log n$ *with* $c \log(\alpha) < 1/8$. *Then the following holds: for any* $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\sup_{\|x\|_2=1, x^\top(B^{(l)}\mathbf{1})=x^\top(B^{(l)}\sigma)=0} \|B^{(l)}x\|_2 \le n^{\varepsilon}\alpha^{l/2} \right) = 1.$$

Theorem 4.4.6 is a key ingredient to prove the bulk eigenvalues of $B^{(l)}$ are $O(n^{\varepsilon}\alpha^{l/2})$ in Theorem 4.6.1. The proof of Theorem 4.4.6 is given in in Section 4.9.

4.5 Coupling with multi-type Poisson hypertrees

Recall the definition of a hypertree from Definition 4.2.4. We construct a hypertree growth process in the following way. The hypertree is designed to obtain a coupling with the local neighborhoods of the random hypergraph H.

- Generate a root ρ with spin $\tau(\rho) = +$, then generate Pois $\left(\frac{\alpha}{d-1}\right)$ many hyperedges that only intersects at ρ. Call the vertices in these hyperedges except ρ to be the *children* of ρ and of generation 1. Call ρ to be their *parent*.
- For $0 \le r \le d-1$, we define a hyperedge is of type r if r many children in the hyperedge has spin $\tau(\rho)$ and $(d-1-r)$ many children has spin $-\tau(\rho)$. We first assign a type for each hyperedge independently. Each hyperedge will be of type $(d-1)$ with probability $\frac{(d-1)a}{\alpha 2^{d-1}}$ and of type r with probability $\frac{(d-1)b\binom{d-1}{r}}{\alpha 2^{d-1}}$ for $0 \le r \le d-2$. Since $\frac{(d-1)a}{\alpha 2^{d-1}} + \sum_{r=0}^{d-2} \frac{(d-1)b\binom{d-1}{r}}{\alpha 2^{d-1}} = 1$, the probabilities of being various types of hyperedges add up to 1. Because the type is chosen i.i.d for each hyperedge, by Poisson thinning, the number of hyperedges of different types are independent and Poisson.

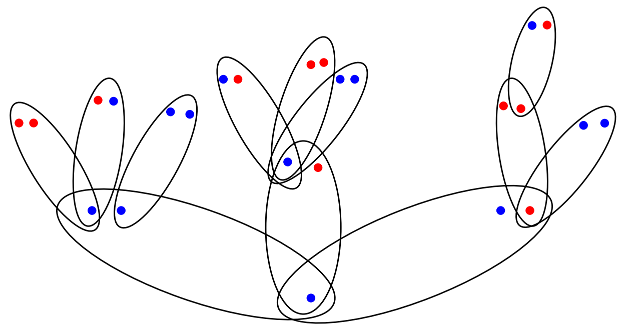

Figure 4.3: A Galton-Watson hypertree with $d = 3$. The vertices with spin $+$ are in blue and vertices with spin $-$ are in red.

- We draw the hypertree in a plane and label each child from left to right. For each type r hyperedge, we uniformly randomly pick r vertices among $d-1$ vertices in the first generation to put spins $\tau(\rho)$, and the rest $d-1-r$ many vertices are assigned with spins $-\tau(\rho)$.

- After defining the first generation, we keep constructing subsequent generations by induction. For each children v with spin $\tau(v)$ in the previous generation, we generate Pois $\left(\frac{\alpha}{d-1}\right)$ many hyperedges that pairwise intersects at v and assign a type to each hyperedge by the same rule with $\tau(\rho)$ replaced by $\tau(v)$. We call such random hypergraphs with spins a *multi-type Galton-Watson hypertree*, denoted by (T, ρ, τ) (see Figure 4.3).

Let $W_t^{\pm}$ be the number of vertices with $\pm$ spins at the t-th generation and $W_t^{(r)}$ be the number of hyperedges which contains exactly r children with spin $+$ in the t-th generation. Let $\mathcal{G}_{t-1} := \sigma(W_k^{\pm}, 1 \le k \le t-1)$ be the σ-algebra generated by $W_k^{\pm}, 1 \le k \le t-1$. From our definition, $W_0^+ = 1, W_0^- = 0$ and $\{W_t^{(r)}\}_{0 \le r \le d-1}$ are independent conditioned on $\mathcal{G}_{t-1}$, and the conditioned laws of $W_t^{(r)}$ are given by

$$\mathcal{L}(W_t^{(d-1)}|\mathcal{G}_{t-1}) = \text{Pois}\left(\frac{a}{2^{d-1}}W_{t-1}^{+} + \frac{b}{2^{d-1}}W_{t-1}^{-}\right), \tag{4.5.1}$$

$$\mathcal{L}(W_t^{(0)}|\mathcal{G}_{t-1}) = \text{Pois}\left(\frac{a}{2^{d-1}}W_{t-1}^{-} + \frac{b}{2^{d-1}}W_{t-1}^{+}\right), \tag{4.5.2}$$

$$\mathcal{L}(W_t^{(r)}|\mathcal{G}_{t-1}) = \text{Pois}\left(\frac{b\binom{d-1}{r}}{2^{d-1}}(W_{t-1}^{-} + W_{t-1}^{+})\right), \quad 1 \le r \le d-2. \tag{4.5.3}$$

We also have

$$W_t^{+} = \sum_{r=0}^{d-1} rW_t^{(r)}, \quad W_t^{-} = \sum_{r=0}^{d-1}(d-1-r)W_t^{(r)}. \tag{4.5.4}$$

Definition 4.5.1. A *rooted hypergraph* is a hypergraph H with a distinguished vertex $i \in V(H)$, denoted by (H,i). We say two rooted hypergraphs (H,i) and (H',i') are *isomorphic* and if and only if there is a bijection $\phi : V(H) \to V(H')$ such that $\phi(i) = i'$ and $e \in E(H)$ if and only if $\phi(e) := \{\phi(j) : j \in e\} \in E(H')$.

Let (H,i,σ) be a rooted hypergraph with root i and each vertex j is given a spin $\sigma(j) \in \{-1,+1\}$. Let (H',i',σ') be a rooted hypergraph with root i' where for each vertex $j \in V(H')$, a spin $\sigma'(j) \in \{-1,+1\}$ is given. We say (H,i,σ) and (H',i',σ') are *spin-preserving isomorphic* and denoted by $(H,i,\sigma) \equiv (H',i',\sigma')$ if and only if there is an isomorphism $\phi : (H,i) \to (H',i')$ with $\sigma(v) = \sigma'(\phi(v))$ for each $v \in V(H)$.

Let $(H,i,\sigma)_t, (T,\rho,\tau)_t$ be the rooted hypergraphs $(H,i,\sigma), (T,\rho,\tau)$ truncated at distance t from i,ρ, respectively, and let $(T,\rho,-\tau)$ be the corresponding hypertree growth process where the root ρ has spin -1. We prove a local weak convergence of a typical neighborhood of a vertex in the hypergraph H to the hypertree process T we described above. In fact, we prove the following stronger statement. The proof of Theorem 4.5.2 is given in Section 4.5.

Theorem 4.5.2. *Let (H,σ) be a random hypergraph H with spin σ sampled from $\mathcal{H}_n$. Let $i \in [n]$*

be fixed with spin σ_i. Let $l = c\log(n)$ with $c\log(\alpha) < 1/4$, the following holds for sufficiently large n.

1. *If $\sigma_i = +1$, there exists a coupling between (H,i,σ) and (T,ρ,τ) such that $(H,i,\sigma)_l \equiv (T,\rho,\tau)_l$ with probability at least $1-n^{-1/5}$.*

2. *If $\sigma_i = -1$, there exists a coupling between (H,i,σ) and $(T,\rho,-\tau)$ such that $(H,i,\sigma)_l \equiv (T,\rho,-\tau)_l$ with probability at least $1-n^{-1/5}$.*

Now we construct two martingales from the Poisson hypertree growth process. Define two processes

$$M_t := \alpha^{-t}(W_t^+ + W_t^-), \quad \Delta_t := \beta^{-t}(W_t^+ - W_t^-).$$

Lemma 4.5.3. *The two processes $\{M_t\}, \{\Delta_t\}$ are $\mathcal{G}_t$-martingales. If $\beta^2 > \alpha > 1$, $\{M_t\}$ and $\{\Delta_t\}$ are uniformly integrable. The martingale $\{\Delta_t\}$ converges almost surely and in L^2 to a unit mean random variable Δ_∞. Moreover, Δ_∞ has a finite variance and*

$$\lim_{t\to\infty} \mathbb{E}|\Delta_t^2 - \Delta_\infty^2| = 0. \tag{4.5.5}$$

The following Lemma will be used in the proof of Theorem 4.1.1 to analyze the correlation between the estimator we construct and the correct labels of vertices based on the random variable Δ_∞. The proof is similar to the proof of Theorem 4.2 in [141], and we include it in Appendix 4.12.

Lemma 4.5.4. *Let $l = c\log n$ with $c\log\alpha < 1/8$. For any $\varepsilon > 0$,*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\left| \frac{1}{n}\sum_{i=1}^n \beta^{-2l} D_l^2(i) - \mathbb{E}[\Delta_\infty^2] \right| > \varepsilon \right) = 0. \tag{4.5.6}$$

Let $y^{(n)} \in \mathbb{R}^n$ be a random sequence of l_2-normalized vectors defined by

$$y_i^{(n)} := \frac{D_l(i)}{\sqrt{\sum_{j=1}^n D_l(j)^2}}, 1 \le i \le n.$$

Let $x^{(n)}$ be any sequence of random vectors in $\mathbb{R}^n$ such that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}(\|x^{(n)} - y^{(n)}\|_2 > \varepsilon) = 0.$$

For all $\tau \in \mathbb{R}$ that is a point of continuity of the distribution of both Δ_∞ and $-\Delta_\infty$, for any $\varepsilon > 0$, one has the following

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\left|\frac{1}{n}\sum_{i\in[n]:\sigma_i=+} \mathbf{1}\left\{x_i^{(n)} \ge \tau/\sqrt{n\mathbb{E}[\Delta_\infty^2]}\right\} - \frac{1}{2}\mathbb{P}(\Delta_\infty \ge \tau)\right| > \varepsilon\right) = 0, \tag{4.5.7}$$

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\left|\frac{1}{n}\sum_{i\in[n]:\sigma_i=-} \mathbf{1}\left\{x_i^{(n)} \ge \tau/\sqrt{n\mathbb{E}[\Delta_\infty^2]}\right\} - \frac{1}{2}\mathbb{P}(-\Delta_\infty \ge \tau)\right| > \varepsilon\right) = 0.$$

4.6 Proof of the main result

Let $\vec{S}_l := (S_l(1), \ldots, S_l(n))$ and $\vec{D}_l := (D_l(1), \ldots, D_l(n))$. We say the the sequence of vectors $\{v_n\}_{\ge 1}$ is *asymptotically aligned* with the sequence of vectors $\{w_n\}_{n\ge 1}$ if

$$\lim_{n\to\infty} \frac{|\langle v_n, w_n\rangle|}{\|v_n\|_2 \cdot \|w_n\|_2} = 1.$$

With all the ingredients in Sections 4.3-4.5, we establish the following weak Ramanujan property of $B^{(l)}$. The proof of Theorem 4.6.1 is given in Section 4.11.

Theorem 4.6.1. *For $l = c\log(n)$ with $c\log(\alpha) < 1/8$, asymptotically almost surely the two leading eigenvectors of $B^{(l)}$ are asymptotically aligned with vectors $\vec{S}_l, \vec{D}_l$, where the first eigenvalue is of order $\Theta(\alpha^l)$ up to some logarithmic factor and the second eigenvalue is of order $\Omega(\beta^l)$. All* *other eigenvalues are of order* $O(n^{\varepsilon}\alpha^{l/2})$ *for any* $\varepsilon > 0$.

Theorem 4.6.1 connects the leading eigenvectors of $B^{(l)}$ with the local structures of the random hypergraph H and shows that the bulk eigenvalues of $B^{(l)}$ are separated from the two top eigenvalues. Equipped with Theorem 4.6.1 and Lemma 4.5.4, we are ready to prove our main result.

Proof of Theorem 4.1.1. Let $x^{(n)}$ be the l_2-normalized second eigenvector of $B^{(l)}$, by Theorem 4.6.1, $x^{(n)}$ is asymptotically aligned with the l_2-normalized vector

$$y_i^{(n)} = \frac{D_l(i)}{\sqrt{\sum_{j=1}^{n} D_l(j)^2}}, 1 \le i \le n$$

asymptotically almost surely. So we have $\|x^{(n)} - y^{(n)}\|_2 \to 0$ or $\|x^{(n)} + y^{(n)}\|_2 \to 0$ asymptotically almost surely. We first assume $\|x^{(n)} - y^{(n)}\|_2 \to 0$. Since $\mathbb{E}\Delta_\infty = 1$, from the proof of Theorem 2.1 in [141], there exists a point $\tau \in \mathbb{R}$, in the set of continuity points of both Δ_∞ and $-\Delta_\infty$, that satisfies $r := \mathbb{P}(\Delta_\infty \ge \tau) - \mathbb{P}(-\Delta_\infty \ge \tau) > 0$. Take $t = \tau/\sqrt{\mathbb{E}(\Delta_\infty^2)}$ and let $\mathcal{N}^+, \mathcal{N}^-$ be the set of vertices with spin $+$ and $-$, respectively. From the definition of $\hat{\sigma}$, we have

$$\begin{aligned}\frac{1}{n}\sum_{i\in[n]} \sigma_i\hat{\sigma}_i =& \frac{1}{n}\sum_{i\in[n]} \sigma_i \left(\mathbf{1}_{\left\{x_i^{(n)} \ge t/\sqrt{n}\right\}} - \mathbf{1}_{\left\{x_i^{(n)} < t/\sqrt{n}\right\}}\right) \\ =& -\frac{1}{n}\sum_{i\in[n]}\sigma_i + \frac{2}{n}\sum_{i\in\mathcal{N}^+} \mathbf{1}_{\left\{x_i^{(n)} \ge \tau/\sqrt{n\mathbb{E}\Delta_\infty^2}\right\}} - \frac{2}{n}\sum_{i\in\mathcal{N}^-} \mathbf{1}_{\left\{x_i^{(n)} \ge \tau/\sqrt{n\mathbb{E}\Delta_\infty^2}\right\}}.\end{aligned} \tag{4.6.1}$$

Note that $\frac{1}{n}\sum_{i\in[n]}\sigma_i \to 0$ in probability by the law of large numbers. From (4.5.7) in Lemma 4.5.4, we have (4.6.1) converges in probability to $\mathbb{P}(\Delta_\infty \ge \tau) - \mathbb{P}(-\Delta_\infty \ge \tau) = r$. If $\|x^{(n)} + y^{(n)}\|_2 \to 0$, similarly we have $\frac{1}{n}\sum_{i\in[n]}\sigma_i\hat{\sigma}_i$ converges to $-r$ in probability. From these two cases, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{H}_n}\left(\{|ov_n(\hat{\sigma},\sigma) - r| > \varepsilon\} \bigcap \{|ov_n(\hat{\sigma},\sigma) + r| > \varepsilon\}\right) = 0.$$

This concludes the proof of Theorem 4.1.1. □

4.7 Proof of Theorem 4.3.1

Proof of (4.3.2) in Theorem 4.3.1

For ease of notation, we drop the index n from l_n in the proof, and it will be clear from the law $\mathcal{H}_n$. For any sequences of real numbers $\{a_t\}_{t=1}^l, \{b_t\}_{t=1}^l$, we have the following expansion identity for $l \geq 2$ (see for example, Equation (15) in [141] and Equation (27) in [40]):

$$\prod_{t=1}^{l}(a_t - b_t) = \prod_{t=1}^{l} a_t - \sum_{m=1}^{l}\left(\prod_{t=1}^{l-m}(a_t - b_t)\right) b_{l-m+1} \prod_{t=l-m+2}^{l} a_t.$$

Therefore the following identity holds.

$$\prod_{t=1}^{l}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) = \prod_{t=1}^{l} A_{i_{t-1}i_t}^{e_{i_t}} - \sum_{m=1}^{l}\left(\prod_{t=1}^{l-m}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})\right)\overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}}.$$

Summing over all $w \in \mathrm{SAW}_{ij}$, $\Delta_{ij}^{(l)}$ can be written as

$$B_{ij}^{(l)} - \sum_{m=1}^{l}\sum_{w\in \mathrm{SAW}_{ij}}\left(\prod_{t=1}^{l-m}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})\right)\overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}}. \tag{4.7.1}$$

Introduce the set Q_{ij}^m of walks w defined by concatenations of two self-avoiding walks w_1, w_2 such that w_1 is a self-avoiding walk of length $l-m$ from i to some vertex k, and w_2 is a self-avoiding walk of length m from k to j for all possible $1 \leq m \leq l$ and $k \in [n]$. Then $\mathrm{SAW}_{ij} \subset Q_{ij}^m$ for all $1 \leq m \leq l$. Let $R_{ij}^m = Q_{ij}^m \setminus \mathrm{SAW}_{ij}$. Define the matrix $\Gamma^{(l,m)}$ as

$$\Gamma_{ij}^{(l,m)} := \sum_{w\in R_{ij}^m}\prod_{t=1}^{l-m}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})\overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}}. \tag{4.7.2}$$

From (4.7.1), $\Delta_{ij}^{(l)}$ can be expanded as

$$B_{ij}^{(l)} - \sum_{m=1}^{l} \sum_{w \in Q_{ij}^m \setminus R_{ij}^m} \left(\prod_{t=1}^{l-m} (A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) \right) \overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}}.$$

It can be further written as

$$B_{ij}^{(l)} - \sum_{m=1}^{l} \sum_{w \in Q_{ij}^m} \prod_{t=1}^{l-m} (A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) \overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}} + \sum_{m=1}^{l} \Gamma_{ij}^{(l,m)}.$$

From the definition of matrix multiplication, we have

$$\begin{aligned} &\sum_{w \in Q_{ij}^m} \prod_{t=1}^{l-m} (A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}}) \overline{A}_{i_{l-m}i_{l-m+1}}^{e_{i_{l-m+1}}} \prod_{t=l-m+2}^{l} A_{i_{t-1}i_t}^{e_{i_t}} \\ = &\sum_{1 \le u,v \le n} \Delta_{iu}^{(l-m)} \overline{A}_{uv} B_{vj}^{(m-1)} = \left(\Delta^{(l-m)} \overline{A} B^{(m-1)} \right)_{ij}. \end{aligned} \tag{4.7.3}$$

Combining the expansion of $\Delta_{ij}^{(l)}$ above and (4.7.3), we obtain

$$\Delta_{ij}^{(l)} = B_{ij}^{(l)} - \sum_{m=1}^{l} (\Delta^{(l-m)} \overline{A} B^{(m-1)})_{ij} + \sum_{m=1}^{l} \Gamma_{ij}^{(l,m)}. \tag{4.7.4}$$

Since (4.7.4) is true for any $i, j \in [n]$, it implies (4.3.2).

Proof of (4.3.3) in Theorem 4.3.1

We first prove the following spectral norm bound on $\Delta^{(l)}$.

Lemma 4.7.1. *For $l = O(\log n)$ and fixed k, we have*

$$\mathbb{E}_{\mathcal{H}_n}[\rho(\Delta^{(l)})^{2k}] = O(n \alpha^{kl} \log^{6k} n). \tag{4.7.5}$$

Proof. Note that $\mathbb{E}_{\mathcal{H}_n}[\rho(\Delta^{(l)})^{2k}] \le \mathbb{E}_{\mathcal{H}_n}[\mathrm{tr}(\Delta^{(l)})^{2k}]$. The estimation is based on a coding argument, and we modify the proof in [141] to count circuits in hypergraphs. Let $W_{2k,l}$ be the set of all circuits of length $2kl$ in the complete hypergraph $K_{n,d}$ which are concatenations of $2k$ many self-avoiding walks of length l. For any circuits $w \in W_{2k,l}$, we denote it by $w = (i_0, e_{i_1}, i_1, \dots e_{i_{2kl}}, i_{2kl})$, with $i_{2kl} = i_0$. From (4.3.1), we have

$$\mathbb{E}_{\mathcal{H}_n}\left[\operatorname{tr}(\Delta^{(l)})^{2k}\right] = \sum_{j_1,\dots,j_{2k}\in[n]} \mathbb{E}_{\mathcal{H}_n}\left[\Delta^{(l)}_{j_1 j_2}\Delta^{(l)}_{j_2 j_3}\cdots\Delta^{(l)}_{j_{2k} j_1}\right] = \sum_{w\in W_{2k,l}} \mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A^{e_{i_t}}_{i_{t-1}i_t} - \overline{A}^{e_{i_t}}_{i_{t-1}i_t})\right]. \tag{4.7.6}$$

For each circuit, the weight it contributes to the sum is the product of $(A^e_{ij} - \overline{A^e_{ij}})$ over all the hyperedges e traversed in the circuits. In order to have an upper bound on $\mathbb{E}_{\mathcal{H}_n}[\operatorname{tr}(\Delta^{(l)})^{2k}]$, we need to estimate how many such circuits are included in the sum and what are the weights they contribute.

We also write $w = (w_1, w_2, \dots w_{2k})$, where each w_i is a self-avoiding walk of length l. Let v and h be the number of distinct vertices and hyperedges traversed by the circuit, respectively. The idea is to bound the number of all possible circuits w in (4.7.6) with given v and h, and then sum over all possible (v,h) pairs.

Fix v and h, for any circuit w we form a labeled multigraph $G(w)$ with labeled vertices $\{1,\dots,v\}$ and labeled multiple edges $\{e_1,\dots,e_h\}$ by the following rules:

- Label the vertices in $G(w)$ by the order they first appear in w, starting from 1. For any pair vertices $i,j \in [v]$, we add an edge between i,j in $G(w)$ whenever a hyperedge appears between the ith and jth distinct vertices in the circuit w. $G(w)$ is a multigraph since it is possible that for some i,j, there exists two distinct hyperedges connecting the ith and jth distinct vertices in w, which corresponds to two distinct edges in $G(w)$ connecting i,j.

- Label the edges in $G(w)$ by the order in which the corresponding hyperedge first appears in w from e_1 to e_h. Note that the number of edges in $G(w)$ is at least h since distinct edges in $G(w)$ can get the same hyperedge labels. At the end we obtain a multigraph

$G(w) = (V(w), E(w))$ with vertex set $\{1, \dots, v\}$ and edge set $E(w)$ with hyperedge labels in $\{e_1, \dots e_h\}$.

It is crucial to see that the labeling of vertices and edges in $G(w)$ is in order, and it tells us how the circuit w is traversed. Consider any edge in $G(w)$ such that its right endpoint (in the order of the traversal of w) is a new vertex that has not been traversed by w. We call it a *tree edge*. Denote by $T(w)$ the tree spanned by those edges. It is clear for the construction that $T(w)$ includes all vertices in $G(w)$, so $T(w)$ is a spanning tree of $G(w)$. Since the labels of vertices and edges are given in $G(w)$, $T(w)$ is uniquely defined. For example, in Figure 4.4, we have

$$w_1 = (1, e_1, 2, e_2, 3, e_3, 4, e_4, 5, e_5, 6),$$

$$w_2 = (6, e_5, 5, e_4, 4, e_6, 7, e_7, 8, e_8, 3),$$

$$w_3 = (3, e_2, 2, e_1, 1, e_9, 9, e_{10}, 10, e_{11}, 11),$$

$$w_4 = (11, e_{12}, 10, e_{10}, 9, e_{13}, 12, e_{14}, 13, e_{15}, 1).$$

Edges that are not included in $T(w)$ are $\{e_8, e_{12}, e_{15}\}$. The triplet sequences associated to the 4 self-avoiding walks $\{w_i\}_{i=1}^4$ are given by

$$(0,6,0); \quad (4,2,3),(0,0,0); \quad (1,3,0); \quad (0,0,10),(9,2,1),(0,0,0),$$

respectively.

For a given $w \in W_{2k,l}$ with distinct hyperedges $e_1, \dots, e_h$, define $\mathrm{end}(e_i)$ to be the set of vertices in $V(w)$ such that they are the endpoints of edges with label e_i in $G(w)$. For example, consider a hyperedge $e_1 = \{1,2,3,4\}$ such that $\{1,2\},\{1,3\}$ are all the edges in $G(w)$ with labels e_1, then $\mathrm{end}(e_1) = \{1,2,3\}$. We consider circuits w in three different cases and estimate their contribution to (4.7.6) separately.

Case (1). We first consider $w \in W_{2k,l}$ such that

- each hyperedge label in $\{e_i\}_{1\leq i\leq h}$ appears exactly once on the edges of $G(w)$;

- vertices in $e_i \setminus \mathrm{end}(e_i)$ are all distinct for $1 \leq i \leq h$, and they are not vertices with labels in $V(w)$.

The first condition implies the number of edges in $G(w)$ is h. The second condition implies that there are exactly $(d-2)h+v$ many distinct vertices in w. We will break each self-avoiding walk w_i into three types of successive sub-walks where each sub-walk is exactly one of the following 3 types, and we encode these sub-walks as follows.

- Type 1: hyperedges with corresponding edges in $G(w) \setminus T(w)$. Given our position in the circuit w, we can encode a hyperedge of this type by its right-end vertex. Hyperedges of Type 1 breaks the walk w_i into disjoint sub-walks, and we partition these sub-walks into Type 2 and 3 below.

- Type 2: sub-walks such that all their hyperedges correspond to edges of $T(w)$ and have been traversed already by $w_1, \dots, w_{i-1}$. Each sub-walk is a part of a self-avoiding walk, and it is a path contained in the tree $T(w)$. Given its initial and its end vertices, there will be exactly one such path in $T(w)$. Therefore these walks can be encoded by the end vertices.

- Type 3: sub-walks such that their hyperedges correspond to edges of $T(w)$ and they are being traversed for the first time. Given the initial vertex of a sub-walk of this type, since it is traversing new edges and knowing in what order the vertices are discovered, we can encode these walks by their length, and from the given length, we know at which vertex the sub-walk ends.

