IMT School for Advanced Studies, Lucca
Lucca, Italy

Spectral Signatures of Breaking of Ensemble Equivalence

PhD Program in Systems Science
Track in Computer Science and Systems Engineering
XXXV Cycle

By

Pierfrancesco Dionigi

2024

The dissertation of Pierfrancesco Dionigi is approved.

PhD Program Coordinator: Prof. Alberto Bemporad, IMT School for Advanced Studies Lucca

Advisor: Prof. Diego Garlaschelli, IMT Lucca & Leiden University

Co-Advisor: Prof. Frank den Hollander, Leiden University

Co-Advisor: Prof. Michel Mandjes, Leiden University

The dissertation of Pierfrancesco Dionigi has been reviewed by:

Dr. A. Knowles, University of Geneva, Switzerland

Dr. I. Kryven, Universiteit Utrecht, The Netherlands

Prof. Dr. F. Merlévède, Université Paris-Est Marne-la-Vallée, France

IMT School for Advanced Studies Lucca
2024

To my grandparents.

Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I extend my deepest gratitude to my doctoral supervisors, Prof. Dr. Frank den Hollander, Prof. Dr. Diego Garlaschelli and Prof. Dr. Michel Mandjes, for their guidance, invaluable advice, and continuous support throughout the course of this journey.

I would like to start thanking Frank for the role model he has been for me, for the nice moments we spent together and the many passions we shared outside work. His expertise and insight have been crucial in shaping both the direction and the outcome of my work. His patience, encouragement, and rigorous standards have not only helped me to elevate my work to its highest potential but have also significantly contributed to my personal and professional growth. Without you and your patience this journey would have not been the same.

My second academical father has been Diego, to whom I will always be grateful for guiding me in this journey between physics and mathematics. Despite the distance, your guidance has always been insightful and an inspiration for what physics intuition should be and what it can bring to mathematics. Besides your academical mentorship I also enjoyed your human companionship. The good moments we spent in Lucca, Leiden and Lipari with the research group are unforgettable and a good reminder of how fun scientific research can and should be.

Next, I would like to express my sincere appreciation to Michel as well. The multiple journeys I made to Amsterdam to update you on the progress of our research have always been a moment of inspiration and learning. Your vision on the problems we faced has always been clear and stimulated me to

find always the easiest and most natural answer to the questions we have been asking.

Even if not formerly my supervisor, I wish to mention Dr. Rajat S. Hazra as one of my most precious mentors in this journey, his contribution to my research has been invaluable and his constant and precious support played a crucial role in the successful outcome of these four years. I am deeply grateful to him, for the time he dedicated me, and for the instructive discussion we had at the blackboard. The moment he arrived in Leiden was a great benefit for me and the whole probability group. Working with you has been a beautiful experience and I hope it will continue in the future. Between the senior staff, I wish to thank Prof. Dr. Luca Avena as well, for the useful discussions and many cheerful moments we spent together (rigorously in Italian). I am also immensely thankful to my colleagues. I would like to mention Alessio, Daoyi, Federico, Francesca, Nandan and Oliver, with whom I spent most of my time at the institute. Thank you for the useful discussions, the laughs together and the support throughout these four years. You helped making this adventure a beautiful experience, professionally and humanly.

To my dear parents, Sabrina and Antonello, and to Andrea no words can express my gratitude for your unconditional love, endless patience, and unwavering support. Your sacrifices, belief in my abilities, and constant encouragement have been the cornerstone of my achievements. Your example has instilled in me the values of hard work, perseverance, and the importance of pursuing one's passions with integrity and dedication. To my dear grandparents too, that unfortunately are no more here and cannot see what I achieved thanks to their sacrifices and love.

I must also express my heartfelt thanks to my friends, both within and outside the academic community, for their under-

standing, support, and encouragement throughout this challenging yet rewarding journey. Your presence, encouragement, and occasional distractions were necessary reprieves that helped maintain my sanity and kept me grounded. In particular I want to mention the friends that this Dutch experience brought in my life, and with whom I shared as well the difficult times of Covid epidemic: Alessandro, Davide, Francesco, Giovanni, Gianluca, Matteo, Niccolò, Nicolò, Valerio, Vladimir. Finally, I would like to acknowledge NETWORKS for their financial support and Leiden University who contributed directly or indirectly to the completion of this thesis. This journey has been a transformative experience, and I am grateful to have been surrounded by such a supportive and inspiring group of people. Thank you all.

Vita

April 7, 1994	Born, Pesaro, Italy
2016	BSc in Physics Final mark: 110/110 cum laude Università di Bologna, Bologna, Italy
2019	MSc in Theoretical Physics Final mark: 110/110 cum laude Università di Bologna, Bologna, Italy

Publications

1. *A spectral signature of breaking of ensemble equivalence for constrained random graphs*, Electronic Communications in Probability 26 (2021): 1-15, DOI: 10.1214/21-ECP432. Joint work with D. Garlaschelli, F. den Hollander, M. Mandjes.
2. *Central limit theorem for the principal eigenvalue and eigenvector of Chung-Lu random graphs*, Journal of Physics: Complexity (2022): 1-23, DOI 10.1088/2632-072X/acb8f7. Joint work with D. Garlaschelli, F. den Hollander, R. Hazra, M. Mandjes.
3. *Largest eigenvalue of configuration model and breaking of ensemble equivalence*, arXiv:2312.07812. Joint work with D. Garlaschelli, F. den Hollander and R. Hazra.

Presentations

1. 13 - 16 September 2023: Talk at "RandNET Workshop on Graph Limits and Networks", Prague, Czech Republic
2. 30 - 31 August 2023: Contributed talk at "Dutch NetSci Summer Symposium 2023", Delft, The Netherlands
3. 5 - 9 June 2023: Poster presentation at "Random Matrices and Applications", RIMS, Kyoto University, Japan
4. 27 - 31 March 2023: Invited speaker at 18th YEP workshop "Spectra of random graphs and related combinatorial problems", EURANDOM Eindhoven University of Technology, NL
5. 6 - 8 February 2023: Talk at 2023 Drafting Workshop, Rényi Institute, Budapest, Hungary
6. 14 December 2022: NETWORKS Research Seminars @ IMT Lucca, Italy
7. 12 May 2022: NETWORKS training week, Asperen, NL
8. 15-17 November 2021 (poster session): 49^{th} annual meeting of the Dutch probability and statistics community, Lunteren
9. 25 June 2021: LCN2 seminar
10. 29 October 2020: NETWORKS training week
11. 23 September 2020: iPod seminar (Leiden University Probability group seminar)
12. 2-13 April 2019 (poster session): Non Local and Fractional Operators (in honor of prof. R. Spigler), 1, Sapienza University, Rome

Abstract

In this thesis we explore the concept of Breaking of Ensemble Equivalence (BEE) within the context of random graph models, focusing on spectral properties of adjacency matrices. Our research aims to identify spectral quantities that can distinguish between different random graph ensembles, thereby providing new insights into the structure and behavior of complex networks. We cover both theoretical aspects and practical implications, including simulations and sampling methods for random graph models.

In Chapter 1 we introduce some basic notions of random graph theory, and discuss how maximum entropy graph models are fundamental in modeling real-world networks. We explain what BEE is, what is its characterization in the context of statistical mechanics, and how it is intimately connected to differences that arise naturally between the canonical versus the microcanonical description of random graph ensembles. In order to do so, we delve into the spectral theory of random graphs and use it to investigate BEE.

In Chapter 2 we formulate a conjecture on the equivalence of measure-BEE and the presence of a gap between the largest non-centered and non-scaled largest eigenvalues of the adjacency matrix in the canonical and the microcanonical ensemble. We prove this conjecture in the setting of homogeneous graphs.

In Chapter 3 we study the same question for Chung-Lu random graphs. In particular, we prove central limit theorems for the largest eigenvalue and its associated eigenvector.

In Chapter 4 we compute the expectation of the largest eigenvalue for the configuration model, which verifies our conjecture in the setting of inhomogeneous graphs as well.

In Chapter 5 we provide numerical evidence for our findings through simulation, after a brief introduction to graph sampling. We formulate the main conclusions of our work and indicate possible further directions of research.

Chapter 1

Introduction

The present thesis deals with random graph models and how to capture their differences via their spectral properties. The first four chapters are of theoretical nature, while the fifth deals with sampling random graphs.

1.1 Background

The rapid development of *Network Science* in past years is part of the rising interest in complex systems encountered in physics, chemistry, biology, the social sciences, the medical sciences, and beyond. Mathematics provides a formidable framework for understanding the complex forms of *interconnectedness* in these systems, empowered by the increasing abundance of real-world data. The presence of these data, which need to be explained and understood, required the development of powerful models, not only to explain what could be extracted from the data, but also to forecast properties of the network that lie hidden. Thus, modelling and testing of graph-like structures were a main driving force of network theory, and in turn led to many new questions of theoretical relevance. These questions gradually gained new territory, making network theory into a vibrant and interdisciplinary research area. Important questions are: What is the best way to model a network-like structure observed in real life? What features need to be included to obtain a faithful

model? How can the functionality of the network be captured properly? Such questions naturally lead to *Random Graph Theory* (RGT).

The mathematical field of graph theory has a long history that traces back to the beginning of the 20th century. The birth of the probabilistic treatment of graphs in RGT can be identified with the seminal paper of Erdős and Rényi [59], where the by now most famous model of a random graph – the Erdős-Rényi random graph (ERRG in the following) – was introduced. The original aim of the authors was to use this probabilistic model to answer some graph-theroretic questions (Ramsey theory, colouring problems, extremal graph theory, etc.). This approach is known today as the *probabilistic method* (see [6] for a survey). Despite the versatility of the ERRG and its successful application to solve some hard problems in discrete mathematics, its simpleness made it unrealistic as a meaningful model for real-world systems. Indeed, real networks are far from being describable as a set of independent random variables. A first question in network theory was how to recreate the specific structures observed in real-world networks and what is the distribution of the *dependent* random variables that form the model. Most real-world networks have a clustering coefficient that is higher than the one arising from ERRG (see [48]). Examples are the networks formed by social interactions, which are naturally transitive (e.g. if A and B are friends and B and C are friends, then A and C are likely to be friends as well) and therefore tend to form triangles between the nodes, a property that is mostly absent in simple network models. Another example is the difficulty to explain higher network structures that appear naturally in society, such as communities, with the help of only a few independent parameters. This complexity led researchers to develop generalizations of ERRG that include inhomogeneities, clustering and other features of real-world networks (see [99] for a review). A powerful and versatile network model was developed in the seminal work [62], which was further developed especially in [105] and led to creation of *Exponential Random Graphs*. This family of models has many important properties, the most striking being the ability to create a probability distribution that favors graphs with a pre-chosen set of features. Of course, this does not come for free: the

more complex are the features, the more difficult are the dependencies hidden in the model. It did not take much time to recognize that this approach is powerful and not dissimilar to an old problem in *Statistical Physics* (SP).

SP was born with the aim of describing the *statistical properties* of physical systems consisting of a large number of interacting particles. By statistical properties we mean the distribution of the relevant functions of the random variables defining the system (e.g. classical quantities such as energy, density, pressure, temperature or magnetization) which usually are linked to measurable macroscopic quantities. The aim of SP is to describe the microscopic equilibrium states of the system (and the fluctuations around these equilibria) when only a handful of these macroscopic quantities are known and are fixed (i.e., measured in real experiments). One way to do this is to create a probability distribution $\mathbb{P}$ whose equilibrium state (i.e., the expectation with respect to $\mathbb{P}$) has the required value of the relevant quantity, but still allows for many microstates whose likelihood is smaller the farther they are from equilibrium. The advantages of this approach are twofold: on the one side, the probability distribution describing the system recreates the measurements that were made; on the other side, we are not imposing any information on the model other than what we actually know. The power of this approach was explained in full generality by Jaynes in [84] and was, in the context of graph theory, further developed in what we call *Maximum-Entropy Networks* (see [110]). Maximum-Entropy Networks offer a principled and versatile framework for modeling probability distributions in a way that balances the need to fit observed data with the desire to avoid unwarranted assumptions. This approach has proven effective in addressing a wide range of problems where traditional models may fall short, making them a valuable tool in the arsenal of probabilistic modeling techniques. Restricting ourself to graph-like systems, the approach just described is different from the approach of just sampling from the set of graphs that have *exactly* a given property (e.g. sampling uniformly from all the graphs with 2028 vertices and 347 triangles). The dichotomy between the two approaches is well known in SP. Sampling according to the uniform distribution from

a set of objects with a prescribed property leads to what is called the *microcanoncal ensemble*, while fixing the average subject to maximal entropy leads to what is called the *canonical ensemble*.

1.2 Comparison of ensembles

One question that might arise at this point is why the first construction, where we let the defining features of our model *fluctuate* (= soft constraint), is preferable over the second construction, where we select *only* those graphs with the desirable property (= hard constraint). Indeed, one could argue that our empirical knowledge of the system under study comes only from what we can measure, and therefore the construction where we pick only the graphs that have exactly the measured feature must be the best one. The reasons why this observation is not accurate are multiple. First, any measurement comes with an error and, given that the microcanonical ensemble selects only graphs with a *hard constraint*, this may lead to a possibly biased description. Moreover, the majority of the networks that we want to analyze vary over time, so sticking to a particular value of the measured feature can be questionable. A further reason comes from the difficulty in sampling from the uniform distribution (see, for example, [73]), which often makes it hard to work with the microcanonical ensemble. A model given by a uniform distribution over a very large set is hard to manipulate mathematically. It is often the case that to say anything about these types of models requires hard combinatorial estimates. Furthermore, the random variables that form the model are typically highly dependent, with correlation patterns that are not easy to capture. These arguments explain why network scientists often prefer to work with the canonical ensemble.

In view of its preferable characteristics, a relevant question is: What is the asymptotic error if we use canonical instead of microcanonical? In mathematical terms this amounts to studying the differences in the expectation, the variance and large deviations of functions of the model with respect to the two different probability distributions. In SP, the widespread belief is that swapping microcanonical to canonical leads to

negligible corrections for very large systems. In other words, it is customary to assume that for very large systems the two ensembles can be used interchangeably with a negligible approximation error. While this can be shown to be true for systems with a short-range interaction Hamiltonian subject to constraints on global quantities like the energy, the relation between the two ensembles is more involved for systems with long-range interactions subject to complicated constraints. Nevertheless, *Ensemble Equivalence in the Thermodynamic Limit* is most of the time taken for granted. The first appearance of systems where ensemble equivalence was failing was in [91], where thermodynamic properties of certain stellar systems were considered. Since then, many studies have appeared where ensemble equivalence was questioned. In particular, in [55] it was concluded that *Breaking of Ensemble Equivalence* (BEE) is deeply connected to the large deviations properties of the two ensembles. In short, *Large Deviation Theory* appears as the proper mathematical setting in which to analyse the problem (see [119] for a review). In a series of papers [118, 120, 121], Touchette showed that BEE can be characterized in three different but equivalent ways: *Thermodynamic BEE, Measure BEE, Macrostate BEE*. While Thermodynamic BEE characterizes BEE in a classical thermodynamic setting, in terms of non-concavity of certain thermodynamic quantities such as entropy and free energy (relating the problem to duality between a function and its Legendre-Fenchel transform), Measure BEE and Macrostate BEE have an SP interpretation that we will describe in Chapter 1.4. Since the work of Touchette, a series of paper by Garlaschelli, den Hollander, Squartini and co-authors [71, 72, 113] has appeared that analyze BEE in random-graph ensembles, with the main focus on Measure BEE. The main contributions in this area up to 2018 are summarized in A. Roccaverde's PhD thesis [103].

1.3 Random matrices

In the study of complex systems, the inherent randomness and complexity often defy traditional analytical approaches. *Random Matrix Theory* (RMT), originating in the mid-20th century, has proven to be an in-

valuable tool for characterizing the statistical behavior of complex matrices that arise in diverse fields of science. This theory offers a unique perspective, focusing on universal properties that transcend specific details of system dynamics, allowing researchers to extract essential features and gain insights into the underlying complexity. The origins of RMT can be traced back to nuclear physics, where it was developed by Wigner [131] to describe the statistical properties of nuclear energy levels of large nuclei. Over time, the scope of RMT has expanded significantly, evolving into a versatile and interdisciplinary tool that has found applications in fields such as quantum mechanics, statistical physics, information theory, and even the analysis of large-scale financial systems, gradually gaining the status of a powerful and versatile mathematical framework for understanding the statistical properties of complex systems in various different settings (see [9, 95] for references and [3] for applications). It did not take long before the interaction between RMT and RGT appeared. The possibility of interpreting a graph as a matrix (adjacency, incidence, Laplacian) suggests that certain features of the graph are well captured by its spectrum (see, for example, the monographs [47, 110]). Soon, spectral graph questions were linked to important graph-theoretic notions, such as *expander graphs* and *stochastic processes* on graphs. It is thus natural to look at how BEE is linked to spectral quantities, which is the main theme of the present PhD thesis.

In the remainder of this introduction we will formally introduce maximum entropy graph models, canonical and microcanonical ensembles, BEE, and the role of RMT in our study. We will close with a summary of the content of this thesis, some conclusions and some directions of future research.

1.4 Maximum entropy graph models and breaking of ensemble equivalence

The fundamental idea behind Maximum Entropy Networks is to construct a probability distribution that is consistent with the observed data, i.e., respects some constraints, while maximizing the Shannon Entropy

(which plays the role of uncertainty in information theory). In other words, these network models seek to find the most unbiased probability distribution that satisfies the available information. By maximizing the entropy, these networks aim to avoid making unnecessary assumptions about the underlying structure of the data, allowing for a more flexible and data-driven modeling approach. These network models have found applications in various fields, including machine learning, statistics, sociology and computational biology. They are particularly useful in situations where the relationships between variables are not well understood or are highly complex. The flexibility of Maximum Entropy Networks makes them valuable for capturing dependencies in diverse data sets, ranging from biological networks to social interactions, and beyond.

1.5 Maximum entropy ensembles and canonical vs. microcanonical

Suppose that we are given a system that can be modeled through a graph G^*. While the full knowledge and reconstruction of G^* is almost never achievable, it is often the case that we can measure different characteristics of G^*. For example, say that we know that our system has size n and that we can measure the degree d_i^* of each vertex. For example, think of a social network in which we can measure how many friends each person has. This information on the degree sequence should be present in the model, but we do not want to force any other information into our probability distribution on $\mathbb{G}_n$, the set of simple graphs of size n. More formally, given a graph function $\vec{C}(G) \to \mathbb{R}^m$, $G \in \mathbb{G}_n$, and a vector of quantities $\vec{C}^* = \{C_i\}_{i=1}^m$ that is *graphical* (i.e., there exist at least one graph in $\mathbb{G}_n$ such that $\vec{C}(G) = \vec{C}^*$), we want to create a probability distribution $\mathbb{P}_n(G)$ on the space $\mathbb{G}_n$ of simple graphs of size n such that $\vec{C}(G)$ is a sufficient statistics (in the example above, $m = n$, $\vec{C} = \{C_i\}_{i=1}^m$, $C_i(G) = d_i$ is the degree of the vertex i) and maximizes the Shannon entropy

$$S[\mathbb{P}] = -\sum_{G \in \mathbb{G}_n} \mathbb{P}(G) \ln \mathbb{P}(G).$$

(In the sequel we will often suppress the dependence of the measure on n.) The Pitman-Koopman-Darmois theorem states that this has to be an exponential family of probability distributions, and its form can be calculated through a maximization problem via the Karush–Kuhn–Tucker theorem. This gives

$$\underset{\mathbb{P}}{\operatorname{argmax}}\ S[\mathbb{P}] \text{ such that } \mathbb{E}_{\mathbb{P}}[C_i] = \sum_{\mathbb{G}_n} \mathbb{P}(G) C_i(G) = C_i^* \ \forall\, 1 \leq i \leq m,$$

where the maximization problem is over the space of probability measures on $\mathbb{G}_n$. This leads to the Lagrangian function

$$\mathcal{L}(\mathbb{P}, \vec{\theta}) = S[\mathbb{P}] + \sum_{i=0}^{m} \theta_i \left(C_i^* - \sum_{G \in \mathbb{G}_n} \mathbb{P}(G) C_i(G) \right), \tag{1.5.1}$$

where $C_0 = 1$ and $C_0^* = 1$ ensure that $\mathbb{P}$ is a probability measure: $\sum_{G \in \mathbb{G}_n} \mathbb{P}(G) = 1$. The solution of the above maximization problem is an exponential family of measures with parameters $\vec{\theta}$, playing the role of Lagrange multipliers, which are fixed $\vec{\theta}^*$ such that

$$\mathbb{E}_{\mathbb{P}, \vec{\theta}}[C_i] = C_i^*, \qquad 1 \leq i \leq m.$$

This solution goes by the name of *canonical Gibbs ensemble*, and takes the form

$$\mathbb{P}_{\text{can}}(G, \vec{\theta}^*) = \frac{\mathrm{e}^{-H(G, \vec{\theta}^*)}}{\sum_{G \in \mathbb{G}_n} \mathrm{e}^{-H(G, \vec{\theta}^*)}} = \frac{\mathrm{e}^{-H(G, \vec{\theta}^*)}}{\mathcal{Z}_{\vec{\theta}}}, \tag{1.5.2}$$

where $H(G, \vec{\theta}^*) = \sum_{i=1}^{m} \theta_i^* C_i(G)$ is the interaction *Hamiltonian* and $\mathcal{Z}_{\vec{\theta}}$ is the *partition function*. It is worth noting that the values of θ_i^* are chosen from the data through the *log-likelihood* maximization principle.

In contrast, the definition of the microcanonical is far easier. Let $\mathbb{G}_n$, $\vec{C}^* = \{C_i\}_{i=1}^m$ and $\vec{C}(G)$ be as above. Define the level set of the function $\vec{C}$

$$\Gamma_{\vec{C}^*} = \{G \in \mathbb{G}_n \colon\ C_i(G) = C_i^* \ \forall\, 1 \leq i \leq m\}, \tag{1.5.3}$$

and let $|\Gamma_{\vec{C}^*}|$ be the cardinality of the above set. The microcanonical ensemble is the probability distribution given by

$$\mathbb{P}_{\text{mic}}(G, \vec{\theta}^*) = \begin{cases} \frac{1}{|\Gamma_{\vec{C}^*}|}, & \text{if } G \in \Gamma_{\vec{C}^*} \\ 0, & \text{otherwise.} \end{cases} \tag{1.5.4}$$

Despite its easy definition, the difficulty of the microcanonical ensemble lies in the definition of (1.5.3), in particular, in the estimation of its cardinality $|\Gamma_{\vec{C}^*}|$. This typically involves hard combinatorial computations that are linked to problems in extremal graph theory (see [14] for an example). Is important to note that $\mathbb{P}_{\text{can}}$ is constant on the level sets of $\vec{C}$, and so every graph with the same value of the constraint is equally likely to be drawn. This fact will play a crucial role in Chapter 2.