We encode any Type 1, Type 2, or Type 3 sub-walk by 0 if the sub-walk is empty. Now we can decompose each w_i into sequences characterizing by its sub-walks:

$$(p_1,q_1,r_1),(p_2,q_2,r_2),\dots,(p_t,q_t,r_t). \tag{4.7.7}$$

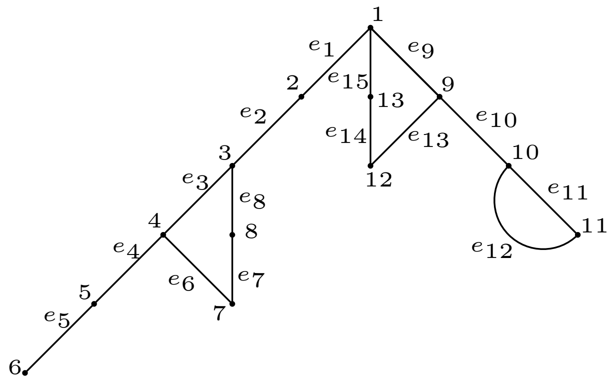

Figure 4.4: A multigraph $G(w)$ associated to a circuit $w = (w_1, \ldots, w_4)$ of length $2kl$ with $k = 2, l = 5$

Here $r_1, \ldots r_{t-1}$ are codes from sub-walks of Type 1. From the way we encode such hyperedges, we have $r_i \in \{1, \ldots v\}$ for $1 \leq i \leq t-1$. Type 2 and Type 3 sub-walks are encoded by $p_1, \ldots, p_t$ and $q_1, \ldots, q_t$, respectively. Since Type 1 hyperedges break w into disjoint pieces, we use (p_t, q_t, r_t) to represent the last piece of the sub-walk and make $r_t = 0$. Each p_i represents the right-end vertex of the Type 2 sub-walk, and $p_i = 0$ if it the sub-walk is empty, hence $p_i \in \{0, \ldots v\}$ for $1 \leq i \leq t$. Each q_i represents the length of Type 3 sub-walks, so $q_i \in \{0, \ldots l\}$ for $1 \leq i \leq t$. From the way we encode these sub-walks, there are at most $(v+1)^2(l+1)$ many possibilities for each triplet (p_j, q_j, r_j).

We now consider how many ways we can concatenate sub-walks encoded by the triplets to form a circuit w. All triples with $r_j \in [v]$ for $1 \leq j \leq t-1$ indicate the traversal of an edge not in $T(w)$. Since we know the number of edges in $G(w) \setminus T(w)$ is $(h - v + 1)$, and within a self-avoiding walk w_i, edges on $G(w)$ can be traversed at most once, the length of the triples in (4.7.7) satisfies $t - 1 \leq h - v + 1$, which implies $t \leq h - v + 2$. Since each hyperedge can be traversed at most $2k$ many times by w due to the constraint that the circuits w of length $2kl$ are formed by self-avoiding walks, so the number of triple sequences for fixed v, h is at most $[(v+1)^2(l+1)]^{2k(2+h-v)}$.

There are multiple w with the same code sequence. However, they must all have the same number of vertices and edges, and the positions where vertices and hyperedges are repeated must be the same. The number of ordered sequences of v distinct vertices is at most n^v. Given the vertex sequence, the number of ordered sequences of h distinct hyperedges in $K_{n,d}$ is at most $\binom{n}{d-2}^h$. Therefore, given v,h, the number of circuits that share the same triple sequence (4.7.7) is at most $n^v\binom{n}{d-2}^h$.

Combining the two estimates, the number of all possible circuits w with fixed v,h in Case (1) is at most

$$n^v\binom{n}{d-2}^h[(v+1)^2(l+1)]^{2k(2+h-v)}. \tag{4.7.8}$$

Now we consider the expected weight of each circuit in the sum (4.7.6). Given σ, if $i,j\in e$, we have $A_{ij}^e\sim\mathrm{Ber}\left(p_{\sigma(e)}\right)$, where $p_{\sigma(e)}=\frac{a}{\binom{n}{d-1}}$ if vertices in e have the same $\pm$ spins and $p_{\sigma(e)}=\frac{b}{\binom{n}{d-1}}$ otherwise. For a given hyperedge appearing in w with multiplicity $m\in\{1,\dots,2k\}$, the corresponding expectation $\mathbb{E}_{\mathcal{H}_n}\left[(A_{ij}^e-\overline{A_{ij}^e})^m\right]$ is 0 if $m=1$. Since $0\leq A_{ij}^e\leq 1$, for $m\geq 2$, we have

$$\mathbb{E}_{\mathcal{H}_n}\left[(A_{ij}^e-\overline{A_{ij}^e})^m\mid\sigma\right]\leq\mathbb{E}_{\mathcal{H}_n}\left[(A_{ij}^e-\overline{A_{ij}^e})^2\mid\sigma\right]\leq p_{\sigma(e)}. \tag{4.7.9}$$

For any hyperedge e corresponding to an edge in $G(w)\setminus T(w)$ we have the upper bound

$$p_{\sigma(e)}\leq\frac{a\vee b}{\binom{n}{d-1}}. \tag{4.7.10}$$

Taking the expectation over σ we have

$$\mathbb{E}_\sigma[p_{\sigma(e)}]=\frac{a+(2^{d-1}-1)b}{2^{d-1}\binom{n}{d-1}}=\frac{\alpha}{(d-1)\binom{n}{d-1}}. \tag{4.7.11}$$

Recall the weight of each circuit in the sum (4.7.6) is given by

$$\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A^{e_{i_t}}_{i_{t-1}i_t}-\overline{A}^{e_{i_t}}_{i_{t-1}i_t})\right].$$

Conditioned on σ, $(A^{e_{i_t}}_{i_{t-1}i_t}-\overline{A}^{e_{i_t}}_{i_{t-1}i_t})$ are independent random variables for distinct hyperedges. Denote these distinct hyperedges by $e_1,\dots e_h$ with multiplicity $m_1,\dots m_h$ and we temporarily order them such that $e_1,\dots e_{v-1}$ are the hyperedges corresponding to edges on $T(w)$. Introduce the random variables $A^{e_i}\sim \mathrm{Ber}\left(p_{\sigma(e_i)}\right)$ for $1\leq i\leq h$ and denote $\overline{A^{e_i}}=\mathbb{E}_{\mathcal{H}_n}[A^{e_i}\mid\sigma]$. Therefore from (4.7.9) we have

$$\begin{aligned}&\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A^{e_{i_t}}_{i_{t-1}i_t}-\overline{A}^{e_{i_t}}_{i_{t-1}i_t})\right]=\mathbb{E}_{\sigma}\left[\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A^{e_{i_t}}_{i_{t-1}i_t}-\overline{A}^{e_{i_t}}_{i_{t-1}i_t})\mid\sigma\right]\right]\\&=\mathbb{E}_{\sigma}\left[\prod_{i=1}^{h}E_{\mathcal{H}_n}\left[(A^{e_i}-\overline{A^{e_i}})^{m_i}\mid\sigma\right]\right]\leq\mathbb{E}_{\sigma}\left[\prod_{i=1}^{h}p_{\sigma(e_i)}\right].\end{aligned}$$

We use the bound (4.7.10) for $p_{\sigma(e_v)},\dots,p_{\sigma(e_h)}$, which implies

$$\mathbb{E}_{\sigma}\left[\prod_{i=1}^{h}p_{\sigma(e_i)}\right]\leq\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{h-v+1}\mathbb{E}_{\sigma}\left[\prod_{i=1}^{v-1}p_{\sigma(e_i)}\right]. \tag{4.7.12}$$

From the second condition for w in Case (1), any two hyperedges among $\{e_1,\dots e_{v-1}\}$ share at most 1 vertex, and $p_{\sigma(e_i)},p_{\sigma(e_j)}$ are pairwise independent for all $1\leq i<j\leq v-1$. Moreover, since the corresponding edges of $e_1,\dots e_{v-1}$ forms the spanning tree $T(w)$, taking any e_j such that the corresponding edge in $T(w)$ is attached to some leaf, we know e_j and $\bigcup_{i\neq j,1\leq i\leq v}e_i$ share exactly one common vertex, therefore $p_{\sigma(e_j)}$ is independent of $\prod_{1\leq i\leq v-1,i\neq j}p_{\sigma(e_i)}$. We then have

$$\mathbb{E}_{\sigma}\left[\prod_{i=1}^{v-1}p_{\sigma(e_i)}\right]=\mathbb{E}_{\sigma}[p_{\sigma(e_j)}]\cdot\mathbb{E}_{\sigma}\left[\prod_{1\leq i\leq v-1,i\neq j}p_{\sigma(e_i)}\right]. \tag{4.7.13}$$

Now the corresponding edges of all hyperedges $\{e_1,\dots e_{v-1}\}\setminus\{e_j\}$ form a tree in $G(w)$ again and the factorization of expectation in (4.7.13) can proceed as long as we have some edge attached to leaves. Repeating (4.7.13) recursively, with (4.7.11), we have

$$\mathbb{E}_\sigma\left[\prod_{i=1}^{v-1} p_{\sigma(e_i)}\right]=\prod_{i=1}^{v-1}\mathbb{E}_\sigma[p_{\sigma(e_i)}]=\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1}. \tag{4.7.14}$$

Since every hyperedge in w must be visited at least twice to make its expected weight nonzero, and w is of length $2kl$, we must have $h\le kl$. In the multigraph $G(w)$, we have the constraint $v\le h+1\le kl+1$. Since the first self-avoiding walk in w of length l takes $l+1$ distinct vertices, we also have $v\ge l+1$. So the possible range of v is $l+1\le v\le kl+1$ and h satisfies $v-1\le h\le kl$.

Putting all the estimates above together, for fixed v,h, the total contribution of self-avoiding walks from Case (1) to the sum is bounded by

$$n^v\binom{n}{d-2}^h[(v+1)^2(l+1)]^{2k(2+h-v)}\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1}\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{h-v+1}.$$

Denote S_1 to be the sum of all contributions from self-avoiding walks in Case (1). Then

$$S_1\le\sum_{v=l+1}^{kl+1}\sum_{h=v-1}^{kl} n^v\left(\frac{d-1}{n-d+2}\right)^h\left(\frac{\alpha}{d-1}\right)^{v-1}[(v+1)^2(l+1)]^{2k(2+h-v)}(a\vee b)^{h-v+1}. \tag{4.7.15}$$

When $l=O(\log n)$ and d,k are fixed, for sufficiently large n, $\left(\frac{n}{n-d+2}\right)^h\le 2$. Then from (4.7.15),

$$\begin{aligned}S_1\le&\sum_{v=l+1}^{kl+1}\sum_{h=v-1}^{kl}2n^{v-h}(d-1)^{h-v+1}[(v+1)^2(l+1)]^{2k(2+h-v)}\alpha^{v-1}(a\vee b)^{h-v+1}\\ \le&2\sum_{v=l+1}^{kl+1}\sum_{h=v-1}^{kl} n\left[\frac{(a\vee b)(d-1)}{n}\right]^{h-v+1}[(kl+2)^2(l+1)]^{2k(2+h-v)}\alpha^{v-1}.\end{aligned}$$

Hence

$$\frac{S_1}{n\alpha^{kl}[(kl+2)^2(l+1)]^{2k}} \leq 2\sum_{v=l+1}^{kl+1}\alpha^{v-1-kl}\sum_{h=v-1}^{kl}\left[n^{-1}(a\vee b)(d-1)((kl+2)^2(l+1))^{2k}\right]^{h-v+1}. \tag{4.7.16}$$

Since for fixed d,k and $l=O(\log n)$, $n^{-1}(a\vee b)(d-1)((kl+2)^2(l+1))^{2k}=o(1)$ for n sufficiently large, the leading term in (4.7.16) is the term with $h=v-1$. For sufficiently large n, we have

$$\frac{S_1}{n\alpha^{kl}[(kl+2)^2(l+1)]^{2k}} \leq 3\sum_{v=l+1}^{kl+1}\alpha^{v-1-kl} = 3\cdot\frac{\alpha-\alpha^{(1-k)l}}{\alpha-1} \leq \frac{3\alpha}{\alpha-1}.$$

It implies that $S_1 = O(n\alpha^{kl}\log^{6k} n)$.

Case (2). We now consider $w \in W_{2k,l}$ such that

- the number of edges in $G(w)$ is greater than h;
- vertices in $e_i \setminus \mathrm{end}(e_i)$ are all distinct for $1 \leq i \leq h$, and they are not vertices with labels in $V(w)$.

Let $\tilde{h}$ be the number of edges in $G(w)$ with $\tilde{h} \geq h+1$. Same as in Case (1), the number of triple sequence is at most $[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}$. Let $s_i, 1 \leq i \leq h$ be the size of $\mathrm{end}(e_i)$. We have $\sum_{i=1}^{h} s_i = 2\tilde{h}$. Note that when $s_i > 3$, there are more than 2 vertices in e_i contained in $V(w)$, therefore given the choices of vertices with labels in $V(w)$, we have fewer possibilities to choose the rest of vertices in e_i. Compared with (4.7.8), the number of all possible circuits in Case (2) with fixed $v,h,\tilde{h}$ is now bounded by

$$[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)} n^v \binom{n}{d-s_1}\cdots\binom{n}{d-s_h}.$$

When k is fixed and $l = O(\log n)$, for large n, the quantity above is bounded by

$$2[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}n^v \left(\frac{d-1}{n}\right)^{2\tilde{h}-h} \binom{n}{d-1}^h.$$

Now we consider the expected weight of each circuit in Case (2). In the spanning tree $T(w)$, we keep edges with distinct hyperedge labels that appear first in the circuit w and remove other edges. This gives us a forest denoted $F(w)$ inside $T(w)$, with at least $v-1-\tilde{h}+h$ many edges. We temporarily label those edges in the forest as $e_1,\dots,e_q$ with $q \geq v-1-\tilde{h}+h$. Then similar to the analysis of (4.7.14) in Case (1), we have

$$\mathbb{E}_\sigma\left[\prod_{i=1}^{q} p_{\sigma(e_q)}\right] = \left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^q,$$

and

$$\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A^{e_{i_t}}_{i_{t-1}i_t} - \overline{A}^{e_{i_t}}_{i_{t-1}i_t})\right] \leq \mathbb{E}_\sigma\left[\prod_{i=1}^{h} p_{\sigma(e_i)}\right] \leq \left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{\tilde{h}-v+1}\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1-\tilde{h}+h}.$$

Since every hyperedge in w must be visited at least twice to make its expected weight nonzero, we must have $l \leq h \leq kl$. In the multigraph $G(w)$, we have the constraint $v \leq \tilde{h}+1$. Since the first self-avoiding walk in w of length l takes $l+1$ distinct vertices, we also have $v \geq l+1$. So the possible range of v is $l+1 \leq v \leq \tilde{h}+1$ and h satisfies $l \leq h \leq kl$. Therefore we have

$$
\begin{aligned}
S_2 \leq & 2\sum_{h=l}^{kl}\sum_{\tilde{h}=h+1}^{2kl}\sum_{v=l+1}^{\tilde{h}+1}[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}n^v\left(\frac{d-1}{n}\right)^{2\tilde{h}-h}\binom{n}{d-1}^h \\
& \cdot\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{\tilde{h}-v+1}\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1-\tilde{h}+h} \\
= & O(\alpha^{kl}\log^{6k} n).
\end{aligned}
$$

Case (3). We now consider $w \in W_{2k,l}$ not included in Cases (1) or Case (2), which satisfies that

- for some $i \neq j$, there are common vertices in $e_i \setminus \mathrm{end}(e_i)$ and $e_j \setminus \mathrm{end}(e_j)$;
- or there are vertices in $e_i \setminus \mathrm{end}(e_i)$ with labels in $V(w)$.

Let $v, h, \tilde{h}$ be defined in the same way as in Case (2). The number of triple sequence is at most $[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}$. Consider the forest $F(w)$ introduced in Case (2) as a subgraph of $T(w)$, which has at least $(v-1-\tilde{h}+h)$ many edges with distinct hyperedge labels. We temporarily denote the edges by $e_1,\dots,e_q$, and the ordering is chosen such that e_1 is adjacent to a leaf in $F(w)$, and each $e_i, i \leq 2$ is adjacent to a leaf in $F(w) \setminus \{e_1,\dots,e_{i-1}\}$. For $1 \leq i \leq q$, we call e_i a *bad* hyperedge if the set $e_i \setminus \mathrm{end}(e_i)$ share a vertex with some set $e_j \setminus \mathrm{end}(e_j)$ for $j > i$, or there are vertices in $e_i \setminus \mathrm{end}(e_i)$ with labels in $V(w)$. In both cases, we have fewer choices for the vertices in e_i.

Suppose among $e_i, 1 \leq i \leq q$, there are t bad hyperedges. Let $s_i, 1 \leq i \leq h$ be the size of $\mathrm{end}(e_i)$ in $G(w)$. Then the number of all possible circuits in Case (3) with fixed $v, h, \tilde{h}$, and t, is bounded by

$$
[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)}n^v\binom{n}{d-s_1-\delta_1}\cdots\binom{n}{d-s_h-\delta_h}, \tag{4.7.17}
$$

where $\delta_i \in \{0,1\}$ and $\delta_i = 1$ if e_i is a bad hyperedge. Note that $\sum_{i=1}^{h} s_h = 2\tilde{h}$ and $\sum_{i=1}^{h} \delta_i = t$. For large n, the number in (4.7.17) is at most

$$2[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)} n^v \left(\frac{d-1}{n}\right)^{2\tilde{h}-h+t} \binom{n}{d-1}^h.$$

After removing the t edges with bad hyperedge labels from the forest $F(w)$, we can do the same analysis as in Case (2). The expected weight of each circuit in Case (3) with given $v, h, \tilde{h}, t$ now satisfies

$$\mathbb{E}_{\mathcal{H}_n}\left[\prod_{t=1}^{2kl}(A_{i_{t-1}i_t}^{e_{i_t}} - \overline{A}_{i_{t-1}i_t}^{e_{i_t}})\right] \leq \left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{\tilde{h}-v+1+t} \left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1-\tilde{h}+h-t}.$$

Let S_3 be the total contribution from circuits in Case (3) to (4.7.6) . Then

$$\begin{aligned}
S_3 \leq & \sum_{h=l}^{kl}\sum_{\tilde{h}=h}^{2kl}\sum_{v=l+1}^{\tilde{h}+1}\sum_{t=0}^{v-1} 2[(v+1)^2(l+1)]^{2k(2+\tilde{h}-v)} n^v \left(\frac{d-1}{n}\right)^{2\tilde{h}-h+t} \binom{n}{d-1}^h \\
& \cdot \left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{\tilde{h}-v+1+t} \left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1-\tilde{h}+h-t} \\
= & O(n\alpha^{kl}\log^{6k} n).
\end{aligned}$$

From the estimates on S_1, S_2 and S_3, Lemma 4.7.1 holds. □

With Lemma 4.7.1, we are able to derive (4.3.3). For any fixed $\varepsilon > 0$, choose k such that $1-2k\varepsilon < 0$, using Markov inequality, we have

$$\mathbb{P}_{\mathcal{H}_n}(\rho(\Delta^{(l)}) \geq n^{\varepsilon}\alpha^{l/2}) \leq \frac{\mathbb{E}_{\mathcal{H}_n}(\rho(\Delta^{(l)})^{2k})}{n^{2k\varepsilon}\alpha^{kl}} = O(n^{1-2k\varepsilon}\log^{6k} n).$$

This implies (4.3.3) in the statement of Theorem 4.3.1.

Proof of (4.3.4) in Theorem 4.3.1

Using a similar argument as in the proof of Lemma 4.7.1, we can prove the following estimate of $\rho(\Gamma^{(l,m)})$. The proof is given in Appendix 4.12.

Lemma 4.7.2. *For $l = O(\log n)$, fixed k, and any $1 \leq m \leq l$, there exists a constant $C > 0$ such that*

$$\mathbb{E}_{\mathcal{H}_n}[\rho(\Gamma^{(l,m)})^{2k}] \leq Cn^{1-2k}\alpha^{k(l+m-2)}\log^{14k} n. \tag{4.7.18}$$

With Lemma 4.7.2, we can apply the union bound and Markov inequality. For any $\varepsilon > 0$, choose $k > 0$ such that $1 - 2k\varepsilon < 0$, we have

$$\begin{aligned}
&\mathbb{P}_{\mathcal{H}_n}\left(\bigcup_{m=1}^{l}\left\{\rho(\Gamma^{(l,m)}) \geq n^{\varepsilon-1}\alpha^{(l+m)/2}\right\}\right) \leq \sum_{m=1}^{l}\mathbb{P}_{\mathcal{H}_n}\left(\rho(\Gamma^{(l,m)}) \geq n^{\varepsilon-1}\alpha^{(l+m)/2}\right) \\
\leq &\sum_{m=1}^{l}\frac{\mathbb{E}_{\mathcal{H}_n}\rho(\Gamma^{(l,m)})^{2k}}{n^{2k(\varepsilon-1)}\alpha^{k(l+m)}} \leq \sum_{m=1}^{l}\frac{C\log^{14k}(n)\cdot n^{1-2k}\alpha^{k(l+m-2)}}{n^{2k(\varepsilon-1)}\alpha^{k(l+m)}} = O\left((\log^{14k+1}(n)\cdot n^{1-2k\varepsilon}\alpha^{-2k}\right).
\end{aligned}$$

This proves (4.3.4) in Theorem 4.3.1.

4.8 Proof of Theorem 4.4.2

Let $n^{\pm}$ be the number of vertices with spin $\pm$, respectively. Consider the event

$$\tilde{\Omega} := \{|n^{\pm} - \frac{n}{2}| \leq \log(n)\sqrt{n}\}. \tag{4.8.1}$$

By Hoeffding's inequality,

$$\mathbb{P}_{\sigma}\left(|n^{\pm} - \frac{n}{2}| \geq \log(n)\sqrt{n}\right) \leq 2\exp(-2\log^2(n)), \tag{4.8.2}$$

which implies $\mathbb{P}_\sigma(\tilde{\Omega}) \geq 1 - 2\exp(-2\log^2(n))$. In the rest of this section we will condition on the event $\tilde{\Omega}$, which will not effect our conclusion and probability bounds, since for any event A, if $\mathbb{P}_{\mathcal{H}_n}(A \mid \tilde{\Omega}) = 1 - O(n^{-\gamma})$ for some $\gamma > 0$, we have

$$\mathbb{P}_{\mathcal{H}_n}(A) = \mathbb{P}_{\mathcal{H}_n}(A \mid \tilde{\Omega})\mathbb{P}_{\mathcal{H}_n}(\tilde{\Omega}) + \mathbb{P}_{\mathcal{H}_n}(A \mid \tilde{\Omega}^c)\mathbb{P}_{\mathcal{H}_n}(\tilde{\Omega}^c) = 1 - O(n^{-\gamma}).$$

The following identity from Equation (38) in [141] will be helpful in the proof.

Lemma 4.8.1. *For any nonnegative integers i, j, n and nonnegative numbers a, b such that $a/n, b/n < 1$, we have*

$$\frac{ai + bj}{n} - \frac{1}{2}\left(\frac{ai + bj}{n}\right)^2 \leq 1 - (1 - a/n)^i(1 - b/n)^j \leq \frac{ai + bj}{n}. \tag{4.8.3}$$

We will also use the following version of Chernoff bound (see [46]):

Lemma 4.8.2. *Let X be a sum of independent random variables taking values in $\{0,1\}$. Let $\mu = \mathbb{E}[X]$. Then for any $\delta > 0$, we have*

$$\mathbb{P}(X \geq (1+\delta)\mu) \leq \exp(-\mu h(1+\delta)), \tag{4.8.4}$$

$$\mathbb{P}(|X - \mu| \leq \delta\mu) \geq 1 - 2\exp(-\mu\tilde{h}(\delta)), \tag{4.8.5}$$

where

$$h(x) := x\log(x) - x + 1, \quad \tilde{h}(x) := \min\{(1+x)\log(1+x) - x, (1-x)\log(1-x) + x\}.$$

For any $t \geq 0$, the number of vertices with spin $\pm$ at distance t (respectively $\leq$) of vertices i is denoted $U_t^{\pm}(i)$ (respectively, $U_{\leq t}^{\pm}(i)$) and we know $S_t(i) = U_t^+(i) + U_t^-(i)$. We will omit index i when considering quantities related to a fixed vertex i. Let $n^{\pm}$ be the number of vertices with

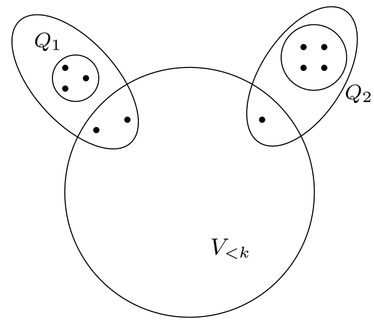

Figure 4.5: $d=5$, Q_1 is a connected 3-subsets in V_k and Q_2 is a connected 4-subsets in V_k.

spin $\pm$ and $\mathcal{N}^{\pm}$ be the set of vertices with spin $\pm$. For a fixed vertex i. Let

$$\mathcal{F}_t := \sigma(U_k^+, U_k^-, k \le t, \sigma_i, 1 \le i \le n) \tag{4.8.6}$$

be the σ-algebra generated by $\{U_k^+, U_k^-, 0 \le k \le t\}$ and $\{\sigma_i, 1 \le i \le n\}$. In the remainder of the section we condition on the spins σ_i of all $i \in [n]$ and assume $\tilde{\Omega}$ holds. We denote $\mathbb{P}(\cdot) := \mathbb{P}_{\mathcal{H}_n}(\cdot \mid \tilde{\Omega})$.

A main difficulty to analyze U_t^+, U_t^- compared to the graph SBM in [141] is that $U_k^{\pm}$ are no longer independent conditioned on $\mathcal{F}_{k-1}$. Instead, we can only approximate $U_k^{\pm}$ by counting subsets connected to V_{k-1}. To make it more precise, we have the following definition for connected-subsets.

Definition 4.8.3. A *connected s-subset* in V_k for $1 \le s \le d-1$ is a subset of size s which is contained in some hyperedge e in H and the rest $d-s$ vertices in e are from V_{k-1} (see Figure 4.5 for an example). Define $U_{k,s}^{(r)}, 0 \le r \le s$ to be the number of *connected s-subsets* in V_k where exactly r many vertices have $+$ spins. For convenience, we write $U_k^{(r)} := U_{k,d-1}^{(r)}$ for $0 \le r \le d-1$. Let $U_{k,s} = \sum_{r=0}^{s} U_{k,s}^{(r)}$ be the number of all connected s-subsets in V_k.

We will show that $\sum_{r=0}^{d-1} rU_k^{(r)}$ is a good approximation of U_k^+ and $\sum_{r=0}^{d-1}(d-1-r)U_k^{(r)}$ is a good approximation of U_k^-, then the concentration of $U_k^{(r)}, 0 \le r \le d-1$ implies the concentration of $U_k^{\pm}$.

Since each hyperedge appears independently, conditioned on $\mathcal{F}_{k-1}$, we know $\{U_k^{(r)}, 0 \leq r \leq d-1\}$ are independent binomial random variables. For $U_k^{(d-1)}$, the number of all possible connected $(d-1)$-subsets with $d-1$ many $+$ signs is $\binom{n^+ - U^+_{\leq k-1}}{d-1}$, and each such subset is included in the hypergraph if and only if it forms a hyperedge with any vertex in V_{k-1}. Therefore each such subset is included independently with probability

$$1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U^+_{k-1}} \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U^-_{k-1}}.$$

Similarly, we have the following distributions for $U_k^{(r)}, 1 \leq r \leq d-1$:

$$U_k^{(d-1)} \sim \mathrm{Bin}\left(\binom{n^+ - U^+_{\leq k-1}}{d-1}, 1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U^+_{k-1}} \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U^-_{k-1}}\right), \tag{4.8.7}$$

$$U_k^{(0)} \sim \mathrm{Bin}\left(\binom{n^- - U^-_{\leq k-1}}{d-1}, 1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U^-_{k-1}} \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U^+_{k-1}}\right), \tag{4.8.8}$$

and for $1 \leq r \leq d-2$,

$$U_k^{(r)} \sim \mathrm{Bin}\left(\binom{n^+ - U^+_{\leq k-1}}{r}\binom{n^- - U^-_{\leq k-1}}{d-1-r}, 1 - \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{S_{k-1}}\right). \tag{4.8.9}$$

For two random variable X, Y, we denote $X \preceq Y$ if X is stochastically dominant by Y, i.e., $\mathbb{P}(X \leq x) \geq \mathbb{P}(Y \leq x)$ for any $x \in \mathbb{R}$. We denote $U_k^* := \sum_{s=1}^{d-2} U_{k,s}$ to be the number of all connected s-subsets in V_k for $1 \leq s \leq d-2$.

For each $1 \leq s \leq d-2$, conditioned on $\mathcal{F}_{k-1}$, the number of possible s-subsets is at most $\binom{n}{s}$, and each subset is included in the hypergraph independently with probability at most

$\left(\frac{a\vee b}{\binom{n}{d-1}}\binom{S_{k-1}}{d-s}\right)\wedge 1$, so we have

$$U_{k,s} \preceq \operatorname{Bin}\left(\binom{n}{s}, \frac{a\vee b}{\binom{n}{d-1}}\binom{S_{k-1}}{d-s}\wedge 1\right). \tag{4.8.10}$$

With the definitions above, we have the following inequality for $U_k^{\pm}$ by counting the number of $\pm$ signs from each type of subsets:

$$U_k^{+} \leq \sum_{r=0}^{d-1} rU_k^{(r)} + (d-2)U_k^{*}, \tag{4.8.11}$$

$$U_k^{-} \leq \sum_{r=0}^{d-1} (d-1-r)U_k^{(r)} + (d-2)U_k^{*}. \tag{4.8.12}$$

To obtain the upper bound of $U_k^{\pm}$, we will show that U_k^{*} is negligible compared to the number of $\pm$ signs from $U_k^{(r)}$. Since $U_k^{(r)}, 1 \leq r \leq d-1$ are independent binomial random variables, we can prove concentration results of these random variables. For the lower bound of $U_k^{\pm}$, we need to show that only a negligible portion of $(d-1)$ connected subsets are overlapped, therefore U_k^{+} is lower bounded by $\sum_{r=0}^{d-1} rU_k^{(r)}$ minus some small term, and we can do it similarly for U_k^{-}. We will extensively use Chernoff bounds in Lemma 4.8.2 to prove the concentration of $U_k^{\pm}$ in the following theorem.

Theorem 4.8.4. *Let $\varepsilon \in (0,1)$, and $l = c\log(n)$ with $c\log(\alpha) < 1/4$. For any $\gamma \in (0, 3/8)$, there exists some constant $K > 0$ and such that the following holds with probability at least $1 - O(n^{-\gamma})$ for all $i \in [n]$.*

1. *Let $T := \inf\{t \leq l : S_t \geq K \log n\}$, then $S_T = \Theta(\log n)$.*

2. *Let $\varepsilon_t := \varepsilon\alpha^{-(t-T)/2}$ for some $\varepsilon > 0$ and*

$$M := \frac{1}{2}\begin{bmatrix} \alpha+\beta & \alpha-\beta \\ \alpha-\beta & \alpha+\beta \end{bmatrix}. \tag{4.8.13}$$

Then for all $t,t' \in \{T,\dots l\}$, $t > t'$, the vector $\vec{U}_t := (U_t^+, U_t^-)^\top$ satisfies the coordinate-wise bounds:

$$U_t^+ \in \left[\prod_{s=t'}^{t-1}(1-\varepsilon_s), \prod_{s=t'}^{t-1}(1+\varepsilon_s)\right](M^{t-t'}\vec{U}_{t'})_1, \tag{4.8.14}$$

$$U_t^- \in \left[\prod_{s=t'}^{t-1}(1-\varepsilon_s), \prod_{s=t'}^{t-1}(1+\varepsilon_s)\right](M^{t-t'}\vec{U}_{t'})_2, \tag{4.8.15}$$

where $(M^{t-t'}\vec{U}_{t'})_j$ is the j-th coordinate of the vector $M^{t-t'}\vec{U}_{t'}$ for $j = 1,2$.

Proof. In this proof, all constants C_i's, C, C' are distinct for different inequalities unless stated otherwise. By the definition of T, $S_{T-1} \le K\log(n)$. Let Z_T be the number of all hyperedges in H that are incident to at least one vertices in V_{T-1}. We have $S_T \le (d-1)Z_T$, and since the number of all possible hyperedges including a vertex in V_{T-1} is at most $S_{T-1}\binom{n}{d-1}$, Z_T is stochastically dominated by

$$\operatorname{Bin}\left(K\log(n)\binom{n}{d-1}, \frac{a\vee b}{\binom{n}{d-1}}\right),$$

which has mean $(a\vee b)K\log(n)$. Let $K_1 = (a\vee b)K$. By (4.8.4) in Lemma 4.8.2, we have for any constant $K_2 > 0$,

$$\mathbb{P}(Z_T \ge K_2\log(n)|\mathcal{F}_{T-1}) \le \exp(-K_1\log(n)h(K_2/K_1)) \tag{4.8.16}$$

Taking $K_2 > K_1$ large enough such that $K_1 h(K_2/K_1) \ge 2+\gamma$, we then have

$$\mathbb{P}(Z_T \ge K_2\log(n)|\mathcal{F}_{T-1}) \le n^{-2-\gamma}. \tag{4.8.17}$$

So with probability at least $1-n^{-2-\gamma}$, for a fixed $i\in[n]$, $S_T \le K_3\log(n)$ with $K_3 = (d-2)K_2$. Taking a union bound over $i\in[n]$, part (1) in Lemma 4.8.4 holds. We continue to prove (4.8.14) and (4.8.15) in several steps.

Step 1: base case. For the first step, we prove (4.8.14) and (4.8.15) for $t = T+1, t' = T$, which is

$$U^{\pm}_{T+1} \in [1-\varepsilon, 1+\varepsilon]\left(\frac{\alpha+\beta}{2}U^{\pm}_T + \frac{\alpha-\beta}{2}U^{\mp}_T\right). \tag{4.8.18}$$

This involves a two-sided estimate of $U^{\pm}_{T+1}$. The idea is to show the expectation of $U^{\pm}_{T+1}$ conditioned on $\mathcal{F}_T$ is closed to $\frac{\alpha+\beta}{2}U^{\pm}_T + \frac{\alpha-\beta}{2}U^{\mp}_T$, and $U^{\pm}_{T+1}$ is concentrated around its mean.

(i) Upper bound. Define the event $\mathcal{A}_T := \{S_T \leq K_3 \log n\}$. We have just shown for a fixed i,

$$\mathbb{P}(\mathcal{A}_T) \geq 1 - n^{-2-\gamma}. \tag{4.8.19}$$

Recall $|n^{\pm} - n/2| \leq \sqrt{n}\log n$ and conditioned on $\mathcal{A}_T$, for some constant $C > 0$,

$$U^{+}_{\leq T} \leq \sum_{t=0}^{T} S_t \leq 1 + TK_3 \log n \leq 1 + lK_3 \log n \leq CK_3 \log^2 n.$$

Conditioned on $\mathcal{F}_T$ and $\mathcal{A}_T$, for sufficiently large n, there exists constants $C_1 > 0$ such that

$$\binom{n^+ - U^{+}_{\leq T}}{d-1} \geq C_1 \binom{\frac{n}{2}}{d-1}.$$

From inequality (4.8.3), there exists constant $C_2 > 0$ such that

$$\begin{aligned}
1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U^{+}_T}\left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U^{-}_T} &\geq \frac{aU^{+}_T + bU^{-}_T}{\binom{n}{d-1}} - \frac{1}{2}\left(\frac{aU^{+}_T + bU^{-}_T}{\binom{n}{d-1}}\right)^2 \\
&\geq \frac{C_2(aU^{+}_T + bU^{-}_T)}{\binom{n}{d-1}} \geq \frac{C_2(a \wedge b)K \log n}{\binom{n}{d-1}}.
\end{aligned}$$

Then from (4.8.7), for some constant $C_3 > 0$,

$$\begin{aligned}\mathbb{E}[U_{T+1}^{(d-1)} \mid \mathcal{F}_T, \mathcal{A}_T] =& \binom{n^+ - U_{\le T}^+}{d-1} \left(1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U_T^+} \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U_T^-}\right) \\ \ge& C_1 \binom{\frac{n}{2}}{d-1} \cdot \frac{C_2(a \wedge b) K \log n}{\binom{n}{d-1}} \ge C_3 K \log n.\end{aligned}$$

We can choose K large enough such that $C_3 K \tilde{h}(\varepsilon/(2d)) \ge 2 + \gamma$, then from (4.8.5) in Lemma 4.8.2, for any given $\varepsilon > 0$ and $\gamma \in (0,1)$,

$$\begin{aligned}&\mathbb{P}\left(|U_{T+1}^{(d-1)} - \mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T]| \le \frac{\varepsilon}{2d}\mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T]\big|\mathcal{F}_T\right) \\ \ge& \mathbb{P}\left(|U_{T+1}^{(d-1)} - \mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T]| \le \frac{\varepsilon}{2d}\mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T]\big|\mathcal{F}_T, \mathcal{A}_T\right)\mathbb{P}(\mathcal{A}_T) \\ \ge& \left[1 - \exp\left(-\mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T, \mathcal{A}_T]\tilde{h}(\varepsilon/2d)\right)\right](1 - n^{-2-\gamma}) \ge (1 - n^{-2-\gamma})^2 \ge 1 - 2n^{-2-\gamma}.\end{aligned}$$

From the symmetry of $\pm$ labels, the concentration of $U_{T+1}^{(0)}$ works in the same way. Similarly, there exists a constant $C_1 > 0$ such that $\mathbb{E}[U_{T+1}^{(r)} \mid \mathcal{F}_T], 1 \le r \le d-2$:

$$\mathbb{E}[U_{T+1}^{(r)} \mid \mathcal{F}_T] = \binom{n^+ - U_{\le T}^+}{r}\binom{n^- - U_{\le T}^-}{d-1-r}\left(1 - \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{S_T}\right) \ge C_1 K \log n.$$

We can choose K large enough such that for all $0 \le r \le d-1$,

$$\mathbb{P}\left(\left|U_{T+1}^{(r)} - \mathbb{E}[U_{T+1}^{(r)} \mid \mathcal{F}_T]\right| \le \frac{\varepsilon}{2d}\mathbb{E}[U_{T+1}^{(r)}|\mathcal{F}_T] \mid \mathcal{F}_T\right) \ge 1 - 2n^{-2-\gamma}.$$

Next, we estimate $U_{T+1}^* = \sum_{s=1}^{d-2} U_{T+1,s}$. Recall from (4.8.10), we have $U_{T+1,s} \preceq Z_{T+1,s}$ where

$$Z_{T+1,s} \sim \mathrm{Bin}\left(\binom{n}{s}, \frac{a \vee b}{\binom{n}{d-1}}\binom{S_T}{d-s}\right).$$

Conditioned on $\mathcal{A}_T$ we know $K\log n \leq S_T \leq K_3 \log n$, and

$$\mathbb{E}[Z_{T+1,s} \mid \mathcal{A}_T, \mathcal{F}_T] = \binom{n}{s} \frac{a \vee b}{\binom{n}{d-1}} \binom{S_T}{d-s} \leq C_2 \log^{d-s}(n) n^{1+s-d}$$

for some constant $C_2 > 0$. Using the fact that $h(x) \geq \frac{1}{2} x \log(x)$ for x large enough, from (4.8.4), we have for any constant $\lambda > 0$, $1 \leq s \leq d-2$, there exists a constant $C_3 > 0$ such that for large n,

$$\begin{aligned}
&\mathbb{P}(U_{T+1,s} \geq \lambda S_T | \mathcal{F}_T, \mathcal{A}_T) \leq \mathbb{P}(Z_{T+1,s} \geq \lambda S_T | \mathcal{F}_T, \mathcal{A}_T) \\
\leq &\exp\left(-\mathbb{E}[Z_{T+1,s} \mid \mathcal{A}_T, \mathcal{F}_T] h\left(\frac{\lambda S_T}{\mathbb{E}[Z_{T+1,s} \mid \mathcal{A}_T, \mathcal{F}_T]}\right)\right) \\
\leq &\exp\left(-\frac{1}{2} \lambda S_T \log\left(\frac{\lambda S_T}{\mathbb{E}[Z_{T+1,s} \mid \mathcal{A}_T, \mathcal{F}_T]}\right)\right) \leq \exp(-\lambda C_3 \log^2 n) \leq n^{-2-\gamma}. \qquad (4.8.20)
\end{aligned}$$

Therefore with (4.8.19) and (4.8.20),

$$\mathbb{P}(U_{T+1,s} < \lambda S_T | \mathcal{F}_T) \geq \mathbb{P}(U_{T+1,s} < \lambda S_T | \mathcal{F}_T, \mathcal{A}_T) \mathbb{P}(\mathcal{A}_T) \geq (1 - n^{-2-\gamma})^2 \geq 1 - 2n^{-2-\gamma}.$$

Taking $\lambda = \frac{(\alpha-\beta)\varepsilon}{4d^2}$, we have $U_{T+1,s} \leq \frac{(\alpha-\beta)\varepsilon}{4d^2} S_T$ with probability at least $1 - 2n^{-2-\gamma}$ for any $\gamma \in (0,1)$.

Taking a union bound over $2 \leq r \leq d-1$, it implies

$$U^*_{T+1} \leq \frac{(\alpha-\beta)\varepsilon}{4d} S_T \qquad (4.8.21)$$

with probability $1 - O(n^{-2-\gamma})$ for any $\gamma \in (0,1)$.

Note that $n^{\pm} = \frac{n}{2} + O(\sqrt{n}\log n)$ and $U^{\pm}_{\leq T} = \sum_{k=1}^{T} S_k = O(\log^2(n))$. From (4.8.3),

$$\left(1 - \frac{aU_T^+ + bU_T^-}{2\binom{n}{d-1}}\right) \frac{aU_T^+ + bU_T^-}{\binom{n}{d-1}} \leq 1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U_T^+} \left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U_T^-} \leq \frac{aU_T^+ + bU_T^-}{\binom{n}{d-1}}.$$

It implies that

$$
\begin{aligned}
\mathbb{E}[U_{T+1}^{(d-1)}|\mathcal{F}_T,\mathcal{A}_T] &= \binom{\frac{n}{2}+O(\sqrt{n}\log n)}{d-1}\left(1+O\left(\frac{\log(n)}{n^{d-1}}\right)\right)\frac{aU_T^{+}+bU_T^{-}}{\binom{n}{d-1}} \\
&= \left(\frac{1}{2^{d-1}}+O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)(aU_T^{+}+bU_T^{-}). \qquad (4.8.22)
\end{aligned}
$$

Similarly, for $1 \leq r \leq d-2$.

$$
\begin{aligned}
\mathbb{E}[U_{T+1}^{(0)}|\mathcal{F}_T,\mathcal{A}_T] &= \left(\frac{1}{2^{d-1}}+O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)(bU_T^{+}+aU_T^{-}), \\
\mathbb{E}[U_{T+1}^{(r)}|\mathcal{F}_T,\mathcal{A}_T] &= \left(\frac{1}{2^{d-1}}+O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)\binom{d-1}{r}(bU_T^{+}+bU_T^{-}).
\end{aligned}
$$

Therefore from the estimations above, with the definition of α,β from (4.1.3),

$$
\begin{aligned}
&\mathbb{E}[\sum_{r=0}^{d-1} rU_{T+1}^{(r)}|\mathcal{F}_T,\mathcal{A}_T] \qquad (4.8.23) \\
&= \left(1+O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)\frac{1}{2^{d-1}}\left((d-1)(aU_T^{+}+bU_T^{-})+\sum_{r=1}^{d-2} r\binom{d-1}{r}b(U_T^{+}+U_T^{-})\right) \\
&= \left(1+O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)\left(\frac{\alpha+\beta}{2}U_T^{+}+\frac{\alpha-\beta}{2}U_T^{-}\right). \qquad (4.8.24)
\end{aligned}
$$

Since we have shown $\sum_{r=0}^{d-1} U_{T+1}^{(r)}$ concentrated around its mean by $\frac{\varepsilon}{2d}$ with probability at least $1-O(n^{-2-\gamma})$, conditioned on $\mathcal{A}_T$, we obtain

$$
\begin{aligned}
\left|\sum_{r=0}^{d-1} rU_{T+1}^{(r)}-\mathbb{E}[\sum_{r=0}^{d-1} rU_{T+1}^{(r)}|\mathcal{F}_T]\right| &\leq \sum_{r=0}^{d-1} r\left|U_{T+1}^{(r)}-\mathbb{E}[U_{T+1}^{(r)} \mid \mathcal{F}_T]\right| \leq \frac{\varepsilon}{2d}\sum_{r=1}^{d-1} r\mathbb{E}[U_{T+1}^{(r)} \mid \mathcal{F}_T] \\
&\leq \frac{\varepsilon}{4}\left(1+O\left(\frac{\log(n)}{\sqrt{n}}\right)\right)\left(\frac{\alpha+\beta}{2}U_T^{+}+\frac{\alpha-\beta}{2}U_T^{-}\right) \qquad (4.8.25)
\end{aligned}
$$

with probability $1-O(n^{-2-\gamma})$. Therefore from (4.8.24), conditioned on $\mathcal{A}_T$, for large n, with probability $1-O(n^{-2-\gamma})$,

$$\sum_{r=0}^{d-1} rU_{T+1}^{(r)} \in \left[1-\frac{\varepsilon}{3}, 1+\frac{\varepsilon}{3}\right]\left(\frac{\alpha+\beta}{2}U_T^+ + \frac{\alpha-\beta}{2}U_T^-\right). \tag{4.8.26}$$

From (4.8.11), (4.8.21) and (4.8.26), conditioned on $\mathcal{A}_T$ and $\mathcal{F}_T$, with probability $1-O(n^{-2-\gamma})$,

$$\begin{aligned} U_{T+1}^+ &\leq \sum_{r=0}^{d-1} rU_{T+1}^{(r)} + (d-2)U_{T+1}^* \leq \sum_{r=0}^{d-1} rU_{T+1}^{(r)} + (d-2)\frac{(\alpha-\beta)\varepsilon S_T}{4d} \\ &\leq (1+\varepsilon)\left(\frac{\alpha+\beta}{2}U_T^+ + \frac{\alpha-\beta}{2}U_T^-\right). \end{aligned}$$

Since $\mathbb{P}(\mathcal{A}_T) = 1-n^{-2-\gamma}$, and by symmetry of $\pm$ labels, with probability $1-O(n^{-2-\gamma})$,

$$U_{T+1}^{\pm} \leq (1+\varepsilon)\left(\frac{\alpha+\beta}{2}U_T^{\pm} + \frac{\alpha-\beta}{2}U_T^{\mp}\right). \tag{4.8.27}$$

(ii) Lower bound. To show (4.8.14), (4.8.15) for $t' = T+1, t = T$, we cannot directly bound $U_{T+1}^{\pm}$ from below by $U_{T+1}^{(r)}, 1 \leq r \leq d-1$ since from our definition of the connected $(d-1)$-subsets, they can overlap with each other, which leads to over-counting of the number vertices with $\pm$ labels. In the following we show the overlaps between different connected $(d-1)$-sets are small, which gives us the desired lower bound.

Let $W_{t+1,i}^{\pm}$ be the set of vertices in $V_{>t}$ with spin $\pm$ and appear in at least i distinct connected $(d-1)$-subsets in $V_{>t}$ for $i \geq 1$. Let $W_{t+1,i} = W_{t+1,i}^+ \cup W_{t+1,i}^-$. From our definition, $W_{T+1,1}^+$ are the vertices with spin $+$ that appear in at least one connected $(d-1)$-subsets, so $|W_{T+1,1}^+| \leq U_{T+1}^+$. By counting the multiplicity of vertices with spin $+$, we have the following relation

$$\sum_{r=1}^{d-1} rU_{T+1}^{(r)} = |W_{T+1,1}^+| + \sum_{i\geq 2}|W_{T+1,i}^+| \leq U_{T+1}^+ + \sum_{i\geq 2}|W_{T+1,i}|. \tag{4.8.28}$$

This implies a lower bound on U_{T+1}^{+}:

$$U_{T+1}^{+} \geq \sum_{r=1}^{d-1} r U_{T+1}^{(r)} - \sum_{i \geq 2} |W_{T+1,i}|. \tag{4.8.29}$$

Next we control $|W_{T+1,2}|$. Let $m = n - |V_{\leq T}|$. We enumerate all vertices in $V_{>T}$ from 1 to m temporarily for the proof of the lower bound. Let $X_i, 1 \leq i \leq m$ be the random variables that $X_i = 1$ if $i \in W_{T+1,2}$ and 0 otherwise, we then have $|W_{T+1,2}| = \sum_{i=1}^{m} X_i$. A simple calculation yields

$$|W_{T+1,2}|^2 - |W_{T+1,2}| = \left(\sum_{i=1}^{m} X_i\right)^2 - \sum_{i=1}^{m} X_i = 2 \sum_{1 \leq i < j \leq m} X_i X_j. \tag{4.8.30}$$

The product $X_i X_j$ is 1 if $i, j \in W_{T+1,2}$ and 0 otherwise.

We further consider 3 events, E_{ij}^{s} for $s = 0, 1, 2$, where E_{ij}^{0} is the event that all $(d-1)$-subsets in $V_{>T}$ containing i, j are not connected to V_T, E_{ij}^{1} is the event that there is only one $(d-1)$-subset in $V_{>T}$ containing i, j connected to V_T and E_{ij}^{2} is the event that there are at least two $(d-1)$-subsets in $V_{>T}$ containing i, j connected to V_T. Now we have

$$\begin{aligned}\mathbb{E}[X_i X_j \mid \mathcal{F}_T, \mathcal{A}_T] &= \mathbb{P}(i, j \in W_{T+1,2} \mid \mathcal{F}_T, \mathcal{A}_T) \\ &= \sum_{r=0}^{2} \mathbb{P}\left(i, j \in W_{T+1,2} \mid E_{ij}^{r}, \mathcal{F}_T, \mathcal{A}_T\right) \mathbb{P}(E_{ij}^{r} \mid \mathcal{F}_T, \mathcal{A}_T). \end{aligned} \tag{4.8.31}$$

We estimate the three terms in the sum separately. Conditioned on E_{ij}^{0}, $\mathcal{F}_T$, and $\mathcal{A}_T$, the two events that $i \in W_{T+1,2}$ and $j \in W_{T+1,2}$ are independent. And the probability that $i \in W_{T+1,2}$ is bounded by

$$\binom{n}{d-2}^2 \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^2 S_T^2 \leq \frac{C_1 \log^2(n)}{n^2}$$

for some constant $C_1 > 0$. So we have

$$\begin{aligned}
&\mathbb{P}\left(i,j \in W_{T+1,2} \mid E^0_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) \mathbb{P}(E^0_{ij} \mid \mathcal{F}_T, \mathcal{A}_T) \le \mathbb{P}\left(i,j \in W_{T+1,2} \mid E^0_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) \\
=&\mathbb{P}\left(i \in W_{T+1,2} \mid E^0_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) \mathbb{P}\left(j \in W_{T+1,2} \mid E^0_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) \le \frac{C_1^2 \log^4 n}{n^4}. \qquad (4.8.32)
\end{aligned}$$

For the term that involves E^1_{ij}, we know for some $C_2 > 0$,

$$\mathbb{P}(E^1_{ij} \mid \mathcal{F}_T, \mathcal{A}_T) \le \binom{n}{d-3} \frac{a \vee b}{\binom{n}{d-1}} S_T \le \frac{C_2 \log n}{n^2},$$

and conditioned on E^1_{ij} and $\mathcal{F}_T, \mathcal{A}_T$, the two events that $i \in W_{T+1,2}$ and $j \in W_{T+1,2}$ are independent again, since we require i, j to be contained in at least 2 connected-subsets. We have

$$\mathbb{P}\left(i \in W_{T+1,2} \mid E^1_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) \le \binom{n}{d-2} S_T \frac{a \vee b}{\binom{n}{d-1}} \le \frac{C_2 \log n}{n}.$$

Therefore we have

$$\begin{aligned}
&\mathbb{P}\left(i,j \in W_{T+1,2} \mid E^1_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) \mathbb{P}(E^1_{ij} \mid \mathcal{F}_T, \mathcal{A}_T) \\
=&\mathbb{P}\left(i \in W_{T+1,2} \mid E^1_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) \mathbb{P}\left(j \in W_{T+1,2} \mid E^1_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) \mathbb{P}(E^1_{ij} \mid \mathcal{F}_T, \mathcal{A}_T) \\
\le&\frac{C_2^2 \log^2 n}{n^2} \cdot \frac{C_2 \log n}{n^2} = \frac{C_2^3 \log^3 n}{n^4}. \qquad (4.8.33)
\end{aligned}$$

Conditioned on E^2_{ij}, i, j have already been included in 2 connected $(d-1)$ subsets, so

$$\mathbb{P}\left(i,j \in W_{T+1,2} \mid E^2_{ij}, \mathcal{F}_T, \mathcal{A}_T\right) = 1.$$

We then have for some $C_3 > 0$,

$$\mathbb{P}\left(i,j \in W_{T+1,2} \mid E_{ij}^2, \mathcal{F}_T, \mathcal{A}_T\right) \mathbb{P}(E_{ij}^2 \mid \mathcal{F}_T, \mathcal{A}_T)$$
$$= \mathbb{P}(E_{ij}^2 \mid \mathcal{F}_T, \mathcal{A}_T) \leq \binom{n}{d-3}^2 S_T^2 \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^2 \leq \frac{C_3 \log^2 n}{n^4}. \tag{4.8.34}$$

Combining (4.8.32)-(4.8.34), we have for some constant $C' > 0$,

$$\mathbb{E}[X_i X_j \mid \mathcal{F}_T, \mathcal{A}_T] \leq \frac{C' \log^4 n}{n^4}. \tag{4.8.35}$$

Taking conditional expectation in (4.8.30), we have

$$\mathbb{E}\left[|W_{T+1,2}|^2 - |W_{T+1,2}| \mid \mathcal{F}_T, \mathcal{A}_T\right] = 2 \sum_{1 \leq i < j \leq m} \mathbb{E}[X_i X_j \mid \mathcal{F}_T, \mathcal{A}_T] \leq \frac{C' \log^4 n}{n^2}.$$