1.6 An example

To give an example, let us take $\vec{C} = \vec{d}$, where $\vec{d} = \{d_1, \dots, d_n\}$ is a given degree sequence that satisfies the Erdős-Gallai criterion ([42]). Consider the Hamiltonian

$$H(G) = \sum_i \theta_i d_i = \sum_i \sum_{j>i} (\theta_i + \theta_j) a_{ij},$$

where a_{ij} is the indicator function of the event that vertices i and j are connected, written $i \sim j$, i.e., a_{ij} is the ij-th entry of the *adjacency matrix* $A(G)$. One can use this precise form of the Hamiltonian to perform a trick (see [99]) and write the partition function as

$$\begin{aligned}\mathcal{Z}_{\vec{\theta}^*} &= \sum_{G \in \mathbb{G}_n} \mathrm{e}^{-H(G,\vec{\theta}^*)} = \sum_{G \in \mathbb{G}_n} \exp\left(- \sum_i \sum_{j>i} (\theta_i + \theta_j) a_{ij} \right) = \prod_{i<j} \sum_{a_{ij}=0}^{1} \exp\left(-(\theta_i + \theta_j\right) \\ &= \prod_i \prod_{i<j} \left(1 + \mathrm{e}^{-(\theta_i+\theta_j)}\right) = \prod_i \prod_{i<j} (1 + x_i x_j),\end{aligned}$$

where we put $x_i = \mathrm{e}^{-\theta_i}$. Thus, putting $p_{ij} = \frac{x_i x_j}{1+x_i x_j}$, we can rewrite the probability of a graph G as

$$\mathbb{P}(G) = \prod_i \prod_{i<j} p_{ij}^{a_{ij}} (1 - p_{ij})^{1-a_{ij}}. \tag{1.6.1}$$

As will be discussed in Chapter 3, for suitable degrees the above model becomes a *Chung-Lu inhomogeneous random graph*, where the denominator $1 + x_i x_j$ gives a lower order correction to the connection probability.

The microcanonical distribution is, in this case, the uniform distribution on all the simple graphs with a given degree sequence $\vec{d}_n$. This model can be described in different ways (see [132] for example). One way is via the so-called *configuration model* conditioned on simplicity (the configuration model produces a multigraph with a positive probability when the degrees are bounded, and with a probability tending to one when the degrees diverge with n).

For the case $d_i \equiv d$, we have a homogeneous model for both the canonical and the microcanonical ensemble. Is not difficult to see that (1.6.1) degenerates to an ERRG with edge probability $p = \mathrm{e}^{-\theta}$ (by symmetry we have only one Lagrange multiplier θ), while the microcanonical becomes an instance of a random regular graph.

1.7 Breaking of Ensemble Equivalence

Breaking of Ensemble Equivalence measures the information-theoretic price we pay asymptotically in exchanging the canonical and the microcanonical ensembles. BEE can be defined in three different ways (in [121] it is proved that all three are equivalent).

• *Thermodynamical BEE.* As can be seen from (1.5.1), the non concavity of $S[\mathbb{P}]$ can lead to problems in the solution of the maximization problem. Indeed, this type of BEE focusses on the duality of two important thermodynamic potentials – the *free energy* and the *entropy* – which play a key role in determining the properties of the canonical and the microcanoncal ensembles, respectively. Under normal circumstances, these two quantities are related by a Legendre-Fenchel transform, but concavity problems that may arise from the Hamiltonian can lead to a failure of this duality, signaling the presence of BEE. This is intimately related to large deviation properties as stated in the Gartner-Ellis theorem (see [123, Chapter V] for further explanations), where entropy can be seen as a rate function and free energy as a scaled cumulant generating function. Nevertheless, the relation between BEE and large deviations are better captured through the next type of BEE.

• *Measure BEE.* This compares the canonical and the microcanonical ensembles in an information-theoretic sense, namely, it measures the price we pay in describing the microcanonical ensemble via the canonical ensemble. To do this, we take the Kullback-Leibler divergence (or relative entropy) of the two probability measures:

$$S_n(\mathbb{P}_{\text{mic}} \mid \mathbb{P}_{\text{can}}) = D_{KL}(\mathbb{P}_{\text{mic}} \mid \mathbb{P}_{\text{can}}) = \sum_{G \in \mathbb{G}_n} \mathbb{P}_{\text{mic}}(G) \log \frac{\mathbb{P}_{\text{mic}}(G)}{\mathbb{P}_{\text{can}}(G)}.$$

Given a sequence $\alpha_n \gg 1$, we say that $\mathbb{P}_{\text{mic}}$ and $\mathbb{P}_{\text{can}}$ are equivalent at scale α_n if

$$\lim_{n\to\infty} s_n^{\alpha_n} = \lim_{n\to\infty} \frac{1}{\alpha_n} S_n(\mathbb{P}_{\text{mic}} \mid \mathbb{P}_{\text{can}}) = 0. \tag{1.7.1}$$

It can of course happen that two ensembles are equivalent on given scale α_n but not on a scale $\beta_n = o(\alpha_n)$. The scale α_n captures the difference in the large deviation behaviour of the tails of $\mathbb{P}_{\text{mic}}$ and $\mathbb{P}_{\text{can}}$, much like Sanov theorem captures the price we pay in describing the empirical distribution of a sample x_i^* by the prior probability distribution $p_n(x_i)$. In a series of papers [71, 72, 113] the scale α_n at which $\lim_{n\to\infty} s_n^{\alpha_n} \neq 0$ was studied. It was found that for non-dense graphs (i.e., with degrees $o(n)$), when the constraint is on the degree sequence, the scale is $\alpha_n = n$ and

$$\frac{1}{n} S_n(\mathbb{P}_{\text{mic}} \mid \mathbb{P}_{\text{can}}) = \Theta(\log n).$$

• *Macrostate BEE.* While Measure BEE deals with the microstate description, i.e., the analysis of every state the system can be in, Macrostate BEE analyzes the differences between their ensembles at their equilibrium. In a probabilistic rephrasing of the previous sentence, macrostate equivalence looks at the *expectations* of functions of the system under study. Indeed, while equivalence at the measure level deals with the differences in the tails of the two distributions, the presence of non-equivalence tells us that for diverging n we can expect some tail events to behave differently, and so there should exist some graph function (i.e., a measurable quantity of our network model) that is different between the two models. For $f(G)$ such a function, we can rephrase Macrostate BEE as

$$\lim_{n\to\infty} |\mathbb{E}_{\text{can}}[f] - \mathbb{E}_{\text{mic}}[f]| > 0. \tag{1.7.2}$$

An important aspect of the above characterization is that it gives no clue on how to choose f. Indeed, the search for a universal quantity signalling BEE is non-trivial. For example, when the constraint is applied to the degree sequence, any linear function of the degree sequence behaves in the same way in the two ensembles, while any non-linear function is difficult to evaluate. Restricting ourselves to the case where the constraint is on the degree sequence (like in the examples above), the main contribution of this thesis is the qualitative and quantitative evidence that a good choice for f is the largest eigenvalue of the adjacency matrix of the random graph.

1.8 Spectral theory of random graphs

RMT aims to characterize the behavior of eigenvalues of large matrix ensembles. The collective behavior of eigenvalues was the main object of study in the work of Wigner [131]. There the *empirical spectral distribution* (ESD) of a class of large matrices was determined. Later works identified it as the universal behaviour for a wide class of symmetric matrices with i.i.d. entries, called *Wigner matrices*. Let A_n be a symmetric matrix of dimension n, and let a_{ij}, $j \geq i$, be its elements, i.i.d.[1] with $\mathbb{E}[a_{ij}] = 0$ and $\mathrm{Var}[a_{ij}] = 1$. Define the ESD as

$$\mu_{\frac{1}{\sqrt{n}}A_n} = \sum_{i=1}^{n} \delta_{\lambda_i\left(\frac{1}{\sqrt{n}}A_n\right)}, \tag{1.8.1}$$

where λ_i, $1 \leq i \leq n$, are the eigenvalues of A_n. Then

$$\lim_{n\to\infty} \mu_{\frac{1}{\sqrt{n}}A_n} \xrightarrow{\text{a.s.}} \mu_{\text{sc}}, \tag{1.8.2}$$

where $\mu_{\text{sc}} = \frac{1}{2\pi}(4 - x^2)_+^{1/2}dx$ is the Wigner semicircle distribution. Interpreting the graph as an adjacency matrix, we can analyze random graph models as a matrix ensemble. For the Erdős-Rényi random graphs with

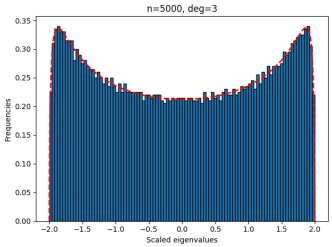
[1] In Wigner matrices the diagonal elements can be chosen independently from a different distribution than the off-diagonal elements without changing the asymptotic behaviour of the ESD.

a mean degree $p(n-1) = d > (\log n)^a$ with $a > 3$, after a proper scaling and centering of the matrix elements, the ESD and many other spectral characteristic were extensively studied in [56, 57]. For the case with fixed d, less is known. This is an active field of research with many open problems. See [11, 46] and reference therein for an overview.

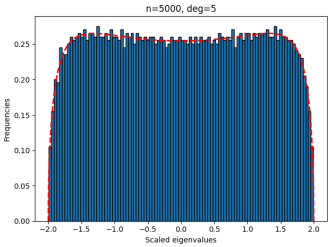
For a random regular graph with degree $d > 3$ a similar result applies, and the convergence is to the Kesten-McKay distribution

$$\mu_{\mathrm{KM}}^{d}(\mathrm{d}x) = \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)}\mathrm{d}x, \qquad |x| \leq 2\sqrt{d-1}, \tag{1.8.3}$$

where the adjacency matrix has been normalized by the square root sof the degree, $\sqrt{d}$.



(a) *In blue, histogram of the eigenvalues of a random regular graph with* $d = 3$ *and* 5000 *nodes. In red, (scaled) Kesten-McKay distribution with* $d = 3$.

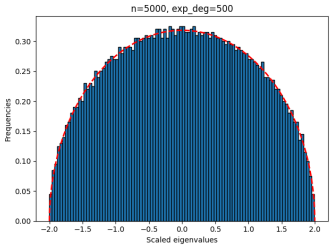


(b) *In blue, histogram of the eigenvalues of a random regular graph with* $d = 5$ *and* 5000 *nodes. In red, (scaled) Kesten-McKay distribution with* $d = 5$.

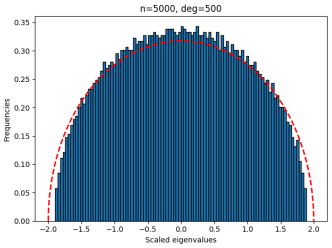
Further properties of spectral statistics of random regular graph with fixed degree were studied in [15, 16]. For a growing $d = d(n) \gg 1$, it was proved in [124] that

$$\lim_{d\to\infty} \mu_{\mathrm{KM}}^{d} = \mu_{\mathrm{sc}}. \tag{1.8.4}$$

By (1.8.4) and the above observations, ESD cannot be the right quantity to look at Macrostate BEE. Indeed, for sufficient large degrees, the ESDs of the microcanonical and the canonical ensemble (i.e., the random regular graph and the ERRG in the homogenous case) are asymptotically equivalent.



(a) *In blue, histogram of the eigenvalues of an Erdős-Rényi random graph with* $\mathbb{E}[d] = 500$ *and* 5000 *nodes. In red, the semicircle distribution.*



(b) *In blue, histogram of the eigenvalues of a random regular graph with* $d = 500$ *and* 5000 *nodes. In red, (scaled) Kesten-McKay distribution with* $d = 500$ *(which is practically indistinguishable from a semicircle law).*

This is no surprise. Indeed, over time it has been understood that convergence to the semicircle law is a universal phenomenon (a type of central limit theorem for matrices) that does not depend on the particular distribution or characteristics of the random variables that form the model. It turns out that the characteristics of the model are better captured by the *non-normalized* largest eigenvalue of the *non-centered* adjacency matrix. This object carries important information on the model, and shows interesting behavior such as a phase transition dependent on the degree of the graph [13]. In what follows we will explore to what extent the principal eigenvalue, λ_1, is a good indicator of breaking of ensemble equivalence, and we will prove the following conjecture in the cases under study:

$$\begin{aligned} \Delta_\infty \neq 0 \quad &\implies \quad \text{BEE}, \\ \text{BEE} \quad &\implies \quad \Delta_\infty \neq 0 \quad \text{apart from exceptional constraints}, \end{aligned} \tag{1.8.5}$$

where

$$\Delta_\infty = \lim_{n\to\infty} \Big(\mathbb{E}_{\text{can}}[\lambda_1(n)] - \mathbb{E}_{\text{mic}}[\lambda_1(n)] \Big). \tag{1.8.6}$$

1.9 Outline of the thesis

Chapters 2–5 deal with the following:

- In Chapter 2 we analyze the homogeneous case, when the degree sequence $\vec{d}$ is constant and equal to d. We will show that, while spectral BEE appears for the pair Erdős-Rényi random graph and random regular graph, it does not for a model where we just fix the total number of edges. The latter is a less strong constraint that does not give rise to measure BEE on scale n and, according to (1.8.5), neither Spectral BEE. To prove this result we relate tail events under $\mathbb{P}_{\text{can}}$ and $\mathbb{P}_{\text{mic}}$, namely, we will show that, for given event $\mathcal{E}$, it is possible to obtain a bound on the tail of this event in $\mathbb{P}_{\text{mic}}$ by just looking at the tail decay of $\mathcal{E}$ in $\mathbb{P}_{\text{can}}$. This trick is possible only when the tail of $\mathbb{P}_{\text{can}}(\mathcal{E})$ goes to zero faster than $\exp\left(-S(\mathbb{P}_{\text{mic}}|\mathbb{P}_{\text{can}})\right)$. It generalizes the method used in [124], and gives a general tool to prove concentration inequalities of matrix ensembles with dependent entries that can be described in the canonical versus microcanonical formalism.

- In Chapter 3 we study the inhomogeneous case, for a non-constant degree sequences $\vec{d}$ with some restrictions on the degree density and the degree inhomogeneity. The resulting model is the one described by (1.6.1), where the connection probability can be further simplified given the density assumptions. For this model, we first show that λ_1 can be expressed as a series expansion in terms of the powers of the centered ajdacency matrix $H = A - \mathbb{E}[A]$. Once this is achieved, we can accurately compute the expectation of λ_1 as a function of the degree sequence, providing the leading and error terms coming from the series expansion. This is a first step in proving spectral BEE in the inhomogeneous case, for which $\mathbb{E}_{\text{can}}[\lambda_1]$ was not known. We also derive a central limit theorem for λ_1, taking advantage of the particular form of the terms that appear in the series expansion. Furthermore, the same formula that produces the expansion of λ_1 gives an analogous result for v_1, the eigenvalue cor-

responding to λ_1. We derive a law of large numbers and a central limit theorem for each component of this normalized eigenvector.

– In Chapter 4 we analyze the configuration model, and compute the expectation of λ_1 conditional on simplicity. This leads to the microcanonical ensemble of the previous chapter. To do so, we need to perform a series expansion of λ_1 similar to the one performed in Chapter 3. A key step to achieve this is to analyze the spectral norm $\|H\|$ of the centered matrix $H = A - \mathbb{E}[A]$, in order to obtain good bounds. In particular, we need $\|H\|$ to be $O(\sqrt{d})$ with a super-polynomial small probability. Once this is solved, $\mathbb{E}_{\text{mic}}[\lambda_1]$ is calculated from the terms of the series expansion. The result obtained, compared to the one obtained in Chapter 3, confirms the conjecture in (1.8.5) for this model, and provides a value of Δ_∞ consistent with the homogeneous case.

– In Chapter 5 we offer a brief discussion of how to properly sample the graphs that are considered in the present thesis, followed by some simulations that helped us to understand the problem under study and that may serve as an inspiration for future research.

1.10 Conclusions

We analyzed breaking of ensemble equivalence from the macrostate perspective and indentified a quantity that is capable to spot this phenomenon, for the classes of random graphs studied in this thesis. Many natural questions remain to be solved.

A first question is how general the conjecture in (1.8.5) is. It is easy to cook up a counterexample where the constraint appearing in the Hamiltonian is the eigenvalue itself. At that point it is natural that $\mathbb{E}_{\text{mic}}[\lambda_1] = \mathbb{E}_{\text{can}}[\lambda_1]$. For constraints different from the pure degree sequence, less is known, starting from the order of divergence of $s_n^{\alpha_n}$ in (1.7.1). It is fair to expect that if the constraint is a function of the graph that forces the ensemble to pick specific degree sequences, then (1.8.5) holds. For instance, ERRGs with an excessive number of triangles have clusters with very

dense vertices (with high degrees), forcing the model to pick realizations with peculiar degree sequences. Given the type of arguments used in our proofs, it is reasonable to expect that something like (1.8.5) happens even in this case, where the heart of the problem is now the difficulty to obtain the connection probabilities of the canonical model in closed form (like in (1.6.1)), in order to allow for explicit calculations.

Another question is whether there exist quantities different from λ_1 that are able to spot BEE at the macrostate level. Arguably, any function of the constraints that contains in its definition the second moment of the constraints will have a discrepancy between the expectations in the two ensembles. This is the case for λ_1, for which the expansion we used to calculate its expectation is composed of simpler quantities and contains a term related to the second moment of the degree sequence. Indeed, while for the microcanonical ensemble the variance of the constraint function $\vec{C}$ is zero, for the canonical it is not. For λ_1 more is true. Every term in the expansion of λ_1 contains a combination of different moments of the degree sequence, so every constraint that affects a moment of the degree distribution in a different way in the two ensembles will be detected at some order. It is therefore difficult to conjecture a quantity other than λ_1 that has the right properties to be a *universal BEE signature*.

A deeper understanding of the relations between measure BEE and macrostate BEE is also needed. Lemma 3.1 links the tail behaviour of events in the microcanonical ensemble to their tails in the canonical ensemble. This convenient approach gives for free an upper bound on the scaling of the tail events of the microcanonical ensemble *if* the tail of the same event goes to zero in the canonical ensemble faster than $\mathrm{e}^{-S_n(\mathbb{P}_{\mathrm{mic}}|\mathbb{P}_{\mathrm{can}})}$. Whether the latter is a necessary condition as well remains an interesting and unanswered question. The combinatorial implications of the above would be substantial, especially in view of the conjecture put forward in [111], where a simple method to calculate the scaling of $s_n^{\alpha_n}$ is described. Indeed, the canonical ensemble is the model with less correlations between its entries, is easier to use for the calculations of tail events, and provides a good way to obtain tail bounds on functions of dependent random variables once the problem is embedded in the

canonical versus microcanonical framework.

A further research line that we are pursuing is to derive the CLT behaviour of λ_1 in the configuration model of Chapter 4, in the same way as this was obtained in Chapter 3 for the Chung-Lu model. Furthermore, it would be interesting to see whether the largest eigenvalues of the models analyzed in Chapter 3 and 4 do behave as a Gaussian process when a suitable dynamics is defined on the respective graph spaces (for example, a switching chain on the configuration model conditional on simplicity).

Chapter 2

A spectral signature of breaking of ensemble equivalence

This chapter is based on:
P. Dionigi, D. Garlaschelli, F. den Hollander, M. Mandjes. *A spectral signature of breaking of ensemble equivalence*. Electronic Communications in Probability, 2021.

Abstract

For random systems subject to a constraint, the *microcanonical ensemble* requires the constraint to be met by every realisation ('hard constraint'), while the *canonical ensemble* requires the constraint to be met only on average ('soft constraint'). It is known that for random graphs subject to topological constraints *breaking of ensemble equivalence* may occur when the size of the graph tends to infinity, signalled by a non-vanishing *specific relative entropy* of the two ensembles. We

investigate to what extent breaking of ensemble equivalence is manifested through the *largest eigenvalue* of the adjacency matrix of the graph. We consider two examples of constraints in the dense regime: (1) fix the degrees of the vertices (= the degree sequence); (2) fix the sum of the degrees of the vertices (= twice the number of edges). Example (1) imposes an extensive number of local constraints and is known to lead to breaking of ensemble equivalence. Example (2) imposes a single global constraint and is known to lead to ensemble equivalence. Our working hypothesis is that breaking of ensemble equivalence corresponds to a *non-vanishing difference of the expected values of the largest eigenvalue* under the two ensembles. We verify that, in the limit as the size of the graph tends to infinity, the difference between the expected values of the largest eigenvalue in the two ensembles does not vanish for (1) and vanishes for (2). A key tool in our analysis is a transfer method that uses relative entropy to determine whether probabilistic estimates can be carried over from the canonical ensemble to the microcanonical ensemble, and illustrates how breaking of ensemble equivalence may prevent this from being possible.

2.1 Introduction

Background. Spectral properties of random graphs have been studied intensively in past years. A non-exhaustive list of key contributions is [7, 18, 38, 51, 52, 56, 57, 66, 88, 133]. Both the adjacency matrix and the Laplacian matrix have been popular. Scaling properties have been derived for the spectral distribution and the largest eigenvalue, with focus on central limit and large deviation behaviour. Most papers deal with random graphs whose edges are drawn *independently*. Different types of behaviour show up in the *dense regime* (where the number of edges is of the order of the square of the number of vertices), in the *sparse regime* (where the number of edges is of the order of the number of vertices),

and in between.

In this paper we focus on the *largest eigenvalue* of the non-normalized and non-centred adjacency matrix for a class of *constrained* random graphs. The largest eigenvalue is a highly non-linear functional of the entries of the adjacency matrix and therefore carries global information about the structure of the graph. Constraints are natural in the framework of statistical mechanics and *Gibbs ensembles*. Typically, they introduce a dependence between the edges that makes the spectral analysis challenging.

Breaking of ensemble equivalence (BEE). One of the interesting phenomena exhibited by certain classes of constrained random graphs is *Breaking of Ensemble Equivalence* (BEE). To understand what this is, we recall that in statistical physics different microscopic descriptions are available for a system that is subjected to a constraint, referred to as *Gibbs ensembles*. In the *microcanonical ensemble* the constraint is *hard*, i.e., each microscopic realisation of the system matches the constraint *exactly*. In the *canonical ensemble* the constraint is *soft*, i.e., is met only *on average*. For finite systems the two ensembles are clearly different, since they represent different physical situations (energetic isolation, respectively, thermal equilibrium with a reservoir at an appropriate temperture). However, the general belief is that this discrepancy vanishes in the thermodynamic limit. This expectation, referred to as *Equivalence of Ensembles* (EE), permeates the theory of Gibbs ensembles. It turns out that for many physical systems EE holds, but not for all. We refer to [117] for more background.

For interacting particle systems, EE has been studied at three different levels: thermodynamic, macrostate and measure. It was shown in [117] that these levels are equivalent. The present paper uses the measure level, which is based on the vanishing of the specific relative entropy. In [50, 70, 107, 108], the phenomenon of BEE was studied for random graphs subject to different types of constraints. It was found that, interestingly, BEE is the rule rather than the exception for constraints that are either *extensive* in the number of vertices or *frustrated*. An overview can be found in [102].

Spectral signature of BEE. Let A be the adjacency matrix of a random graph on n vertices, i.e., $A = \{a_{ij}\}_{i,j\in[n]}$ with $a_{ij} = 1_{\{i\sim j\}}$. Let $\lambda_1(n)$ denote its largest eigenvalue. For $i \in [n]$, let d_i be the degree of vertex i. Write $\mathbb{E}_{\rm can}$ and $\mathbb{E}_{\rm mic}$ to denote expectation with respect to the canonical, respectively, microcanonical ensemble. Put

$$\Delta_\infty = \lim_{n\to\infty} \Big(\mathbb{E}_{\rm can}[\lambda_1(n)] - \mathbb{E}_{\rm mic}[\lambda_1(n)]\Big). \tag{2.1.1}$$

Our *working hypothesis* is that

$$\begin{array}{lcll} \Delta_\infty \neq 0 & \Longrightarrow & \text{BEE}, & \\ \text{BEE} & \Longrightarrow & \Delta_\infty \neq 0 & \text{apart from exceptional constraints.} \end{array} \tag{2.1.2}$$

The goal of the present paper and future work is to verify when this working hypothesis is valid and to identify what are the exceptional constraints (see Remark 2.1.4 below).