By Markov's inequality, there exists a constant $C > 0$ such that for any constant $\lambda > 0$ and sufficiently large n,

$$\mathbb{P}\left(|W_{T+1,2}| > \lambda S_T \mid \mathcal{F}_T, \mathcal{A}_T\right) \leq \mathbb{P}\left(|W_{T+1,2}|(|W_{T+1,2}| - 1) > \lambda S_T(\lambda S_T - 1) \mid \mathcal{F}_T, \mathcal{A}_T\right) \tag{4.8.36}$$
$$\leq \frac{\mathbb{E}[|W_{T+1,2}|(|W_{T+1,2}| - 1) \mid \mathcal{F}_T, \mathcal{A}_T]}{\lambda S_T(\lambda S_T - 1)} \leq \frac{C \log^2 n}{\lambda^2 n^2},$$

where in the last inequality we use the fact that $S_T \geq K \log n$. Taking $\lambda = \frac{(\alpha-\beta)\varepsilon}{4}$, we have for all large n and for any $\gamma \in (0,1)$,

$$\mathbb{P}\left(|W_{T+1,2}| > \frac{(\alpha-\beta)\varepsilon}{4} S_T \mid \mathcal{F}_T, \mathcal{A}_T\right) = O\left(\frac{\log^2 n}{n^2}\right) \leq n^{-1-\gamma}. \tag{4.8.37}$$

For a fixed vertex $j \in V_{>T}$, the probability that $j \in W_{T+1,i}$ is at most $\binom{n}{d-2}^i S_T^i \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^i$, then we have for sufficiently large n,

$$\mathbb{E}[|W_{T+1,i}| \mid \mathcal{F}_T, \mathcal{A}_T] \leq n\binom{n}{d-2}^i S_T^i \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^i \leq n\left(\frac{C_4 \log n}{n}\right)^i$$

for some $C_4 > 0$. For the rest of the terms in (4.8.28), we have for some constant $C > 0$,

$$\mathbb{E}\left[\sum_{i \geq 3} |W_{T+1,i}| \,\Big|\, \mathcal{F}_T, \mathcal{A}_T\right] \leq n \sum_{i=3}^{\infty} \left(\frac{C_4 \log n}{n}\right)^i \leq \frac{C \log^3(n)}{n^2}.$$

By Markov's inequality,

$$\mathbb{P}\left(\sum_{i \geq 3} |W_{T+1,i}| \geq \frac{(\alpha-\beta)\varepsilon}{4} S_T \mid \mathcal{F}_T, \mathcal{A}_T\right) \leq \frac{C \log^2(n)}{n^2} \leq n^{-1-\gamma}.$$

Together with (4.8.37), we have conditioned on $\mathcal{A}_T$, $\sum_{i \geq 2} |W^+_{T+1,2}| \leq \frac{(\alpha-\beta)\varepsilon}{2} S_T$ with probability at least $1 - 2n^{-1-\gamma}$ for any $\gamma \in (0,1)$ and all large n. Note that

$$\frac{(\alpha-\beta)\varepsilon}{2} S_T \leq \frac{\varepsilon}{2}\left(\frac{\alpha+\beta}{2} U_T^+ + \frac{\alpha-\beta}{2} U_T^-\right).$$

With (4.8.26), (4.8.29), and (4.8.19), we have

$$U_{T+1}^+ \geq \sum_{r=1}^{d-1} r U_{T+1}^{(r)} - \frac{\varepsilon}{2}\left(\frac{\alpha+\beta}{2} U_T^+ + \frac{\alpha-\beta}{2} U_T^-\right) \geq (1-\varepsilon)\left(\frac{\alpha+\beta}{2} U_T^+ + \frac{\alpha-\beta}{2} U_T^-\right)$$

with probability $1 - O(n^{-1-\gamma})$. By symmetry, the argument works for U_{T+1}^-, therefore with probability $1 - O(n^{-1-\gamma})$ for any $\gamma \in (0,1)$, we have

$$U_{T+1}^{\pm} \geq (1-\varepsilon)\left(\frac{\alpha+\beta}{2} U_T^{\pm} + \frac{\alpha-\beta}{2} U_T^{\mp}\right). \tag{4.8.38}$$

From (4.8.27) and (4.8.38), we have with probability $1 - O(n^{-1-\gamma})$ for any $\gamma \in (0,1)$, (4.8.18) holds.

Step 2: Induction. It remains to extend this estimate in Step 1 for all $T \le t' < t \le l$. We now define the event

$$\mathcal{A}_t := \left\{ U_t^{\pm} \in [1-\varepsilon_{t-1}, 1+\varepsilon_{t-1}] \left(\frac{\alpha+\beta}{2} U_{t-1}^{\pm} + \frac{\alpha-\beta}{2} U_{t-1}^{\pm} \right) \right\} \tag{4.8.39}$$

for $T+1 \le t \le l$, and recall $\varepsilon_t = \varepsilon\alpha^{-(t-T)/2}, \mathcal{A}_T = \{S_T \le K_3 \log n\}$.

From the proof above, we have shown $\mathcal{A}_{T+1}$ holds with probability $1 - O(n^{-1-\gamma})$. Conditioned on $\mathcal{A}_T$, $\mathcal{A}_{T+1}, \cdots, \mathcal{A}_t$ for some fix t with $T+2 \le t \le l$, the vector $\vec{U}_t = (U_t^+, U_t^-)$ satisfies (4.8.14), (4.8.15) for any $T \le t' < t$.

Set $t' = T+1$. From [141], for any integer $k > 0$, $M^k = \frac{1}{2}\begin{bmatrix} \alpha^k + \beta^k & \alpha^k - \beta^k \\ \alpha^k - \beta^k & \alpha^k + \beta^k \end{bmatrix}$. (4.8.14) implies that

$$\begin{aligned} U_t^{\pm} &\ge \left(\prod_{s=T+1}^{t-1} (1-\varepsilon_s) \right) \left(\frac{\alpha^{t-T-1} + \beta^{t-T-1}}{2} U_{T+1}^{\pm} + \frac{\alpha^{t-T-1} - \beta^{t-T-1}}{2} U_{T+1}^{\mp} \right) \\ &\ge (1-O(\varepsilon)) \frac{\alpha^{t-T-1}}{2} (1-\varepsilon) \left(\frac{\alpha+\beta}{2} U_T^{\pm} + \frac{\alpha-\beta}{2} U_T^{\mp} \right) \\ &\ge (1-O(\varepsilon)) \alpha^{t-T} \frac{(1-\varepsilon)(\alpha-\beta)}{4\alpha} S_T \ge C_1 \alpha^{t-T} \log(n), \end{aligned} \tag{4.8.40}$$

for some constant $C_1 > 0$. For any t with $T \le t$, conditioned on $\mathcal{A}_T$, $\mathcal{A}_{T+1}, \cdots, \mathcal{A}_t$, since $\beta < \alpha$,

$$\begin{aligned} U_t^{\pm} &\le \left(\prod_{s=T}^{t-1} (1+\varepsilon_s) \right) \left(\frac{\alpha^{t-T} + \beta^{t-T}}{2} U_T^{\pm} + \frac{\alpha^{t-T} - \beta^{t-T}}{2} U_T^{\mp} \right) \\ &\le (1+O(\varepsilon)) \frac{\alpha^{t-T} + \beta^{t-T}}{2} S_T \le (1+O(\varepsilon)) \alpha^{t-T} K_3 \log(n) \le C_2 \alpha^{t-T} \log n \end{aligned} \tag{4.8.41}$$

for some $C_2 > 0$. Combining lower and upper bounds on $U_t^{\pm}$, we obtain

$$S_t = U_t^+ + U_t^- = \Theta(\alpha^{t-T} \log n). \tag{4.8.42}$$

We now show by induction that $\mathcal{A}_{t+1}$ holds with high probability conditioned on $\{\mathcal{A}_j, T \le j \le t\}$.

(i) Upper bound. Note that $\alpha^l = o(n^{1/4})$, for some constant $C > 0$

$$U_{\le t}^{+} \le \sum_{i=1}^{t} S_i \le C\alpha^{t-T}\log^2 n \le C\alpha^l \log n = o(n^{1/4}\log n).$$

Recall $|n^{\pm} - \frac{n}{2}| \le \sqrt{n}\log n$. From (4.8.7)-(4.8.9), similar to the case for $t = T$, we have

$$\begin{aligned}
&\mathbb{E}[U_{t+1}^{(d-1)} | \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] = \binom{n^+ - U_{\le t}^{+}}{d-1}\left(1 - \left(1 - \frac{a}{\binom{n}{d-1}}\right)^{U_t^+}\left(1 - \frac{b}{\binom{n}{d-1}}\right)^{U_t^-}\right) \\
&= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log n}{\sqrt{n}}\right)\right)(aU_t^+ + bU_t^-),
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[U_{t+1}^{(0)} | \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] &= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log n}{\sqrt{n}}\right)\right)(bU_t^+ + aU_t^-), \\
\mathbb{E}[U_{t+1}^{(r)} | \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] &= \left(\frac{1}{2^{d-1}} + O\left(\frac{\log n}{\sqrt{n}}\right)\right)\binom{d-1}{r}(bU_t^+ + bU_t^-),
\end{aligned}$$

for $1 \le r \le d-2$. Hence there exists a constant $C_0 > 0$ such that for all $0 \le r \le d-1$,

$$\mathbb{E}[U_{t+1}^{(r)} | \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] \ge C_0 S_t.$$

From (4.8.5) in Lemma 4.8.2, for any $0 \le r \le d-1$, to show

$$\mathbb{P}\left(\left|U_{t+1}^{(r)} - \mathbb{E}[U_{t+1}^{(r)} \mid \cap_{j=T}^{t}\mathcal{A}_j, \mathcal{F}_t]\right| \le \frac{\varepsilon}{2d}\mathbb{E}[U_{t+1}^{(r)} \mid \cap_{j=T}^{t}\mathcal{A}_j, \mathcal{F}_t] \;\middle|\; \cap_{j=T}^{t}\mathcal{A}_j, \mathcal{F}_t\right) \ge 1 - n^{-2-\gamma}, \tag{4.8.43}$$

it suffices to have

$$C_0 S_t \tilde{h}\left(\frac{\varepsilon_t}{2d}\right) \geq (2+\gamma)\log n. \tag{4.8.44}$$

From (4.8.5), by a second-order expansion of $\tilde{h}$ around 0, $\tilde{h}(x) \geq x^2/3$ when $x > 0$ is small. For $\gamma \in (0,1)$, the left hand side in (4.8.44) is lower bounded by

$$C_1 K \alpha^{t-T} \log(n) \tilde{h}\left(\frac{\varepsilon_t}{2d}\right) \geq C_2 \alpha^{t-T} K \log(n) \varepsilon_t^2 = C_2 K \log n \geq (2+\gamma)\log n,$$

by taking K large enough. Therefore (4.8.43) holds.

We also have

$$U_{t+1,s} \preceq Z_{t+1,s}, \quad Z_{t+1,s} \sim \operatorname{Bin}\left(\binom{n}{s}, \frac{a \vee b}{\binom{n}{d-1}} \binom{S_t}{d-s}\right),$$

and $Z_{t+1,s}$ has mean $\binom{n}{s} \frac{a\vee b}{\binom{n}{d-1}} \binom{S_t}{d-s} = \Theta\left(\frac{\alpha^{(d-s)(t-T)} \log^{d-s}(n)}{n^{d-1-s}}\right)$. For $1 \leq s \leq d-2$, using the fact that $h(x) \geq \frac{1}{2} x \log(x)$ for x large enough, similar to (4.8.20), there are constants $C_1, C_2, C_3, C_4 > 0$ such that for any $\lambda > 0$,

$$\begin{aligned}
&\mathbb{P}(U_{t+1,s} \geq \lambda S_t \mid \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t) \leq \mathbb{P}(Z_{t+1,s} \geq \lambda S_t \mid \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t) \\
\leq& \exp\left(-C_1 \lambda \alpha^{t-T} \log(n) \log\left(\frac{C_2 \lambda \alpha^{t-T} \log(n)}{C_3 \alpha^{(d-s)(t-T)} \log^{d-s}(n) n^{1+s-d}}\right)\right).
\end{aligned}$$

Taking $\lambda = \frac{(\alpha-\beta)\varepsilon_t}{4d^2} = \frac{(\alpha-\beta)\varepsilon\alpha^{-(t-T)/2}}{4d^2}$, we have

$$\begin{aligned}
&\mathbb{P}\left(U_{t+1,s} \geq \frac{(\alpha-\beta)\varepsilon_t}{4d^2} S_t \mid \cap_{j=T}^t \mathcal{A}_j, \mathcal{F}_t\right) \\
\leq& \exp\left(-C_1' \alpha^{(t-T)/2} \log(n) \cdot \log(C_2' \alpha^{(s-d+\frac{1}{2})(t-T)} \log^{1+s-d}(n) n^{d-1-s})\right).
\end{aligned}$$

Since for some constants $C_4, C_5, C_6 > 0$,

$$\begin{aligned}&\log(C_2'\alpha^{(s-d+\frac{1}{2})(t-T)}\log^{1+s-d}(n)n^{d-1-s})\\ \geq &C_4 - C_5(t-T)\log(\alpha) + \log(\log^{1+s-d}(n)) + (d-1-s)\log n \geq C_6 \log n,\end{aligned}$$

we have for all $1 \leq s \leq d-2$,

$$\mathbb{P}(U_{t+1,s} \geq \frac{(\alpha-\beta)\varepsilon_t}{4d^2}S_t \mid \cap_{j=T}^{t}\mathcal{A}_j, \mathcal{F}_t) \leq \exp\left(-C_1' C_6 \log^2 n\right) \leq n^{-2-\gamma} \tag{4.8.45}$$

for any $\gamma \in (0,1)$. Recall for sufficiently large n,

$$\varepsilon_t = \varepsilon\alpha^{-(t-T)/2} \geq \varepsilon\alpha^{-l/2} > n^{-1/8}.$$

Therefore $\frac{\log n}{\sqrt{n}} = o(\varepsilon_t)$. From (4.8.45), conditioned on $\mathcal{A}_T, \dots, A_t$ and $\mathcal{F}_t$,

$$U_{t+1}^{+} \leq \sum_{r=1}^{d-1} rU_{t+1}^{(r)} + (d-2)U_{t+1}^{*} \leq (1+\varepsilon_t)\left(\frac{\alpha+\beta}{2}U_t^{+} + \frac{\alpha-\beta}{2}U_t^{-}\right)$$

with probability at least $1-O(n^{-2-\gamma})$. A similar bound works for U_{t+1}^{-}, which implies conditioned on $\mathcal{A}_T, \dots, A_t$,

$$U_{t+1}^{\pm} \leq (1+\varepsilon_t)\left(\frac{\alpha+\beta}{2}U_t^{\pm} + \frac{\alpha-\beta}{2}U_t^{\pm}\right) \tag{4.8.46}$$

with probability $1-O(n^{-2-\gamma})$ for any $\gamma \in (0,1)$.

(ii) Lower bound. We need to show that conditioned on $\mathcal{A}_T, \dots, \mathcal{A}_t$,

$$U_{t+1}^{\pm} \geq (1-\varepsilon_t)\left(\frac{\alpha+\beta}{2}U_t^{\pm} + \frac{\alpha-\beta}{2}U_t^{\pm}\right)$$

with probability $1-O(n^{-1-\gamma})$ for some $\gamma \in (0,1)$. This part of the proof is very similar to the case for $t = T$. Same as (4.8.29), we have the following lower bound on U_{t+1}^{+}:

$$U_{t+1}^{+} \geq \sum_{r=1}^{d-1} r U_{t+1}^{(r)} - \sum_{i \geq 2} |W_{t+1,i}|.$$

Next we control $|W_{t+1,2}|$. Let $m = n - |V_{\leq t}|$ and we enumerate all vertices in $V_{>t}$ from 1 to m. Let $X_1, \dots X_m$ be the random variable that $X_i = 1$ if $i \in W_{t+1,2}$ and 0 otherwise. Same as (4.8.30),

$$|W_{t+1,2}|^2 - |W_{t+1,2}| = 2 \sum_{1 \leq i < j \leq m} X_i X_j. \tag{4.8.47}$$

Let E_{ij}^{s} for $s = 0, 1, 2$, be the similar events as in (4.8.31) before, now we have

$$\begin{aligned}
&\mathbb{E}[X_i X_j \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] = \mathbb{P}\left(i, j \in W_{t+1,2} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right) \\
&= \sum_{r=0}^{2} \mathbb{P}\left(i, j \in W_{t+1,2} \mid E_{ij}^{r}, \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right) \mathbb{P}(E_{ij}^{r} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t).
\end{aligned}$$

The three terms in the sum can be estimated separately in the same way as before. By using the upper bound $C\alpha^{t-T} \log n \leq S_t \leq C_0 \alpha^{t-T} \log n$ for some $C, C_0 > 0$, and use the same argument for the case when $t = T$, we have the following three inequalities for some constants $C_1, C_2, C_3 > 0$:

$$\begin{aligned}
\mathbb{P}\left(i, j \in W_{t+1,2} \mid E_{ij}^{0}, \mathcal{F}_t\right) \mathbb{P}(E_{ij}^{0} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t) &\leq \frac{C_1^2 \alpha^{4(t-T)} \log^4 n}{n^4}, \\
\mathbb{P}\left(i, j \in W_{t+1,2} \mid E_{ij}^{1}, \mathcal{F}_t\right) \mathbb{P}(E_{ij}^{1} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t) &\leq \frac{C_2^3 \alpha^{3(t-T)} \log^3 n}{n^4}, \\
\mathbb{P}\left(i, j \in W_{t+1,2} \mid E_{ij}^{2}, \mathcal{F}_t\right) \mathbb{P}(E_{ij}^{2} \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t) &\leq \frac{C_3 \alpha^{2(t-T)} \log^2 n}{n^4}.
\end{aligned}$$

This implies $\mathbb{E}[X_i X_j \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] \leq \frac{C' \alpha^{4(t-T)} \log^4 n}{n^4}$ for some $C' > 0$. Taking conditional expectation in (4.8.47), we have

$$\mathbb{E}\left[|W_{t+1,2}|^2 - |W_{t+1,2}| \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right] \le \frac{C'\alpha^{4(t-T)}\log^4 n}{n^2}.$$

Then by Markov inequality and (4.8.42), similar to (4.8.36), there exists a constant $C > 0$ such that for any $\lambda = \Omega(\alpha^{-(t-T)})$,

$$\mathbb{P}\left(|W_{t+1,2}| > \lambda S_t \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right) \le \frac{C\alpha^{2(t-T)}\log^2 n}{\lambda^2 n^2}.$$

Take $\lambda = \frac{(\alpha-\beta)\varepsilon_t}{4}$. Since $c\log(\alpha) < 1/4$, we have $\alpha^l < n^{1/4}$, and

$$\mathbb{P}\left(|W_{t+1,2}| > \frac{(\alpha-\beta)\varepsilon_t}{4} S_t \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right) \le \frac{C\alpha^{2(t-T)}\log^2 n}{n^2} \le n^{-1-\gamma}$$

for any $\gamma \in (0, 1/2)$.

For each $|W_{t+1,i}|$ for $i \ge 3$, we have for sufficiently large n, there exists a constant $C_4 > 0$

$$\mathbb{E}[|W_{t+1,i}| \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t] \le n \binom{n}{d-2}^i S_t^i \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^i \le n \left(\frac{C_4 \alpha^{t-T}\log n}{n}\right)^i.$$

For the rest of the terms, we have for some constant $C_4' > 0$,

$$\mathbb{E}\left[\sum_{i\ge 3} |W_i| \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right] \le n \sum_{i=3}^{\infty} \left(\frac{C_4 \alpha^{t-T}\log n}{n}\right)^i \le \frac{C_4' \alpha^{3(t-T)}\log^3(n)}{n^2}.$$

By Markov's inequality,

$$\mathbb{P}\left(\sum_{i\ge 3} |W_i| \ge \frac{(\alpha-\beta)\varepsilon_t}{4} S_t \mid \cap_{j=T}^{t} \mathcal{A}_j, \mathcal{F}_t\right) \le \frac{C_5 \alpha^{2.5(t-T)}\log^2(n)}{n^2} \le n^{-1-\gamma}$$

for any $\gamma \in (0,3/8)$. Together with the estimate on $W_{t+1,2}$, we have

$$\sum_{i\geq 2} |W^+_{t+1,2}| \leq \frac{(\alpha-\beta)\varepsilon_t}{2} S_t \leq \frac{\varepsilon_t}{2}\left(\frac{\alpha+\beta}{2}U_t^+ + \frac{\alpha-\beta}{2}U_t^-\right)$$

with probability $1-2n^{-1-\gamma}$ for any $\gamma \in (0,3/8)$.

With (4.8.29) and (4.8.26), $U^+_{t+1} \geq (1-\varepsilon_t)\left(\frac{\alpha+\beta}{2}U_t^+ + \frac{\alpha-\beta}{2}U_t^-\right)$ with probability $1-O(n^{-1-\gamma})$. By symmetry, the argument works for U^-_{t+1}. Therefore conditioned on $\mathcal{A}_T,\dots,\mathcal{A}_t$, with probability $1-O(n^{-1-\gamma})$ for any $\gamma \in (0,3/8)$,

$$U^{\pm}_{t+1} \geq (1-\varepsilon_t)\left(\frac{\alpha+\beta}{2}U_t^{\pm} + \frac{\alpha-\beta}{2}U_t^{\mp}\right). \tag{4.8.48}$$

This finishes the proof the lower bound part of Step 2. Recall (4.8.39). With (4.8.48) and (4.8.46), we have shown that conditioned on $\mathcal{A}_T,\dots,\mathcal{A}_t$, with probability $1-O(n^{-1-\gamma})$, $\mathcal{A}_{t+1}$ holds. This finishes the induction step. Finally, for fixed $i\in[n]$ and $\gamma \in (0,3/8)$,

$$\begin{aligned}\mathbb{P}\left(\bigcap_{t=T}^{l}\mathcal{A}_t\right) &= \mathbb{P}(\mathcal{A}_T)\prod_{t=T+1}^{l}\mathbb{P}(\mathcal{A}_t \mid \mathcal{A}_{t-1},\dots,\mathcal{A}_T)\\ &\geq (1-Cn^{-2-\gamma})(1-Cn^{-1-\gamma})^l \geq 1-C_6\log(n)n^{-1-\gamma},\end{aligned}$$

for some constant $C_6>0$. Taking a union bound over $i\in[n]$, we have shown $\mathcal{A}_t$ holds for all $T\leq t\leq l$ and all $i\in[n]$ with probability $1-O(n^{-\gamma})$ for any $\gamma\in(0,3/8)$. This completes the proof of Theorem 4.8.4. □

With Theorem 4.8.4, the rest of the proof of Theorem 4.4.2 follows similarly from the proof of Theorem 2.3 in [141]. We include it for completeness.

Proof of Theorem 4.4.2. Assume all the estimates in statement of Theorem 4.8.4 hold. For $t\leq l$, if $t \leq T$, from the definition of T, we have $S_t, |D_t| = O(\log n)$. For $t > T$, from [141], M satisfies

$$M^k = \frac{1}{2}\begin{bmatrix} \alpha^k + \beta^k & \alpha^k - \beta^k \\ \alpha^k - \beta^k & \alpha^k + \beta^k \end{bmatrix}.$$

Using (4.8.14) and (4.8.15), we have for $t > t' \geq T$,

$$S_t \leq \left(\prod_{s=t'}^{t-1}(1+\varepsilon_s)\right)(1,1)M^{t-t'}\vec{U}_{t'} \leq \left(\prod_{s=t'}^{t-1}(1+\varepsilon_s)\right)\alpha^{t-t'}S_{t'}, \tag{4.8.49}$$

$$S_t \geq \left(\prod_{s=t'}^{t-1}(1-\varepsilon_s)\right)(1,1)M^{t-t'}\vec{U}_{t'} \geq \left(\prod_{s=t'}^{t-1}(1-\varepsilon_s)\right)\alpha^{t-t'}S_{t'}. \tag{4.8.50}$$

Setting $t' = T$ in (4.8.49), we obtain

$$S_t \leq \left(\prod_{s=T}^{t-1}(1+\varepsilon_s)\right)\alpha^{t-T}S_T = O(\alpha^{t-T}\log n) = O(\alpha^t \log n).$$

Therefore (4.4.1) holds. Let $t = l$ in (4.8.49) and (4.8.50), we have for all $T \leq t' < l$,

$$\left(\prod_{s=t'}^{l-1}(1-\varepsilon_s)\right)\alpha^{l-t'}S_{t'} \leq S_l \leq \left(\prod_{s=t'}^{l-1}(1+\varepsilon_s)\right)\alpha^{l-t'}S_{t'}.$$

And it implies

$$\left(\prod_{s=t'}^{l-1}(1-\varepsilon_s)\right)S_{t'} \leq \alpha^{t'-l}S_l \leq \left(\prod_{s=t'}^{l-1}(1+\varepsilon_s)\right)S_{t'}. \tag{4.8.51}$$

Note that

$$\max\left\{\prod_{s=t'}^{l-1}(1+\varepsilon_s) - 1, 1 - \prod_{s=t'}^{l-1}(1-\varepsilon_s)\right\} = O(\varepsilon_{t'}) = O(\alpha^{-t'/2}).$$

Together with (4.8.51), we have for all $T \leq t' < l$,

$$|S_{t'} - \alpha^{t'-l}S_l| \leq O(\alpha^{-t'/2})S_{t'} = O(\alpha^{t'/2}\log n). \tag{4.8.52}$$

On the other hand, for $t \leq T$, we know $S_t = O(\log n)$. Let $t' = T$ in (4.8.52), we have

$$|S_T - \alpha^{T-l} S_l| = O(\alpha^{T/2} \log n). \tag{4.8.53}$$

So for $1 \leq t \leq T$,

$$\begin{aligned} |S_t - \alpha^{t-l} S_l| &= O(\log n) + \alpha^{t-T}(S_T + O(\log(n)\alpha^{T/2})) \\ &= O(\log n) + O(\alpha^{t-T/2} \log n) = O(\alpha^{t/2} \log n). \end{aligned} \tag{4.8.54}$$

The last inequality comes from the inequality $t - T/2 \leq t/2$. Combining (4.8.52) and (4.8.54), we have proved (4.4.3) holds for all $1 \leq t \leq l$.

Using (4.8.14) and (4.8.15), we have

$$D_{t+1} = U_{t+1}^{+} - U_{t+1}^{-} \leq \beta(U_t^{+} - U_t^{-}) + \alpha\varepsilon_t(U_t^{+} + U_t^{-}) = \beta D_t + \alpha\varepsilon_t S_t.$$

Similarly, $\beta D_t - \alpha\varepsilon_t S_t \leq D_{t+1} \leq \beta D_t + \alpha\varepsilon_t S_t$. By iterating, we have for $l \geq t > t' \geq T$,

$$|D_t - \beta^{t-t'} D_{t'}| \leq \sum_{s=t'}^{t-1} \alpha\beta^{t-1-s} \varepsilon_s S_s. \tag{4.8.55}$$

Recall $S_s = O(\log(n)\alpha^{s-T})$, $|D_T| = O(\log n)$, and $\varepsilon_s = \alpha^{-(s-T)/2}$. Taking $t' = T$ in (4.8.55), for $t > T$,

$$|D_t| = O\left(\log(n)\beta^t\right) + O\left(\sum_{s=T}^{t-1} \alpha\beta^{t-1-s} \log(n)\alpha^{(s-T)/2}\right).$$

Since $1 < \alpha < \beta^2$, it follows that

$$\begin{aligned}\sum_{s=T}^{t-1} \alpha\beta^{t-1-s} \log(n)\alpha^{(s-T)/2} =& \beta^{t-1}\alpha^{1-T/2}\log(n)\sum_{s=T}^{t-1}\left(\frac{\alpha}{\beta^2}\right)^{s/2} \\ =& \beta^{t-1}\alpha^{1-T/2}\log(n)O(\alpha^{T/2}\beta^{-T}) = O(\log(n)\beta^t).\end{aligned}$$

So we have $|D_t| = O(\log n\beta^t)$. The right side of (4.8.55) is of order

$$\sum_{s=t'}^{t-1} \alpha\beta^{t-1-s}\alpha^{(s-T)/2}\log(n) = O(\log(n)\beta^{t-t'}\alpha^{t'/2}).$$

Thus setting $t = l$ in (4.8.55), for $l > t' \geq T$, we obtain $D_l - \beta^{l-t'}D_{t'} = O(\log(n)\beta^{l-t'}\alpha^{t'/2})$. Therefore $D_{t'} = \beta^{t'-l}D_l + O(\log(n)\alpha^{t'/2})$ holds for all $T \leq t' < l$. For $t' < T$, we have $D_{t'} = O(\log n)$ and

$$\begin{aligned}|D_{t'} - \beta^{t'-l}D_l| \leq& O(\log n) + \beta^{t'-T}(|D_T| + O(\log(n)\alpha^{T/2})) \\ =& O(\log n) + O(\beta^{t'-T}\alpha^{T/2}\log n) = O(\alpha^{t'/2}\log n),\end{aligned}$$

where the last estimate is because $\beta^{t'-T} < \alpha^{(t'-T)/2}$ under the condition that $t' < T$. Altogether we have shown (4.4.4) holds for all $1 \leq t' \leq l$. This completes the proof of Theorem 4.4.2. $\square$

4.9 Proof of Theorem 4.4.6

We first state the following lemma before proving Theorem 4.4.6. The proof is included in Appendix 4.12.

Lemma 4.9.1. *For all $m \in \{1, \ldots, l\}$ with $l = c\log n$, $c\log\alpha < 1/4$, it holds asymptotically almost* *surely that*

$$\sup_{\|x\|_2=1, x^\top B^{(l)}\mathbf{1}=x^\top B^{(l)}\sigma=0} \|\mathbf{1}^\top B^{(m-1)}x\|_2 = O(\sqrt{n}\alpha^{(m-1)/2}\log n), \tag{4.9.1}$$

$$\sup_{\|x\|_2=1, x^\top B^{(l)}\mathbf{1}=x^\top B^{(l)}\sigma=0} \|\sigma^\top B^{(m-1)}x\|_2 = O(\sqrt{n}\alpha^{(m-1)/2}\log n). \tag{4.9.2}$$

Proof of Theorem 4.4.6. Using matrix expansion identity (4.3.2) and the estimates in Theorem 4.3.1, for any l_2-normalized vector x with $x^\top B^{(l)}\mathbf{1} = x^\top B^{(l)}\sigma = 0$, we have for sufficiently large n, asymptotically almost surely

$$\begin{aligned}
\|B^{(l)}x\|_2 &= \left\| \Delta^{(l)}x + \sum_{m=1}^{l} (\Delta^{(l-m)}\overline{A}B^{(m-1)})x - \sum_{m=1}^{l} \Gamma^{(l,m)}x \right\|_2 \\
&\leq \rho(\Delta^{(l)}) + \sum_{m=1}^{l} \rho(\Delta^{(l-m)})\|\overline{A}B^{(m-1)}x\|_2 + \sum_{m=1}^{l} \rho(\Gamma^{(l,m)}) \\
&\leq 2n^{\varepsilon}\alpha^{l/2} + \sum_{m=1}^{l} n^{\varepsilon}\alpha^{(l-m)/2}\|\overline{A}B^{(m-1)}x\|_2,
\end{aligned} \tag{4.9.3}$$

where $\overline{A} = \mathbb{E}_{\mathcal{H}_n}[A \mid \sigma]$. We have the following expression for entries of $\overline{A}$. If $i \neq j$ and $\sigma_i = \sigma_j = +1$,

$$\overline{A}_{ij} = \frac{a}{\binom{n}{d-1}}\binom{n^+-2}{d-2} + \frac{b}{\binom{n}{d-1}}\left(\binom{n-2}{d-2} - \binom{n^+-2}{d-2}\right) =: \tilde{a}_n^+.$$

If $i \neq j$ and $\sigma_i = \sigma_j = -1$,

$$\overline{A}_{ij} = \frac{a}{\binom{n}{d-1}}\binom{n^--2}{d-2} + \frac{b}{\binom{n}{d-1}}\left(\binom{n-2}{d-2} - \binom{n^--2}{d-2}\right) =: \tilde{a}_n^-.$$

If $\sigma_i \neq \sigma_j$,

$$\overline{A}_{ij} = \frac{b}{\binom{n}{d-1}}\binom{n-2}{d-2} := \tilde{b}_n.$$

We then have $\tilde{a}_n^+, \tilde{a}_n^-, \tilde{b}_n = O(1/n)$. Conditioned on the event $\{|n^{\pm} - n/2| \leq \log(n)\sqrt{n}\}$, we obtain

$$\tilde{a}_n^- - \tilde{a}_n^+ = \frac{a-b}{\binom{n}{d-1}}\left(\binom{n^- - 2}{d-2} - \binom{n^+ - 2}{d-2}\right) = O\left(\frac{\log n}{n^{3/2}}\right).$$

Let R be a $n\times n$ matrix such that

$$R_{ij} = \begin{cases} 1 & \sigma_i = \sigma_j = -1 \text{ and } i \neq j, \\ 0 & \text{otherwise.} \end{cases}$$

We then have $\|R\|_2 \le \sqrt{\sum_{ij} R_{ij}^2} \le n$. The following decomposition of $\overline{A}$ holds:

$$\begin{aligned} \overline{A} &= \tilde{a}_n^+\left[\frac{1}{2}(\mathbf{1}\cdot\mathbf{1}^\top + \sigma\sigma^\top) - I\right] + \frac{\tilde{b}_n}{2}(\mathbf{1}\cdot\mathbf{1}^\top - \sigma\sigma^\top) + (\tilde{a}_n^- - \tilde{a}_n^+)R & (4.9.4) \\ &= \frac{\tilde{a}_n^+ + \tilde{b}_n}{2}\mathbf{1}\cdot\mathbf{1}^\top + \frac{\tilde{a}_n^+ - \tilde{b}_n}{2}\sigma\sigma^\top + \left((\tilde{a}_n^- - \tilde{a}_n^+)R - \tilde{a}_n^+ I\right). & (4.9.5) \end{aligned}$$

Since

$$\|(\tilde{a}_n^- - \tilde{a}_n^+)R - \tilde{a}_n^+ I\|_2 \le |\tilde{a}_n^- - \tilde{a}_n^+|\cdot\|R\|_2 + |\tilde{a}_n^+| = O(\log n/\sqrt{n}),$$

by (4.9.5), we have

$$\|\overline{A}B^{(m-1)}x\|_2 = O\left(\frac{1}{n}\right)\|\mathbf{1}\cdot\mathbf{1}^\top B^{(m-1)}x\|_2 + O\left(\frac{1}{n}\right)\|\sigma\sigma^\top B^{(m-1)}x\|_2 + O\left(\frac{\log n}{\sqrt{n}}\right)\|B^{(m-1)}x\|_2.$$

By Cauchy inequality,

$$\|\mathbf{1}\cdot\mathbf{1}^\top B^{(m-1)}x\|_2 \le \sqrt{n}\|\mathbf{1}^\top B^{(m-1)}x\|_2, \quad \|\sigma\sigma^\top B^{(m-1)}x\|_2 \le \sqrt{n}\|\sigma^\top B^{(m-1)}x\|_2.$$

Therefore,

$$\|\overline{A}B^{(m-1)}x\|_2 = O(n^{-1/2})(\|\sigma^\top B^{(m-1)}x\|_2 + \|\mathbf{1}^\top B^{(m-1)}x\|_2) + O(\log n/\sqrt{n})\|B^{(m-1)}x\|_2.$$

Using (4.9.1) and (4.9.2), the right hand side in the expression above is upper bounded by

$$O(\alpha^{(m-1)/2}\log n) + O(\|B^{(m-1)}x\|_2 \cdot \log n/\sqrt{n}). \tag{4.9.6}$$

Since $B^{(m-1)}$ is a nonnegative matrix, the spectral norm is bounded by the maximum row sum (see Theorem 8.1.22 in [115]), we have that

$$\|B^{(m-1)}x\|_2 \le \rho(B^{(m-1)}) \le \max_i \sum_{j=1}^{n} B_{ij}^{(m-1)}.$$

By (4.4.1), (4.4.5) and (4.4.7), the right hand side above is $O(\alpha^{m-1}\log n)$. Combing (4.9.6) and noting that $\alpha^{m-1}/\sqrt{n} = o(n^{-1/4})$, it implies

$$\|\overline{A}B^{(m-1)}x\|_2 = O(\alpha^{(m-1)/2}\log n) + O(\alpha^{m-1}\log^2 n/\sqrt{n}) = O(\alpha^{(m-1)/2}\log n). \tag{4.9.7}$$

Taking (4.9.7) into (4.9.3), we have for any $\varepsilon > 0$, with high probability, $\|B^{(l)}x\|_2 = O(n^{\varepsilon}\alpha^{l/2}\log^2 n) \le n^{2\varepsilon}\alpha^{l/2}$ for n sufficiently large. This completes the proof. □

4.10 Proof of Theorem 4.5.2

The proof in this section is a generalization of the method in [146] for sparse random graphs. We now prove the case where $\sigma_i = +1$, and the case for $\sigma_i = -1$ can be treated in the same way. Recall the definition of V_t from Definition 4.4.1. Let A_t be the event that no vertex in V_t is connected by two distinct hyperedges to V_{t-1}. Let B_t be the event that there does not exist two vertices in V_t that are contained in a hyperedge $e \subset \binom{V_t}{d}$.

We can construct the multi-type Poisson hypertree (T, ρ, τ) in the following way. For a vertex $v \in T$, Let $Y_v^{(r)}, 0 \le r \le d-1$ be the number of hyperedges incident to v which among the remaining $d-1$ vertices, r of them have the same spin with $\tau(v)$. We have

$$Y_v^{(d-1)} \sim \text{Pois}\left(\frac{a}{2^{d-1}}\right), \quad Y_v^{(r)} \sim \text{Pois}\left(\frac{\binom{d-1}{r}b}{2^{d-1}}\right), 0 \le r \le d-2.$$

Note that (T,ρ,τ) can be entirely reconstructed from the label of the root and the sequence $\{Y_v^{(r)}\}$ for $v \in V(T), 0 \le r \le d-1$.

We define similar random variables for (H,i,σ). For a vertex $v \in V_t$, let $X_v^{(r)}$ be the number of hyperedges incident to v, where all the remaining $d-1$ vertices are in V_{t+1} such that r of them have spin $\sigma(v)$. Then we have

$$\begin{aligned} X_v^{(d-1)} &\sim \text{Bin}\left(\binom{|V_{>t}^{\sigma(v)}|}{d-1}, \frac{a}{\binom{n}{d-1}}\right), \\ X_v^{(r)} &\sim \text{Bin}\left(\binom{|V_{>t}^{\sigma(v)}|}{r}\binom{|V_{>t}^{-\sigma(v)}|}{d-1-r}, \frac{b}{\binom{n}{d-1}}\right), \quad 0 \le r \le d-2 \end{aligned}$$

and conditioned on $\mathcal{F}_t$ (recall the definition of $\mathcal{F}_t$ from (4.8.6)) they are independent. Recall Definition 4.5.1. We have the following lemma on the spin-preserving isomorphism. The proof of Lemma 4.10.1 is given in Appendix 4.12.

Lemma 4.10.1. *Let $(H,i,\sigma)_t, (T,\rho,\tau)_t$ be the rooted hypergraph truncated at distance t from i, ρ, respectively. If*

1. *there is a spin-preserving isomorphism ϕ such that $(H,i,\sigma)_{t-1} \equiv (T,\rho,\tau)_{t-1}$,*
2. *for every $v \in V_{t-1}$, $X_v^{(r)} = Y_{\phi(v)}^{(r)}$ for $0 \le r \le d-1$,*
3. *A_t, B_t hold,*

then $(H,i,\sigma)_t \equiv (T,\rho,\tau)_t$.

To make our notation simpler, for the rest of this section, we will identify v with $\phi(v)$. Recall the event $\Omega_t(i) = \{S_t(i) \le C\log(n)\alpha^t\}$ where the constant C is the same one as in Theorem

4.4.2. Now define a new event

$$C_t := \bigcap_{s \le t} \Omega_s(i). \tag{4.10.1}$$

From the proof of Theorem 4.4.2, for all $t \le l$, $\mathbb{P}_{\mathcal{H}_n}(C_t) = 1 - O(n^{-1-\gamma})$ for any $\gamma \in (0, 3/8)$. Note that conditioned on C_t, there exists $C' > 0$ such that

$$|V_{\le t}| \le \sum_{s \le t} C \log(n) \alpha^t \le C' \log^2(n) \alpha^t. \tag{4.10.2}$$

We now estimate the probability of event A_t, B_t conditioned on C_t. The proof is included in Appendix 4.12.

Lemma 4.10.2. *For any $t \ge 1$,*

$$\mathbb{P}(A_t | C_t) \ge 1 - o(n^{-1/2}), \quad \mathbb{P}(B_t | C_t) \ge 1 - o(n^{-1/2}).$$

Before proving Theorem 4.5.2, we also need the following bound on the total variation distance between binomial and Poisson random variables, see for example Lemma 4.6 in [146].

Lemma 4.10.3. *Let m, n be integers and c be a positive constant. The following holds:*

$$\left\| \mathrm{Bin}\left(m, \frac{c}{n}\right) - \mathrm{Pois}(c) \right\|_{\mathrm{TV}} = O\left(\frac{1 \vee |m - n|}{n}\right).$$

Proof of Theorem 4.5.2. Fix t and suppose that C_t holds, and $(T, \rho)_t \equiv (H, i)_t$. Then for each $v \in V_t$, recall

$$X_v^{(d-1)} \sim \mathrm{Bin}\left(\binom{|V_{>t}^{\sigma(v)}|}{d-1}, \frac{a}{\binom{n}{d-1}}\right), \quad X_v^{(r)} \sim \mathrm{Bin}\left(\binom{|V_{>t}^{\sigma(v)}|}{r}\binom{|V_{>t}^{-\sigma(v)}|}{d-1-r}, \frac{b}{\binom{n}{d-1}}\right)$$

and

$$Y_v^{(d-1)} \sim \text{Pois}\left(\frac{a}{2^{d-1}}\right), \quad Y_v^{(r)} \sim \text{Pois}\left(\frac{\binom{d-1}{r}b}{2^{d-1}}\right), \quad 0 \le r \le d-2.$$

Recall $|n^{\pm} - n/2| \le \sqrt{n}\log n$. We have the following bound for $V_{>t}^{\pm}$:

$$\begin{aligned}
|V_{>t}^{\pm}| &\ge n^{\pm} - |V_{\le t}| \ge \frac{n}{2} - \sqrt{n}\log(n) - O(\log^2(n)\alpha^{2t}) \ge \frac{n}{2} - 2\sqrt{n}\log(n), \\
|V_{>t}^{\pm}| &\le n^{\pm} \le \frac{n}{2} + \sqrt{n}\log(n).
\end{aligned}$$

Therefore $|V_{>t}^{\pm} - \frac{n}{2}| \le 2\sqrt{n}\log n$. Then from Lemma 4.10.3,

$$\begin{aligned}
\|X_v^{(d-1)} - Y_v^{(d-1)}\|_{\text{TV}} &\le C \frac{\left|\binom{|V_{>t}^{\sigma(v)}|}{d-1} - \frac{1}{2^{d-1}}\binom{n}{d-1}\right|}{\frac{1}{2^{d-1}}\binom{n}{d-1}} = O(n^{-1/2}\log n), \\
\|X_v^{(r)} - Y_v^{(r)}\|_{\text{TV}} &= O(n^{-1/2}\log n), \quad 0 \le r \le d-2.
\end{aligned}$$

We can couple $X_v^{(r)}$ with $Y_v^{(r)}, 0 \le r \le d-1$ such that $\mathbb{P}\left(X_v^{(r)} \ne Y_v^{(r)}\right) = O(n^{-1/2}\log n)$. Taking a union bound over all $v \in V_t$, and $0 \le r \le d-1$ and recall (4.10.2), we can find a coupling such that with probability at least

$$1 - O(\log^3(n)\alpha^l n^{-1/2}) \ge 1 - o(n^{-1/4}),$$

$X_v^{(r)} = Y_v^{(r)}$ for every $v \in V_t$ and $0 \le r \le d-1$. Lemma 4.10.2 implies A_t, B_t, C_t hold simultaneously with probability at least $1 - o(n^{-1/4})$. Altogether we have that assumptions (2),(3) in Lemma 4.10.1 hold with probability $1 - o(n^{-1/4})$, which can be written as

$$\mathbb{P}\left((H,i,\sigma)_{t+1} \equiv (T,\rho,\tau)_{t+1}, C_{t+1} \,\middle|\, (H,i,\sigma)_t \equiv (T,\rho,\tau)_t, C_t\right) \ge 1 - o(n^{-1/4}).$$

Since we can certainly couple i with ρ from our construction, $\mathbb{P}((H,i,\sigma)_0 \equiv (T,\rho,\tau)_0, C_0) = 1$. Therefore for large n,

$$
\begin{aligned}
&\mathbb{P}((H,i,\sigma)_l \equiv (T,\rho,\tau)_l) \\
=&\prod_{t=1}^{l} \mathbb{P}\left((H,i,\sigma)_t \equiv (T,\rho,\tau)_t, C_t \;\middle|\; (H,i,\sigma)_{t-1} \equiv (T,\rho,\tau)_{t-1}, C_{t-1}\right) \\
&\cdot \mathbb{P}((H,i,\sigma)_0 \equiv (T,\rho,\tau)_0, C_0) \\
\geq&(1-o(n^{-1/4}))^l \geq 1-n^{-1/5}.
\end{aligned}
$$

This completes the proof. □

4.11 Proof of Theorem 4.6.1

The proof of the following Lemma 4.11.1 follows in a similar way as Lemma 4.4 in [141], and we include it in Appendix 4.12.

Lemma 4.11.1. *For $l = c\log(n), c\log(\alpha) < 1/4$, the following hold asymptotically almost surely*

$$\|B^{(l)}\mathbf{1} - \vec{S}_l\|_2 = o(\|B^{(l)}\mathbf{1}\|_2), \tag{4.11.1}$$

$$\|B^{(l)}\sigma - \vec{D}_l\|_2 = o(\|B^{(l)}\sigma\|_2), \tag{4.11.2}$$

$$\langle B^{(l)}\mathbf{1}, B^{(l)}\sigma\rangle = o\left(\|B^{(l)}\mathbf{1}\|_2 \cdot \|B^{(l)}\sigma\|_2\right). \tag{4.11.3}$$

The next lemma estimate $\|B^{(l)}x\|_2$ when $x = B^{(l)}\sigma$ and $B^{(l)}\mathbf{1}$. The proof of Lemma 4.11.2 is provided in Appendix 4.12.

Lemma 4.11.2. *Assume $\beta^2 > \alpha > 1$ and $l = c\log(n)$ with $c\log(\alpha) < 1/8$. Then for some fixed*

$\gamma > 0$ *asymptotically almost surely one has*

$$\Omega(\alpha^l)\|B^{(l)}\mathbf{1}\|_2 \leq \|B^{(l)}B^{(l)}\mathbf{1}\|_2 \leq O(\alpha^l \log n)\|B^{(l)}\mathbf{1}\|_2, \tag{4.11.4}$$

$$\Omega(\beta^l)\|B^{(l)}\sigma\|_2 \leq \|B^{(l)}B^{(l)}\sigma\|_2 \leq O(n^{-\gamma}\alpha^l)\|B^{(l)}\sigma\|_2. \tag{4.11.5}$$

Together with Lemma 4.11.1 and Lemma 4.11.2, we are ready to prove Theorem 4.6.1.

Proof of Theorem 4.6.1. From Theorem 4.4.6 and Lemma 4.11.2, the top two eigenvalues of $B^{(l)}$ will be asymptotically in the span of $B^{(l)}\mathbf{1}$ and $B^{(l)}\sigma$. By the lower bound in (4.11.4) and the upper bound in (4.11.5), the largest eigenvalue of $B^{(l)}$ will be $\Theta(\alpha^l)$ up to a logarithmic factor, and the first eigenvector is asymptotically aligned with $B^{(l)}\mathbf{1}$.

From (4.11.1), $B^{(l)}\mathbf{1}$ is also asymptotically aligned with $\vec{S}_l$, therefore our statement for the first eigenvalue and eigenvector holds. Since $B^{(l)}\mathbf{1}$ and $B^{(l)}\sigma$ are asymptotically orthogonal from (4.11.3), together with (4.11.5), the second eigenvalue of $B^{(l)}$ is $\Omega(\beta^l)$ and the second eigenvector is asymptotically aligned with $B^{(l)}\sigma$. From (4.11.2), $B^{(l)}\sigma$ is asymptotically aligned with $\vec{D}_l$. So the statement for the second eigenvalue and eigenvector holds. The order of other eigenvalues follows from Theorem 4.4.6 and the Courant minimax principle (see [115]). $\square$

4.12 Proof of auxiliary lemmas

Proof of Lemma 4.4.3

Proof. The two sequences $(U_k^{\pm}(i))_{k\leq l}$, $(U_k^{\pm}(j))_{k\leq l}$ are independent conditioned on the event $\{V_{\leq l}(i)\cap V_{\leq l}(j)=\emptyset\}$. It remains to estimate $\mathbb{P}_{\mathcal{H}_n}(\{V_{\leq l}(i)\cap V_{\leq l}(j)=\emptyset\})$. Introduce the events

$$\mathcal{J}_k := \bigcap_{t\leq k}\{S_t(i)\vee S_t(j)\leq C\log(n)\alpha^t\}, \quad \mathcal{L}_k := \{V_{\leq k}(i)\bigcap V_{\leq k}(j)=\emptyset\},$$

where the constant C is the same one as in the statement of Theorem 4.4.2. For any vertex $v \in [n] \setminus (V_{\leq k}(i) \cup V_{\leq k}(j))$, Conditioned on $\mathcal{L}_k$ and $\mathcal{J}_k$, there are two possible situations where v is included in $V_{k+1}(i) \cap V_{k+1}(j)$:

(1) There is a hyperedge containing v and a vertex in $V_k(i)$, and a different hyperedge containing v and a vertex in $V_k(j)$.

(2) There is a hyperedge containing v, one vertex in $V_k(i)$, and another vertex in $V_k(j)$.

There exists a constant $C_1 > 0$ such that Case (1) happens with probability at most

$$S_k(i)S_k(j)\binom{n}{d-2}^2\left(\frac{a \vee b}{\binom{n}{d-1}}\right)^2 \leq C_1 \log^2(n)\alpha^{2k}/n^2,$$

and Case (2) happens with probability at most

$$S_k(i)S_k(j)\binom{n}{d-3}\frac{a \vee b}{\binom{n}{d-1}} = C_1 \log^2(n)\alpha^{2k}/n^2.$$

Since $\alpha^{2l} = n^{2c\log\alpha} = o(n^{1/2})$, we have for large n,

$$\mathbb{P}_{\mathcal{H}_n}(v \in V_{k+1}(i) \cap V_{k+1}(j) \mid \mathcal{J}_k, \mathcal{L}_k) \leq 2C_1 \log^2(n)\alpha^{2l}/n^2 < n^{-1.5}.$$

Taking a union bound over all possible v, we have for some constant $C_3 > 0$,

$$\mathbb{P}_{\mathcal{H}_n}(V_{k+1}(i) \cap V_{k+1}(j) = \emptyset \mid \mathcal{J}_k, \mathcal{L}_k) \geq 1 - C_3 n^{-1/2}.$$

From the proof of Theorem 4.4.2, for all $0 \leq k \leq l$, $\mathbb{P}_{H_n}(\mathcal{J}_k) = 1 - O(n^{-1-\gamma})$ for any $\gamma \in (0, 3/8)$. We then have

$$\mathbb{P}_{\mathcal{H}_n}(V_{k+1}(i) \cap V_{k+1}(j) = \emptyset \mid \mathcal{L}_k) \geq \mathbb{P}_{\mathcal{H}_n}(V_{k+1}(i) \cap V_{k+1}(j) = \emptyset \mid \mathcal{J}_k, \mathcal{L}_k)\, \mathbb{P}_{\mathcal{H}_n}(\mathcal{J}_k) \geq 1 - O(n^{-1/2}).$$

Finally, for large n,

$$\begin{aligned}\mathbb{P}_{\mathcal{H}_n}(\{V_{\le l}(i)\cap V_{\le l}(j)=\emptyset\}) =&\mathbb{P}_{\mathcal{H}_n}(\mathcal{L}_l)\ge \mathbb{P}_{\mathcal{H}_n}(V_l(i)\cap V_l(j)=\emptyset \mid \mathcal{L}_{l-1})\mathbb{P}_{\mathcal{H}_n}(\mathcal{L}_{l-1})\\ \ge&\mathbb{P}_{\mathcal{H}_n}(\mathcal{L}_0)\prod_{k=0}^{l-1}\mathbb{P}_{\mathcal{H}_n}(V_{k+1}(i)\cap V_{k+1}(j)=\emptyset \mid \mathcal{L}_k)\\ \ge&(1-O(n^{-1/2}))^l\ge 1-n^{-1/3}.\end{aligned}$$

This completes the proof. □

Proof of Lemma 4.4.4

Proof. Consider the exploration process of the neighborhood of a fixed vertex i. Conditioned on $\mathcal{F}_{k-1}$, there are two ways to create new cycles in $V_{\ge k-1}(i)$:

1. Type 1: a new hyperedge $e\subset V_{\ge k-1}(i)$ containing two vertices in $V_{k-1}(i)$ may appear, which creates a cycle including two vertices in $V_{k-1}(i)$.