We will verify the working hypothesis for two specific examples of constraints: (1) fix the degrees of the vertices (= the degree sequence); (2) fix the sum of the degrees of the vertices (= twice the number of edges). Example (1) corresponds to the so-called *configuration model*. We consider the particular case where all the degrees are fixed at a common value $d(n)$, in which case the microcanonical ensemble becomes the $d(n)$-regular random graph, for which $\lambda_1(n) = d(n)$ with probability 1. For this case, BEE is known to occur for all choices of $d(n) \neq \{0, n-1\}$, and we will see that $\Delta_\infty \neq 0$ except in the ultra-dense regime where $\lim_{n\to\infty} n^{-1}d(n) = 1$. The failure of our working hypothesis in this regime is a consequence of the saturation of the adjacency matrix. Indeed, the largest eigenvalue becomes ineffective in detecting BEE when the two ensembles asymptotically concentrate around the complete graph, for which the largest eigenvalue achieves the maximal value $n-1$. In contrast, relative entropy manages to detect BEE because the two ensembles still look different in the ultra-dense regime, where the number of achievable graphs scales as the exponential of n^2. For Example (2) we will see that no BEE occurs and that $\Delta_\infty = 0$. For both examples the canonical ensemble coincides with the Erdős-Rényi random graph with an appropriate retention probability [108].

For Erdős-Rényi random graphs, $\lambda_1(n)$ was studied for various different regimes in [57, 66, 88]. Throughout the sequel we consider the regime

$$\exists\, \beta \in (6, \infty): \quad n^{-1}(\log n)^{\beta} \leq p(n) < 1 - n^{-1}(\log n)^{\beta}. \tag{2.1.3}$$

Theorem 2.1.1. [57, Theorem 6.2] *Let $G(n, p(n))$ be the Erdős-Rényi random graph on n vertices with retention probability $p(n)$ satisfying* (2.1.3). *Let $\lambda_1(n)$ be the largest eigenvalue of the adjacency matrix of $G(n, p(n))$. Then*

$$\mathbb{E}_{G(n,p(n))}[\lambda_1(n)] = (n-1)p(n) + (1-p(n)) + O\left(\frac{(1-p(n))^{3/2}}{q(n)\sqrt{(n-1)p(n)}}\right), \tag{2.1.4}$$

where $q(n) = \sqrt{(n-1)p(n)}$ when $p(n) \leq c < 1$, and $q(n) = \sqrt{(n-1)(1-p(n))}$ when $p(n) = 1 - o(1)$.

To state (2.1.4), we removed diagonal entries so as to get simple graphs, as explained in Chapter 2.5.1. Theorem 2.1.1 shows that the largest eigenvalue of the Erdős-Rényi random graph is a perturbative correction around the mean degree $d(n) = (n-1)p(n)$. In the dense regime $p(n) \equiv p \in (0, 1)$ we get the classical result from [66]. In the ultra-dense regime, where the complementary graph is sparse, we can still use [57, Definition 2.1]. The lower bound on $p(n)$ in (2.1.3) implies that we do not capture the sparse regime below the connectivity threshold: a crossover in the scaling behaviour of $\lambda_1(n)$ occurs when $d(n) \asymp \log n$, as proved in [7].

Theorem 2.1.1 leads us to our main result.

Theorem 2.1.2. *Let $p(n)$ satisfy* (2.1.3).
(1) *Let the constraint be $d_i = d(n)$, $i \in [n]$, with $nd(n)$ even and $\lim_{n\to\infty}[n^{-1}d(n)]/p(n) = 1$. Then*

$$\Delta_\infty = \begin{cases} 1-p, & \text{if } p(n) \equiv p \in (0,1), \\ 1, & \text{if } p(n) = o(1), \\ 0, & \text{if } p(n) = 1 - o(1). \end{cases} \tag{2.1.5}$$

(2) *Let the constraint be $\frac{1}{2}\sum_{i\in[n]} d_i = L(n)$ with $\lim_{n\to\infty}[2n^{-2}L(n)]/p(n) = 1$. Then*

$$\Delta_\infty = 0. \tag{2.1.6}$$

The restriction that $nd(n)$ is even is needed to make the constraint *graphical*, i.e., there exist simple graphs that meet the constraint. Note the remarkable fact that both $\mathbb{E}_{\mathrm{mic}}[\lambda_1(n)]$ and $\mathbb{E}_{\mathrm{can}}[\lambda_1(n)]$ tend to infinity as $n \to \infty$ while their difference remains bounded.

As shown in [70, 108], BEE occurs in example (1) and EE in example (2), and hence Theorem 2.1.2 supports our working hypothesis that BEE corresponds to a non-vanishing difference of the expected largest eigenvalues under the two ensembles.

Remark 2.1.3. *In [88] a general technique is used that also covers the regime $0 < p(n) < n^{-1}(\log n)^{\beta}$. However, as stated by the authors in their conclusions, their method does not allow for a derivation of the asymptotics of $\mathbb{E}[\lambda_1(n)]$. Nevertheless, it is worth mentioning that when $p(n) = \frac{c}{n}$, $c \in (0, \infty)$, the asymptotic behaviour of $\lambda_1(n)$ in the Erdős-Rényi model $G(n, p(n))$ is*

$$\lim_{n\to\infty} \left(\lambda_1(n) - \sqrt{\frac{\log n}{\log\log n}} \right) = 0 \tag{2.1.7}$$

with high probability. Interestingly, in view of the results in Section 2.4, this suggests that (2.1.5) *may have limit ∞ in this regime.*

Remark 2.1.4. *In [70] it is shown that BEE occurs for three regimes of constant degree $d(n)$: (I) $d(n) = o(\sqrt{n})$ (sparse regime); (II) $\delta n \leq d(n) \leq 1-\delta$ for some $\delta \in (0, \frac{1}{2}]$ (dense regime); (III) $d(n) = n - o(\sqrt{n})$ (ultra dense regime). The scaling of the specific relative entropy is n for regimes (I) and (II), and $n \log n$ for regime (II). Theorem 2.1.2(1) shows that our working hypothesis holds in regime (I) (subject to $d(n) \geq (\log n)^{\beta}$) and (II), but fails in regime (III). The reason is that, while the specific relative entropy is invariant under the map where edges are replaced by non-edges and vice versa, the same is not true for the largest eigenvalue. In the ultra dense regime, other spectral quantities may be better candidates to look at than the maximal eigenvalue. This is no surprise: in [117] it was shown that the relative entropy is the most sensitive global quantity to detect BEE, while other global quantities may detect BEE in certain settings and fail to do so in others. For instance, if the constraint is that the maximal eigenvalue takes a prescribed value, then clearly $\Delta_{\infty} = 0$ while BEE may still be possible.*

Outline. The remainder of this paper is organised as follows. In Section 2.2 we recall the definition of the microcanonical and the canonical

ensemble in the setting of constrained random graphs. Section 2.3 describes our main tool: a transfer method based on relative entropy, which carries over estimates on rare events from the canonical ensemble to the microcanonical ensemble, and describe its role in the general framework of BEE. In Section 2.4 we prove Theorem 2.1.2(1), in Section 2.5 we prove Theorem 2.1.2(2).

2.2 Gibbs ensembles for constrained random graphs

Consider the discrete probability space $(\mathbb{G}_n, \mathcal{B}, \mathbb{P})$, with $\mathbb{G}_n$ the set of all simple graphs on n vertices, $\mathcal{B} = 2^{\mathbb{G}_n}$ the power set of $\mathbb{G}_n$ consisting of all the subsets of $\mathbb{G}_n$, and $\mathbb{P}$ a probability measure.

A *constraint* is defined to be a vector-valued function $\vec{C}\colon\ \mathbb{G}_n \to \mathbb{R}^d$. Fix a value $\vec{C}^\star$ that is *graphical*, i.e., $\vec{C}(g) = \vec{C}^\star$ for at least one $g \in \mathbb{G}_n$. Define

$$\Gamma_{\vec{C}^\star} = \left\{ g \in \mathbb{G}_n \colon\ \vec{C}(g) = \vec{C}^\star \right\}. \tag{2.2.1}$$

The *microcanonical ensemble* is the uniform probability distribution on $\Gamma_{\vec{C}^\star}$:

$$\mathbb{P}_{\text{mic}}(g) = \begin{cases} 1/|\Gamma_{\vec{C}^\star}|, & \text{if } g \in \Gamma_{\vec{C}^\star}, \\ 0, & \text{otherwise.} \end{cases} \tag{2.2.2}$$

The *canonical ensemble* is defined via the Hamiltonian $H(g, \vec{\theta}) = \langle \vec{\theta}, \vec{C}(g) \rangle$ (where $\langle \cdot, \cdot \rangle$ denotes the scalar product), namely,

$$\mathbb{P}_{\text{can}}(g) = \frac{1}{Z_{\vec{\theta}^\star}} \, \mathrm{e}^{-H(g, \vec{\theta}^\star)}, \qquad g \in \mathbb{G}_n, \tag{2.2.3}$$

with the normalising factor $Z_{\vec{\theta}^\star} = \sum_{g \in \mathbb{G}_n} \exp[-H(g, \vec{\theta}^\star)]$, called the *partition function*. Note that both $\mathbb{P}_{\text{mic}}$ and $\mathbb{P}_{\text{can}}$ depend on n, but we suppress this dependence. The parameter $\vec{\theta}$ is set to the particular value $\vec{\theta}^\star$ that realises the constraint:

$$\mathbb{E}_{\text{can}}\left[\vec{C}\right]\Big|_{\theta = \theta^\star} = \vec{C}^\star. \tag{2.2.4}$$

The constraint $\vec{C}^\star$, apart from being graphical, must also be *irreducible*, i.e., no subset of the constraint is redundant [107]. Once these conditions

are met, the value of $\vec{\theta}^\star$ that solves (2.2.4) is unique, and so the canonical ensemble is well defined (see the appendices in [107] for further details).

The relative entropy of $\mathbb{P}_{\text{mic}}$ w.r.t. $\mathbb{P}_{\text{can}}$ is defined as

$$\begin{aligned} S_n(\mathbb{P}_{\text{mic}} \parallel \mathbb{P}_{\text{can}}) &= \sum_{g \in \mathbb{G}_n} \mathbb{P}_{\text{mic}}(g) \log \frac{\mathbb{P}_{\text{mic}}(g)}{\mathbb{P}_{\text{can}}(g)} = \frac{1}{|\Gamma_{\vec{C}^\star}|} \sum_{g \in \Gamma_{\vec{C}^\star}} \log \frac{\mathbb{P}_{\text{mic}}(g)}{\mathbb{P}_{\text{can}}(g)} \\ &= -\frac{1}{|\Gamma_{\vec{C}^\star}|} \log \left[|\Gamma_{\vec{C}^\star}| \, \mathbb{P}_{\text{can}}(g^\star) \right] \sum_{g \in \Gamma_{\vec{C}^\star}} 1 = -\log \mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star}) \end{aligned} \tag{2.2.5}$$

where we use the convention $0 \log 0 = 0$ and $g^\star$ is any graph in $\Gamma_{\vec{C}^\star}$. EE in the measure sense is defined as the vanishing of the relative entropy density, i.e., $\lim_{n\to\infty} n^{-1} S_n(\mathbb{P}_{\text{mic}} \parallel \mathbb{P}_{\text{can}}) = 0$ (see [117]).

2.3 Transfer method

Comparison of the two ensembles. The additional freedom in the canonical ensemble implies that there is less dependence between the constituent random variables. In our case these random variables are the edges of the graph. For example, if the constraint is on the degree sequence, then the microcanonical ensemble corresponds to the *hard configuration model* (which in the case of constant degrees becomes the regular random graph), while the canonical ensemble corresponds to the *soft configuration model* (which is a special case of the generalized random graph model). The former requires an algorithm that randomly pairs half-edges and creates dependencies, while the latter is constructed via a sequence of independent random trials (which results in a multivariate Poisson-Binomial distribution for the degrees of the vertices [70]). Consequently, in the canonical ensemble calculations are carried out more easily. For example, a lot is known about spectral properties of adjacency matrices of random graphs under the canonical ensemble: because the entries of the adjacency matrix are independent, powerful tools from random matrix theory can be used. The challenge is to transfer properties from the canonical ensemble to the microcanonical ensemble without performing elaborate combinatorial computations.

Transfer principle. We start by noting that

$$\mathbb{P}_{\text{mic}}(B) = \frac{\mathbb{P}_{\text{can}}(B)}{\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star})}, \qquad B \subseteq \Gamma_{\vec{C}^\star}. \tag{2.3.1}$$

The latter holds because $g \mapsto H(g, \vec{\theta}^\star)$ and $g \mapsto \mathbb{P}_{\text{can}}(g)$ are constant on the support of $\mathbb{P}_{\text{mic}}$, i.e., all microcanonical realisations have the same probability under the canonical ensemble. In particular,

$$\mathbb{P}_{\text{can}}(B \mid \Gamma_{\vec{C}^\star}) = \mathbb{P}_{\text{mic}}(B), \qquad B \in \mathcal{B}, \tag{2.3.2}$$

where again $\mathcal{B} = 2^{\mathbb{G}_n}$. Consequently, we have the following *transfer principle*.

Lemma 2.3.1. *For every $B \in \mathcal{B}$, if $\lim_{n\to\infty} \mathbb{P}_{\text{can}}(B \mid \Gamma_{\vec{C}^\star}) = 0$, then $\lim_{n\to\infty} \mathbb{P}_{\text{mic}}(B) = 0$.*

Distinguishing sets. Let $\mathcal{E}_{\mathcal{P}} \in \mathcal{B}$ be the subset of $\mathbb{G}_n$ given by

$$\mathcal{E}_{\mathcal{P}} = \{g \in \mathbb{G}_n : \ g \text{ has property } \mathcal{P}\}. \tag{2.3.3}$$

Write $[\mathcal{E}_{\mathcal{P}}]^c$ to denote the complementary event. The crucial step in the argument underlying the transfer method is to find the right event $[\mathcal{E}_{\mathcal{P}}]^c$ that asymptotically implies failure of the property $\mathcal{P}$ that we want to transfer from the canonical ensemble to the microcanonical ensemble.

For the remainder, two events are important: $\mathcal{E}_{\mathcal{P}} \cap \Gamma_{\vec{C}^\star}$ and $[\mathcal{E}_{\mathcal{P}}]^c \cap \Gamma_{\vec{C}^\star}$. These represent the sets that are in the support of $\mathbb{P}_{\text{mic}}$ for which property $\mathcal{P}$ holds and fails, respectively. Our focus will be on replacing $\mathbb{P}_{\text{can}}([\mathcal{E}_{\mathcal{P}}]^c \cap \Gamma_{\vec{C}^\star})$ by $\mathbb{P}_{\text{can}}([\mathcal{E}_{\mathcal{P}}]^c)$. Since $\mathbb{P}_{\text{mic}}([\mathcal{E}_{\mathcal{P}}]^c \cap \Gamma_{\vec{C}^\star}) \leq \mathbb{P}_{\text{mic}}([\mathcal{E}_{\mathcal{P}}]^c)$, if we are able to prove that $\lim_{n\to\infty} \mathbb{P}_{\text{mic}}([\mathcal{E}_{\mathcal{P}}]^c) = 0$, then we also have $\lim_{n\to\infty} \mathbb{P}_{\text{mic}}([\mathcal{E}_{\mathcal{P}}]^c \cap \Gamma_{\vec{C}^\star}) = 0$, and we say that the property defining the set $\mathcal{E}_{\mathcal{P}}$ holds with high probability as $n \to \infty$. As explained in Section 2.2,

$$\mathbb{P}_{\text{can}}\big([\mathcal{E}_{\mathcal{P}}]^c \mid \Gamma_{\vec{C}^\star}\big) = \frac{\mathbb{P}_{\text{can}}([\mathcal{E}_{\mathcal{P}}]^c \cap \Gamma_{\vec{C}^\star})}{\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star})} \leq \frac{\mathbb{P}_{\text{can}}([\mathcal{E}_{\mathcal{P}}]^c)}{\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star})}, \tag{2.3.4}$$

and so if we manage to prove that $\mathbb{P}_{\text{can}}([\mathcal{E}_{\mathcal{P}}]^c) = o(\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star}))$, then we obtain
$\lim_{n\to\infty} \mathbb{P}_{\text{mic}}([\mathcal{E}_{\mathcal{P}}]^c) = 0$.

Role of relative entropy and BEE. Equation (2.3.4) sets the scale at which the transfer method is effective. This scale is given by the denominator $\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star})$. Indeed, if it happens that $\mathbb{P}_{\text{can}}([\mathcal{E}_\mathcal{P}]^c) \neq o(\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star}))$, then (2.3.4) is ineffective. Importantly, from (2.2.5) we have

$$\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star}) = \mathrm{e}^{-S_n(\mathbb{P}_{\text{mic}} \,\|\, \mathbb{P}_{\text{can}})}. \tag{2.3.5}$$

This leads to an interesting connection between BEE and the transferability of a property $\mathcal{P}$: if $\mathbb{P}_{\text{can}}([\mathcal{E}_\mathcal{P}]^c) = o(\mathrm{e}^{-S_n(\mathbb{P}_{\text{mic}} \,\|\, \mathbb{P}_{\text{can}})})$, then $\lim_{n\to\infty} \mathbb{P}_{\text{mic}}([\mathcal{E}_\mathcal{P}]^c) = 0$. Since EE coincides with $S_n(\mathbb{P}_{\text{mic}} \parallel \mathbb{P}_{\text{can}}) = o(n)$, when the ensembles are equivalent it is easier to transfer. Our proof of Theorem 2.1.2(2) makes use of precisely this fact, and $\mathcal{P}$ is a certain concentration inequality for the largest eigenvalue of the adjacency matrix. By contrast, BEE makes the transfer more difficult. Indeed, Theorem 2.1.2(1) can be seen as an example where the same concentration inequality $\mathcal{P}$ cannot be transferred because the relative entropy is of higher order, namely, $S_n(\mathbb{P}_{\text{mic}} \parallel \mathbb{P}_{\text{can}}) = \Theta(n \log n)$ [70, 108].

Largest eigenvalue. We know from the results in [117] that ,whenever BEE occurs, there must exist quantities whose macrostate expectation is different under the two ensembles. Clearly, not all macroscopic quantities are good candidates for this. For instance, any linear combination of the constraints necessarily has the same expected value under the two ensembles. What we propose as a candidate is the largest eigenvalue of the adjacency matrix of the graph, because this is a highly nonlinear function of the imposed constraints and is sensitive to the global structure of the graph. In Sections 2.4–2.5 we will consider two examples of constraints in the dense regime: (1) fix the degrees of all the vertices; (2) fix the total number of edges. For the former we focus on the special case where all the degrees are equal.

Remark 2.3.2. *Since $\lambda_1(A) = \sup_{\|x\|=1} x^T A x$, Jensen's inequality implies that $\lambda_1(A)$ is a convex function of the entries of the matrix A, which means that for both ensembles*

$$\lambda_1\left(\mathbb{E}_{(\cdot)}[A]\right) \leq \mathbb{E}_{(\cdot)}[\lambda_1(A)]. \tag{2.3.6}$$

Taking into account the results of Theorem 2.1.1 and Section 2.4, we get

$$\lambda_1(\mathbb{E}_{mic}[A]) = \mathbb{E}_{mic}[\lambda_1(A)] = \lambda_1(\mathbb{E}_{can}[A]) \leq \mathbb{E}_{can}[\lambda_1(A)]. \tag{2.3.7}$$

If, on top of the constraint on the degree sequence, we add more (compatible) constraints, then by exchangeability we still have $\lambda_1(\mathbb{E}_{mic}[A]) = \mathbb{E}_{mic}[\lambda_1(A)] = \lambda_1(\mathbb{E}_{can}[A])$. *Applying* (2.3.6)*, we therefore still expect that* $\mathbb{E}_{mic}[\lambda_1(A)] \leq \mathbb{E}_{can}[\lambda_1(A)]$. *This shows that* λ_1 *is particularly sensitive to the moments of the underlying degree sequence (as can also be seen from the* power method *used in [57, 66]; see* (2.5.4) *and* (2.5.35) *below). We may therefore expect that our working hypothesis holds in all those cases where BEE forces the degree sequence to assume either a different mean of a different variance in the two ensembles, as in the case under study.*

2.4 Proof of main Theorem part I: constraint on the degree sequence

In what follows we suppress the n-dependence from $p(n), d(n), \lambda_1(n)$, writing p, d, λ_1. The d-regular random graph with n vertices, written $G_{n,d}$, coincides with the microcanonical ensemble with constraint $\vec{C}^\star = (d, \ldots, d)$ on the degree sequence, where we allow $d = d(n)$. The largest eigenvalue of the adjacency matrix of $G_{n,d}$ equals d, irrespective of n. The Erdős-Rényi random graph with retention probability $p = d/(n-1)$ coincides with the canonical ensemble with the same constraint.

In order to understand the difference in behaviour of λ_1 under the two ensembles, we need Theorem 2.1.1. Indeed, the result in (2.1.4), which actually holds for a generic symmetric random matrix subject to specific regularity conditions, can be interpreted as follows. The adjacency matrix A associated with $G(n,p)$ consists of elements $\{a_{ij}\}_{i,j\in[n]}$ that are identically 0 when $i = j$ and *Bernoulli random trials* ($a_{ij} = 0, 1$) with success probability p when $i \neq j$. The largest eigenvalue of the deterministic matrix $\bar{A}$ whose entries are $\bar{a}_{ij} = \mathbb{E}_{\text{can}}[a_{ij}] = p$ when $i \neq j$ and $\bar{a}_{ij} = 0$ when $i = j$ is given by $\lambda_1(\bar{A}) = (n-1)p$. Hence, compared to $\lambda_1(\bar{A})$, λ_1 is shifted by a random variable whose expected value is $(1-p)$ and is distributed as $\mathcal{N}(1-p, 2p(1-p))$ under certain conditions on d (see [57, equation 6.10]) plus an error term of order dependent on the

considered regime ($O(1/\sqrt{n})$ for p constant). It is important to note that the parameters of this shift depend on p only. In [57, 66] it is shown that (2.1.4) relies on the fact that in the canonical ensemble the eigenvector $\vec{v}_1$ corresponding to the largest eigenvalue λ_1 is very close to the vector $\vec{\mathbb{1}} = (1, \dots, 1)$ (i.e., the norm of the projection of $\vec{v}_1$ onto $\vec{\mathbb{1}}$ is much larger than the norm of the projection of $\vec{v}_1$ onto the perpendicular space $\vec{\mathbb{1}}^{\perp}$).

It was shown in [70] that BEE holds in the all regimes covered in Theorem 2.1.2(1), namely the delta tame regime, which corresponds to $\delta \leq p = d/(n-1) \leq 1 - \delta$ with $\delta \in (0, \frac{1}{2}]$ (see [70, Definition 1.1]) and the sparse regime ($d = o(\sqrt{n})$). Hence the claim in Theorem 2.1.2(1) follows.

2.5 Proof of main Theorem part II: constraint on the total number of edges

Consider the case where the constraint is on the total number of edges: $\vec{C}(g) = \vec{C}^{\star} = \binom{n}{2}p$ for some $p \in (0, 1)$. Then the canonical ensemble is still the Erdős-Rényi random graph with parameter p. It was proved in [108] that the two ensembles are asymptotically equivalent on scale n. In particular, it was shown that $S_n(\mathbb{P}_{\text{mic}} \parallel \mathbb{P}_{\text{can}}) = \log n + \Theta(1)$. The canonical probability of drawing a microcanonical realization is given by (2.3.5):

$$\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^{\star}}) = \mathrm{e}^{-S_n(\mathbb{P}_{\text{mic}} \parallel \mathbb{P}_{\text{can}})} = \mathrm{e}^{-\log n + \Theta(1)} = \Theta(n^{-1}). \tag{2.5.1}$$

Together with (2.3.4), this tells us that if we can find an event $[\mathcal{E}_{\mathcal{P}}]^c$ such that $\mathbb{P}_{\text{can}}([\mathcal{E}_{\mathcal{P}}]^c) = o(n^{-1})$, then we know that $\lim_{n\to\infty} \mathbb{P}_{\text{mic}}(\mathcal{E}_{\mathcal{P}}) = 1$. Our goal is to use the results in [57] to apply (2.3.4) with (2.5.1).

In Section 2.5.1 we show how our results follow from [57, Theorem 6.2] both in the dense and the non-dense regime. In Section 2.5.2 and 2.5.3 we focus on the dense regime and show how our results follow by making the concentration inequalities used in [66] tighter. In particular, we will find that the approach heavily depends on the ability of identifying good concentration inequalities for the degree sequence, which is a special case of the bounds presented in [57]. The heavy dependence on

the degree sequence is further evidence of what was said in Remark 2.3.2. In Section 2.5.2 we prove a concentration inequality for the degrees under the canonical ensemble (Lemma 2.5.1) that is of independent interest. In Section 2.5.3 we use this to prove a concentration inequality for a functional of the degrees that approximates the largest eigenvalue well in the dense regime (Lemma 2.5.3). In Section 2.5.4 we transfer the results from the previous sections to the microcanonical ensemble (Lemma 2.5.4), and show that this leads to a negligible shift of the expected largest eigenvalue.

2.5.1 Proof of main Theorem via known results

In [57, Chapter 6] the largest eigenvalue of matrices of the form $A = A_0 + \mathbb{E}[A] = A_0 + f|\vec{e}\rangle\langle\vec{e}|$ is studied, where A_0 is a matrix with mean-zero entries, $\vec{e} = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)^T$ and $1 + \varepsilon_0 \leq f \leq N^C$, with $\varepsilon_0, C \in (0, \infty)$ constants. In order to consider only adjacency matrices A_s of simple graphs, we have to get rid of the diagonal of A. This can be easily done by considering $A_s = A - pI = A_0 + f|\vec{e}\rangle\langle\vec{e}| - pI$, where $p = p(n)$ is the retention probability that appears Theorem 2.1.1 subject to (2.1.3) and $f = np$. We note that if λ_1 is the largest eigenvalue of A, then $\lambda_1 - p$ is the largest eigenvalue of A_s, so it suffices to study the largest eigenvalue of A.

Let $\tilde{A}$ be the normalized version of A, defined by $\tilde{A} = A/\sqrt{np(1-p)}$. This scaling is needed in order to have $\|H\| = O(1)$ with high probability. Let $\tilde{A}_0$ be the centered version of $\tilde{A}$, i.e., $\tilde{A}_0 = \tilde{A} - \mathbb{E}_{\text{can}}\left[\tilde{A}\right]$. It is easy to see that $\mathbb{E}_{\text{can}}\left[\tilde{A}\right]$ can be expressed as $\sqrt{np/(1-p)}|\vec{e}\rangle\langle\vec{e}|$, where again $\vec{e} = \frac{1}{\sqrt{n}}(1, 1, \ldots, 1)^T$. Following [57, Theorem 6.2], we say that an event $\mathcal{E}$ holds with (ξ, ν)-high probability when

$$\mathbb{P}(\mathcal{E}^c) \leq \mathrm{e}^{-\nu(\log n)^\xi}, \tag{2.5.2}$$

where ν and ξ can be two positive n-dependent constants bounded from below by $\nu > 0$ and $\xi > 1$. Note that $e^{-\nu(\log n)^\xi} = o(n^{-1})$ whenever $\nu > 0$ and $\xi > 1$. Thus, if an event $\mathcal{E}_\mathcal{P}$ of the type described in (2.3.3) holds with (ξ, ν)-high probability under $\mathbb{P}_{\text{can}}$, then by (2.3.4) and (2.5.1) $\mathcal{E}_\mathcal{P}$ it holds

also under $\mathbb{P}_{\text{mic}}$. Starting from the equation

$$\left(\mathbb{I} - \frac{\tilde{A}_0}{\lambda_1}\right)\lambda_1 \vec{v} = \sqrt{\frac{np}{1-p}}\langle \vec{e}, \vec{v}\rangle \vec{e}, \tag{2.5.3}$$

where $\vec{v}$ is the eigenvector associated with λ_1 and $\mathbb{I}$ is the identity matrix, after multiplying by $(\mathbb{I} - \frac{\tilde{A}_0}{\lambda_1})^{-1}$ and projecting on $\vec{e}$ we obtain the following series for λ_1:

$$\lambda_1 = \sqrt{\frac{np}{1-p}} \sum_{k\in\mathbb{N}_0} \left\langle \vec{e}, \left(\frac{\tilde{A}_0}{\lambda_1}\right)^k \vec{e} \right\rangle. \tag{2.5.4}$$

We see that for the series to converge we need $\left\|\tilde{A}_0\right\|/\lambda_1 < 1$. From [57, Lemma 4.3] (see also [4, 106, 116, 130]) and the leading order of (2.5.4) (see also [57, Eq.(6.5)]) we have that $\left\|\tilde{A}_0\right\|/\lambda_1 < 1$ with (ξ,ν)-high probability (which also holds for the microcanonical ensemble). Iterating (2.5.4), we get that with (ξ,ν)-high probability

$$\begin{aligned}\lambda_1 = \sqrt{\tfrac{np}{1-p}} + \left\langle \vec{e}, \tilde{A}_0\vec{e}\right\rangle + \left(\left\langle \vec{e}, \tilde{A}_0^2\vec{e}\right\rangle - \left\langle \vec{e}, \tilde{A}_0\vec{e}\right\rangle^2\right)\left(\sqrt{\tfrac{np}{1-p}}\right)^{-1} \\ + \left(\left\langle \vec{e}, \tilde{A}_0\vec{e}\right\rangle^3 - 3\left\langle \vec{e}, \tilde{A}_0\vec{e}\right\rangle\left\langle \vec{e}, \tilde{A}_0^2\vec{e}\right\rangle\right)\left(\sqrt{\tfrac{np}{1-p}}\right)^{-2} + O\left(\left(\sqrt{\tfrac{np}{1-p}}\right)^{-3} + \left(\tfrac{(np)q}{1-p}\right)^{-1}\right),\end{aligned} \tag{2.5.5}$$

where q is the parameter defined in Theorem 2.1.1. Taking expectations, using [57, Lemma 6.5] and scaling back, we get (2.1.4) for $\mathbb{E}[\lambda_1]-p$, the expected eigenvalue of A_s. Note that all the bounds hold with (ξ,ν)-high probability. We can therefore conclude via (2.3.4) that (2.1.4) approximates λ_1 with a vanishing error also in the microcanonical ensemble, where the constraint is on the total number of edges. Together with the result of Lemma 2.5.3, we conclude that $\lim_{n\to\infty}\left(\mathbb{E}_{\text{can}}\left[\lambda_1\right] - \mathbb{E}_{\text{mic}}\left[\lambda_1\right]\right) = 0$.

2.5.2 Concentration for the degrees under the dense canonical ensemble

For the remainder of the paper we take $p \in (0,1)$ constant and A to be the unnormalized adjacency matrix. For $i \neq j$, $\mathbb{E}_{\text{can}}[a_{ij}] = p$ and

$\mathrm{Var}_{\mathrm{can}}[a_{ij}] = p(1-p)$. In what follows we abbreviate $\mu = p$ and $\sigma^2 = p(1-p)$. We write $\vec{\mathbb{1}} = \vec{v}_1 + \vec{r}$ with $\vec{r} \in \vec{\mathbb{1}}^{\perp}$, $\langle \vec{v}_1, \vec{r} \rangle = 0$ and $A\vec{v}_1 = \lambda_1 \vec{v}_1$. Following the power method in [96], we define

$$\vec{d} = A\vec{\mathbb{1}} = A(\vec{v}_1 + \vec{r}) = \lambda_1 \vec{v}_1 + A\vec{r}, \tag{2.5.6}$$

which is the vector of row sums of the matrix A, i.e., the vector of degrees of the vertices (the degree sequence). Centering $\vec{d}$ by $\Theta\vec{\mathbb{1}}$ with $\Theta = \mathbb{E}[d_i] = (n-1)p$ and using $\vec{\mathbb{1}} = \vec{v}_1 + \vec{r}$, we get

$$\vec{d} - \Theta\vec{\mathbb{1}} = (\lambda_1 - \Theta)\vec{v}_1 + (A\vec{r} - \Theta\vec{r}). \tag{2.5.7}$$

Our key step is the following lemma.

Lemma 2.5.1. *With σ^2 denoting $p(1-p)$, there exist two constants $c_1, c_2 \in (0, \infty)$ such that*

$$\mathbb{P}_{\mathrm{can}}\left(\left| \sum_{i=1}^{n} (d_i - \Theta)^2 - \sigma^2 n(n-1) \right| \geq t \right) \leq c_2 \mathrm{e}^{-c_1 t / n^{3/2}}. \tag{2.5.8}$$

Proof. The term $\sum_{i=1}^{n} (d_i - \Theta)^2$ can be written as

$$\sum_{i=1}^{n} \left(\sum_{j=1}^{n} (a_{ij} - \mathbb{E}_{\mathrm{can}}[a_{ij}]) \right)^2 = \sum_{i=1}^{n} \left(\sum_{j=1}^{n} b_{ij} \right)^2 = \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} b_{ij} b_{ik}, \tag{2.5.9}$$

where

$$b_{ij} = a_{ij} - \mathbb{E}_{\mathrm{can}}[a_{ij}] = \begin{cases} a_{ij} - p, & \text{if } i \neq j, \\ 0, & \text{if } i = j, \end{cases} \tag{2.5.10}$$

are the centred entries of the adjacency matrix. Note that

$$\mathbb{E}_{\mathrm{can}} \left[\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} b_{ij} b_{ik} \right] = \sigma^2 n(n-1). \tag{2.5.11}$$

Straightforward counting shows that the sum in (2.5.9) contains $O(n^3)$ different terms. Let us represent $b_{ij} = b_{ji}$ by a variable X_α, $\alpha \in \left[\binom{n}{2}\right]$. Then (2.5.9) can be rewritten in the form

$$\sum_{\alpha, \beta \in \left[\binom{n}{2}\right]} h_{\alpha\beta} X_\alpha X_\beta, \tag{2.5.12}$$

which is the quadratic form of the matrix $H = \{h_{\alpha\beta}\}_{\alpha,\beta\in\left[\binom{n}{2}\right]}$. Because there is a one-to-one correspondence between the terms in (2.5.12) and (2.5.9), we can conclude that H has $O(n^3)$ entries, whose values are either 1 (off-diagonal) or 2 (diagonal). We can apply to (2.5.12) the *Hanson-Wright* inequality (see [75] or [2, Theorem 1.4, item 6]).

Theorem 2.5.2. *Let $X = (X_1, \dots X_N)$ be mean-zero square-integrable random variables taking values in $\mathbb{R}$, and let $\xi > 0$ be such that*

$$\|X\|_{\psi_2} = \inf\left\{t > 0 : \mathbb{E}\left[\exp\left(\|X\|_2^2/t^2\right)\right] \le 2\right\} \le \xi. \tag{2.5.13}$$

Let $H = (h_{\alpha\beta})_{\alpha\beta\in[N]}$ be a real symmetric matrix. Then $Y = \sum_{\alpha,\beta\in[N]} h_{\alpha\beta}X_\alpha X_\beta$ satisfies

$$\mathbb{P}\left(|Y - \mathbb{E}[Y]| \ge t\right) \le 2\exp\left(-\frac{1}{C}\min\left\{\frac{t^2}{\xi^4\|H\|_{\mathrm{HS}}^2}, \frac{t}{\xi^2\|H\|_{\ell_2^N\to\ell_2^N}}\right\}\right), \qquad t > 0, \tag{2.5.14}$$

where C is a suitable constant, $\|H\|_{\mathrm{HS}}^2 = \sum_{\alpha,\beta\in[N]} h_{\alpha\beta}^2$ is the Hilbert-Schmidt norm of H, and

$$\|H\|_{\ell_2^N\to\ell_2^N}^2 = \sup\left\{\sum_{\alpha,\beta\in[N]} h_{\alpha\beta}x_\alpha y_\beta : \sum_{\alpha\in[N]} x_\alpha^2 \le 1, \sum_{\alpha\in[N]} y_\alpha^2 \le 1\right\} \tag{2.5.15}$$

is the $\ell_2^N \to \ell_2^N$ norm of H.

In our setting, $N = \binom{n}{2}$. Since $|X_\alpha| < 1$, we have $\|X\|_{\psi_2} \le 1/\log 2$, so that (2.5.13) applies with $\xi = 1/\log 2$. Since H has bounded entries, we have $\|H\|_{\mathrm{HS}}^2 = O(n^3)$. Moreover, by the Cauchy-Schwarz inequality we have

$$\|H\|_{\ell_2^N\to\ell_2^N}^2 = \sup\left\{\|Hx\|_2 \colon \|x\|_2 \le 1\right\} = \|H\|_{\mathrm{op}}, \tag{2.5.16}$$

where the latter is the operator norm of H. But

$$\|H\|_{\mathrm{HS}}^2 = \mathrm{Tr}(H^\dagger H) \ge \lambda_{\max}(H^\dagger H) = \|H\|_{\mathrm{op}}^2, \tag{2.5.17}$$

and so the exponent in the right-hand side of (2.5.14) is bounded below by

$$\min\left\{\frac{t^2}{\xi^4 n^3}, \frac{t}{\xi^2 n^{3/2}}\right\} \ge \frac{c_3 t}{n^{3/2}}, \tag{2.5.18}$$

where c_3 is a suitable constant. Taking $c_1 \le c_3/C$, with C the constant appearing in (2.5.14), we obtain (2.5.8).

We end this section with an immediate consequence of Lemma 2.5.1. Picking $t = \sigma^2 n^2$ and using that, for appropriately chosen constants C_1, C_2, C_3, C_4,

$$\frac{\sigma^4 n^4}{\|H\|_{\mathrm{HS}}^2} \geq \frac{\sigma^4 n^4}{C_1 n^3} \geq C_2 n, \qquad \frac{\sigma^2 n^2}{L^2 \|H\|_{\mathrm{op}}} \geq \frac{\sigma^2 n^2}{C_3 \|H\|_{\mathrm{HS}}} \geq C_4 \sqrt{n}, \quad (2.5.19)$$

we find that there are constants $\tilde{c} \leq C_4/C$ and $\tilde{C}$ such that

$$\begin{aligned} &\mathbb{P}_{\mathrm{can}}\left(\left| \sum_{i=1}^n (d_i - \Theta)^2 - \sigma^2 n^2 \right| \geq 2\sigma^2 n^2 \right) \\ &\quad \leq 2 \exp\left(-\frac{1}{C} \min\left\{ \frac{4\sigma^4 n^4}{\|H\|_{HS}^2}, \frac{2\sigma^2 n^2}{\|H\|_{op}} \right\} \right) \leq \tilde{C} \mathrm{e}^{-\tilde{c}\sqrt{n}}. \end{aligned} \quad (2.5.20)$$

2.5.3 Concentration for the largest eigenvalue under the dense canonical ensemble

After applying A once to $\vec{\mathbb{1}}$, we must find a suitable normalization in order to isolate λ_1. This is given by

$$\frac{\sum_{i=1}^n d_i^2}{\sum_{i=1}^n d_i} = \frac{\langle \vec{d}, \vec{d} \rangle}{\langle \vec{\mathbb{1}}, \vec{d} \rangle} = \frac{\|A\vec{\mathbb{1}}\|}{\langle \vec{\mathbb{1}}, A\vec{\mathbb{1}} \rangle} = \lambda_1 + \frac{\|A\vec{r}\|^2 - \lambda_1 \langle \vec{r}, A\vec{r} \rangle}{\sum_{i=1}^n d_i}. \quad (2.5.21)$$

In [66], it was shown that $\sum_{i=1}^n d_i^2 / \sum_{i=1}^n d_i$ approximates λ_1 with high probability, in the sense that for any $x > 0$,

$$\mathbb{P}_{\mathrm{can}}\left(\left| \frac{\sum_{i=1}^n d_i^2}{\sum_{i=1}^n d_i} - \frac{\sum_{i=1}^n d_i}{n} - \frac{\sigma^2}{\mu} \right| \geq \frac{3\sigma^2 x}{\sqrt{n}} \right) \leq \frac{1}{x^2}, \quad (2.5.22)$$

which with the choice $x = \sqrt{n}$ leads to an upper bound of order $1/n$. As it turns out, however, in order to transfer the estimates to the micro-canonical ensemble via (2.3.4), we need the upper bound to hold with probability $o(1/n)$. This result is covered by the following lemma.

Lemma 2.5.3. *Let $\vec{d}$ be as before, and $\mu = p$, $\sigma^2 = p(1-p)$. For any $\gamma > 0$ there exist $\gamma', \gamma_1, \gamma_2$ satisfying $c_1\gamma_1, \gamma_2 > 1$, with c_1 the constant in* (2.5.8)*, such that*

$$\mathbb{P}_{\mathrm{can}}\left(\left| \frac{\sum_{i=1}^n d_i^2}{\sum_{i=1}^n d_i} - \frac{\sum_{i=1}^n d_i}{n} - \frac{\sigma^2}{\mu} \right| \geq \frac{\gamma}{\sqrt{n}} \right) \leq \frac{\gamma'}{n^{\min\{c_1\gamma_1, \gamma_2\}}}. \quad (2.5.23)$$

Proof. First note that

$$\mathbb{E}_{\text{can}}\left[\frac{\sum_{i=1}^n d_i}{n}\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}_{\text{can}}[d_i] = (n-1)p = \Theta \tag{2.5.24}$$

and write

$$\frac{\sum_{i=1}^n d_i^2}{\sum_{i=1}^n d_i} - \frac{\sum_{i=1}^n d_i}{n} = \frac{\sum_{i=1}^n (d_i-\Theta)^2}{\sum_{i=1}^n d_i} - \frac{\left(n^{-1}\sum_{i=1}^n d_i - \Theta\right)^2}{n^{-1}\sum_{i=1} d_i}. \tag{2.5.25}$$

To analyse the first ratio in (2.5.25), note that

$$\sum_{i=1}^n d_i = \sum_{i,j\in[n]} a_{ij} = 2\sum_{i,j\in[n],j>i} a_{ij}. \tag{2.5.26}$$

Applying Hoeffding's inequality (see e.g. [23, 29]), we have

$$\mathbb{P}_{\text{can}}\left(\left|\sum_{i,j\in[n],j>i} a_{ij} - \frac{n(n-1)}{2}\mu\right| \geq t\right) \leq 2\exp\left(-\frac{4t^2}{n(n-1)}\right). \tag{2.5.27}$$

Take $t = n\sqrt{\gamma_2 \log n}$ in (2.5.27) with $\gamma_2 > 1$ and apply Lemma 2.5.1 with $t = n^{3/2}\gamma_1 \log n$, with $\gamma_1 c_1 > 1$ and c_1 the constant in the exponential bound of (2.5.8). Then, for some $\gamma > 0$,

$$\frac{\sum_{i=1}^n (d_i-\Theta)^2}{\sum_{i=1}^n d_i} \leq \frac{n(n-1)\sigma^2 + n^{3/2}\gamma_1 \log n}{n(n-1)\mu + n\sqrt{\gamma_2 \log n}} \leq \frac{\sigma^2}{\mu} + \frac{\gamma}{\sqrt{n}} \tag{2.5.28}$$

with probability at least $1 - 1/n^{\gamma_1 c_1} - 1/n^{\gamma_2}$. Similarly, the probability of

$$\frac{\sum_{i=1}^n (d_i-\Theta)^2}{\sum_{i=1}^n d_i} \geq \frac{\sigma^2}{\mu} - \frac{\gamma}{\sqrt{n}} \tag{2.5.29}$$

is bounded from below by $1 - 1/n^{\gamma_1 c_1} - 1/n^{\gamma_2}$. Hence

$$\mathbb{P}_{\text{can}}\left(\left|\frac{\sum_{i=1}^n (d_i-\Theta)^2}{\sum_{i=1}^n d_i} - \frac{\sigma^2}{\mu}\right| \geq \frac{\gamma}{\sqrt{n}}\right) \leq \frac{\gamma'}{n^{\min\{\gamma_1 c_1, \gamma_2\}}}. \tag{2.5.30}$$

To analyse the second ratio in (2.5.25), we write

$$\left(n^{-1}\sum_{i=1}^n d_i - \Theta\right)^2 = \frac{1}{n^2}\left(2\sum_{i,j\in[n],j>i} (a_{ij} - \mathbb{E}_{\text{can}}[a_{ij}])\right)^2, \tag{2.5.31}$$

and apply Hoeffding's inequality with $t = O(n^2)$ twice. This gives

$$\mathbb{P}_{\text{can}}\left(\frac{\left(n^{-1}\sum_{i=1}^n d_i - \Theta\right)^2}{n^{-1}\sum_{i=1}^n d_i} > \frac{\tilde{\gamma}}{n}\right) \leq \tilde{\gamma}_2 e^{-\tilde{\gamma}_1 n^2}, \qquad \tilde{\gamma} > 0, \tag{2.5.32}$$

where $\tilde{\gamma}_2$ and $\tilde{\gamma}_1$ are suitable constants. Applying the union bound to the complementary events, we obtain (2.5.23).