2. Type 2: two vertices in $V_{k-1}(i)$ may be connected to the same vertex in $V_{\ge k}(i)$ by two new distinct hyperedges.

Define the event

$$\Omega_{k-1}(i) := \{S_{k-1}(i)\le C\log(n)\alpha^{k-1}\}, \tag{4.12.1}$$

where the constant C is the same one as in Theorem 4.4.2. From the proof of Theorem 4.4.2, $\mathbb{P}_{\mathcal{H}_n}(\Omega_k(i))=1-O(n^{-1-\gamma})$ for some $\gamma\in(0,3/8)$. Let $E_k^{(1)}(i)$ be the number of hyperedges of type 1. Conditioned on $\mathcal{F}_{k-1}$, $E_k^{(1)}(i)$ is stochastically dominated by $\mathrm{Bin}\left(\binom{S_{k-1}(i)}{2}\binom{n}{d-2}, \frac{a\vee b}{\binom{n}{d-1}}\right)$. Then for some constant $C_1>0$,

$$\mathbb{E}_{\mathcal{H}_n}[E_k^{(1)}(i)\mid \Omega_{k-1}(i)]\le C_1\log^2(n)\alpha^{2k-2}/n\le C_1\log^2(n)\alpha^{2l}/n.$$

By Markov's inequality,

$$\begin{aligned}\mathbb{P}_{\mathcal{H}_n}(\{E_k^{(1)}(i) \geq 1\}) \leq & \mathbb{P}_{\mathcal{H}_n}(\{E_k^{(1)}(i) \geq 1\} \mid \Omega_{k-1}(i)) + \mathbb{P}_{\mathcal{H}_n}(\Omega_{k-1}^c(i)) \\ \leq & \mathbb{E}_{\mathcal{H}_n}[E_k^{(1)}(i) \mid \Omega_{k-1}(i)] + O(n^{-1-\gamma}) = O(\log^2(n)\alpha^{2l}/n).\end{aligned}$$

Taking the union bound, the probability that there is a type 1 hyperedge in the l-neighborhood of i is

$$\mathbb{P}_{\mathcal{H}_n}\left(\bigcup_{k=1}^{l}\{E_k^{(1)}(i) \geq 1\}\right) \leq \sum_{k=1}^{l} \mathbb{P}_{\mathcal{H}_n}(\{E_k^{(1)}(i) \geq 1\}) = O(\log^3(n)\alpha^{2l}/n).$$

The number of hyperedge pair (e_1, e_2) of Type 2 is stochastically dominated by

$$\text{Bin}\left(nS_{k-1}^2\binom{n}{d-2}^2, \left(\frac{a \vee b}{\binom{n}{d-1}}\right)^2\right),$$

which conditioned on $\Omega_{k-1}(i)$ has expectation $O(\log^2(n)\alpha^{2l}/n)$. By a Markov's inequality and a union bound, in the same way as the proof for Type 1, we have the probability there is a type 2 hyperedge pair in the l-neighborhood of i is $O(\log^2(n)\alpha^{2l}/n)$. Altogether the probability that there are at least one cycles within the $l-$neighborhood of i is $O(\log^3(n)\alpha^{2l}/n)$.

Let Z_i be the random variable such that $Z_i = 1$ if l-neighborhood of i contains one cycle and $Z_i = 0$ otherwise. From the analysis above, we have $\mathbb{E}[Z_i] = O(\log^3(n)\alpha^{2l}/n)$. By Markov's inequality,

$$\mathbb{P}_{\mathcal{H}_n}\left(\sum_{i\in[n]} Z_i \geq \alpha^{2l}\log^4(n)\right) \leq \frac{\sum_i \mathbb{E}[Z_i]}{\log^4(n)\alpha^{2l}} = \frac{O(\log^3(n)\alpha^{2l})}{\alpha^{2l}\log^4(n)} = O(\log^{-1}(n)).$$

Then asymptotically almost surely the number of vertices whose l-neighborhood contains one cycle at most $\log^4(n)\alpha^{2l}$. It remains to show H is l-tangle free asymptotically almost surely. For a fixed vertex $i \in [n]$, there are several possible cases where there can be two cycles in $V_{\leq l}(i)$.

(1) There is one hyperedge of Type 1 or a hyperedge pair of Type 2 which creates more than one cycles. We discuss in the following cases conditioned on the event $\cap_{t=1}^{l}\Omega_t(i)$.

(a) The number of hyperedge of the first type which connects to more than two vertices in V_{k-1} is stochastically dominated by $\text{Bin}\left(\binom{S_{k-1}}{3}\binom{n}{d-3}, \frac{a\vee b}{\binom{n}{d-1}}\right)$. The expectation is at most $O(\alpha^{3l}\log^3(n)/n^2)$.

(b) If the intersection of the hyperedge pair of Type 2 contains 2 vertices in $V_{\geq k}$, it will create two cycles. The number of such hyperedge pairs is stochastically dominated by

$$\text{Bin}\left(\binom{n}{2}S_{k-1}^2\binom{n}{d-3}^2, \left(\frac{a\vee b}{\binom{n}{d-1}}\right)^2\right)$$

with mean $O(\log^2(n)\alpha^{2l}/n^2)$.

Then by Markov's inequality and a union bound, asymptotically almost surely, there is no $V_{\leq l}(i)$ such that its neighborhood contains Type 1 hyperedges or Type 2 hyperedge pairs which create more than one cycles.

(2) The remaining case is that there is a $V_{\leq l}(i)$ where two cycles are created by two Type 1 hyperedges or two Type 2 hyperedge pairs or one Type 1 hyperedge and another hyperedge pairs. By the same argument, under the event $\cap_{t=1}^{l}\Omega_t(i)$, the probability that such event happens is $O(\log^6(n)\alpha^{4l}/n^2)$. Since $\alpha^{4l}=o(n)$, by taking a union bound over $i\in[n]$, we have H is l-tangle-free asymptotically almost surely. □

Proof of Lemma 4.4.5

Proof. Let $i\notin\mathcal{B}$ whose l-neighborhood contains no cycles. For any $k\in[n]$ and any $m\leq l$, there is a unique self-avoiding walk of length m from i to k if and only if $d(i,k)=m$, so we have

$B_{ik}^{(m)} = \mathbf{1}_{d(i,k)=m}$. For such i we have

$$(B^{(m)}\mathbf{1})_i = S_m(i), \quad (B^{(m)}\sigma)_i = D_m(i).$$

Then (4.4.5), (4.4.6) follows from Theorem 4.4.2. By Lemma 4.4.4, asymptotically almost surely all vertices in $\mathcal{B}$ have only one cycle in l-neighborhood. For any $m \leq l, i \in \mathcal{B}$, since $(B^{(m)}\mathbf{1})_i = \sum_{k\in[n]} B_{ik}^{(m)}$, and only vertices at distance at most m from i can be reached by a self-avoiding walk of length m from i, which will be counted in $(B^{(m)}\mathbf{1})_i$. Moreover, for any $k \in [n]$ with $B_{ik}^{(m)} \neq 0$, since the l-neighborhood of i contains at most one cycle, there are at most 2 self-avoiding walks of length m between i and k. Altogether we know

$$\sum_{k\in[n]} B_{ik}^{(m)} \leq 2\sum_{t=0}^{m} S_t(i) = O(\alpha^m \log n)$$

asymptotically almost surely. Then (4.4.7) follows. $\square$

Proof of Lemma 4.5.3

Proof. Recall the definitions of α, β from (4.1.3). From (4.5.1)-(4.5.3),

$$\begin{aligned}
&\mathbb{E}(W_{t+1}^{+}|\mathcal{G}_t) \\
=&\sum_{r=0}^{d-1} r\mathbb{E}(W_{t+1}^{(r)}|\mathcal{G}_t) = \sum_{r=1}^{d-2} r\left(\frac{b\binom{d-1}{r}}{2^{d-1}}(W_t^{-}+W_t^{+})\right) + (d-1)\left(\frac{a}{2^{d-1}}W_t^{+} + \frac{b}{2^{d-1}}W_t^{-}\right) \\
=&\frac{\alpha+\beta}{2}W_t^{+} + \frac{\alpha-\beta}{2}W_t^{-} = \frac{\alpha^{t+1}}{2}M_t + \frac{\beta^{t+1}}{2}\Delta_t.
\end{aligned}$$

Similarly, $\mathbb{E}[W_{t+1}^-|\mathcal{G}_t] = \frac{\alpha^{t+1}}{2}M_t - \frac{\beta^{t+1}}{2}\Delta_t$. Therefore

$$\mathbb{E}[M_{t+1} \mid \mathcal{G}_t] = \alpha^{-t-1}\mathbb{E}[W_{t+1}^+ + W_{t+1}^- \mid \mathcal{G}_t] = M_t,$$
$$\mathbb{E}[\Delta_{t+1} \mid \mathcal{G}_t] = \beta^{-t-1}\mathbb{E}[W_{t+1}^+ - W_{t+1}^- \mid \mathcal{G}_t] = \Delta_t.$$

It follows that $\{M_t\}, \{\Delta_t\}$ are martingales with respect to $\mathcal{G}_t$. From (4.5.1)-(4.5.4),

$$\begin{aligned}\mathrm{Var}(M_t|\mathcal{G}_{t-1}) &= \mathrm{Var}(\alpha^{-t}(W_t^+ + W_t^-)|\mathcal{G}_{t-1}) = \alpha^{-2t}\mathrm{Var}\left((d-1)\sum_{r=0}^{d-1} W_t^{(r)}|\mathcal{G}_{t-1}\right)\\ &= (d-1)^2\alpha^{-2t} \cdot \frac{\alpha}{d-1}(W_{t-1}^+ + W_{t-1}^-) = (d-1)\alpha^{-t}M_{t-1}.\end{aligned}$$

Sine $\mathbb{E}M_0 = 1$, by conditional variance formula,

$$\mathrm{Var}(M_t) = \mathrm{Var}(\mathbb{E}[M_t|\mathcal{G}_{t-1}]) + \mathbb{E}\mathrm{Var}(M_t|\mathcal{G}_{t-1}) = \mathrm{Var}(M_{t-1}) + (d-1)\alpha^{-t}.$$

Since $\mathrm{Var}(M_0) = 0$, we have for $t \geq 0$, $\mathrm{Var}(M_t) = (d-1)\frac{1-\alpha^{-t}}{\alpha-1}$. So $\{M_t\}$ is uniformly integrable for $\alpha > 1$. Similarly,

$$\begin{aligned}\mathrm{Var}(\Delta_t|\mathcal{G}_{t-1}) &= \mathrm{Var}(\beta^{-t}(W_t^+ - W_t^-)|\mathcal{G}_{t-1}) = \beta^{-2t}\sum_{r=0}^{d-1}(2r-d+1)^2\mathrm{Var}(W_t^{(r)}|\mathcal{G}_{t-1})\\ &= (\alpha/\beta^2)^t M_{t-1}(d-1)\alpha^{-1} \cdot \frac{(d-1)a + (2^{d-1}+1-d)b}{2^{d-1}} =: \kappa(\alpha/\beta^2)^t M_{t-1},\end{aligned}$$

where $\kappa := \frac{(d-1)(a-b)+2^{d-1}b}{a+(2^{d-1}-1)b}$. And we also have the following recursion:

$$\mathrm{Var}(\Delta_t) = \mathrm{Var}(\mathbb{E}[\Delta_t|\mathcal{G}_{t-1}]) + \mathbb{E}\mathrm{Var}(\Delta_t|\mathcal{G}_{t-1}) = \mathrm{Var}(\Delta_{t-1}) + \kappa\beta^{-2t}\alpha^t.$$

Since $\mathrm{Var}(\Delta_0) = 0$, we have for $t > 0$,

$$\mathrm{Var}(\Delta_t) = \kappa \cdot \frac{1 - (\beta^2/\alpha)^{-t}}{\beta^2/\alpha - 1}. \tag{4.12.2}$$

So $\{\Delta_t\}$ is uniformly integrable if $\beta^2 > \alpha$. From the martingale convergence theorem, $\mathbb{E}\Delta_\infty = \Delta_0 = 1$, $\mathrm{Var}(\Delta_\infty) = \frac{\kappa}{\beta^2/\alpha - 1}$, and (4.5.5) holds. This finishes the proof. $\square$

Proof of Lemma 4.5.4

Proof. From Theorem 4.5.2, For each $i \in [n]$, there exists a coupling such that with probability $1 - O(n^{-\varepsilon})$ for some positive ε, $\beta^{-l}\sigma(i)D_l(i) = \Delta_l$ and we denote this event by $\mathcal{C}$. When the coupling fails, by Theorem 4.4.2, $\beta^{-l}\sigma(i)D_l(i) = O(\log(n))$ with probability $1 - O(n^{-\gamma})$ for some $\gamma > 0$. Recall the event

$$\Omega_{k-1}(i) := \{S_{k-1}(i) \le C\log(n)\alpha^{k-1}\}. \tag{4.12.3}$$

We define $\Omega := \bigcap_{i=1}^n \Omega(i), \Omega(i) := \bigcap_{k \le l} \Omega_k(i)$. We have

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^n \beta^{-2l} D_l^2(i) \mid \Omega\right) = O(\log^2(n))n^{-\varepsilon} + \mathbb{E}(\Delta_l^2 \mathbf{1}_{\mathcal{C}} \mid \Omega). \tag{4.12.4}$$

Moreover,

$$\begin{aligned} |\mathbb{E}(\Delta_l^2 \mathbf{1}_{\mathcal{C}}|\Omega) - \mathbb{E}(\Delta_\infty^2)| &= \left|\frac{\mathbb{E}(\Delta_l^2 \mathbf{1}_{\mathcal{C}} - \mathbb{E}(\Delta_l^2 \mathbf{1}_{\mathcal{C}} \mathbf{1}_{\overline{\Omega}}) - \mathbb{P}(\Omega)\mathbb{E}(\Delta_\infty^2)}{\mathbb{P}(\Omega)}\right| \\ &\le \frac{|\mathbb{E}(\Delta_l^2 - \Delta_\infty^2)|}{\mathbb{P}(\Omega)} + \frac{1 - \mathbb{P}(\Omega)}{\mathbb{P}(\Omega)}\mathbb{E}(\Delta_\infty^2) + \frac{|\mathbb{E}(\Delta_l^2 \mathbf{1}_{\overline{\mathcal{C}}}) - \mathbb{E}(\Delta_l^2 \mathbf{1}_{\mathcal{C} \cap \overline{\Omega}})|}{\mathbb{P}(\Omega)}. \end{aligned} \tag{4.12.5}$$

Since we know $\mathbb{P}(\Omega \cap \mathcal{C}) \to 1$ and (4.5.5), the first two terms in (4.12.5) converges to 0.

The third term also converges to 0 by dominated convergence theorem. So we have

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\beta^{-2l}D_l^2(i) \mid \Omega\right) \to \mathbb{E}(\Delta_\infty^2).$$

We then estimate the second moment. Note that

$$\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\beta^{-2l}D_l^2(i) \mid \Omega\right)^2 = \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n}\beta^{-4l}D_l^4(i) \mid \Omega\right) + \frac{2}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(D_l(i)^2 D_l^2(j) \mid \Omega), \tag{4.12.6}$$

and from Theorem 4.4.2, the first term is $O(\log^4(n)/n) = o(1)$. Next, we show the second term satisfies

$$\frac{2}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(D_l(i)^2 D_l^2(j) \mid \Omega) = \frac{2}{n^2}\sum_{i<j}\beta^{-4l}\frac{1}{\mathbb{P}(\Omega)}\mathbb{E}(\mathbf{1}_\Omega D_l(i)^2 D_l^2(j)) = o(1). \tag{4.12.7}$$

Since $\mathbb{P}(\Omega) = 1 - O(n^{-\gamma})$, it suffices to show

$$\frac{2}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(\mathbf{1}_\Omega D_l(i)^2 D_l^2(j)) = o(1).$$

Consider $\beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)\cap\Omega(j)}D_l^2(i)D_l^2(j))$. From Lemma 4.4.3, when $i \neq j$, $D_l(i), D_l(j)$ are asymptotically independent. On the event that the coupling with independent copies fails (recall the failure probability is $O(n^{-\gamma})$), we bound $D_l^2(i)D_l^2(j)$ by $O(\beta^{4l}\log^4(n))$. When the coupling succeeds,

$$\beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)\cap\Omega(j)}D_l(i)^2 D_l^2(j)) = \beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)}D_l(i)^2)\mathbb{E}(\mathbf{1}_{\Omega(j)}D_l(j)^2).$$

Then from (4.5.6),

$$
\begin{aligned}
&\frac{2}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)\cap\Omega(j)}D_l(i)^2D_l^2(j))\\
=&O\left(\frac{1}{n^2}\sum_{i<j}\beta^{-4l}\mathbb{E}(\mathbf{1}_{\Omega(i)}D_l(i)^2)\mathbb{E}(\mathbf{1}_{\Omega(j)}D_l(j)^2)+O(n^{-2\gamma}\log^4 n)\right)\\
=&O\left((\mathbb{E}(\Delta_\infty^2))^2\right)=O(1). \qquad (4.12.8)
\end{aligned}
$$

Therefore from (4.12.6), (4.12.7), and (4.12.8),

$$
\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n}\beta^{-2l}D_l^2(i)\mid\Omega\right)^2=O(1).
$$

With (4.12.4), by Chebyshev's inequality, conditioned on Ω, in probability we have

$$
\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\beta^{-2l}D_l^2(i)=\mathbb{E}(\Delta_\infty^2).
$$

Since $\mathbb{P}(\Omega)\to 1$, (4.5.6) follows.

We now establish (4.5.7). Without loss of generality, we discuss the case of $+$ sign. Since τ is a continuous point of the distribution of Δ_∞, for any fixed $\delta>0$, we can find two bounded K-Lipschitz function f,g for some constant $K>0$ such that

$$
f(x)\le(\mathbf{1}_{x\ge\tau})\le g(x), x\in\mathbb{R},\quad 0\le\mathbb{E}(g(\Delta_\infty)-f(\Delta_\infty))\le\delta.
$$

Consider the empirical sum $\frac{1}{n}\sum_{i\in\mathcal{N}^+}f(x_i^{(n)}\sqrt{n\mathbb{E}(\Delta_\infty^2)})$, we have

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i\in\mathcal{N}^+}f(x_i^{(n)}\sqrt{n\mathbb{E}\Delta_\infty^2})-\frac{1}{n}\sum_{i\in\mathcal{N}^+}f(\beta^{-l}D_l(i))\right|\\
\le&\frac{K}{n}\sum_{i\in\mathcal{N}^+}|(x_i^{(n)}-y_i^{(n)})\sqrt{n\mathbb{E}\Delta_\infty^2}|+\frac{K}{n}\sum_{i\in\mathcal{N}^+}|y_i^{(n)}\sqrt{n\mathbb{E}\Delta_\infty^2}-\beta^{-l}D_l(i)|.
\end{aligned}
$$

The first term converges to 0 by the assumption that $\|x-y\|_2 \to 0$ in probability. The second term converges to 0 in probability from (4.5.6). Moreover, $\frac{1}{n}\sum_{i\in\mathcal{N}^+} f(\beta^{-l}D_l(i))$ converges in probability to $\frac{1}{2}\mathbb{E}f(\Delta_\infty)$. So we have

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i\in\mathcal{N}^+} f(x_i^{(n)}\sqrt{n\mathbb{E}\Delta_\infty^2}) = \frac{1}{2}\mathbb{E}f(\Delta_\infty),$$

and the same holds for g. If follows that

$$\limsup_{n\to\infty}\left|\frac{1}{n}\sum_{i\in[n]:\sigma_i=+}\mathbf{1}_{\left\{x_i^{(n)}\geq\tau/\sqrt{n\mathbb{E}[\Delta_\infty^2]}\right\}} - \frac{1}{2}\mathbb{P}(\Delta_\infty\geq\tau)\right| \leq \delta$$

for any $\delta > 0$. Therefore (4.5.7) holds. $\square$

Proof of Lemma 4.7.2

Proof. For any $n\times n$ real matrix M, we have $\rho(M)^{2k} \leq \mathrm{tr}[(MM^\top)^k]$, therefore

$$\begin{aligned}\mathbb{E}_{\mathcal{H}_n}[\rho(\Gamma^{(l,m)})^{2k}] &\leq \mathbb{E}_{\mathcal{H}_n}\left[\mathrm{tr}\left(\Gamma^{(l,m)}\Gamma^{(l,m)\top}\right)^k\right] \\ &= \sum_{i_1,\dots,i_{2k}\in[n]}\mathbb{E}_{\mathcal{H}_n}\left[\Gamma^{(l,m)}_{i_1i_2}\Gamma^{(l,m)}_{i_3i_2}\dots\Gamma^{(l,m)}_{i_{2k-1}i_{2k}}\Gamma^{(l,m)}_{i_1i_{2k}}\right].\end{aligned} \tag{4.12.9}$$

Recall the definition of $\Gamma^{(l,m)}_{ij}$ from (4.7.2), the sum in (4.12.9) can be expanded to be the sum over all circuits $w=(w_1,\dots w_{2k})$ of length $2kl$ which are obtained by concatenation of $2k$ walks of length l, and each $w_i, 1\leq i\leq 2k$ is a concatenation of two self-avoiding walks of length $l-m$ and $m-1$. The weight that each hyperedge in the circuit contributes can be either $A^e_{ij}-\overline{A^e_{ij}}, \overline{A^e_{ij}}$ or A^e_{ij}. For all circuits w in (4.12.9) with nonzero expected weights, there is an extra constraint that each w_i intersects with some other w_j, otherwise the expected weight that w_i contributes to the sum (4.12.9) will be 0. We want to bound the number of such circuits with nonzero expectation.

Let v,h denoted the number of distinct vertices and hyperedges traversed by the circuit. Here we don't count the hyperedges that are weighted by $\overline{A_{ij}^{e}}$. We associate a multigraph $G(w)$ for each w as before, but the hyperedges with weight $\overline{A_{ij}^{e}}$ are not included. Since $\mathbb{E}_{\mathcal{H}_n}[\Gamma_{ij}^{(l,m)}] = 0$ for any $i, j \in [n]$, if the expected weight of w is nonzero, the corresponding graph $G(w)$ must be connected.

We detail the proof for circuits in Case (1), where

- each hyperedge label in $\{e_i\}_{1\leq i\leq h}$ appears exactly once on $G(w)$;
- vertices in $e_i \setminus \mathrm{end}(e_i)$ are all distinct for $1 \leq i \leq h$, and they are not vertices with labels in $V(w)$,

and the cases for other circuits follow similarly from the proof of Lemma 4.7.1.

Let m be fixed. For each circuit w, there are $4k$ self-avoiding walks, and each w_i is broken into two self-avoiding walks of length $m-1$ and $l-m$, respectively. We adopt the way of encoding each self-avoiding walk as before, except that we must also include the labels of the endpoint j after the traversal of an edge e with weight from $\overline{A_{ij}^{e}}$, which gives us the initial vertex of the self-avoiding walk of length $l-m$ within each w_i. These extra labels tell us how to concatenate the two self-avoiding walks of length $m-1$ and $l-m$ into the walk w_i of length l. For each w_i, label is encoded by a number from $\{1,\dots,v\}$. So all possible such labels can be bounded by v^{2k}. Then the upper bound on the number of valid triplet sequences with extra labels for fixed v,h is now given by $v^{2k}[(v+1)^2(l+1)]^{4k(2+h-v)}$.

The total number of circuits that have the same triplet sequences with extra labels is at most $n^v\binom{n}{d-2}^{h+2k}$ where $h+2k$ is the total number of distinct hyperedges we can have in w, including the hyperedges with weights from $\overline{A_{ij}^{e}}$.

We also need to bound the possible range of v,h. There are overall $2k(l-1)$ hyperedges traversed in w (remember we don't count the edges with weights from $\overline{A_{ij}^{e}}$). Out of these, $2k(l-m)$ hyperedges (with multiplicity) with weights coming from $A_{ij}^{e} - \overline{A_{ij}^{e}}$ must be at least doubled for the expectation not to vanish. Then the number of distinct hyperedges in w excluding the hyperedge weighted by some $\overline{A^e_{ij}}$, satisfies $h \leq k(l-m)+(2k(l-1)-2k(l-m)) = k(l+m-2)$. We have $v \geq \max\{m, l-m+1\}$ since each self-avoiding walk of length $m-1$ or $l-m$ has distinct vertices. Moreover, since $G(w)$ is connected, $h \geq v-1$, so we have $v-1 \leq h \leq k(l+m-2)$. And the range of v is then given by $\max\{m, l-m+1\} \leq v \leq k(l+m-2)+1$.

The expected weight that a circuit contributes can be estimated similarly as before. From (4.7.14), the expected weights from $v-1$ many hyperedges that corresponds to edges on $T(w)$ is bounded by $\left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1}$. Similar to (4.7.10), the expected weights from $h-v+1+2k$ many hyperedges that corresponds to edges on $G(w)\setminus T(w)$ together with hyperedges whose weights are from $\overline{A^e_{ij}}$ is bounded by $\left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{h-v+1+2k}$.

Putting all estimates together, for fixed v, h, the total contribution to the sum is bounded by

$$\begin{aligned} & n^v \binom{n}{d-2}^{h+2k} v^{2k}[(v+1)^2(l+1)]^{4k(2+h-v)} \left(\frac{\alpha}{(d-1)\binom{n}{d-1}}\right)^{v-1} \left(\frac{a\vee b}{\binom{n}{d-1}}\right)^{h-v+1+2k} \\ =& n^v \left(\frac{\alpha}{d-1}\right)^{v-1} \left(\frac{d-1}{n-d+2}\right)^{h+2k} v^{2k}\, Q(k,l,v,h), \end{aligned}$$

where $Q(k,l,v,h) := [(v+1)^2(l+1)]^{4k(2+h-v)}\,(a\vee b)^{h-v+1+2k}$.