2.5.4 Transfer to the dense microcanonical ensemble

Next we use the transfer method to pass the property characterised by the event in (2.5.23) to the microcanonical ensemble. Indeed, using the notation of Section 2.3, we identify

$$\left|\frac{\sum_{i=1}^n d_i^2}{\sum_{i=1}^n d_i} - \frac{\sum_{i=1}^n d_i}{n} - \frac{\sigma^2}{\mu}\right| \geq \frac{\gamma}{\sqrt{n}} \tag{2.5.33}$$

as the event $\mathcal{E}_{\mathcal{P}}^c$, i.e., the set of graphs that do not possess the property that we would like to pass on. The fact that $\mathbb{P}_{\text{can}}(\mathcal{E}_{\mathcal{P}}^c)$ tends to zero faster than $\mathbb{P}_{\text{can}}(\Gamma_{\vec{C}^\star})$ (as $n \to \infty$, that is) tells us that also $\mathbb{P}_{\text{mic}}(\mathcal{E}_{\mathcal{P}}^c)$ tends to zero, and implies that

$$\lim_{n\to\infty} \mathbb{P}_{\text{mic}}\left(\left|\frac{\sum_{i=1}^n d_i^2}{\sum_{i=1}^n d_i} - \frac{\sum_{i=1}^n d_i}{n} - \frac{\sigma^2}{\mu}\right| \leq \frac{\gamma}{\sqrt{n}}\right) = 1. \tag{2.5.34}$$

Thus, in the microcanonical ensemble $\sum_{i=1}^n d_i^2 / \sum_{i=1}^n d_i$ concentrates around the sum $n^{-1}\sum_{i=1}^n d_i + \sigma^2/\mu$ with an error of order $1/\sqrt{n}$. However, we need to also see what $n^{-1}\sum_{i=1}^n d_i + \sigma^2/\mu$ is in the microcanonical ensemble. The term σ^2/μ, a constant equal to $1-p$, is in accordance with the constraint in the microcanonical ensemble. For the other term we have $n^{-1}\sum_{i=1}^n d_i = (n-1)p$. The two together give precisely the expected value in the canonical ensemble, as follows from Proposition 2.1.1. Hence we only need to show that $\sum_{i=1}^n d_i^2 / \sum_{i=1}^n d_i$ concentrates around λ_1 also in the microcanonical ensemble, for which we can once more use the transfer method.

Lemma 2.5.4. *For any $\eta > 0$, there exist ζ and Λ such that*

$$\mathbb{P}_{\text{can}}\left(\left|\frac{\sum_{i=1}^n d_i^2}{\sum_{i=1}^n d_i} - \lambda_1\right| \geq \frac{\eta}{\sqrt{n}}\right) \leq \Lambda e^{-\zeta\sqrt{n}}. \tag{2.5.35}$$

Proof. We need to show that the last term in (2.5.21),

$$\frac{\|A\vec{r}\|^2 - \lambda_1 \langle \vec{r}, A\vec{r} \rangle}{\sum_{i=1}^n d_i}, \tag{2.5.36}$$

is small. First we show that $\|\vec{r}\|$ is bounded in probability. Indeed,

$$\sum_{i=1}^n (d_i - \Theta)^2 = (\lambda_1 - \Theta)^2 \|\vec{v_1}\|^2 + \|A\vec{r} - \Theta\vec{r}\|^2. \tag{2.5.37}$$

Since $(\lambda_1 - \Theta)^2 \|\vec{v_1}\|^2 \geq 0$, we have $\|A\vec{r} - \Theta\vec{r}\|^2 < \sum_{i=1}^n (d_i - \Theta)^2$. By the Courant-Fisher theorem [116, Theorem 1.3.2], we get that $\|A\vec{r} - \Theta\vec{r}\| \geq |\Theta - \lambda_2| \|\vec{r}\|$ (indeed, $|\Theta - \lambda_i| \geq |\Theta - \lambda_2|$ for $i > 2$). Next, we need a concentration inequality for λ_2. Use [4, Theorem 1] plus the fact that the largest eigenvalue of a centred matrix is of order $O(\sigma\sqrt{n})$ almost surely [106], [116, Theorem 2.3.24], [130, Theorem 1.3]. Again use the Courant-Fisher theorem to pass to the non-centred case [66, Lemma 1]. We find that, for any $\beta > 2$ and for $\tilde{n}$ large enough,

$$\mathbb{P}_{\text{can}} \left(\max_{i>1} |\lambda_i| \geq \beta\sigma\sqrt{n} \right) \leq 4e^{-\zeta_1 n}, \quad n > \tilde{n} \tag{2.5.38}$$

where ζ_1 is a suitable constant. Since $\max_{i>1}|\lambda_i| \geq \lambda_2 \geq 0$, we can bound $\lambda_2 \leq \beta\sigma\sqrt{n}$ with high probability. Using (2.5.20), we have

$$\mathbb{P}_{\text{can}} \left(\|\vec{r}\|^2 < \frac{(d_i - \Theta)^2}{(\Theta - \lambda_2)^2} < \frac{n^2\sigma^2}{(\mu n - \beta\sigma\sqrt{n})^2} < \frac{4\sigma^2}{\mu^2} \right) \geq 1 - 4e^{-\zeta_1 n} - \tilde{C}e^{-\tilde{c}\sqrt{n}}, \tag{2.5.39}$$

as a consequence of the union bound applied to the last term of $\mathbb{P}\left(\cap_n \mathcal{E}_n\right) = 1 - \mathbb{P}(\cup_n [\mathcal{E}_n]^c)$, with $[\mathcal{E}_n]^c$ denoting the events described by Lemma 2.5.1 and (2.5.38). Thus, we have

$$\mathbb{P}_{\text{can}} \left(\|\vec{r}\|^2 \geq \frac{4\sigma^2}{\mu^2} \right) \leq \tilde{C}_1 e^{-\tilde{c}_1\sqrt{n}}, \tag{2.5.40}$$

where $\tilde{C}_1$ and $\tilde{c}_1$ are suitable constants.

All the other terms in (2.5.36) can be obtained by repeatedly using (2.5.40), (2.5.38) and (2.5.20). Note that in order to get (2.5.40) we have used both (2.5.38) and (2.5.20), and the events that these inequalities identify. Thus, using (2.5.38) twice, we obtain

$$\mathbb{P}_{\text{can}} \left(\|A\vec{r}\|^2 \leq \lambda_2^2 \|\vec{r}\|^2 \leq \frac{50\,\sigma^4}{\mu^2} n \right) \geq 1 - 4e^{-\zeta_1 n} - \tilde{C}e^{-\tilde{c}\sqrt{n}}. \tag{2.5.41}$$

Therefore

$$\mathbb{P}_{\text{can}}\left(\|A\vec{r}\|^2 \geq \frac{50\,\sigma^4}{\mu^2}n\right) \leq \tilde{C}_2 e^{-\tilde{c}_2\sqrt{n}}, \tag{2.5.42}$$

where $\tilde{C}_2$ and $\tilde{c}_2$ are suitable constants. In the same way we can bound $|\langle\vec{r}, A\vec{r}\rangle| \leq \|\vec{r}\|\|A\vec{r}\|$, which yields

$$\mathbb{P}_{\text{can}}\left(|\langle\vec{r}, A\vec{r}\rangle| \geq \frac{2\sqrt{50}\sigma^3}{\mu^2}\sqrt{n}\right) \leq \tilde{C}_3 e^{-\tilde{c}_3\sqrt{n}}. \tag{2.5.43}$$

Now, using the trivial deterministic bound $\lambda_1 \leq \max_i \sum_j |a_{ij}| < n$ and Hoeffding's inequality on $\sum_{i=1}^n d_i = 2\sum_{j>i} a_{ij}$, we can conclude that, for any $\eta > 0$,

$$\mathbb{P}_{\text{can}}\left(\left|\frac{\|A\vec{r}\|^2 - \lambda_1\langle\vec{r}, A\vec{r}\rangle}{\sum_{i=1}^n d_i}\right| \geq \frac{\eta}{\sqrt{n}}\right) \leq \Lambda e^{-\zeta\sqrt{n}}, \tag{2.5.44}$$

where ζ and Λ are suitable constants. Thus, recalling (2.5.21), we have settled (2.5.35).

We thus find that the probability in the canonical ensemble of the event in (2.5.35) is $o\,(1/n)$, which confirms the results of Section 2.5.1. In particular, we have shown that the central object is the ratio $\sum_{i=1}^n d_i^2 / \sum_{i=1}^n d_i$.

Remark 2.5.5. The constants in the right-hand side of (2.5.23) can be chosen freely. By Lemma 2.5.4, this means that for any choice of constraint for which $S_n(\mathbb{P}_{\text{mic}} \parallel \mathbb{P}_{\text{can}}) = O(\log n)$ and the canonical ensemble is the Erdős-Rényi random graph, we have that λ_1 is close to $n^{-1}\sum_{i=1}^n d_i + \frac{\sigma^2}{\mu}$ in both ensembles. If the constraint does not prevent $\mathbb{E}_{\text{mic}}[n^{-1}\sum_{i=1}^n d_i + \frac{\sigma^2}{\mu}]$ to take the value $(n-1)p+(1-p)$, then we have the same result as in Theorem 2.1.2(2), which supports the working hypothesis put forward in Section 2.1. Indeed, as shown in Section 2.3, $S_n(\mathbb{P}_{\text{mic}} \parallel \mathbb{P}_{\text{can}}) = o(n)$ is the condition for EE. Instead, in the sparse regime we have to rely on events that hold with (ξ, ν)-high probability, where ξ is in principle allowed to vary with n, and the condition has to be checked for the specific value of $p(n)$.

Chapter 3

Central limit theorem for the principal eigenvalue and eigenvector of Chung-Lu random graphs

This chapter is based on:
P. Dionigi, D. Garlaschelli, R.S. Hazra, F. den Hollander, M. Mandjes. *Central limit theorem for the principal eigenvalue and eigenvector of Chung-Lu random graphs*. Journal of Physics: Complexity, 2023.

Abstract

A Chung-Lu random graph is an inhomogeneous Erdős-Rényi random graph in which vertices are assigned average degrees, and pairs of vertices are connected by an edge with a probability that is proportional to the product of their average degrees, independently for different edges. We derive a central limit theorem for the principal eigenvalue and the components of the principal eigenvector of the adjacency matrix of

a Chung-Lu random graph. Our derivation requires certain assumptions on the average degrees that guarantee connectivity, sparsity and bounded inhomogeneity of the graph.

3.1 Introduction, main results and discussion

3.1.1 Introduction

The spectral properties of adjacency matrices play an important role in various areas of network science. In the present paper we consider an inhomogeneous version of the Erdős-Rényi random graph called the *Chung-Lu random graph* and we derive a central limit theorem for the principal eigenvalue and eigenvector of its adjacency matrix.

Setting

Recall that the homogeneous Erdős-Rényi random graph has vertex set $[n] = \{1, \dots, n\}$, and each edge is present with probability p and absent with probability $1-p$, independently for different edges, where $p \in (0,1)$ may depend on n (in what follows we often suppress the dependence on n from the notation; the reader is however warned that most quantities depend on n). The average degree is the same for every vertex and equals $(n-1)p$ when self-loops are not allowed, and np when self-loops are allowed (and are considered to contribute to the degrees of the vertices). In [43] the following generalisation of the Erdős-Rényi random graph is considered, called the Chung-Lu random graph, with the goal to accommodate general degrees. Given a sequence of degrees $\vec{d}_n = (d_i)_{i\in[n]}$, consider the random graph $\mathcal{G}_n(\vec{d}_n)$ in which to each pair i, j of vertices an edge is assigned independently with probability $p_{ij} = d_i d_j / m_1$, where $m_1 = \sum_{i=1}^n d_i$ (for computational simplicity we allow self-loops). Here, the degrees can act as vertex weights. Vertices with low weights are more likely to have less neighbours than vertices with high weights which act as hubs (see [77, Chapter 6] for a general introduction to generalised random graphs). If $m_\infty^2 \leq m_1$ with $m_\infty = \max_{i\in[n]} d_i$, then $p_{ij} \leq 1$ for all $i, j \in [n]$, and the sequence $\vec{d}_n$ is *graphical*. Note that in $\mathcal{G}_n(\vec{d}_n)$ the *ex-*

pected degree of vertex i is d_i. The classical Erdős-Rényi random graph (with self-loops) corresponds to $d_i = np$ for all $i \in [n]$.

Principal eigenvalue and eigenvector

The largest eigenvalue of the adjacency matrix A and its corresponding eigenvector, written as (λ_1, v_1), contain important information about the random graph. Several community detection techniques depend on a proper understanding of these quantities [126], [90], [1], which in turn play an important role for various measures of network centrality [92], [97] and for the properties of dynamical processes (such as the spread of an epidemic) taking place on networks [37, 100]. For Erdős-Rényi random graphs, it was shown in [89] that *with high probability* (whp in the following) λ_1 scales like

$$\lambda_1 \sim \max\{\sqrt{D_\infty}, np\}, \qquad n \to \infty, \tag{3.1.1}$$

where D_∞ is the maximum degree. This result was partially extended to $\mathcal{G}_n(\vec{d}_n)$ in [44], and more recently to a class of inhomogeneous Erdős-Rényi random graphs in [19], [20]. For a related discussion on the behaviour of (λ_1, v_1) in real-world networks, see [37, 100]. In the present paper we analyse the fluctuations of (λ_1, v_1). We will be interested specifically in the case where λ_1 is detached from the bulk, which for Erdős-Rényi random graphs occurs when $\lambda_1 \sim np$ whp, and for Chung-Lu random graphs when $\lambda_1 \sim m_2/m_1$, where $m_2 = \sum_{i\in[n]} d_i^2$. Note that the quotient m_2/m_1 arises from the fact that the average adjacency matrix is rank one and that its only non-zero eigenvalue is m_2/m_1. Such rank-one perturbations of a symmetric matrix with independent entries became prominent after the work in [13]. Later studies extended this work to finite-rank perturbations [12], [21], [35], [36], [60], [61]. Erdős-Rényi random graphs differ, in the sense that perturbations live on a scale different from $\sqrt{n}$. For Chung-Lu random graphs we assume that $m_2/m_1 \to \infty$.

In the setting of inhomogeneous Erdős-Rényi random graphs, finite-rank perturbations were studied in [39]. In that paper the connection probability between between i and j is given by $p_{ij} = \varepsilon_n f(i/n, j/n)$, where $f\colon [0,1]^2 \to [0,1]$ is almost everywhere continuous and of finite

rank, $\varepsilon_n \in [0,1]$ and $n\varepsilon_n \gg (\log n)^8$. However, for a Chung-Lu random graph with a given degree sequence it is not always possible to construct an almost everywhere continuous function f independent of n such that $\varepsilon_n f(i/n, j/n) = d_i d_j/m_1$. In the present paper we extend the analysis in [39] to Chung-Lu random graphs by focussing on (λ_1, v_1). For Erdős-Rényi random graphs it was shown in [57], [56] that λ_1 satisfies a central limit theorem (CLT) and that v_1 aligns with the unit vector. These papers extend the seminal work carried out in [67].

Chung-Lu random graphs

In the present paper, subject to mild assumptions on $\vec{d}_n$, we extend the CLT for λ_1 from Erdős-Rényi random graphs to Chung-Lu random graphs, and derive a pointwise CLT for v_1 as well. It was shown in [44] that if $m_2/m_1 \gg \sqrt{m_\infty}\,(\log n)$, then $\lambda_1 \sim m_2/m_1$ whp, while if $\sqrt{m_\infty} \gg (m_2/m_1)(\log n)^2$, then $\lambda_1 = m_\infty$ whp. In fact, examples show that a result similar to (3.1.1) does not hold, and that λ_1 does not scale like $\max\{m_2/m_1, \sqrt{m_\infty}\}$. These facts clearly show that the behaviour of λ_1 is controlled by subtle assumptions on the degree sequence. In what follows we stick to a bounded inhomogeneity regime where $m_2/m_1 \asymp m_\infty$.

The behaviour of v_1 is interesting and challenging, and is of major interest for applications. One of the crucial properties to look for in eigenvectors is the phenomenon of localization versus delocalization. An eigenvector is called localized when its mass concentrates on a small number of vertices, and delocalized when its mass is approximately uniformly distributed on the vertices. The complete delocalization picture for Erdős-Rényi random graphs was given in [57]. In fact, it was proved that λ_1 is close to the scaled unit vector in the ℓ_∞-norm. In the present paper we do not study localization versus delocalization for Chung-Lu random graphs in detail, but we do show that in a certain regime there is strong evidence for delocalization because v_1 is close to the scaled unit vector. In [32, Corollary 1.3] the authors found that the eigenvectors of a generalized Wigner matrix are distributed according to a Haar measure on the orthogonal group, and the coordinates have Gaussian fluctuations after appropriate scaling. Our work shows that the coordinate-wise fluc-

tuations hold as well for the principal eigenvector of the non-centered Chung-Lu adjacency matrix and that they are Gaussian after appropriate centering and scaling.

Outline

In Section 3.1.2 we define the Chung-Lu random graph, state our assumption on the degree sequence, and formulate two main theorems: a CLT for the largest eigenvalue and a CLT for its associated eigenvector. In Section 3.1.3 we discuss these theorems and place them in their proper context. Section 3.2 contains the proof of the CLT of the eigenvalue and Section 3.3 studies the properties of the principal eigenvector.

3.1.2 Main results

Set-up

Let $\mathbb{G}_n$ be the set of simple graphs with n vertices. Let $\vec{d}_n = (d_i)_{i\in[n]}$ be a sequence of degrees, such that $d_i \in \mathbb{N}$ for all $i \in [n]$ and abbreviate

$$m_k = \sum_{i\in[n]} (d_i)^k, \qquad m_\infty = \max_{i\in[n]} d_i, \qquad m_0 = \min_{i\in[n]} d_i,$$

Note that these numbers depend on n, but in the sequel we will suppress this dependence. For each pair of vertices i, j (not necessarily distinct), we add an edge independently with probability

$$p_{ij} = \frac{d_i d_j}{m_1}. \tag{3.1.2}$$

The resulting random graph, which we denote by $\mathcal{G}_n(\vec{d}_n)$, is referred to in the literature as the Chung-Lu random graph. In [43] it was assumed that $m_\infty^2 \leq m_1$ to ensure that $p_{ij} \leq 1$. In the present paper we need sharper restrictions.

Assumption 3.1.1. Throughout the paper we need two assumptions on $\vec{d}_n$ as $n \to \infty$:

(D1) **Connectivity and sparsity:** There exists a $\xi > 2$ such that

$$(\log n)^{2\xi} \ll m_\infty \ll n^{1/2}.$$

(D2) **Bounded inhomogeneity:** $m_0 \asymp m_\infty$.

♠

The lower bound in Assumption 3.1.1(D1) guarantees that the random graph is connected whp and that it is not too sparse. The upper bound is needed in order to have $m_\infty = \mathrm{o}(\sqrt{m_1})$, which implies that (3.1.2) is well defined. Assumption 3.1.1(D2) is a restriction on the inhomogeneity of the model and requires that the smallest and the largest degree are comparable.

Remark 3.1.2. The lower bound on m_∞ in Assumption 3.1.1(D1) can be seen as an adaptation to our setting of the main condition in [44, Theorem 2.1] for the asymptotics of λ_1. As mentioned in Section 3.1.1, under the assumption

$$\frac{m_2}{m_1} \gg \sqrt{m_\infty}\,(\log n)^\xi,$$

[44] shows that $\lambda_1 = [1+\mathrm{o}(1)]\, m_2/m_1$ whp. It is easy to see that the above condition together with Assumption 3.1.1(D2) gives the lower bound in Assumption 3.1.1(D1). ♠

Remark 3.1.3. When $m_\infty \ll n^{1/6}$, [77, Theorem 6.19] implies that our results also hold for the *Generalized Random Graph* (GRG) model with the same average degrees. This model is defined by choosing connection probabilities of the form

$$p_{ij} = \frac{d_i d_j}{m_1 + d_i d_j},$$

and arises in statistical physics as the *canonical ensemble* constrained on the expected degrees, which is also called the *canonical configuration model*. Note that in the above connection probability, d_i plays the role of a hidden variable, or a Lagrange multiplier controlling the expected degree of vertex i, but does not in general coincide with the expected degree itself. However, under the assumptions considered here, d_i does coincide with the expected degree asymptotically. The reader can find more about GRG and their use in [77, Chapter 6], and about their role in statistical physics in [109]. In the corresponding *microcanonical ensemble* the degrees are not only fixed in their expectation but they take a precise deterministic value, which corresponds to the *microcanonical configuration model*.

The two ensembles were found to be *nonequivalent* in the limit as $n \to \infty$ [113]. This result was shown to imply a finite difference between the expected values of the largest eigenvalue λ_1 in the two models [53] when the degree sequence was chosen to be constant ($d_i = d$ for all $i \in [n]$). In this latter case the canonical ensemble reduces to the Erdős-Rényi random graph with $p = d/n$, while the microcanonical ensemble reduces to the d-regular random graph model. Although ensemble nonequivalence is not our main focus here, we will briefly relate some of our results to this phenomenon. ♠

Notation

Let A be the adjacency matrix of $\mathcal{G}_n(\vec{d}_n)$ and $\mathbb{E}[A]$ its expectation. The (i,j)-th entry of $\mathbb{E}[A]$ equals to p_{ij} in (3.1.2). The (i,j)-th entry of $A - \mathbb{E}[A]$ is an independent centered Bernoulli random variable with parameter p_{ij}. Let $\lambda_1 \geq \ldots \geq \lambda_n$ be the eigenvalues of A and let $v_1, \ldots, v_n$ be the corresponding eigenvectors. The vector e will be the n-dimensional column vector

$$e = \frac{1}{\sqrt{m_1}}(d_1, \cdots, d_n)^t, \tag{3.1.3}$$

where t stands for transpose. It is easy to see that $\mathbb{E}[A] = ee^t$.

Definition 3.1.4. Following [57], we say that an event $\mathcal{E}$ holds with (ξ, ν)-*high probability* (written (ξ, ν)-hp) when there exist $\xi > 2$ and $\nu > 0$ such that

$$\mathbb{P}(\mathcal{E}^c) \leq \mathrm{e}^{-\nu(\log n)^\xi}. \tag{3.1.4}$$

♠

Note that this is different from the classical notion of whp, because it comes with a specific rate.

Remark 3.1.5. Our results hold for any $\nu > 0$ as soon as $\xi > 2$ (think of $\nu = 1$). The role of ν becomes important when we consider specific subsets $\mathcal{S}$ of the event space and split into $\mathcal{S} \cap \mathcal{E}$ and $\mathcal{S} \cap \mathcal{E}^c$ (see e.g. [57]). ♠

We write $\xrightarrow{w}$ to denote weak convergence as $n \to \infty$, and use the symbols o, O to denote asymptotic order for sequences of real numbers.

CLT for the principal eigenvalue

Our first theorem identifies two terms in the expectation of the largest eigenvalue, and shows that the largest eigenvalue follows a central limit theorem.

Theorem 3.1.6. *Under Assumption 3.1.1, the following hold:*

(I)
$$\mathbb{E}[\lambda_1] = \frac{m_2}{m_1} + \frac{m_1 m_3}{m_2^2} + \mathrm{o}(1), \qquad n \to \infty.$$

(II)
$$\frac{m_2}{m_1}\left(\frac{\lambda_1 - \mathbb{E}[\lambda_1]}{\sigma_1}\right) \xrightarrow{w} \mathcal{N}(0,2), \qquad n \to \infty,$$

where

$$\sigma_1^2 = \sum_{i,j} (p_{ij})^3 (1 - p_{ij}) \sim \frac{m_3^2}{m_1^3}, \qquad n \to \infty.$$

CLT for the principal eigenvector

Our second theorem shows that the principal eigenvector is parallel to the normalised degree vector, and is close to this vector in ℓ^∞-norm. It also identifies the expected value of the components of the principal eigenvector, and shows that the components follow a central limit theorem.

Theorem 3.1.7. *Let $\tilde{e} = e\sqrt{m_1/m_2}$ be the ℓ^2-nomalized degree vector. Let v_1 be the eigenvector corresponding to λ_1 and let $v_1(i)$ denote the i-th coordinate of v_1. Under Assumption 3.1.1, the following hold:*

(I) $\langle v_1, \tilde{e}\rangle = 1 + \mathrm{o}(1)$ *as* $n \to \infty$ *with* (ξ, ν)*-hp .*

(II) $\|v_1 - \tilde{e}\|_\infty \leq \mathrm{O}\left(\frac{(\log n)^\xi}{\sqrt{n m_\infty}}\right)$ *as* $n \to \infty$ *with* (ξ, ν)*-hp .*

(III) $\mathbb{E}[v_1(i)] = \frac{d_i}{\sqrt{m_2}} + \mathrm{O}\left(\frac{(\log n)^{2\xi}}{\sqrt{m_2}}\right)$ *as* $n \to \infty$.

Moreover, if the lower bound in Assumption 3.1.1(D1) is strengthened to $(\log n)^{4\xi} \ll m_\infty$*, then for all* $i \in [n]$*,*

(IV)

$$\frac{m_2^{3/2}}{m_1}\left(\frac{v_1(i)-d_i/\sqrt{m_2}}{s_1(i)}\right) \xrightarrow{w} \mathcal{N}(0,1), \qquad n\to\infty,$$

where

$$s_1^2(i)=\sum_j d_j^2 p_{ij}(1-p_{ij}) \sim d_i\frac{m_3}{m_1}, \qquad n\to\infty.$$

3.1.3 Discussion

We place the theorems in their proper context.

1. Theorems 3.1.6–3.1.7 provide a CLT for λ_1, v_1. We note that m_2/m_1 is the leading order term in the expansion of λ_1, while m_1m_3/m_2^2 is a correction term. We observe that Theorem 3.1.6(I) does not follow from the results in [44], because the largest eigenvalue need not be uniformly integrable and also the second order expansion is not considered there. We also note that in Theorem 3.1.6(II) the centering of the largest eigenvalue, $\mathbb{E}[\lambda_1]$, cannot be replaced by its asymptotic value as the error term is not compatible with the required variance.

2. The lower bound in Assumption 3.1.1(D1) is needed to ensure that the random graph is connected, and is crucial because the largest eigenvalue is very sensitive to connectivity properties. Assumption 3.1.1(D2) is needed to control the inhomogeneity of the random graph. It plays a crucial role in deriving concentration bounds on the central moments $\langle e,(A-\mathbb{E}[A])^k e\rangle$, $k\in\mathbb{N}$, with the help of a result from [57]. Further refinements may come from different tools, such as the non-backtracking matrices used in [19], [20]. While Assumption 3.1.1(D1) appears to be close to optimal, Assumption 3.1.1(D2) is far from optimal. It would be interesting to allow for empirical degree distributions that converge to a limiting degree distribution with a power law tail.

3. As already noted, if the expected degrees are all equal to each other, i.e., $d_i = d$ for all $i\in[n]$, then the Chung-Lu random graph, or canonical configuration model, reduces to the homogeneous Erdős-Rényi random graph with $p = d/n$, while the corresponding microcanonical configuration model reduces to the homogeneous d-regular random graph

model (here, all models allow for self-loops). This implies that, for the homogeneous Erdős-Rényi random graph with connection probability $p \gg (\log n)^{2\xi}/n$, $\xi > 2$, Theorem 3.1.6(I) reduces to

$$\mathbb{E}[\lambda_1] = np + 1 + \mathrm{o}(1), \qquad n \to \infty,$$

while Theorem 3.1.6(II) reduces to

$$\frac{1}{\sqrt{p}} \left(\lambda_1 - \mathbb{E}[\lambda_1]\right) \xrightarrow{w} \mathcal{N}(0, 2), \qquad n \to \infty.$$

Both these properties were derived in [56] for homogeneous Erdős-Rényi random graphs and also for rank-1 perturbations of Wigner matrices. In [53], the fact that $\mathbb{E}[\lambda_1]$ in the canonical ensemble differs by a finite amount from the corresponding expected value (here, $d = np$) in the microcanonical ensemble (d-regular random graph) was shown to be a signature of ensemble nonequivalence.

4. In case $d_i = d$ for all $i \in [n]$, Theorem 3.1.7(III) reduces to the following CLT, which was not covered by [56] and [53].

Corollary 3.1.8. *For the Erdős-Rényi random graph with* $(\log n)^{4\xi}/n \ll p \ll n^{-1/2}$ *for some* $\xi > 2$,

$$n\sqrt{\frac{p}{1-p}} \left(v_1(i) - \frac{1}{\sqrt{n}}\right) \xrightarrow{w} \mathcal{N}(0, 1), \qquad n \to \infty.$$

Note that, in the corresponding microcanonical ensemble (d-regular random graph), v_1 coincides with the constant vector where $v_1(i) = 1/\sqrt{n}$ for all $i \in [n]$. Therefore in the canonical ensemble each coordinate $v_1(i)$ has Gaussian fluctuations around the corresponding deterministic value for the microcanonical ensemble. This behaviour is similar to the degrees having, in the canonical configuration model, either Gaussian (in the dense setting) or Poisson (in the sparse setting) fluctuations around the corresponding deterministic degrees for the microcanonical configuration model [71].

5. One way to satisfy Assumption 3.1.1 is to specify functions $\omega, c_1, \ldots, c_n$, satisfying $(\log n)^{2\xi} \ll \omega(n) \ll \sqrt{n}$ and $c \leq c_1(n) \leq \ldots \leq c_n(n) \leq C$ with

$c, C \geq 0$, such that

$$d_i(n) = c_i(n)\omega(n), \qquad p_{ij} = \frac{c_i c_j}{\frac{1}{n}\sum_k c_k}\frac{\omega}{n}.$$

The reason why we avoid such a description is that our setting is potentially broader. The concentration estimate in Lemma 3.2.4 requires us to assume homogeneous degree sequences as above, while Theorem 3.1.6(I) holds for much more general degree sequences. A further refinement of Lemma 3.2.4 may be possible. The advantage of the above description is that it makes the scale $\omega(n)$ on which the degrees live explicit. However, most of the bounds in our proofs depend on some power of m_∞, up to some multiplicative constant. This means that, in the bounded inhomogeneity setting, expressing the asymptotics through $\omega(n)$ or m_∞ are equivalent. Bounds expressed through $\omega(n)$ would cease to be meaningful as soon as we manage to push beyond the bounded inhomogeneity setting of our model, while the skeleton of our proof would still hold.

6. In [41] the empirical spectral distribution of A was considered under the assumption that

$$(m_\infty)^2/m_1 \ll 1 \ll n(m_\infty)^2/m_1,$$

which is weaker than Assumption 3.1.1. It was shown that if $\mu_n \xrightarrow{w} \mu$ with $\mu_n = n^{-1}\sum_{i=1}^n \delta_{d_i/m_\infty}$ and μ some probability distribution on $\mathbb{R}$, then

$$\mathrm{ESD}\left(\frac{A}{\sqrt{n(m_\infty)^2/m_1}}\right) \xrightarrow{w} \mu \boxtimes \mu_{\mathrm{sc}}$$

with μ_{sc} the Wigner semicircle law and $\boxtimes$ the free multiplicative convolution. Since $\mu \boxtimes \mu_{\mathrm{sc}}$ is compactly supported, this shows that the scaling for the largest eigenvalue and the spectral distribution are different.

3.2 Proof of Theorem 3.1.6

In what follows we use the well-known method of writing the largest eigenvalue of a matrix as a rank-1 perturbation of the centered matrix. This method was previously successfully employed in [57, 67, 101].

Given the adjacency matrix A of our graph G, we can write $A = H + \mathbb{E}[A]$ with $H = A - \mathbb{E}[A]$. Let v_1 be the eigenvector associated with the eigenvalue λ_1. Then

$$Av_1 = \lambda_1 v_1, \quad (H + \mathbb{E}[A])v_1 = \lambda_1 v_1, \quad (\lambda_1 I - H)v_1 = \mathbb{E}[A]v_1.$$

Using that $\mathbb{E}[A] = ee^t$, we have $(\lambda_1 I - H)v_1 = \langle e, v_1 \rangle e$, where I is the $n \times n$ identity matrix. It follows that if λ_1 is not an eigenvalue of H, then the matrix $(\lambda_1 I - H)$ is invertible, and so

$$v_1 = \langle e, v_1 \rangle (\lambda_1 I - H)^{-1} e. \tag{3.2.1}$$

Eliminating the eigenvector v_1 from the above equation, we get

$$1 = \left\langle e, (\lambda_1 I - H)^{-1} e \right\rangle,$$

where we use that $\langle e, v_1 \rangle \neq 0$ (since λ_1 is not an eigenvalue of H). Note that this can be expressed as

$$\lambda_1 = \left\langle e, \left(I - \frac{H}{\lambda_1} \right)^{-1} e \right\rangle = \sum_{k=0}^{\infty} \left\langle e, \left(\frac{H}{\lambda_1} \right)^{k} e \right\rangle \quad \textit{with } (\xi, \nu)\textit{-hp}\,, \tag{3.2.2}$$

where the validity of the series expansion will be an immediate consequence of Lemma 3.2.2 below.

Section 3.2.1 derives bounds on the spectral norm of H. Section 3.2.2 analyses the expansion in (3.2.2) and prove the scaling of $\mathbb{E}[\lambda_1]$. Section 3.2.3 is devoted to the proof of the CLT for λ_1, Section 3.3 to the proof of the CLT for v_1. In the expansion we distinguish three ranges: (i) $k = 0, 1, 2$; (ii) $3 \leq k \leq L$; (iii) $L < k < \infty$, where

$$L = \lfloor \log n \rfloor.$$

We will show that (i) controls the mean and the variance in both CLTs, while (ii)-(iii) are negligible error terms.

3.2.1 The spectral norm

In order to study λ_1, we need good bounds on the spectral norm of H. The spectral norm of matrices with inhomogeneous entries has been studied in a series of papers [19], [20], [8] for different density regimes.

An important role is played by $\lambda_1(\mathbb{E}[A])$. In recent literature this quantity has been shown to play a prominent role in the so-called BBP-transition [13]. Given our setting (3.1.2), it is easy to see that

$$\lambda_1(\mathbb{E}[A]) = \frac{m_2}{m_1}, \tag{3.2.3}$$

while all other eigenvalues of $\mathbb{E}[A]$ are zero.

Remark 3.2.1. *Since* $m_0 \leq \frac{m_2}{m_1} \leq m_\infty$, Assumption 3.1.1(D2) implies that

$$\frac{m_2}{m_1} \asymp m_\infty. \tag{3.2.4}$$

♠

We start with the following lemma, which ensures concentration of λ_1 and is a direct consequence of the results in [20] (which matches Assumption 3.1.1). In particular, we use [20, Theorem 3.2] to check that the boundaries of the bulk of the spectral distribution live on a scale smaller than the scale of λ_1.

Lemma 3.2.2. *Under Assumption 3.1.1, with* (ξ, ν)*-hp*

$$\left| \frac{\lambda_1(A) - \lambda_1(\mathbb{E}[A])}{\lambda_1(\mathbb{E}[A])} \right| = \mathrm{O}\left(\frac{1}{\sqrt{m_\infty}} \right), \qquad n \to \infty,$$

and consequently

$$\frac{\lambda_1(A)}{\lambda_1(\mathbb{E}[A])} \xrightarrow{\mathbb{P}} 1, \qquad n \to \infty.$$

Proof. In the proof it is understood that all statements hold with (ξ, ν)-hp in the sense of (3.1.4). Let $A = H + \mathbb{E}[A]$. Due to Weyl's inequality, we have that

$$\lambda_1(\mathbb{E}[A]) - \|H\| \leq \lambda_1(A) \leq \lambda_1(\mathbb{E}[A]) + \|H\|.$$

From [20, Theorem 3.2] we know that there is a universal constant $C > 0$ such that

$$\mathbb{E}\left[\|A - \mathbb{E}[A]\|\right] = \mathbb{E}\left[\|H\|\right] \leq \sqrt{m_\infty} \left(2 + \frac{C}{q} \sqrt{\frac{\log n}{1 \vee \log\left(\frac{\sqrt{\log n}}{q} \right)}} \right),$$

where

$$q = \sqrt{m_\infty} \wedge n^{1/10}\kappa^{-1/9}$$

with κ defined by

$$\kappa = \max_{ij} \frac{p_{ij}}{m_\infty/n} = \frac{n m_\infty}{m_1}.$$

Thanks to Assumption 3.1.1(D2), we have $\kappa = \mathrm{O}(1)$. By Remark 3.1 of [20, Remark 3.1] (which gives us that $q = \sqrt{m_\infty}$ for n large enough) and Assumption 3.1.1, we get that

$$\mathbb{E}\left[\|H\|\right] \leq \begin{cases} \sqrt{m_\infty}\left(2 + \frac{C\sqrt{\log n}}{\sqrt{m_\infty}}\right), & (\log n)^{2\xi} \leq \sqrt{m_\infty} \leq n^{1/10}\kappa^{-1/9}, \\ \sqrt{m_\infty}\left(2 + \frac{C'\sqrt{\log n}}{n^{1/10}}\right), & \sqrt{m_\infty} \geq n^{1/10}\kappa^{-1/9}. \end{cases} \tag{3.2.5}$$

Using [30, Example 8.7] or [20, Equation 2.4] (the Talagrand inequality), we know that there exists a universal constant $c > 0$ such that

$$\mathbb{P}\left(|\|H\| - \mathbb{E}[\|H\|]| > t\right) \leq 2\mathrm{e}^{-ct^2}.$$

For $t = \sqrt{\nu(\log n)^\xi}$,

$$\mathbb{E}[\|H\|] - \sqrt{\nu}(\log n)^{\xi/2} \leq \|H\| \leq \mathbb{E}[\|H\|] + \sqrt{\nu}(\log n)^{\xi/2}. \tag{3.2.6}$$

Thus, we have

$$|\lambda_1(A) - \lambda_1(\mathbb{E}[A])| \leq \|H\| \leq \sqrt{m_\infty}(2 + \mathrm{o}(1)) + \sqrt{\nu}(\log n)^{\xi/2}. \tag{3.2.7}$$

Using that $\lambda_1(\mathbb{E}[A]) = m_2/m_1$, we have that with (ξ, ν)-hp the following bound holds:

$$\left|\frac{\lambda_1(A) - \lambda_1(\mathbb{E}[A])}{\lambda_1(\mathrm{E}[A])}\right| \leq \frac{\sqrt{m_\infty}}{m_2/m_1}(2 + \mathrm{o}(1)) + \frac{\sqrt{\nu}(\log n)^{\xi/2}}{m_2/m_1} = \mathrm{O}\left(\frac{\sqrt{m_\infty}}{m_2/m_1}\right).$$

Via Assumption 3.1.1 and (3.2.4) the claim follows.

Remark 3.2.3.

(a) The proof of Lemma 3.2.2 works well if we replace Assumption 3.1.1(D2) by a milder condition. Indeed, the former is directly linked to the parameter κ that appears in the proof of Lemma 3.2.2 and in the proof of [20, Theorem 3.2], which contains a more general condition on the inhomogeneity of the degrees.

(b) Note that a consequence of proof of Lemma 3.2.2 is that with (ξ, ν)-hp

$$\frac{\|H\|}{\lambda_1(A)} \leq 1 - C_0 \tag{3.2.8}$$

for some $C_0 \in (0,1)$. This allows us to claim that with (ξ, ν)-hp the inverse

$$\left(I - \frac{H}{\lambda_1(A)}\right)^{-1} \tag{3.2.9}$$

exists.

♠

Lemma 3.2.4. *Let $1 \leq k \leq L$. Then, under Assumption 3.1.1, with (ξ, ν)-hp*

$$\left|\left\langle e, H^k e\right\rangle - \mathbb{E}\left[\left\langle e, H^k e\right\rangle\right]\right| \leq C \frac{m_2}{m_1} \frac{m_\infty^{\frac{k}{2}} (\log n)^{k\xi}}{\sqrt{n}},$$

i.e.,

$$\max_{1 \leq k \leq L} \mathbb{P}\left(\left|\left\langle e, H^k e\right\rangle - \mathbb{E}\left[\left\langle e, H^k e\right\rangle\right]\right| > \frac{C(\log n)^{k\xi} m_\infty^{\frac{k}{2}}}{\sqrt{n}} \frac{m_2}{m_1}\right) \leq \mathrm{e}^{-\nu(\log n)^\xi}, \qquad n \geq n_1($$

Lemma 3.2.4 is a generalization to the inhomogeneous setting of [57, Lemma 6.5]. We skip the proof because it requires a straightforward modification of the arguments in [57].

Lemma 3.2.5. *Under Assumption 3.1.1, for $2 \leq k \leq L$, there exists a constant $C > 0$ such that*

$$\mathbb{E}\left[\left\langle e, H^k e\right\rangle\right] \leq \frac{m_2}{m_1} (C m_\infty)^{k/2}. \tag{3.2.10}$$

Proof. Let $\mathcal{E}$ be the high probability event defined by (3.2.6), i.e.,

$$\|H\| \leq \mathrm{E}[\|H\|] + \sqrt{\nu}(\log n)^{\xi/2} \leq m_\infty \left(1 + \mathrm{O}\left(\frac{(\log n)^{\xi/2}}{m_\infty}\right)\right).$$

Due to Assumption 3.1.1(D1) we can bound the right-hand side by $C m_\infty$. Since $\|e\|_2^2 = m_2/m_1$, on this event we have

$$\mathbb{E}\left[\left(\left\langle e, H^k e\right\rangle\right) \mathbf{1}_{\mathcal{E}}\right] \leq \|e\|_2^2 \, \mathbb{E}[\|H\|^k \mathbf{1}_{\mathcal{E}}] \leq \frac{m_2}{m_1} (C m_\infty)^{k/2}.$$

We show that the expectation when evaluated on the complementary event is negligible. Indeed, observe that

$$\mathbb{E}\left[\left\langle e, H^k e\right\rangle\right] = \mathbb{E}\left(\sum_{i_1,\dots,i_{k+1}=1}^{n} e_{i_1} e_{i_{k+1}} \prod_{j=1}^{k} H(i_j, i_{j+1})\right)^2$$
$$\leq \left(\frac{n^{k+1} m_\infty^2}{m_1}\right)^2 \leq C\mathrm{e}^{(2k+2)\log n} \leq \mathrm{e}^{2(\log n)^2},$$

where in the last inequality we use that $m_\infty = \mathrm{o}(\sqrt{m_1})$. This, combined with the exponential decay of the event $\mathcal{E}^c$, gives

$$\mathbb{E}\left[\left\langle e, H^k e\right\rangle \mathbf{1}_{A^c}\right] \leq C\mathrm{e}^{-\nu(\log n)^\xi},$$

and so the claim follows.

3.2.2 Expansion for the principal eigenvalue

We denote the event in Lemma 3.2.2 by $\mathcal{E}$, which has high probability. As noted in Remark 3.2.3(b), $I - \frac{H}{\lambda_1}$ is invertible on $\mathcal{E}$. Hence, expanding on $\mathcal{E}$, we get

$$\lambda_1 = \sum_{k=0}^{\infty} \left\langle e, \frac{H^k}{\lambda_1^k} e \right\rangle.$$

We split the sum into two parts:

$$\lambda_1 = \sum_{k=0}^{L} \frac{\left\langle e, H^k e\right\rangle}{\lambda_1^k} + \sum_{k=L+1}^{\infty} \frac{\left\langle e, H^k e\right\rangle}{\lambda_1^k}. \tag{3.2.11}$$

First we show that we may ignore the second sum. To that end we observe that, by Assumption 3.1.1 (D1), on the event $\mathcal{E}$ we can estimate

$$\left|\sum_{k=L+1}^{\infty} \frac{\left\langle e, H^k e\right\rangle}{\lambda_1^k}\right| \leq \sum_{k=L+1}^{\infty} \frac{\|e\|_2^2 \|H\|^k}{\lambda_1^k} \leq \sum_{k=L+1}^{\infty} \frac{m_2}{m_1} \frac{m_\infty^{k/2}}{(Cm_2/m_1)^k}$$
$$\leq \sum_{k=L+1}^{\infty} \frac{C'}{m_\infty^{k/2-1}} = \mathrm{O}\left(\mathrm{e}^{-c\log\sqrt{n}}\right). \tag{3.2.12}$$

Because of (3.2.12) and the fact that $\mathbb{E}(\langle e, He\rangle) = 0$, (3.2.11) reduces to

$$\lambda_1 = \sum_{k=3}^{L} \frac{\mathbb{E}\left[\langle e, H^k e\rangle\right]}{\lambda_1^k} + \sum_{k=3}^{L} \frac{\langle e, H^k e\rangle - \mathbb{E}\left[\langle e, H^k e\rangle\right]}{\lambda_1^k} + \langle e, e\rangle + \frac{1}{\lambda_1}\langle e, He\rangle + \frac{1}{\lambda_1^2}\langle e, H^2 e\rangle + \mathrm{o}(1).$$

Next, we estimate the second sum in the above equation. Using Lemma 3.2.2, we get

$$\left|\sum_{k=3}^{L} \frac{\langle e, H^k e\rangle - \mathbb{E}\left[\langle e, H^k e\rangle\right]}{\lambda_1^k}\right| \leq \sum_{k=3}^{L} \frac{C m_\infty^{\frac{k}{2}} (\log n)^{k\xi}}{\sqrt{n}(m_2/m_1)^{k-1}} \leq \sum_{k=3}^{L} \frac{C(\log n)^{k\xi}}{\sqrt{n} m_\infty^{k/2-1}} \leq \mathrm{O}\left(\frac{C(\log n)^{\xi+1}}{\sqrt{n m_\infty}}\right) = \mathrm{o}(1).$$

From Lemma 3.2.5 we have

$$\sum_{k=3}^{L} \frac{\mathbb{E}\langle e, H^k e\rangle}{\lambda_1^k} \leq \sum_{k=3}^{L} \frac{\frac{m_2}{m_1}(C m_\infty)^{k/2}}{(m_2/m_1)^k} = \mathrm{O}\left(\frac{1}{\sqrt{m_\infty}}\right) = \mathrm{o}(1),$$

where the last estimate follows from Assumption 3.1.1(D1). Hence, on $\mathcal{E}$,

$$\lambda_1 = \langle e, e\rangle + \frac{1}{\lambda_1}\langle e, He\rangle + \frac{\langle e, H^2 e\rangle}{\lambda_1^2} + \mathrm{o}(1).$$

Iterating the expression for λ_1 in the right-hand side, we get

$$\lambda_1 = \langle e, e\rangle + \langle e, He\rangle \left(\langle e, e\rangle + \frac{1}{\lambda_1}\langle e, He\rangle + \frac{1}{\lambda_1^2}\langle e, H^2 e\rangle + \mathrm{o}(1)\right)^{-1} + \langle e, H^2 e\rangle \left(\langle e, e\rangle + \frac{1}{\lambda_1}\langle e, He\rangle + \frac{1}{\lambda_1^2}\langle e, H^2 e\rangle + \mathrm{o}(1)\right)^{-2} + \mathrm{o}(1),$$