Let S_1 be the contribution of circuits in Case (1) to the sum in (4.12.9). We have

$$S_1 \leq \sum_{v=m\vee(l-m+1)}^{k(l+m-2)+1} \sum_{h=v-1}^{k(l+m-2)} n^v \left(\frac{\alpha}{d-1}\right)^{v-1} \left(\frac{d-1}{n-d+2}\right)^{h+2k} v^{2k}\, Q(k,l,v,h). \tag{4.12.10}$$

Taking $l = O(\log n)$, similar to the discussion in (4.7.16), the leading term in (4.12.10) is given by the term with $h = v-1$. So for any $1 \leq m \leq l$, and sufficiently large n, there are constants

$C_1, C_2 > 0$ such that

$$
\begin{aligned}
S_1 \leq & \sum_{v=m\vee(l-m+1)}^{k(l+m-2)+1} 2n^{1-2k}((d-1)v)^{2k}[(v+1)^2(l+1)]^{4k}\alpha^{v-1}\,(a\vee b)^{2k} \\
\leq & C_2 \log^{14k}(n)\cdot n^{1-2k}\alpha^{k(l+m-2)}.
\end{aligned}
$$

For circuits not in Case (1), similar to the proof of Lemma 4.7.1, their total contribution is bounded by $C_2' n^{1-2k}\alpha^{k(l+m-2)}\log^{14k} n$ for a constant $C_2' > 0$. This completes the proof of Lemma 4.7.2. $\square$

Proof of Lemma 4.9.1

Proof. Let $\mathcal{B}$ be the set of vertices such that their l-neighborhood contains a cycle. Let x be a normed vector such that $x^\top B^{(l)}\mathbf{1} = 0$. We then have

$$
\begin{aligned}
\mathbf{1}^\top B^{(m-1)}x &= \sum_{i\in[n]} x_i (B^{(m-1)}\mathbf{1})_i = \sum_{i\notin\mathcal{B}} x_i S_{m-1}(i) + \sum_{i\in\mathcal{B}} x_i (B^{m-1}\mathbf{1})_i \\
&= \sum_{i\in[n]} x_i(\alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{\frac{m-1}{2}}\log n)) \\
&\quad - \sum_{i\in\mathcal{B}} x_i(\alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{\frac{m-1}{2}}\log n)) + \sum_{i\in\mathcal{B}} x_i (B^{(m-1)}\mathbf{1})_i.
\end{aligned}
\tag{4.12.11}
$$

Since we have $\mathbf{1}^\top B^{(l)}x = 0$, the first term in (4.12.11) satisfies

$$
\left|\sum_{i\in[n]} x_i(\alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{\frac{m-1}{2}}\log n))\right| = \left|\sum_{i\in[n]} x_i O(\alpha^{\frac{m-1}{2}}\log n)\right| = O(\sqrt{n}\alpha^{\frac{m-1}{2}}\log n),
$$

where the last inequality above is from Cauchy inequality.

From Lemma 4.4.4, $|\mathcal{B}| = O(\alpha^{2l}\log^4 n)$. For the second term in (4.12.11), recall from

(4.4.7), for $m \le l$, $|(B^{(m)}\mathbf{1})_i| = O(\alpha^m \log n)$, then by Cauchy inequality

$$\left|\sum_{i\in\mathcal{B}} x_i(\alpha^{m-1-l}(B^{(l)}\mathbf{1})_i + O(\alpha^{\frac{m-1}{2}}\log n))\right| \le \sqrt{|\mathcal{B}|}O(\alpha^{m-1}\log n) = O(\alpha^{l+m-1}\log^3 n).$$

Similarly, the third term satisfies

$$|\sum_{i\in\mathcal{B}} x_i(B^{(m-1)}\mathbf{1})_i| = O(\alpha^{l+m-1}\log^3 n).$$

Note that $\alpha^{l+m-1} = o(n^{1/2})$, altogether we have

$$|\mathbf{1}^\top B^{(m-1)}x| = O(\sqrt{n}\alpha^{\frac{m-1}{2}}\log n + \alpha^{l+m-1}\log^3 n) = O(\sqrt{n}\alpha^{\frac{m-1}{2}}\log n). \tag{4.12.12}$$

(4.9.1) then follows. Using the property $x^\top B^{(l)}\sigma = 0$ instead of $x^\top B^{(l)}\mathbf{1} = 0$ and following the same argument, (4.9.2) holds. □

Proof of Lemma 4.10.1

Proof. Conditioned on $(H,i,\sigma)_{t-1} \equiv (T,\rho,\tau)_{t-1}$, if A_t holds, it implies that hyperedges generated from vertices in V_{t-1} do not overlap (except for the parent vertices in V_{t-1}). If B_t holds, vertices in V_t that are in different hyperedges generated from H_{t-1} do not connect to each other. If both A_t B_t holds, $(H,i,\sigma)_t$ is still a hypertree. Since $X_v^{(r)} = Y_{\phi(v)}^{(r)}$ for $v \in V_{t-1}$, we can extend the hypergraph isomorphism ϕ by mapping the children of $v \in V_t$ to the corresponding vertices in the t-th generation of children of ρ in T, which keeps the hypertree structure and the spin of each vertex. □

Proof of Lemma 4.10.2

Proof. First we fix $u, v \in V_t$. For any $w \in V_{>t}$, the probability that $(u,w),(v,w)$ are both connected is $O(n^{-2})$. We know $|V_{>t}| \leq n$ and $|V_{\leq t}| = O(\log^2(n)\alpha^t)$ conditioned on C_t. Since $\alpha^{2t} \leq \alpha^{2l} = o(n^{1/2})$, taking a union bound over all u, v, w we have

$$\mathbb{P}(A_t|C_t) \geq 1 - O(\log^4(n)\alpha^{2t}n^{-1}) = 1 - o(n^{-1/2}). \tag{4.12.13}$$

For the second claim, the probability of having an edge between $u, v \in V_t$ is $O(n^{-1})$. Taking a union bound over all pairs of $u, v \in V_t$ implies

$$\mathbb{P}(B_t|C_t) \geq 1 - O(\log^4(n)\alpha^{2t}n^{-1}) = 1 - o(n^{-1/2}). \tag{4.12.14}$$

□

Proof of Lemma 4.11.1

Proof. In (4.11.1), the coordinates of two vectors on the left hand side agree at i if the l-neighborhood of l contains no cycle. Recall $\mathcal{B}$ is the set of vertices whose l-neighborhood contains a cycle, from Lemma 4.4.4, and (4.4.7), we have asymptotically almost surely,

$$\|B^{(l)}\mathbf{1} - \vec{S}_l\|_2 \leq \sqrt{|\mathcal{B}|}O(\log(n)\alpha^l) = O(\log^3(n)\alpha^{2l}) = o(\sqrt{n}). \tag{4.12.15}$$

From (4.5.6) we have

$$\|\vec{D}_l\|_2 = \Theta(\sqrt{n}\beta^l) \tag{4.12.16}$$

asymptotically almost surely, and $\|B^{(l)}\mathbf{1}\|_2 \geq \|\vec{D}_l\|_2$, therefore (4.11.1) follows. Similar to

(4.12.15), we have

$$\|B^{(l)}\sigma - \vec{D}_l\|_2 = o(\sqrt{n}), \quad \|B^{(l)}\sigma\|_2 = \|\vec{D}_l\|_2 + o(\sqrt{n}) = \Theta(\sqrt{n}\beta^l). \tag{4.12.17}$$

Then (4.11.2) follows. It remains to show (4.11.3). Using the same argument as in Theorem 4.5.4, we have the following convergence in probability

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} \alpha^{-2l} S_l^2(i) = \mathbb{E}M_\infty^2, \tag{4.12.18}$$

where M_∞ is the limit of the martingale M_t. Similarly, the following convergences in probability hold

$$\begin{aligned}\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} \alpha^{-l}\beta^{-l} S_l(i) D_l(i) &= \lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathcal{N}^+} \alpha^{-l}\beta^{-l} S_l(i) D_l(i) + \lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathcal{N}^-} \alpha^{-l}\beta^{-l} S_l(i) D_l(i)\\ &= \frac{1}{2}\mathbb{E}M_\infty D_\infty - \frac{1}{2}\mathbb{E}M_\infty D_\infty = 0.\end{aligned}$$

Thus $\langle \vec{S}_l, \vec{D}_l\rangle = o(n\alpha^l\beta^l)$ asymptotically almost surely. From (4.12.18) we have

$$\|\vec{S}_l\|_2 = \Theta(\sqrt{n}\alpha^l), \tag{4.12.19}$$

therefore together with (4.12.16), we have $\|\vec{S}_l\|_2 \cdot \|\vec{D}_l\|_2 = \Theta(n\alpha^l\beta^l)$. With (4.11.1) and (4.11.2), (4.11.3) holds. $\square$

Proof of Lemma 4.11.2

Proof. For the lower bound in (4.11.4), note that $B^{(l)}$ is symmetric, we have

$$\|B^{(l)}\mathbf{1}\|_2^2 = \langle B^{(l)}\mathbf{1}, B^{(l)}\mathbf{1}\rangle = \langle \mathbf{1}, B^{(l)}B^{(l)}\mathbf{1}\rangle \le \|\mathbf{1}\|_2 \|B^{(l)}B^{(l)}\mathbf{1}\|_2. \tag{4.12.20}$$

Therefore from (4.12.19) and (4.11.1),

$$\|B^{(l)}B^{(l)}\mathbf{1}\|_2 \geq \frac{\|B^{(l)}\mathbf{1}\|_2^2}{\|\mathbf{1}\|_2} = \Theta(\alpha^l)\|B^{(l)}\mathbf{1}\|_2. \tag{4.12.21}$$

For the upper bound in (4.11.4), from (4.4.1) and (4.4.7), the maximum row sum of $B^{(l)}$ is $O(\alpha^l \log n)$. Since $B^{(l)}$ is nonnegative, the spectral norm $\rho(B^{(l)})$ is bounded by the maximal row sum, hence (4.11.4) holds. The lower bound in (4.11.5) can be proved similarly as in (4.11.4), from the inequality $\|B^{(l)}\sigma\|_2^2 \leq \|\sigma\|_2\|B^{(l)}B^{(l)}\sigma\|_2$ together with (4.12.16) and (4.11.2). Recall $\mathcal{B}$ is the set of vertices whose l-neighborhood contains cycles. Let $\overline{\mathcal{B}} = [n] \setminus \mathcal{B}$. Since

$$\left(B^{(l)}B^{(l)}\sigma\right)_i = \sum_{j\in[n]} B_{ij}^{(l)}(B^{(l)}\sigma)_j,$$

we can decompose the vector $B^{(l)}B^{(l)}\sigma$ as a sum of three vectors $z+z'+z''$, where

$$\begin{aligned}
&z_i := \mathbf{1}_{\overline{\mathcal{B}}}(i)\sum_{j:d(i,j)=l} D_l(j)\mathbf{1}_{\overline{\mathcal{B}}}(j), \quad z_i' := \mathbf{1}_{\overline{\mathcal{B}}}(i)\sum_{j:d(i,j)=l} O(\alpha^l \log n)\mathbf{1}_{\mathcal{B}}(j),\\
&z_i'' := \mathbf{1}_{\mathcal{B}}(i)O(\alpha^{2l}\log^2 n).
\end{aligned}$$

The decomposition above depends on whether $i, j \in \mathcal{B}$ and the estimation follows from (4.4.7). From Lemma 4.4.4, $\mathcal{B} = O(\alpha^{2l}\log^4(n))$ asymptotically almost surely, so one has

$$\begin{aligned}
\|z'\|_2^2 &= \sum_{i=1}^n (z_i')^2 = \sum_{i\in\overline{\mathcal{B}}}\sum_{j:d(i,j)=l}\sum_{j':d(i,j')=l} O(\alpha^{2l}\log^2 n)\mathbf{1}_{\mathcal{B}}(j)\mathbf{1}_{\mathcal{B}}(j')\\
&= \sum_{j\in\mathcal{B}}\sum_{j'\in\mathcal{B}}\sum_{\substack{i\in\overline{\mathcal{B}}\\ d(i,j)=d(i,j')=l}} O(\alpha^{2l}\log^2 n) = \sum_{j,j'\in\mathcal{B}} O(\alpha^{3l}\log^3 n) = O(\alpha^{7l}\log^{11} n),
\end{aligned}$$

which implies $\|z'\|_2 = O(\alpha^{7l/2}\log^{11/2} n)$. And similarly $\|z''\|_2 = O(\alpha^{3l}\log^2 n)$. We know from (4.12.17), $\|B^{(l)}\sigma\|_2 = \Theta(\beta^l\sqrt{n})$, and since $c\log\alpha < 1/8$, we have $\alpha^{5l/2} = n^{-\gamma'}\sqrt{n}$ for some $\gamma' > 0$, therefore

$$\|z'+z''\|_2 = O(\alpha^{7l/2}\log^{11/2} n) = o(\alpha^{5l/2}\beta^{2l}) = O(n^{-\gamma}\beta^l \|B^{(l)}\sigma\|_2). \tag{4.12.22}$$

It remains to upper bound $\|z\|_2$. Assume the $2l$-neighborhood of i is cycle-free, then the i-th entry of $B^{(l)}B^{(l)}\sigma$, denoted by X_i, can be written as

$$\begin{aligned} X_i := (B^{(l)}B^{(l)}\sigma)_i &= \sum_{k=1}^{n} B_{ik}^{(l)}(B^{(l)}\sigma)_k = \sum_{k=1}^{n} \mathbf{1}_{d(i,k)=l} \sum_{j=1}^{n} \mathbf{1}_{d(j,k)=l}\sigma_j \\ &= \sum_{h=0}^{l} \sum_{j:d(i,j)=2h} \sigma_j |\{k : d(i,k) = d(j,k) = l\}|. \end{aligned} \tag{4.12.23}$$

We control the magnitude of X_i in the corresponding hypertree growth process. Since $2l = 2c\log n$ and $2c\log(\alpha) < 1/4$, the coupling result in Theorem 4.5.2 can apply. Let C_i be the event that coupling between $2l$-neighborhood of i with the Poisson Galton-Watson hypertree has succeeded and $n^{-\varepsilon}$ be the failure probability of the coupling. When the coupling succeeds, $z_i = X_i$, therefore

$$\begin{aligned} \mathbb{E}(\|z\|_2^2 \mid \Omega) &= \sum_{i\in[n]} n^{-\varepsilon} O(\alpha^{2l}\beta^{2l}\log^2 n) + \sum_{i\in[n]} \mathbb{E}(X_i^2 \mathbf{1}_{C_i} \mid \Omega) \\ &= n^{1-\varepsilon} O(\alpha^{2l}\beta^{2l}\log^2 n) + \sum_{i\in[n]} \mathbb{E}(X_i^2 \mathbf{1}_{C_i} \mid \Omega). \end{aligned} \tag{4.12.24}$$

For any $i,j \in [n], t \in [l]$, define $D_{i,j}^{(t)} := |\{k : d(i,k) = d(j,k) = t\}|$. From (4.12.23), we have

$$X_i^2 = \sum_{h,h'=0}^{l} \sum_{j:d(i,j)=2h} \sum_{j':d(i,j')=2h'} \sigma_j \sigma_{j'} D_{i,j}^{(l)} D_{i,j'}^{(l)}. \tag{4.12.25}$$

We further classify the pair j, j' in (4.12.25) according to their distance. Let $d(j,j') =$

$2(h+h'-\tau)$ for $\tau = 0,\dots,2(h\wedge h')$. This yields

$$X_i^2 = \sum_{h,h'=0}^{l} \sum_{\tau=0}^{2(h\wedge h')} \sum_{j:d(i,j)=2h} \sum_{j':d(i,j')=2h'} \mathbf{1}_{d(j,j')=2(h+h'-\tau)} \sigma_j \sigma'_j D_{i,j}^{(l)} D_{i,j'}^{(l)}.$$

Conditioned on Ω and $\mathcal{C}_i$, similar to the analysis in Appendix H in [141], we have the following holds

$$|\{k : d(i,k) = d(j,k) = l\}| = O(\alpha^{l-h}\log n), \tag{4.12.26}$$

$$|\{k' : d(i,k') = d(j',k') = l\}| = O(\alpha^{l-h'}\log n), \tag{4.12.27}$$

$$|\{j : d(i,j) = 2h\}| = O(\alpha^{2h}\log n), \tag{4.12.28}$$

$$|\{j' : d(i,j') = 2h', d(j,j') = 2(h+h'-\tau)\}| = O(\alpha^{2h'-\tau}\log n). \tag{4.12.29}$$

We claim that

$$\mathbb{E}[\sigma_j\sigma_{j'}|\mathcal{C}_i] \le \left(\frac{\beta}{\alpha}\right)^{d(j,j')-1}, \tag{4.12.30}$$

and prove (4.12.30) in Cases (a)-(d).

(a) Assume j is the parent of j' in the hypertree growth process. Then $d(j,j') = 1$. Let $\mathcal{T}_r$ be the event that the hyperedge containing j' is of type r. Given $\mathcal{T}_r$, by our construction of the hypertree process, the spin of j' is assigned to be σ_j with probability $\frac{r}{d-1}$ and $-\sigma_j$ with probability $\frac{d-1-r}{d-1}$, so we have

$$\mathbb{E}[\sigma_j\sigma_{j'} \mid \mathcal{C}_i] = \sum_{r=0}^{d-1} \mathbb{E}[\sigma_j\sigma'_j \mid \mathcal{T}_r, \mathcal{C}_i]\mathbb{P}[\mathcal{T}_r \mid \mathcal{C}_i] = \sum_{r=0}^{d-1} \left(\frac{r}{d-1} - \frac{d-1-r}{d-1}\right)\mathbb{P}[\mathcal{T}_r \mid \mathcal{C}_i].$$

Recall $\mathbb{P}[\mathcal{T}_{d-1} \mid \mathcal{C}_i] = \frac{(d-1)a}{\alpha 2^{d-1}}$ and $\mathbb{P}[\mathcal{T}_r \mid \mathcal{C}_i] = \frac{(d-1)b\binom{d-1}{r}}{\alpha 2^{d-1}}$ for $0 \le r \le d-2$. A simple calculation implies $\mathbb{E}[\sigma_j\sigma_{j'} \mid \mathcal{C}_i] = \frac{\beta}{\alpha} \le 1$.

(b) Suppose $d(j,j') = t$ and there is a sequence of vertices $j, j_1, \ldots, j_{t-1}, j'$ such that j_1 is a child of j, j_i is a child of j_{i-1} for $1 \le i \le t$, and j' is a child of j_{t-1}. We show by induction that for $t \ge 1$, $\mathbb{E}[\sigma_j \sigma_{j'} \mid \mathcal{C}_i] = \left(\frac{\beta}{\alpha}\right)^t$. When $t = 1$ this has been approved in part (a). Assume it is true for all j, j' with distance $\le t-1$. Then when $d(j,j') = t$, we have

$$
\begin{aligned}
\mathbb{E}[\sigma_j \sigma_{j'} \mid \mathcal{C}_i] =& \mathbb{E}[\sigma_j \sigma_{j'} \mid \sigma_{j_1} = \sigma_j, \mathcal{C}_i] \mathbb{P}(\sigma_{j_1} = \sigma_j \mid \mathcal{C}_i) \\
&+ \mathbb{E}[\sigma_j \sigma_{j'} \mid \sigma_{j_1} = -\sigma_j, \mathcal{C}_i] \mathbb{P}(\sigma_{j_1} = -\sigma_j \mid \mathcal{C}_i) \\
=& \left(\frac{\beta}{\alpha}\right)^{t-1} \mathbb{P}(\sigma_{j_1} = \sigma_j \mid \mathcal{C}_i) - \left(\frac{\beta}{\alpha}\right)^{t-1} \mathbb{P}(\sigma_{j_1} = -\sigma_j \mid \mathcal{C}_i) \\
=& \left(\frac{\beta}{\alpha}\right)^{t-1} \frac{\alpha+\beta}{2\alpha} - \left(\frac{\beta}{\alpha}\right)^{t-1} \frac{\alpha-\beta}{2\alpha} = \left(\frac{\beta}{\alpha}\right)^{t}.
\end{aligned}
$$

Therefore $\mathbb{E}[\sigma_j \sigma_{j'} \mid \mathcal{C}_i] \le \left(\frac{\beta}{\alpha}\right)^{d(j,j')} \le \left(\frac{\beta}{\alpha}\right)^{d(j,j')-1}$. This completes the proof for part (b).

(c) Suppose j, j' are not in the same hyperedge and there exists a vertex k such that j, k satisfies the assumption in Case (b) with $d(j,k) = t_1$, and j', k satisfy the assumption in Case (b) with $d(j',k) = t_2$. Conditioned on σ_k, we know σ_j and σ_j' are independent. Then we have

$$
\begin{aligned}
\mathbb{E}[\sigma_j \sigma_{j'} \mid \mathcal{C}_i] = \mathbb{E}[\mathbb{E}[\sigma_j \sigma_{j'} \sigma_k^2 \mid \sigma_k, \mathcal{C}_i] \mid \mathcal{C}_i] &= \mathbb{E}\left[\mathbb{E}[\sigma_j \sigma_k \mid \sigma_k, \mathcal{C}_i] \cdot \mathbb{E}[\sigma_{j'} \sigma_k \mid \sigma_k, \mathcal{C}_i] \mid \mathcal{C}_i\right] \\
= \left(\frac{\beta}{\alpha}\right)^{t_1+t_2} \le \left(\frac{\beta}{\alpha}\right)^{d(j,j')-1},
\end{aligned}
$$

where the last line follows from the triangle inequality $d(j,k) + d(j',k) \ge d(j,j')$ and the condition $\beta < \alpha$.

(d) If j, j' are in the same hyperedge, then $d(j,j') = 1$ and (4.12.30) holds trivially.

Combining Cases (a)-(d), (4.12.30) holds. From (4.12.30) and (4.12.26)-(4.12.29), we have

$$
\begin{aligned}
\mathbb{E}[X_i^2 \mathbf{1}_\Omega \mid \mathcal{C}_i] &\leq \sum_{h,h'=0}^{l} \sum_{\tau=0}^{2(h\wedge h')} \sum_{j:d(i,j)=2h} \sum_{j':d(i,j')=2h'} \mathbf{1}_{d(j,j')=2(h+h'-\tau)} \mathbb{E}[\sigma_j \sigma'_j \mid \mathcal{C}_i] R_{i,j}^{(l)} R_{i,j'}^{(l)} \\
&\leq \sum_{h,h'=0}^{l} \sum_{\tau=0}^{2(h\wedge h')} \sum_{j:d(i,j)=2h} O(\alpha^{2h'-\tau}\log n)\left(\frac{\beta}{\alpha}\right)^{2(h+h'-\tau)-1} \cdot O(\alpha^{2l-h-h'}\log^2 n) \\
&= \sum_{h,h'=0}^{l} \sum_{\tau=0}^{2(h\wedge h')} O(\alpha^{2l+h+h'-\tau}\log^4 n)\left(\frac{\beta}{\alpha}\right)^{2(h+h'-\tau)-1} \\
&= \sum_{h,h'=0}^{l} \sum_{\tau=0}^{2(h\wedge h')} O(\alpha^{2l}\log^4 n)\cdot(\beta^2/\alpha)^{h+h'-\tau} = O(\beta^{4l}\log^4 n). \qquad (4.12.31)
\end{aligned}
$$

From (4.12.24) and (4.12.31), we have for some $\varepsilon > 0$,

$$\mathbb{E}(\|z\|_2^2 \mid \Omega) = n^{1-\varepsilon} O(\alpha^{2l}\beta^{2l}\log^2 n) + O(n\beta^{4l}\log^2 n).$$

Then by Chebyshev's inequality, asymptotically almost surely,

$$\|z\|_2 = O(n^{1/2-\varepsilon/2}\alpha^l\beta^l\log^2 n) + O(n^{1/2}\beta^{2l}\log^2 n) = (\sqrt{n}\beta^l\log^2 n)\cdot O(\beta^l \vee \alpha^l n^{-\varepsilon/2}).$$

Recall $l = c\log n$. We have $\beta^l = n^{c\log\beta}, \alpha^l = n^{c\log\alpha}$. So $\beta^l = n^{-\varepsilon'}\alpha^l$ for some constant $\varepsilon' > 0$. Since from (4.12.17), $\|B^{(l)}\sigma\|_2 = \Theta(\sqrt{n}\beta^l)$, we have

$$\|z\|_2 = O(n^{-\gamma''}\alpha^l \|B^{(l)}\sigma\|_2) \qquad (4.12.32)$$

for some constant $\gamma'' > 0$. Combining (4.12.22) with (4.12.32), it implies for some constant $\gamma > 0$,

$$\|B^{(l)}B^{(l)}\sigma\|_2 = \|z + z' + z''\|_2 = O(n^{-\gamma}\alpha^l)\|B^{(l)}\sigma\|_2.$$

Then the upper bound on $\|B^{(l)}B^{(l)}\sigma\|_2$ in (4.11.5) holds. □

4.13 Acknowledgment

Chapter 4 is extracted from “Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, (2021): 1-57.” The thesis author is the co-author of this paper.

Bibliography

[1] Emmanuel Abbe. Community detection and stochastic block models: Recent developments. *Journal of Machine Learning Research*, 18(177):1–86, 2018.

[2] 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.

[3] Emmanuel Abbe, Enric Boix-Adserà, Peter Ralli, and Colin Sandon. Graph powering and spectral robustness. *SIAM Journal on Mathematics of Data Science*, 2(1):132–157, 2020.

[4] Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. *Annals of Statistics*, 48(3):1452–1474, 2020.

[5] Emmanuel Abbe and Colin Sandon. Proof of the achievability conjectures for the general stochastic block model. *Communications on Pure and Applied Mathematics*, 71(7):1334–1406, 2018.

[6] Sameer Agarwal, Jongwoo Lim, Lihi Zelnik-Manor, Pietro Perona, David Kriegman, and Serge Belongie. Beyond pairwise clustering. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 2, pages 838–845. IEEE, 2005.

[7] O. Ajanki, L. Erdős, and T. Krüger. *Quadratic Vector Equations on Complex Upper Half-Plane*. Memoirs of the American Mathematical Society. American Mathematical Society, 2019.

[8] Oskari H Ajanki, László Erdős, and Torben Krüger. Universality for general Wigner-type matrices. *Probability Theory and Related Fields*, pages 1–61, 2015.

[9] Oskari H Ajanki, László Erdős, and Torben Krüger. Singularities of solutions to quadratic vector equations on the complex upper half-plane. *Communications on Pure and Applied Mathematics*, 2016.

[10] David Aldous and J Michael Steele. The objective method: probabilistic combinatorial optimization and local weak convergence. In *Probability on discrete structures*, pages 1–72. Springer, 2004.

[11] Dan Alistarh, Jennifer Iglesias, and Milan Vojnovic. Streaming min-max hypergraph partitioning. In *Advances in Neural Information Processing Systems*, pages 1900–1908, 2015.

[12] Noga Alon, Itai Benjamini, Eyal Lubetzky, and Sasha Sodin. Non-backtracking random walks mix faster. *Communications in Contemporary Mathematics*, 9(04):585–603, 2007.

[13] Johannes Alt. Singularities of the density of states of random Gram matrices. *Electron. Commun. Probab.*, 22:13 pp., 2017.

[14] Johannes Alt, Raphaël Ducatez, and Antti Knowles. Extremal eigenvalues of critical Erdős-Rényi graphs. *arXiv preprint arXiv:1905.03243*, 2019.

[15] Johannes Alt, Raphael Ducatez, and Antti Knowles. Delocalization transition for critical Erdős–Rényi graphs. *arXiv preprint arXiv:2005.14180*, 2020.

[16] Johannes Alt, László Erdős, and Torben Krüger. Local law for random Gram matrices. *Electronic Journal of Probability*, 22, 2017.

[17] Johannes Alt, László Erdős, Torben Krüger, and Yuriy Nemish. Location of the spectrum of Kronecker random matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 55(2):661–696, 2019.

[18] Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2010.

[19] Greg W Anderson and Ofer Zeitouni. A CLT for a band matrix model. *Probability Theory and Related Fields*, 134(2):283–338, 2006.

[20] Maria Chiara Angelini, Francesco Caltagirone, Florent Krzakala, and Lenka Zdeborová. Spectral detection on sparse hypergraphs. In *Communication, Control, and Computing (Allerton), 2015 53rd Annual Allerton Conference on*, pages 66–73. IEEE, 2015.

[21] Konstantin Avrachenkov, Laura Cottatellucci, and Arun Kadavankandy. Spectral properties of random matrices for stochastic block model. In *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt), 2015 13th International Symposium on*, pages 537–544. IEEE, 2015.

[22] Zhidong Bai and Jack W Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Annals of Probability*, pages 553–605, 2004.

[23] Zhidong Bai and Jack W Silverstein. *Spectral analysis of large dimensional random matrices*, volume 20. Springer, 2010.

[24] Zhidong Bai, Xiaoying Wang, and Wang Zhou. Functional CLT for sample covariance matrices. *Bernoulli*, 16(4):1086–1113, 2010.

[25] Zhidong Bai and Jianfeng Yao. On the convergence of the spectral empirical process of Wigner matrices. *Bernoulli*, 11(6):1059–1092, 2005.

[26] Zhidong Bai and Yong Q Yin. Convergence to the semicircle law. *The Annals of Probability*, pages 863–875, 1988.

[27] Roland Bauerschmidt, Jiaoyang Huang, Antti Knowles, and Horng-Tzer Yau. Bulk eigenvalue statistics for random regular graphs. *The Annals of Probability*, 45(6A):3626–3663, 2017.

[28] Roland Bauerschmidt, Jiaoyang Huang, Antti Knowles, and Horng-Tzer Yau. Edge rigidity and universality of random regular graphs of intermediate degree. *Geometric and Functional Analysis*, pages 1–77, 2020.

[29] Roland Bauerschmidt, Jiaoyang Huang, and Horng-Tzer Yau. Local Kesten–Mckay law for random regular graphs. *Communications in Mathematical Physics*, 369(2):523–636, 2019.

[30] Gérard Ben Arous and Kim Dang. On fluctuations of eigenvalues of random permutation matrices. *Annales de l'IHP Probabilités et statistiques*, 51(2):620–647, 2015.

[31] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. *The Annals of Probability*, 47(3):1653–1676, 2019.

[32] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. Spectral radii of sparse random matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 56(3):2141–2161, 2020.

[33] Florent Benaych-Georges, Alice Guionnet, and Camille Male. Central limit theorems for linear statistics of heavy tailed random matrices. *Communications in Mathematical Physics*, 329(2):641–686, 2014.

[34] Itai Benjamini and Oded Schramm. Recurrence of distributional limits of finite planar graphs. *Electronic Journal of Probability*, 6, 2001.

[35] Vladimir Blinovsky and Catherine Greenhill. Asymptotic enumeration of sparse uniform linear hypergraphs with given degrees. *The Electronic Journal of Combinatorics*, 23(3):P3–17, 2016.

[36] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[37] 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.

[38] Charles Bordenave, Simon Coste, and Raj Rao Nadakuditi. Detection thresholds in very sparse matrix completion. *arXiv preprint arXiv:2005.06062*, 2020.

[39] Charles Bordenave and Marc Lelarge. Resolvent of large random graphs. *Random Structures & Algorithms*, 37(3):332–352, 2010.

[40] Charles Bordenave, Marc Lelarge, and Laurent Massoulié. Nonbacktracking spectrum of random graphs: Community detection and nonregular Ramanujan graphs. *Annals of probability*, 46(1):1–71, 2018.

[41] Charles Bordenave, Arnab Sen, and Bálint Virág. Mean quantum percolation. *Journal of the European Mathematical Society*, 19(12):3679–3707, 2017.

[42] Christian Borgs, Jennifer Chayes, Henry Cohn, and Yufei Zhao. An L^p theory of sparse graph convergence I: Limits, sparse random graph models, and power law distributions. *Transactions of the American Mathematical Society*, 2019.

[43] Christian Borgs, Jennifer T. Chayes, Henry Cohn, and Yufei Zhao. An L^p theory of sparse graph convergence II: LD convergence, quotients and right convergence. *The Annals of Probability*, 46(1):337–396, 01 2018.

[44] Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing. *Advances in Mathematics*, 219(6):1801–1851, 2008.

[45] Christian Borgs, Jennifer T Chayes, László Lovász, Vera T Sós, and Katalin Vesztergombi. Convergent sequences of dense graphs II. multiway cuts and statistical physics. *Annals of Mathematics*, 176(1):151–219, 2012.

[46] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[47] 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.

[48] Gerandy Brito, Ioana Dumitriu, and Kameron Decker Harris. Spectral gap in random bipartite biregular graphs and its applications. *arXiv preprint arXiv:1804.07808*, 2018.

[49] Andrei Z Broder, Alan M Frieze, Stephen Suen, and Eli Upfal. Optimal construction of edge-disjoint paths in random graphs. *SIAM Journal on Computing*, 28(2):541–573, 1998.

[50] Samuel R Bulò and Marcello Pelillo. A game-theoretic approach to hypergraph clustering. In *Advances in neural information processing systems*, pages 1571–1579, 2009.

[51] Shantanu Prasad Burnwal and Mathukumalli Vidyasagar. Deterministic completion of rectangular matrices using asymmetric Ramanujan graphs: Exact and stable recovery. *IEEE Transactions on Signal Processing*, 68:3834–3848, 2020.

[52] Thierry Cabanal-Duvillard. Fluctuations de la loi empirique de grandes matrices aléatoires. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 37(3):373–402, 2001.

[53] E Rodney Canfield, Catherine Greenhill, and Brendan D McKay. Asymptotic enumeration of dense 0–1 matrices with specified line sums. *Journal of Combinatorial Theory, Series A*, 115(1):32–66, 2008.

[54] E Rodney Canfield and Brendan D McKay. Asymptotic enumeration of dense 0-1 matrices with equal row sums and equal column sums. *the Electronic Journal of Combinatorics*, 12(1):R29, 2005.

[55] Arijit Chakrabarty, Sukrit Chakraborty, and Rajat Subhra Hazra. Eigenvalues outside the bulk of inhomogeneous Erdős–Rényi random graphs. *Journal of Statistical Physics*, 181(5):1746–1780, 2020.

[56] Arijit Chakrabarty, Rajat Subhra Hazra, Frank den Hollander, and Matteo Sfragara. Spectra of adjacency and Laplacian matrices of inhomogeneous Erdős–Rényi random graphs. *Random matrices: Theory and applications*, 10(01):2150009, 2021.

[57] Arijit Chakrabarty, Rajat Subhra Hazra, Frank den Hollander, and Matteo Sfragara. Large deviation principle for the maximal eigenvalue of inhomogeneous Erdős–Rényi random graphs. *arXiv preprint arXiv:2008.08367*, 2020.

[58] Sourav Chatterjee. Fluctuations of eigenvalues and second order Poincaré inequalities. *Probability Theory and Related Fields*, 143(1-2):1–40, 2009.

[59] Sourav Chatterjee, Persi Diaconis, and Elizabeth Meckes. Exchangeable pairs and poisson approximation. *Probability Surveys*, 2:64–106, 2005.

[60] Binbin Chen and Guangming Pan. CLT for linear spectral statistics of normalized sample covariance matrices with the dimension much larger than the sample size. *Bernoulli*, 21(2):1089–1133, 2015.

[61] I Chien, Chung-Yi Lin, and I-Hsiang Wang. Community detection in hypergraphs: Optimal statistical limit and efficient algorithms. In *International Conference on Artificial Intelligence and Statistics*, pages 871–879, 2018.

[62] I Eli Chien, Chung-Yi Lin, and I-Hsiang Wang. On the minimax misclassification ratio of hypergraph community detection. *IEEE Transactions on Information Theory*, 65(12):8095–8118, 2019.

[63] Fan Chung. *Spectral Graph Theory*. Number 92 in CBMS Regional Conference Series. American Mathematical Society, 1997.

[64] Fan Chung, Linyuan Lu, and Van Vu. Spectra of random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 100(11):6313–6318, 2003.

[65] Sam Cole and Yizhe Zhu. Exact recovery in the hypergraph stochastic block model: A spectral algorithm. *Linear Algebra and its Applications*, 593:45–73, 2020.

[66] Nicholas Cook. The circular law for random regular digraphs. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 55(4):2111–2167, 2019.

[67] Nicholas Cook, Larry Goldstein, and Tobias Johnson. Size biased couplings and the spectral gap for random regular graphs. *The Annals of Probability*, 46(1):72–125, 2018.

[68] Simon Coste and Yizhe Zhu. Eigenvalues of the non-backtracking operator detached from the bulk. *Random Matrices: Theory and Applications*, page 2150028, 2020.

[69] Romain Couillet and Merouane Debbah. *Random matrix methods for wireless communications*. Cambridge University Press, 2011.

[70] 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.

[71] Payam Delgosha and Venkat Anantharam. Load balancing in hypergraphs. *Journal of Statistical Physics*, 173(3-4):546–625, 2018.

[72] Jian Ding, Allan Sly, and Nike Sun. Satisfiability threshold for random regular nae-sat. *Communications in Mathematical Physics*, 341(2):435–489, 2016.

[73] Xue Ding. Spectral analysis of large block random matrices with rectangular blocks. *Lithuanian Mathematical Journal*, 54(2):115–126, 2014.

[74] Andrzej Dudek, Alan Frieze, Andrzej Ruciński, and Matas Šileikis. Approximate counting of regular hypergraphs. *Information Processing Letters*, 113(19-21):785–788, 2013.

[75] Ioana Dumitriu and Tobias Johnson. The Marčenko-Pastur law for sparse random bipartite biregular graphs. *Random Structures & Algorithms*, 48(2):313–340, 2016.

[76] Ioana Dumitriu, Tobias Johnson, Soumik Pal, and Elliot Paquette. Functional limit theorems for random regular graphs. *Probability Theory and Related Fields*, 156(3-4):921–975, 2013.

[77] Ioana Dumitriu and Soumik Pal. Sparse regular random graphs: spectral density and eigenvectors. *The Annals of Probability*, 40(5):2197–2235, 2012.

[78] Ioana Dumitriu and Yizhe Zhu. Sparse general Wigner-type matrices: Local law and eigenvector delocalization. *Journal of Mathematical Physics*, 60(2):023301, 2019.

[79] Ioana Dumitriu and Yizhe Zhu. Spectra of random regular hypergraphs. *arXiv preprint arXiv:1905.06487*, 2019.

[80] Ioana Dumitriu and Yizhe Zhu. Global eigenvalue fluctuations of random biregular bipartite graphs. *arXiv preprint arXiv:2008.11760*, 2020.

[81] Nathanaël Enriquez and Laurent Ménard. Spectra of large diluted but bushy random graphs. *Random Structures & Algorithms*, 49(1):160–184, 2016.

[82] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős-Rényi Graphs II: Eigenvalue spacing and the extreme eigenvalues. *Communications in Mathematical Physics*, 314(3):587–640, 2012.

[83] László Erdős, Antti Knowles, Horng-Tzer Yau, and Jun Yin. Spectral statistics of Erdős–Rényi graphs I: local semicircle law. *The Annals of Probability*, 41(3B):2279–2375, 2013.

[84] László Erdős and Peter Mühlbacher. Bounds on the norm of Wigner-type random matrices. *Random Matrices: Theory and Applications*, page 1950009, 2018.

[85] László Erdős, Sandrine Péché, José A Ramírez, Benjamin Schlein, and Horng-Tzer Yau. Bulk universality for Wigner matrices. *Communications on Pure and Applied Mathematics*, 63(7):895–925, 2010.

[86] László Erdős, Horng-Tzer Yau, and Jun Yin. Bulk universality for generalized Wigner matrices. *Probability Theory and Related Fields*, pages 1–67, 2012.

[87] László Erdős, Horng-Tzer Yau, and Jun Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Advances in Mathematics*, 229(3):1435–1515, 2012.

[88] William Evans, Claire Kenyon, Yuval Peres, and Leonard J Schulman. Broadcasting on trees and the ising model. *Annals of Applied Probability*, pages 410–433, 2000.

[89] Reza Rashidi Far, Tamer Oraby, Wlodek Bryc, and Roland Speicher. On slow-fading MIMO systems with nonseparable correlation. *IEEE Transactions on Information Theory*, 54(2):544–553, 2008.

[90] Uriel Feige and Eran Ofek. Spectral techniques applied to sparse random graphs. *Random Structures & Algorithms*, 27(2):251–275, 2005.

[91] Keqin Feng and Wen-Ching Winnie Li. Spectra of hypergraphs and applications. *Journal of number theory*, 60(1):1–22, 1996.

[92] Laura Florescu and Will Perkins. Spectral thresholds in the bipartite stochastic block model. In *Conference on Learning Theory*, pages 943–959, 2016.

[93] Péter Frenkel. Convergence of graphs with intermediate density. *Transactions of the American Mathematical Society*, 370(5):3363–3404, 2018.

[94] Joel Friedman. *A Proof of Alon's Second Eigenvalue Conjecture and Related Problems*. Memoirs of the American Mathematical Society. American Mathematical Society, 2008.

[95] Joel Friedman and Avi Wigderson. On the second eigenvalue of hypergraphs. *Combinatorica*, 15(1):43–65, 1995.

[96] David Gamarnik, Quan Li, and Hongyi Zhang. Matrix completion from $O(n)$ samples in linear time. In *Conference on Learning Theory*, pages 940–947, 2017.

[97] Chao Gao and Zongming Ma. Minimax rates in network analysis: Graphon estimation, community detection and hypothesis testing. *Statistical Science*, 36(1):16–33, 2021.

[98] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral partitioning of uniform hypergraphs under planted partition model. In *Advances in Neural Information Processing Systems*, pages 397–405, 2014.

[99] Debarghya Ghoshdastidar and Ambedkar Dukkipati. A provable generalized tensor spectral method for uniform hypergraph partitioning. In *International Conference on Machine Learning*, pages 400–409, 2015.

[100] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Spectral clustering using multilinear SVD: Analysis, approximations and applications. In *AAAI*, pages 2610–2616, 2015.

[101] Debarghya Ghoshdastidar and Ambedkar Dukkipati. Consistency of spectral hypergraph partitioning under planted partition model. *The Annals of Statistics*, 45(1):289–315, 2017.

[102] V Girko, W Kirsch, and A Kutzelnigg. A necessary and sufficient conditions for the semicircle law. *Random Operators and Stochastic Equations*, 2(2):195–202, 1994.

[103] Viacheslav Leonidovich Girko. *Theory of stochastic canonical equations*, volume 2. Springer Science & Business Media, 2001.

[104] 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.

[105] W Hachem, Philippe Loubaton, and J Najim. The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 42(6):649–670, 2006.

[106] Walid Hachem, Philippe Loubaton, and Jamal Najim. A CLT for information-theoretic statistics of Gram random matrices with a given variance profile. *The Annals of Applied Probability*, 18(6):2071–2130, 2008.

[107] 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.

[108] Yukun He, Antti Knowles, and Matteo Marcozzi. Local law and complete eigenvector delocalization for supercritical Erdős–Rényi graphs. *Annals of Probability*, 47(5):3278–3302, 2019.

[109] Matthias Hein, Simon Setzer, Leonardo Jost, and Syama Sundar Rangapuram. The total variation on hypergraphs-learning on hypergraphs revisited. In *Advances in Neural Information Processing Systems*, pages 2427–2435, 2013.

[110] J William Helton, Reza Rashidi Far, and Roland Speicher. Operator-valued semicircular elements: solving a quadratic matrix equation with positivity constraints. *International Mathematics Research Notices*, 2007, 2007.

[111] Christian Hennig, Marina Meila, Fionn Murtagh, and Roberto Rocci. *Handbook of cluster analysis*. CRC Press, 2015.

[112] Christopher J Hillar and Lek-Heng Lim. Most tensor problems are NP-hard. *Journal of the ACM (JACM)*, 60(6):45, 2013.

[113] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

[114] Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–561, 2006.

[115] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[116] Jiaoyang Huang, Benjamin Landon, and Horng-Tzer Yau. Bulk universality of sparse random matrices. *Journal of Mathematical Physics*, 56(12):123301, 2015.

[117] Jiaoyang Huang, Benjamin Landon, and Horng-Tzer Yau. Transition from Tracy–Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs. *Annals of Probability*, 48(2):916–962, 2020.

[118] Kurt Johansson. On fluctuations of eigenvalues of random Hermitian matrices. *Duke mathematical journal*, 91(1):151–204, 1998.

[119] Tobias Johnson. Exchangeable pairs, switchings, and random regular graphs. *The Electronic Journal of Combinatorics*, 22(1):P1–33, 2015.

[120] Dag Jonsson. Some limit theorems for the eigenvalues of a sample covariance matrix. *Journal of Multivariate Analysis*, 12(1):1–38, 1982.

[121] Paul Jung and Jaehun Lee. Delocalization and limiting spectral distribution of Erdős-Rényi graphs with constant expected degree. *Electronic Communications in Probability*, 23, 2018.

[122] Raghunandan H Keshavan, Andrea Montanari, and Sewoong Oh. Matrix completion from a few entries. *IEEE transactions on information theory*, 56(6):2980–2998, 2010.

[123] Alexei M Khorunzhy, Boris A Khoruzhenko, and Leonid A Pastur. Asymptotic properties of large random matrices with independent entries. *Journal of Mathematical Physics*, 37(10):5033–5060, 1996.

[124] Chiheon Kim, Afonso S Bandeira, and Michel X Goemans. Stochastic block model for hypergraphs: Statistical limits and a semidefinite programming approach. *arXiv preprint arXiv:1807.02884*, 2018.

[125] 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.

[126] Dávid Kunszenti-Kovács, László Lovász, and Balázs Szegedy. Measures on the square as sparse graph limits. *Journal of Combinatorial Theory, Series B*, 2019.

[127] Greg Kuperberg, Shachar Lovett, and Ron Peled. Probabilistic existence of regular combinatorial structures. *Geometric and Functional Analysis*, 27(4):919–972, 2017.

[128] Timothy Kusalik, James A Mingo, and Roland Speicher. Orthogonal polynomials and fluctuations of random matrices. *Journal für die reine und angewandte Mathematik*, 2007(604):1–46, 2007.

[129] Can M Le, Elizaveta Levina, and Roman Vershynin. Concentration and regularization of random graphs. *Random Structures & Algorithms*, 51(3):538–561, 2017.

[130] Marius Leordeanu and Cristian Sminchisescu. Efficient hypergraph clustering. In *Artificial Intelligence and Statistics*, pages 676–684, 2012.

[131] Pan Li and Olgica Milenkovic. Inhomogeneous hypergraph clustering with applications. In *Advances in Neural Information Processing Systems*, pages 2308–2318, 2017.

[132] Wen-Ching Winnie Li and Patrick Solé. Spectra of regular graphs and hypergraphs and orthogonal polynomials. *European Journal of Combinatorics*, 17(5):461–477, 1996.

[133] Chung-Yi Lin, I Eli Chien, and I-Hsiang Wang. On the fundamental statistical limit of community detection in random hypergraphs. In *Information Theory (ISIT), 2017 IEEE International Symposium on*, pages 2178–2182. IEEE, 2017.

[134] Qingshan Liu, Yuchi Huang, and Dimitris N Metaxas. Hypergraph with sampling for image retrieval. *Pattern Recognition*, 44(10-11):2255–2262, 2011.

[135] László Lovász. *Large networks and graph limits*, volume 60. American Mathematical Society Providence, 2012.

[136] László Lovász and Balázs Szegedy. Limits of dense graph sequences. *Journal of Combinatorial Theory, Series B*, 96(6):933–957, 2006.

[137] Linyuan Lu and Xing Peng. Loose Laplacian spectra of random hypergraphs. *Random Structures & Algorithms*, 41(4):521–545, 2012.

[138] Camille Male. Traffic distributions and independence I: permutation invariant random matrices and the three notions of independence. *Memoirs of the American Mathematical Society*, 2020.

[139] Camille Male and Sandrine Péché. Uniform regular weighted graphs with large degree: Wigner's law, asymptotic freeness and graphons limit. *arXiv preprint arXiv:1410.8126*, 2014.

[140] Vladimir A Marčenko and Leonid Andreevich Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

[141] 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.

[142] Brendan D McKay. The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, 40:203–216, 1981.

[143] Brendan D McKay. Subgraphs of random graphs with specified degrees. In *Congressus Numerantium*, volume 33, pages 213–223, 1981.

[144] Brendan D McKay, Nicholas C Wormald, and Beata Wysocka. Short cycles in random regular graphs. *The Electronic Journal of Combinatorics*, pages R66–R66, 2004.

[145] Elchanan Mossel, Joe Neeman, and Allan Sly. Consistency thresholds for the planted bisection model. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 69–75, 2015.

[146] Elchanan Mossel, Joe Neeman, and Allan Sly. Reconstruction and estimation in the planted partition model. *Probability Theory and Related Fields*, 162(3-4):431–461, 2015.

[147] Elchanan Mossel, Joe Neeman, and Allan Sly. Belief propagation, robust reconstruction and optimal recovery of block models. *The Annals of Applied Probability*, 26(4):2211–2256, 2016.

[148] Elchanan Mossel, Joe Neeman, and Allan Sly. A proof of the block model threshold conjecture. *Combinatorica*, 38(3):665–708, 2018.

[149] Alexandru Nica, Dimitri Shlyakhtenko, and Roland Speicher. Operator-valued distributions. I. characterizations of freeness. *International Mathematics Research Notices*, 2002(29):1509–1538, 2002.

[150] Nathan Noiry. Spectral asymptotic expansion of Wishart matrices with exploding moments. *ALEA*, 15(2):897–911, 2018.

[151] Soumik Pal and Yizhe Zhu. Community detection in the sparse hypergraph stochastic block model. *Random Structures & Algorithms*, pages 1–57, 2020.

[152] David A Papa and Igor L Markov. Hypergraph partitioning and clustering. In *Handbook of Approximation Algorithms and Metaheuristics*, pages 959–978. Chapman and Hall/CRC, 2007.

[153] Guillem Perarnau and Giorgis Petridis. Matchings in random biregular bipartite graphs. *The Electronic Journal of Combinatorics*, pages P60–P60, 2013.

[154] Subhabrata Sen. Optimization on sparse random hypergraphs and spin glasses. *Random Structures & Algorithms*, 53(3):504–536, 2018.

[155] Mariya Shcherbina and Brunello Tirozzi. Central limit theorem for fluctuations of linear eigenvalue statistics of large random graphs. *Journal of Mathematical Physics*, 51(2):023523, 2010.

[156] Mariya Shcherbina and Brunello Tirozzi. Central limit theorem for fluctuations of linear eigenvalue statistics of large random graphs: Diluted regime. *Journal of Mathematical Physics*, 53(4):043501, 2012.

[157] Dimitri Shlyakhtenko. Random Gaussian band matrices and freeness with amalgamation. *International Mathematics Research Notices*, 1996(20):1013–1025, 1996.

[158] Galen R Shorack. *Probability for Statisticians*. Springer Texts in Statistics. Springer International Publishing, 2017.

[159] Ya Sinai and Alexander Soshnikov. Central limit theorem for traces of large random symmetric matrices with independent matrix elements. *Boletim da Sociedade Brasileira de Matemática-Bulletin/Brazilian Mathematical Society*, 29(1):1–24, 1998.

[160] Michael Sipser and Daniel A Spielman. Expander codes. *IEEE transactions on Information Theory*, 42(6):1710–1722, 1996.

[161] Sasha Sodin. Random matrices, nonbacktracking walks, and orthogonal polynomials. *Journal of Mathematical Physics*, 48(12):123503, 2007.

[162] Philippe Sosoe and Percy Wong. Regularity conditions in the CLT for linear eigenvalue statistics of Wigner matrices. *Advances in Mathematics*, 249:37–87, 2013.

[163] Ludovic Stephan and Laurent Massoulié. Robustness of spectral methods for community detection. In *Conference on Learning Theory*, pages 2831–2860. PMLR, 2019.

[164] Ludovic Stephan and Laurent Massoulié. Non-backtracking spectra of weighted inhomogeneous random graphs. *arXiv preprint arXiv:2004.07408*, 2020.

[165] R Tanner. A recursive approach to low complexity codes. *IEEE Transactions on information theory*, 27(5):533–547, 1981.

[166] Terence Tao and Van Vu. Random matrices: universality of local eigenvalue statistics. *Acta mathematica*, 206(1):127–204, 2011.

[167] Audrey Terras. *Zeta functions of graphs: a stroll through the garden*, volume 128. Cambridge University Press, 2010.

[168] Konstantin Tikhomirov and Pierre Youssef. The spectral gap of dense random regular graphs. *The Annals of Probability*, 47(1):362–419, 2019.

[169] Linh V Tran. Local law for eigenvalues of random regular bipartite graphs. *Bulletin of the Malaysian Mathematical Sciences Society*, 43(2):1517–1526, 2020.

[170] Linh V Tran, Van H Vu, and Ke Wang. Sparse random graphs: Eigenvalues and eigenvectors. *Random Structures & Algorithms*, 42(1):110–134, 2013.

[171] Lloyd N Trefethen. *Approximation Theory and Approximation Practice*, volume 128. SIAM, 2013.

[172] Alexei Vazquez. Finding hypergraph communities: a bayesian approach and variational solution. *Journal of Statistical Mechanics: Theory and Experiment*, 2009(07):P07006, 2009.

[173] Eugene P Wigner. Characteristic vectors of bordered matrices with infinite dimensions. *Annals of Mathematics*, pages 548–564, 1955.

[174] David P Woodruff. Sketching as a tool for numerical linear algebra. *Foundations and Trends® in Theoretical Computer Science*, 10(1–2):1–157, 2014.

[175] Kevin Yang. Local Marchenko-Pastur law for random bipartite graphs. *arXiv preprint arXiv:1704.08672*, 2017.

[176] Denny Zhou, Jiayuan Huang, and Bernhard Schölkopf. Learning with hypergraphs: Clustering, classification, and embedding. In *Advances in neural information processing systems*, pages 1601–1608, 2007.

[177] Zhixin Zhou and Yizhe Zhu. Sparse random tensors: Concentration, regularization and applications. *Electronic Journal of Statistics*, 15(1):2483–2516, 2021.

[178] Yizhe Zhu. A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Structures & Algorithms*, 56(1):251–279, 2020.

[179] Yizhe Zhu. On the second eigenvalue of random bipartite biregular graphs. *arXiv preprint arXiv:2005.08103*, 2020.