Expanding the second and third term we get,

$$\begin{aligned}\lambda_1 &= \langle e, e\rangle + \frac{\langle e, He\rangle}{\langle e, e\rangle}\left(1 - \frac{\langle e, He\rangle}{\lambda_1\langle e, e\rangle} - \frac{\langle e, H^2 e\rangle}{\lambda_1^2\langle e, e\rangle} + \mathrm{o}(1)\right) \\ &\quad + \frac{\langle e, H^2 e\rangle}{(\langle e, e\rangle)^2}\left(1 - \frac{2\langle e, He\rangle}{\lambda_1\langle e, e\rangle} - \frac{2\langle e, H^2 e\rangle}{\lambda_1^2\langle e, e\rangle} + \mathrm{o}(1)\right) + \mathrm{o}(1), \\ &= \langle e, e\rangle + \frac{\langle e, He\rangle}{\langle e, e\rangle} - \frac{\langle e, He\rangle^2}{\lambda_1\langle e, e\rangle^2} + \frac{\langle e, H^2 e\rangle}{\langle e, e\rangle^2} + \mathrm{o}(1).\end{aligned}$$

Here we use that $\langle e, e\rangle = m_2/m_1 \to \infty$, and we ignore several other terms because they are small with (ξ, ν)-hp , for example,

$$\frac{\langle e, He\rangle \langle e, H^2 e\rangle}{\lambda_1^2 \langle e, e\rangle^2} = \mathrm{O}\left(\frac{m_\infty^{3/2}}{(m_2/m_1)^4}\right) = \mathrm{o}(1).$$

One more iteration gives

$$\begin{aligned}\lambda_1 &= \langle e, e\rangle + \frac{\langle e, He\rangle}{\langle e, e\rangle} + \frac{\langle e, H^2 e\rangle}{\langle e, e\rangle^2} \\ &\quad - \frac{\langle e, He\rangle^2}{\langle e, e\rangle^2}\left(\langle e, e\rangle + \frac{1}{\lambda_1}\langle e, He\rangle + \frac{1}{\lambda_1^2}\langle e, H^2 e\rangle + \mathrm{o}(1)\right)^{-1} + \mathrm{o}(1) \\ &= \langle e, e\rangle + \frac{\langle e, He\rangle}{\langle e, e\rangle} + \frac{\langle e, H^2 e\rangle}{\langle e, e\rangle^2} - \frac{\langle e, He\rangle^2}{\langle e, e\rangle^3} + \frac{\langle e, H^2 e\rangle^2 \langle e, He\rangle}{\lambda_1 \langle e, e\rangle^3} + \frac{\langle e, H^2 e\rangle^3}{\lambda_1^2 \langle e, e\rangle^3} + \mathrm{o}(1).\end{aligned}$$

Proof of Theorem 3.1.6 (I). Since the probability of $\mathcal{E}^c$ decays exponentially with n, taking the expectation of the above term and using that $\mathbb{E}[\langle e, He\rangle] = 0$, we obtain

$$\mathbb{E}[\lambda_1] = \langle e, e\rangle + \frac{\mathbb{E}[\langle e, H^2 e\rangle]}{\langle e, e\rangle^2} - \frac{\mathbb{E}[\langle e, He\rangle^2]}{\langle e, e\rangle^3} + \mathrm{o}(1) = \frac{m_2}{m_1} + \frac{m_1 m_3}{m_2^2} - \frac{m_3^2}{m_2^3} + \mathrm{o}(1).$$

Note that

$$\frac{m_3^2}{m_2^2} \le \frac{m_\infty^2}{n} = \mathrm{o}(1), \qquad \frac{m_1 m_3}{m_2^2} \le \left(\frac{m_\infty}{m_0}\right)^4 = \mathrm{O}(1),$$

and so we can write

$$\mathbb{E}[\lambda_1] = \frac{m_2}{m_1} + \frac{m_1 m_3}{m_2^2} + \mathrm{o}(1). \tag{3.2.13}$$

3.2.3 CLT for the principal eigenvalue

Again consider the high probability event on which (3.2.9) holds. Recall that from the series decomposition in (3.2.11) we have

$$\lambda_1 = \frac{\langle e, He\rangle}{\lambda_1} + \sum_{k=0}^{L} \frac{\mathbb{E}\langle e, H^k e\rangle}{\lambda_1^k} + \sum_{k=2}^{L} \frac{\langle e, H^k e\rangle - \mathbb{E}\langle e, H^k e\rangle}{\lambda_1^k} + \sum_{k>L} \frac{\langle e, H^k e\rangle}{\lambda_1^k}. \tag{3.2.14}$$

Lemma 3.2.6. *The equation*

$$x = \sum_{k=0}^{L} \frac{\mathbb{E}\left\langle e, H^k e\right\rangle}{x^k} \tag{3.2.15}$$

has a solution x_0 satisfying

$$\lim_{n\to\infty} \frac{x_0}{m_2/m_1} = 1.$$

Proof. Define the function $h\colon (0,\infty) \to \mathbb{R}$ by

$$h(x) = \sum_{k=0}^{\log n} \frac{\mathbb{E}\left\langle e, H^k e\right\rangle}{x^k}.$$

Since $\mathbb{E}[e'He] = 0$, we have

$$h\left(\frac{xm_2}{m_1}\right) = \frac{m_2}{m_1} + \sum_{k=2}^{\log n} \frac{\mathbb{E}\left\langle e, H^k e\right\rangle}{(xm_2/m_1)^k}.$$

For $x > 0$,

$$\begin{aligned}\left|\sum_{k=2}^{\log n} \frac{\mathbb{E}[\langle e, H^k e\rangle]}{(xm_2/m_1)^k}\right| &\le \sum_{k=2}^{\infty} \frac{1}{(xm_2/m_1)^k} \frac{m_2}{m_1} (Cm_\infty)^{k/2} \\ &= \mathrm{o}\left(\frac{m_2}{m_1} \sum_{k=2}^{\infty} \frac{1}{x^k (\log n)^{k\xi}}\right) = \mathrm{o}\left(\frac{m_2}{m_1} x^{-2}\right).\end{aligned}$$

This shows that

$$\lim_{n\to\infty} \frac{1}{m_2/m_1} \sum_{k=0}^{\log n} \frac{\mathbb{E}\left\langle e, H^k e\right\rangle}{(xm_2/m_1)^k} = 1.$$

Hence, for any $0 < \delta < 1$,

$$\lim_{n\to\infty} \frac{1}{m_2/m_1} \left[\frac{m_2}{m_1}(1+\delta) - h\left((1+\delta)\frac{m_2}{m_1}\right)\right] = \delta.$$

So, for large enough n,

$$h\left((1+\delta)\frac{m_2}{m_1}\right) < \frac{m_2}{m_1}(1+\delta).$$

Similarly, for any $0 < \delta < 1$,

$$h\left((1-\delta)\frac{m_2}{m_1}\right) > \frac{m_2}{m_1}(1-\delta).$$

This shows that there is a solution for (3.2.15), which lies in the interval $[\frac{m_2}{m_1}(1-\delta), \frac{m_2}{m_1}(1-\delta)]$.

Lemma 3.2.7. *Let x_0 be a solution for* (3.2.15)*. Define*

$$R_n = \lambda_1 - x_0 - \frac{\langle e, He\rangle}{m_2/m_1}.$$

Then

$$R_n = \mathrm{o}_{\mathbb{P}}\left(\frac{m_3}{m_2\sqrt{m_1}}\right), \qquad \mathbb{E}\left[|R_n|\right] = \mathrm{o}\left(\frac{m_3}{m_2\sqrt{m_1}}\right).$$

Proof of Theorem 3.1.6 (II). From the previous lemmas we have

$$\lambda_1 = x_0 + \frac{\langle e, He\rangle}{m_2/m_1} + R_n.$$

Therefore

$$\mathbb{E}[\lambda_1] = x_0 + \mathbb{E}[R_n]$$

and

$$\lambda_1 - \mathbb{E}[\lambda_1] = \frac{\langle e, He\rangle}{m_2/m_1} + \mathrm{o}\left(\frac{m_3}{m_2\sqrt{m_1}}\right).$$

Hence

$$\frac{m_2}{m_1}\left(\lambda_1 - \mathbb{E}[\lambda_1]\right) = \langle e, H, e\rangle + \mathrm{o}\left(\frac{m_3}{m_1^{3/2}}\right). \tag{3.2.16}$$

Observe that

$$\langle e, He\rangle = \sum_{i,j=1}^{N} h_{i,j}\frac{d_i d_j}{m_1} = 2\sum_{i\le j} h_{i,j}\frac{d_i d_j}{m_1}$$

Let

$$\sigma_1^2 = \sum_{i\le j}\mathrm{Var}\left(\frac{2}{m_1}h_{i,j}d_i d_j\right) = \sum_{i\le j}\frac{4d_i^3 d_j^3}{m_1^3}\left(1-\frac{d_i d_j}{m_1}\right) \sim 2\frac{m_3^2}{m_1^3}\left(1+\mathrm{O}\left(\frac{m_\infty^2}{n}\right)\right),$$

where we use the symmetry of the expression in the last equality. We can apply Lyapunov's central limit theorem, because $\{h_{i,j}\colon\ i \le j\}$ is an independent collection of random variables and Lyapunov's condition is satisfied, i.e.,

$$\lim_{n\to\infty} \frac{1}{\sigma_n^3} \sum_{i>j} \mathbb{E}\left[|H(i,j)d_i d_j|^3\right] \le K \lim_{n\to\infty} \frac{m_1^{3/2}}{m_3^3} \frac{m_4^2}{m_1} = 0,$$

where K is a constant that does not depend on n. Hence

$$\frac{m_1^{3/2} \langle e, He\rangle}{\sqrt{2} m_3} \xrightarrow{w} N(0,1).$$

Returning to the eigenvalue equation in (3.2.16) and dividing by σ_1, we have

$$\frac{\sqrt{m_1} m_2}{m_3} (\lambda_1 - \mathrm{E}[\lambda_1]) = \frac{m_1^{3/2} \langle e, He\rangle}{m_3} + \mathrm{o}(1) \xrightarrow{w} N(0,2).$$

We next prove Lemma 3.2.7, on which the proof of the central limit theorem relied.

Proof. Note that by (3.2.14) and (3.2.15) we can write

$$\lambda_1 - x_0 = \frac{\langle e, He\rangle}{\lambda_1} + \sum_{k=2}^{L} \mathbb{E}\langle e, H^k e\rangle \left(\frac{1}{\lambda_1^k} - \frac{1}{x_0^k}\right) + L_n, \tag{3.2.17}$$

where

$$L_n = \sum_{k=2}^{L} \frac{\langle e, H^k e\rangle - \mathbb{E}\langle e, H^k e\rangle}{\lambda_1^k} + \sum_{k>L} \frac{\langle e, H^k e\rangle}{\lambda_1^k}.$$

Thanks to Lemma 3.2.2, Lemma 3.2.4 and (3.2.12) we have

$$L_n = \mathrm{O}\left(\frac{m_\infty (\log n)^{2\xi}}{\sqrt{n} m_2/m_1}\right).$$

Note that $L_n = \mathrm{o}(\frac{m_3}{m_2\sqrt{m_1}})$. Indeed, using $m_3 \ge n m_0^3$ and Assumption 3.1.1(D1), we get

$$\frac{m_\infty (\log n)^{2\xi} m_2 \sqrt{m_1}}{\sqrt{n}(m_2/m_1) m_3} \le \frac{m_\infty^{5/2} n^{3/2} (\log n)^{2\xi}}{\sqrt{n} m_0 n m_0^3 (\log n)^{\xi}} = \frac{m_\infty^{5/2} (\log n)^{\xi}}{m_0^4} = \mathrm{O}\left(\frac{(\log n)^{\xi}}{m_0^{3/2}}\right).$$

Observe that (3.2.17) can be rearranged as

$$(\lambda_1 - x_0) = \frac{\langle e, He\rangle}{\lambda_1} - \sum_{k=2}^{L}(\lambda_1 - x_0)\mathbb{E}\left\langle e, H^k e\right\rangle \lambda_1^{-k} x_0^{-k} \sum_{j=0}^{k-1} x_0^{k-1-j} + L_n.$$

Hence, bringing the second term from the right to the left, we have

$$(\lambda_1 - x_0)\left[1 + \sum_{k=2}^{L}\mathbb{E}\left\langle e, H^k e\right\rangle \lambda_1^{-k} x_0^{-k} \sum_{j=0}^{k-1} x_0^{k-1-j}\right] = \frac{\langle e, He\rangle}{\lambda_1} + L_n.$$

Using the bounds on λ_1 and x_0, we get

$$\left|\sum_{k=2}^{L}\mathbb{E}\left\langle e, H^k e\right\rangle \lambda_1^{-k} x_0^{-k} \sum_{j=0}^{k-1} x_0^{k-1-j}\right| \leq \sum_{k=2}^{L}\frac{k}{(m_2/m_1)^{k+1}}\mathbb{E}\left\langle e, H^k e\right\rangle$$

$$\leq \sum_{k=2}^{L}\frac{k}{(m_2/m_1)^{k+1}}\frac{m_2}{m_1}(Cm_\infty)^{k/2} = \mathrm{O}\left(\frac{m_\infty}{(m_2/m_1)^2(\log n)^{2\xi-1}}\right) = \mathrm{o}(1).$$

We can therefore write

$$\lambda_1 - x_0 = \frac{\langle e, He\rangle}{\lambda_1} + L_n,$$

where $L_n = \mathrm{o}_P(\frac{m_3}{m_2\sqrt{m_1}})$. Finally, to go to R_n, note that

$$R_n = \lambda_1 - x_0 - \frac{\langle e, He\rangle}{m_2/m_1} = \langle e, He\rangle\left(\frac{1}{\lambda_1} - \frac{1}{m_2/m_1}\right) + L_n. \qquad (3.2.18)$$

To bound R_n, it is enough to show that the first term on the right-hand side is with (ξ,ν)-hp bounded by $\frac{m_3}{m_2\sqrt{m_1}}$. Using Lemma 3.2.4 (for $k=1$) and (3.2.7), we have with (ξ,ν)-hp

$$\frac{|\langle e, He\rangle|\,|\lambda_1 - m_2/m_1|}{\lambda_1 m_2/m_1} \leq \frac{\sqrt{m_\infty}(\log n)^\xi}{\sqrt{n}}\frac{\sqrt{m_\infty}}{(m_2/m_1)}. \qquad (3.2.19)$$

Using again Assumption 3.1.1(D1), $m_3 \geq nm_0^3$, $m_1 \leq nm_\infty$ and $m_2 \leq nm_\infty^2$, we get that

$$\frac{m_\infty(\log n)^\xi}{\sqrt{n}(m_2/m_1)}\frac{m_2\sqrt{m_1}}{m_3} \leq \left(\frac{m_\infty}{m_0}\right)^3\frac{c}{\sqrt{m_\infty}} = \mathrm{o}(1).$$

This controls the right-hand side of (3.2.19), and hence $R_n = \mathrm{o}(\frac{m_3}{m_2\sqrt{m_1}})$ with (ξ,ν)-hp .

We want to show that the latter is negligible both pointwise and in expectation. We already have that this is so with (ξ,ν)-hp on R_n. We want to show that the same bound holds in expectation. Let $\mathcal{A}$ be the high probability event of Lemma 3.2.2 and 3.2.4, and write

$$\mathbb{E}[|R_n|] = \mathbb{E}[|R_n|\mathbf{1}_{\mathcal{A}^c}] + \mathbb{E}[|R_n|\mathbf{1}_{\mathcal{A}}],$$

where $\mathbf{1}_{\mathcal{A}}$ is the indicator function of the event $\mathcal{A}$. Since all the bounds hold on the high probability event $\mathcal{A}$, it is immediate that

$$\mathbb{E}[|R_n|\mathbf{1}_{\mathcal{A}}] = \mathrm{o}\left(\frac{m_3}{\sqrt{m_1}m_2}\right).$$

The remainder can be bounded via the Cauchy-Schwarz inequality, namely,

$$\mathbb{E}[|R_n|\mathbf{1}_{\mathcal{A}^c}] \leq \left(\mathbb{E}[|R_n|^2]\mathbb{E}[\mathbf{1}_{\mathcal{A}^c}]\right)^{\frac{1}{2}} \leq \left(\mathbb{E}\left[|R_n|^2\right] e^{-\nu(\log n)^{\xi}}\right)^{\frac{1}{2}}.$$

We see that if $\mathbb{E}[|R_n|^2] = \mathrm{o}(\mathrm{e}^{-\nu(\log n)^{\xi}})$, then we are done. Expanding, we see that

$$\mathbb{E}[|R_n|^2] = \mathbb{E}\left[\left|\lambda_1 - x_0 - \frac{\langle e, He\rangle}{m_2/m_1}\right|^2\right] \leq n^C$$

for some $C > 0$, where we use that

$$\mathbb{E}[(\lambda_1^2)] \leq \mathbb{E}[\mathrm{Tr}\, A^2] = \sum_{i,j=1}^{N} \mathbb{E}[(A(i,j))^2] \leq m_\infty n$$

and the trivial bound $|\langle e, He\rangle| \leq n^{C_*}$ for some $C_* < C$. Hence we have $\left(\mathbb{E}[|R_n|^2]\mathbb{E}[\mathbf{1}_{A^c}]\right)^{\frac{1}{2}} \leq \mathrm{e}^{-\nu(\log n)^{\xi}}$ and

$$\mathbb{E}[|R_n|] = \mathrm{o}\left(\frac{m_3}{\sqrt{m_1}m_2}\right).$$

3.3 Proof of Theorem 3.1.7

In this section we study the properties of the principal eigenvector. Let v_1 be the normalized principal eigenvector, i.e., $\|v_1\| = 1$, and let e be as

defined in (3.1.3). Recall from (3.2.1) that

$$\lambda_1 \left(1 - \frac{H}{\lambda_1}\right) v_1 = e\langle e, v_1\rangle,$$

and after inversion (which is possible on the high probability event) we have

$$v_1 = \frac{\langle e, v_1\rangle}{\lambda_1}(1 - H/\lambda_1)^{-1} e.$$

If K denotes the normalization factor, then we can rewrite the above equation with (ξ, ν)-hp as the series

$$v_1 = \frac{K}{\lambda_1} \sum_{k=0}^{\infty} \frac{H^k e}{\lambda_1^k}. \tag{3.3.1}$$

Our first step is to determine the value of K in (3.3.1). We adapt the results from [57] to derive a component-wise central limit theorem in the inhomogeneous setting described by (3.1.2) under Assumption 3.1.1. By the normalization of v,

$$1 = \langle v_1, v_1\rangle = \frac{K^2}{\lambda_1^2} \left\langle \sum_{k=0}^{\infty} \frac{H^k}{\lambda_1^k} e, \sum_{\ell=0}^{\infty} \frac{H^\ell}{\lambda_1^\ell} e \right\rangle = \frac{K^2}{\lambda_1^2} \sum_{k=0}^{\infty} \frac{(k+1)\left\langle e, H^k e\right\rangle}{\lambda_1^k}, \tag{3.3.2}$$

where we use the symmetry of H.

The following lemma settles Theorem 3.1.7(I).

Lemma 3.3.1. *Under Assumption 3.1.1, and with* $\tilde{e} = e\sqrt{\frac{m_1}{m_2}}$*, with* (ξ, ν)*-hp*

$$\langle \tilde{e}, v_1\rangle = 1 + \mathrm{o}(1). \tag{3.3.3}$$

Proof. Recall that $L = \lfloor \log n \rfloor$. We rewrite (3.3.2) as

$$\begin{aligned}\left(\frac{\lambda_1}{K}\right)^2 = \sum_{k=0}^{L} \frac{(k+1)}{\lambda_1^k} \mathbb{E}\left[\left\langle e, H^k e\right\rangle\right] + \sum_{k=1}^{L} \frac{(k+1)}{\lambda_1^k} \left|\left\langle e, H^k e\right\rangle - \mathbb{E}\left[\left\langle e, H^k e\right\rangle\right]\right| \\ + \sum_{k=L+1}^{\infty} \frac{(k+1)}{\lambda_1^k} \left\langle e, H^k e\right\rangle.\end{aligned} \tag{3.3.4}$$

We first show that the last two parts are negligible and then show that the main term of the first part is the term with $k = 0$, i.e., $\langle e, e\rangle = m_2/m_1$.

The last term in (3.3.4) is dealt with as follows. Using (3.2.8), we have with (ξ, ν)-hp

$$\sum_{k=L+1}^{\infty} \frac{(k+1)}{\lambda^k} \langle e, H^k e\rangle \leq \sum_{k=L+1}^{\infty} (k+1) \frac{\|e\|^2 \|H\|^k}{(m_2/m_1)^k} \leq \frac{m_2}{m_1} \sum_{k=L+1}^{\infty} (k+1)(1-C_0)^k$$
$$\leq \frac{m_2}{m_1} (\log n + 2) \mathrm{e}^{-c' \log n} \frac{1}{C_0^2}$$

with $c' = -\log(1 - C_0)$, where we use that $\sum_{k=0}^{\infty} (k+1)(1-c)^k = 1/c^2$ for $|1 - c| < 1$.

We tackle the second sum in (3.3.4) by using Lemma 3.2.4. Indeed, with (ξ, ν)-hp we have

$$\sum_{k=1}^{L} \frac{(k+1)}{\lambda_1^k} \left| \langle e, H^k e\rangle - \mathbb{E}\left[\langle e, H^k e\rangle\right] \right| \leq \sum_{k=1}^{L} (k+1) \frac{C m_\infty^{k/2} (\log n)^{k\xi}}{\sqrt{n}} \left(\frac{m_2}{m_1}\right)^{1-k}$$
$$\leq \frac{C' \sqrt{m_\infty} (\log n)^{\xi} (\log n + 1)}{\sqrt{n}} \leq \frac{C' \sqrt{m_\infty} (\log n)}{\sqrt{n}}$$

where the constant varies in each step. By Assumption 3.1.1(D1), the last term goes to zero.

As to the first term, note that by (3.2.5) for $k \geq 3$ we have

$$\sum_{k=3}^{L} \frac{(k+1)}{\lambda_1^k} \mathbb{E}\left[\langle e, H^k e\rangle\right] \leq \sum_{k=3}^{L} (k+1) \left(\frac{m_2}{m_1}\right)^{-k+1} (C m_\infty)^{k/2}$$
$$\leq \sum_{k=3}^{L} \frac{C m_\infty^{k/2}}{(m_2/m_1)^{(k-1)}} = \mathrm{O}\left(\frac{1}{\sqrt{m_\infty}}\right).$$

The term with $k = 1$ is zero, while for $k = 2$ we have

$$3 \frac{\mathbb{E}\langle e, H^2 e\rangle}{\lambda_1^2} \leq c \frac{m_1 m_3}{m_2^2} = \mathrm{O}(1)$$

for some constant c. After substituting these results into (3.3.4), we find

$$\left(\frac{\lambda_1}{K}\right)^2 = \frac{m_2}{m_1} \left(1 + \mathrm{O}\left(\frac{1}{m_2/m_1}\right)\right) \tag{3.3.5}$$

and the proof follows by normalizing the vector e and using (3.3.1).

The following lemma is an immediate consequence of (3.3.1) and Lemma 3.3.1.

Lemma 3.3.2. *Under Assumptions 3.1.1, with (ξ, ν)-hp*

$$v_1 = \left(1 + \mathrm{O}\left(\frac{m_1}{m_2}\right)\right)\sqrt{\frac{m_1}{m_2}}\sum_{k=0}^{\infty}\frac{H^k}{\lambda_1^k}e. \tag{3.3.6}$$

In order to estimate how the components of v_1 concentrate, we need the following lemma.

Lemma 3.3.3. *For $1 \le k \le L$, with (ξ, ν)-hp*

$$|H^k e(i)| = \left|\frac{1}{\sqrt{m_1}}\sum_{i_1,\dots,i_k} h_{ii_1}h_{i_1i_2}\dots h_{i_{k-1}i_k}d_{i_k}\right| \le \frac{m_\infty}{\sqrt{m_1}}\left((\log n)^\xi\sqrt{m_\infty}\right)^k.$$

The proof of this lemma is a direct consequence of Lemma 3.2.4, is similar to [57, Lemma 7.10] and therefore we skip it. An immediate corollary of the above estimate is the delocalized behaviour of the largest eigenvector stated in Theorem 3.1.7(II).

Lemma 3.3.4. *Let v_1 be the normalized principal eigenvector, and $\tilde{e} = e\sqrt{\frac{m_1}{m_2}}$. Then with (ξ, ν)-hp*

$$\|v_1 - \tilde{e}\|_\infty \le \mathrm{O}\left(\frac{(\log n)^\xi}{\sqrt{nm_\infty}}\right).$$

Proof. Recall from (3.3.4) that

$$v_1(i) = \frac{K}{\lambda_1}\sum_{k=0}^{\infty}\frac{H^k e(i)}{\lambda_1^k} = \frac{K}{\lambda_1}e(i) + \frac{K}{\lambda_1}\sum_{k=1}^{L}\frac{H^k e(i)}{\lambda_1^k} + \frac{K}{\lambda_1}\sum_{k=L+1}^{\infty}\frac{H^k e(i)}{\lambda_1^k}.$$

The last term is negligible with (ξ, ν)-hp , because it is the tail sum of a geometrically decreasing sequence. For the sum over $1 \le k \le L$ fwe can use Lemma 3.3.3 and the fact that $K/\lambda_1 = \sqrt{\frac{m_1}{m_2}} + \mathrm{o}(1)$ with (ξ, ν)-hp . So we have

$$\frac{K}{\lambda_1}\sum_{k=1}^{L}\frac{H^k e(i)}{\lambda_1^k} \le \frac{m_\infty}{\sqrt{n}m_0}\frac{(\log n)^\xi}{\sqrt{m_\infty}} \le \mathrm{O}\left(\frac{(\log n)^\xi}{\sqrt{nm_\infty}}\right).$$

The first term with (ξ,ν)-hp is

$$\frac{K}{\lambda_1}e(i) = \tilde{e}(i) + \mathrm{o}(1)$$

and the error is uniform over all i. Indeed, with (ξ,ν)-hp

$$\left|\frac{K}{\lambda_1}e(i) - \frac{K}{m_2/m_1}e(i)\right| \le \frac{Kd_i}{\sqrt{m_1}}\frac{|\lambda_1 - m_2/m_1|}{(m_2/m_1)^2} \le \sqrt{\frac{m_2}{m_1}}\frac{cm_\infty^{3/2}}{\sqrt{m_0 n}}\frac{c'}{m_\infty^2} = \mathrm{O}\left(\frac{1}{\sqrt{nm_\infty}}\right), \tag{3.3.7}$$

where we use Assumption 3.1.1, Remark 3.2.1 and (3.2.7). Since the detailed computations are similar to the previous arguments, we skip the details.

We next prove the central limit theorem for the components of the eigenvector stated in Theorem 3.1.7(IV).

Theorem 3.3.5. *Under Assumption 3.1.1, with the extra assumtion* $m_\infty \gg (\log n)^{4\xi}$,

$$\sqrt{\frac{m_2^3}{d_i m_3 m_1}}\Big(v_1(i) - \frac{d_i}{\sqrt{m_2}}\Big) \xrightarrow{w} \mathcal{N}(0,1).$$

Proof. First we compute $\mathbb{E}[v_1(i)]$, and afterwards we show that the CLT holds componentwise.

We use the law of total expectation. Conditioning on the high probability event $\mathcal{E}$ in Lemma 3.2.2, we can write the expectation of the normalized eigenvector v_1 as

$$\mathbb{E}[v_1(i)] = \mathbb{E}[v_1(i)|\mathcal{E}]\,\mathbb{P}(\mathcal{E}) + \mathbb{E}[v_1(i)|\mathcal{E}^c]\,\mathbb{P}(\mathcal{E}^c).$$

Because the components of a normalized n-dimensional vector are bounded, we know that

$$\mathbb{E}[v_1(i)] = \mathbb{E}[v_1(i)|\mathcal{E}]\,\mathbb{P}(\mathcal{E}) + \mathrm{O}\left(e^{-c_\nu(\log n)^\xi}\right)$$

for some suitable constant $c_\nu > 0$, dependent on ν and on the the bound on $v_1(i)$. On $\mathcal{E}$, we can expand v_1 as

$$v_1(i) = \frac{K}{\lambda_1}\left(e(i) + \frac{(He)(i)}{\lambda_1} + \frac{(H^2e)(i)}{\lambda_1^2} + \sum_{k=3}^{\infty}\frac{(H^ke)(i)}{\lambda_1^k}\right).$$

Using the notation $\mathbb{E}_{\mathcal{E}}$ for the conditional expectation on the event $\mathcal{E}$, we have

$$\mathbb{E}_{\mathcal{E}}[v_1(i)] = \mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}e(i)\right] + \mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\frac{(He)(i)}{\lambda_1}\right] + +\mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\sum_{k=2}^{\infty}\frac{(H^k e)(i)}{\lambda_1^k}\right].$$

For the first term we have, using (3.3.5),

$$\mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}e_i\right] = \mathbb{E}_{\mathcal{E}}\left[\frac{1}{\sqrt{m_2/m_1}}e_i\right] + \mathrm{O}\left(\frac{d_i}{\sqrt{m_1}(m_2/m_1)^{3/2}}\right) = \frac{d_i}{\sqrt{m_2}} + \mathrm{O}\left(\frac{d_i}{\sqrt{m_1}(m_2/m_1}\right.$$

For the term corresponding to $k = 1$, we know that $\mathbb{E}[(He)(i)] = 0$ by construction on the whole space. However, under the event $\mathcal{E}$ we can show that its contribution is exponentially negligible. We have

$$\mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\frac{(He)(i)}{\lambda_1}\right] = \mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\frac{\sum_j h_{ij}d_j}{\sqrt{m_1}\lambda_1}\right] = \mathbb{E}_{\mathcal{E}}\left[\frac{\left(1+\mathrm{O}\left(\frac{1}{m_2/m_1}\right)\right)}{\sqrt{m_2/m_1}}\left(\frac{\sum_j h_{ij}d_j}{\sqrt{m_1}(m_2/m_1)}\right.\right.$$
$$\left.\left.+\frac{\sum_j h_{ij}d_j}{\sqrt{m_1}}\frac{|\lambda_1-(m_2/m_1)|}{(m_2/m_1)^2}\right)\right].$$

Since $m_2/m_1 \to \infty$, there exists a constant $\tilde{C}$ such that

$$\frac{(1+\mathrm{O}\left(1/(m_2/m_1)\right))}{\sqrt{m_2/m_1}} \leq \tilde{C}\frac{1}{\sqrt{m_2/m_1}}.$$

We can therefore write

$$\begin{aligned}\mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\frac{\sum_j h_{ij}d_j}{\sqrt{m_1}\lambda_1}\right] &\leq \tilde{C}\frac{1}{\sqrt{m_2/m_1}}\mathbb{E}_{\mathcal{E}}\left[\frac{\sum_j h_{ij}d_j}{\sqrt{m_1}(m_2/m_1)} + \frac{\sum_j h_{ij}d_j}{\sqrt{m_1}}\frac{|\lambda_1-(m_2/m_1)|}{(m_2/m_1)^2}\right]\\ &\leq \mathbb{E}_{\mathcal{E}}\left[\frac{\sum_j h_{ij}d_j}{\sqrt{m_1}(m_2/m_1)}\right] + \mathbb{E}_{\mathcal{E}}\left[\frac{\sum_j h_{ij}d_j}{\sqrt{m_1}}\frac{|\lambda_1-(m_2/m_1)|}{(m_2/m_1)^2}\right]\\ &\leq \mathbb{E}_{\mathcal{E}}\left[\sum_j h_{ij}d_j\right]\left(\frac{1}{\sqrt{m_1}(m_2/m_1)} + \frac{\sqrt{m_\infty}}{\sqrt{m_1}(m_2/m_1)}\right).\end{aligned}$$

Here we use (3.2.7) to bound the difference $|\lambda_1 - (m_2/m_1)|$. Next, write

$$0 = \mathbb{E}\left[\sum_j h_{ij} d_j\right] = \mathbb{E}_{\mathcal{E}}\left[\sum_j h_{ij} d_j\right] \mathbb{P}(\mathcal{E}) + \mathbb{E}_{\mathcal{E}^c}\left[\sum_j h_{ij} d_j\right] \mathbb{P}(\mathcal{E}^c)$$
$$\leq \mathbb{E}_{\mathcal{E}}\left[\sum_j h_{ij} d_j\right] \mathbb{P}(\mathcal{E}) + m_1 \mathbb{P}(\mathcal{E}^c) = \mathbb{E}_{\mathcal{E}}\left[\sum_j h_{ij} d_j\right] \mathbb{P}(\mathcal{E}) + \mathrm{O}\left(e^{-c_\nu (\log n)^\xi}\right),$$

where c_ν is a constant depending on ν, and we use that $|h_{ij}| \leq 1$ and $m_1 = \mathrm{O}\left(e^{3/2 \log n}\right)$. We can therefore conclude that

$$\mathbb{E}_{\mathcal{E}}\left[\frac{(He)(i)}{\lambda_1}\right] = \mathrm{O}\left(e^{-c'_\nu (\log n)^\xi}\right),$$

where $c'_\nu > 0$ is a suitable constant depending on ν, and possibly different from c_ν.

To bound the remaining expectation terms, we use Lemma 3.3.3, which gives a bound on $(H^k e)(i)$ on the event $\mathcal{E}$. As before, we break up the sum into two contributions:

$$\mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\sum_{k=2}^{\infty}\frac{(H^k e)(i)}{\lambda_1^k}\right] = \mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\sum_{k=2}^{L}\frac{(H^k e)(i)}{\lambda_1^k}\right] + \mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\sum_{k=L}^{\infty}\frac{(H^k e)(i)}{\lambda_1^k}\right].$$

For the second term we have

$$\sum_{k=L+1}^{\infty}\frac{\left(H^k e\right)(i)}{\lambda_1^k} \leq C\sqrt{\frac{m_2}{m_1}}\,\mathrm{e}^{-C_c \log n}, \tag{3.3.8}$$

where we use (3.2.8) and $C_c = |\log(1 - C_0)|$. The first term can be bounded via Lemma 3.3.3, which gives

$$\sum_{k=2}^{L}\frac{(H^k e)(i)}{\lambda_1^k} \leq \sum_{k=2}^{L}\frac{m_\infty\left((\log n)^\xi\sqrt{m_\infty}\right)^k}{\sqrt{m_1}(m_2/m_1)^k} = \mathrm{O}\left(\frac{(\log n)^{2\xi}}{\sqrt{m_1}}\right). \tag{3.3.9}$$

Using the above bounds, taking expectations and using (3.3.5), we get

$$\mathbb{E}_{\mathcal{E}}\left[\frac{K}{\lambda_1}\sum_{k=2}^{\infty}\frac{(H^k e)(i)}{\lambda_1^k}\right] = \mathrm{O}\left(\frac{(\log n)^{2\xi}}{\sqrt{m_2}}\right).$$

Thus, we have obtained that

$$\mathbb{E}[v_1(i)] = \frac{d_i}{\sqrt{m_2}} + \mathrm{O}\left(\frac{(\log n)^{2\xi}}{\sqrt{m_2}}\right),$$

which settles Theorem 3.1.7(III).

We can write

$$v_1(i) - \frac{d_i}{\sqrt{m_2}} = \frac{\left(1 + \mathrm{O}\left(\frac{1}{m_2/m_1}\right)\right) e(i)}{\sqrt{m_2/m_1}} - \frac{d_i}{\sqrt{m_2}} + \frac{K}{\lambda_1}\frac{(He)(i)}{\lambda_1} + \mathrm{O}\left(\frac{(\log n)^{2\xi}}{\sqrt{m_2}}\right),$$

where we replace the last terms of the expansion of v_1 by the bounds derived above (note that these bounds are of the same order as the ones obtained for the same terms in expectation). The first term of the centered quantity $v_1(i) - d_i/\sqrt{m_2}$ is given by

$$\frac{\left(1 + \mathrm{O}\left(\frac{1}{m_2/m_1}\right)\right) e(i)}{\sqrt{m_2/m_1}} = \mathrm{O}\left(\frac{d_i}{\sqrt{m_1}(m_2/m_1)^{3/2}}\right).$$

This last error can be easily seen to be $\mathrm{o}\left(\frac{(\log n)^{2\xi}}{\sqrt{m_2}}\right)$. We can therefore write

$$v_1(i) - \mathbb{E}[v_1(i)] = \frac{K}{\lambda_1}\frac{(He)(i)}{\lambda_1} + \mathrm{O}\left(\frac{(\log n)^{2\xi}}{\sqrt{m_2}}\right).$$

We proceed to show that the first term on the right-hand side of the above equality gives a CLT when the expression is rescaled by an appropriate quantity, and the error term goes to zero. It turns out that

$$s_n^2(i) = \mathrm{Var}\left(\sum_j h_{ij} d_j\right) = \sum_j \frac{d_i d_j^3}{m_1}\left(1 + \mathrm{O}\left(\frac{1}{m_0}\right)\right) \sim \frac{d_i m_3}{m_1}.$$

Multiplying by $\sqrt{\frac{m_2^3}{d_i m_3 m_1}}$, we have

$$\sqrt{\frac{m_2^3}{d_i m_3 m_1}}\Big(v_1(i) - \langle \tilde{e}, v_1\rangle \tilde{e}(i)\Big) = \frac{1}{s_n}\sum_j h_{ij} d_j + \mathrm{O}\left(\sqrt{\frac{m_2^2 (\log n)^{4\xi}}{d_i m_3 m_1}}\right).$$

The error term is

$$\sqrt{\frac{m_2^2 (\log n)^{4\xi}}{d_i m_3 m_1}} = \mathrm{O}\left(\frac{(\log n)^{2\xi}}{\sqrt{m_0}}\right) = \mathrm{o}(1),$$

where last inequality follows from the assumption that $m_0 \gg (\log n)^{4\xi}$. We now apply Lindeberg's CLT to the term $\frac{\sum_j h_{ij} d_j}{s_n}$. The Lindeberg condition for the CLT reads

$$\lim_{n\to\infty} \frac{1}{s_n^2(i)} \sum_j^n \mathbb{E}\left[(h_{ij}d_j)^2 \mathbf{1}_{\{|h_{ij}d_j| \geq \epsilon s_n(i)\}}\right] = 0. \tag{3.3.10}$$

Defining $\sigma_j^2(i) = \mathrm{Var}(h_{ij}d_j)$, we note that

$$\lim_{n\to\infty} \frac{\sigma_j^2(i)}{s_n^2(i)} = \lim_{n\to\infty} \frac{d_i d_j^3 m_1}{m_1 m_3 d_i} \leq \lim_{n\to\infty} \frac{m_\infty^3}{m_3} \leq \lim_{n\to\infty} \frac{m_\infty^3}{n m_0^3} = 0.$$

Let us finally examine the event

$$|h_{ij}d_j| \geq \epsilon s_n(i) = \epsilon\sqrt{\frac{d_i m_3}{m_1}} \iff |h_{ij}| \geq \epsilon\sqrt{\frac{m_3}{m_1}\frac{d_i}{d_j^2}}.$$

By definition, $|h_{ij}| < 1$. If we show that

$$\lim_{n\to\infty} \sqrt{\frac{m_3}{m_1}\frac{d_i}{d_j^2}} = \infty,$$

then for all $\epsilon > 0$ there exists n_ϵ such that the event

$$\epsilon\sqrt{\frac{m_3}{m_1}\frac{d_i}{d_j^2}} > 1 > |h_{ij}|$$

has probability 1. Indeed,

$$\lim_{n\to\infty} \epsilon\sqrt{\frac{m_3}{m_1}\frac{d_i}{d_j^2}} > \lim_{n\to\infty} \epsilon\sqrt{\frac{n m_0^4}{n m_\infty^3}} \geq \lim_{n\to\infty} \epsilon\, C\sqrt{m_0} = \infty$$

for a suitable constant C. Thus, (3.3.10) holds.

Chapter 4

Largest eigenvalue of the configuration model and breaking of ensemble equivalence

This chapter is based on:
P. Dionigi, D. Garlaschelli, R.S. Hazra, F. den Hollander. *Largest eigenvalue of the configuration model and breaking of ensemble equivalence*. arXiv:2312.07812, 2023.

Abstract

We analyse the largest eigenvalue of the adjacency matrix of the configuration model with large degrees, where the latter are treated as hard constraints. In particular, we compute the expectation of the largest eigenvalue for degrees that diverge as the number of vertices n tends to infinity, uniformly on a scale between 1 and $\sqrt{n}$, and show that a weak law of large

numbers holds. We compare with what was derived in Chapter 3 for the Chung-Lu model, which in the regime considered represents the corresponding configuration model with soft constraints, and show that the expectation is shifted down by 1 asymptotically. This shift is a signature of breaking of ensemble equivalence between the hard and soft (also known as micro-canonical and canonical) versions of the configuration model. The latter result generalizes the previous finding in 2 obtained in the case when all degrees are equal.

4.1 Introduction and main results

Motivation. Spectral properties of adjacency matrices in random graphs play a crucial role in various areas of network science. The largest eigenvalue is particularly sensitive to the graph architecture, making it a key focus. In this paper we focus on a random graph with a hard constraint on the degrees of nodes. In the homogeneous case (all degrees equal to d), it reduces to a random d-regular graph. In the heterogeneous case (different degrees), it is known as the configuration model. Our interest is characterizing the expected largest eigenvalue of the configuration model and comparing it with the same quantity for a corresponding random graph model where the degrees are treated as soft constraints.

The set of d-regular graphs on n vertices, with $d = d(n)$, is nonempty when $1 \leq d \leq n-1$ and dn is even. Selecting a graph uniformly at random from this set results in what is known as the *random d-regular graph*, denoted by $G_{n,d}$. The spectral properties of $G_{n,d}$ are well-studied for $d \geq 2$ (for $d = 1$, the graph is trivially a set of disconnected edges). For instance, all eigenvalues of its adjacency matrix fall in the interval $[-d, d]$, with the largest eigenvalue being d. The computation of $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$, where λ_2 and λ_n are the second-largest and the smallest eigenvalue, respectively, has been challenging. It is well known that for fixed d the empirical distribution function of the eigenvalues of $G_{n,d}$ converges to the so-called Kesten-McKay law [93], the density of

which is given by

$$f_{\mathrm{KM}}(x) = \begin{cases} \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)} & \text{if } |x| \leq 2\sqrt{d-1}, \\ 0 & \text{otherwise.} \end{cases}$$

From the convergence of the spectral distribution a lower bound of the type $\lambda \geq 2\sqrt{d-1} + o(1)$ trivially follows. An explicit dependence on n and the degree d of the error term was later found in the celebrated Alon-Boppana theorem in [5], stating that $\lambda \geq 2\sqrt{d-1} + O_d(\log^{-2}(n))$, where the constant in the error term only depends on d. In the same paper an upper bound of the type $\lambda \leq 2\sqrt{d-1} + o(1)$ was conjectured. This conjecture was later proven in the pioneering work [65] (for a shorter proof, see [27]), and improved in the recent work [80]. When $d = d(n) \to \infty$, the spectral distribution converges to the semi-circular law [125]. It was proven in [34] that $\lambda = O(\sqrt{d})$ with high probability when $d = o(\sqrt{n})$. This result was extended in [17] to $d = o(n^{2/3})$. For recent results and open problems, we refer the reader to [129].

There has not been much work on the inhomogeneous setting where not all the degrees of the graph are the same. The natural extension of the regular random graph is the configuration model, where a degree sequence $\vec{d_n} = (d_1, \ldots, d_n)$ is prescribed. Unless otherwise specified, we assume the degrees to be hard constraints, i.e. realized exactly on each configuration of the model (this is sometimes called the 'microcanonical' configuration model, as opposed to the 'canonical' one where the degrees are soft constraints realized only as ensemble averages [71, 114]). The empirical spectral distribution of the configuration model is known under some assumptions on the growth of the sum of the degrees. When $\sum_{i=1}^n d_i = O(n)$, the graph locally looks like a tree. It has been shown that the empirical spectral distribution exists and that the limiting measure is a functional of a size-biased Galton-Watson tree [28]. When $\sum_{i=1}^n d_i$ grows polynomially with n, the local geometry is no longer a tree. This case was recently studied in [49], where it is shown that the appropriately scaled empirical spectral distribution of the configuration model converges to the free multiplicative convolution of the semi-circle law with a measure that depends on the empirical degree sequence. The

scaling of the largest and the second-largest eigenvalues has not yet been studied in full generality.

For the configuration model the behaviour of the largest eigenvalue is non-trivial. In the present paper, we consider the configuration model with large degrees, compute the expectation of the largest eigenvalue of its adjacency matrix, and prove a weak law of large numbers. When $d \to \infty$, the empirical distribution of $G_{n,d}$ (appropriately scaled) converges to the semi-circle law. Also, if we look at Erdős-Rényi random graphs on n vertices with connection probability d/n, then the appropriately scaled empirical distribution converges to the semi-circle law. It is known that the latter two random graphs exhibit different behaviour for the expected largest eigenvalue [53]. This can be understood in the broader perspective of *breaking ensemble equivalence* [71, 114], which we discuss later. In the inhomogeneous setting, the natural graph to compare the configuration model with is the Chung-Lu model. For the case where the sum of the degrees grows with n, it was shown in [40] that the empirical spectral distribution converges to the free multiplicative product of the semi-circle law with a measure that depends on the degree sequence, similar to [49]. In [54], we investigated the largest eigenvalue and its expectation for the Chung-Lu model (and even derived a central limit theorem). In the present paper we complete the picture by showing that, when the degrees are large, there is a *gap* between the expected largest eigenvalue of the Chung-Lu model and the configuration model. We refer the reader to the references listed in [53, 54] for more background.

Main theorem. For $n \in \mathbb{N}$, let $\mathrm{CM}(\vec{d}_n)$ be the random graph on n vertices generated according to the *configuration model* with degree sequence $\vec{d}_n = (d_i)_{i\in[n]} \in \mathbb{N}^n$ [79, Chapter 7.2]. Define

$$m_0 = \min_{i\in[n]} d_i, \qquad m_\infty = \max_{i\in[n]} d_i, \qquad m_k = \sum_{i\in[n]} (d_i)^k, \quad k \in \mathbb{N}.$$

Throughout the paper we need the following assumptions on $\vec{d}_n$ as $n \to \infty$.

Assumption 4.1.1.
(D1) **Bounded inhomogeneity:** $m_0 \asymp m_\infty$.
(D2) **Connectivity and sparsity:** $1 \ll m_\infty \ll \sqrt{n}$. ♠

Under these assumptions, $\mathrm{CM}(\vec{d}_n)$ is with high probability *non-simple* [79, Chapter 7.4]. We write $\mathbb{P}$ and $\mathbb{E}$ to denote probability and expectation with respect to the law of $\mathrm{CM}(\vec{d}_n)$ *conditional on being simple*, suppressing the dependence on the underlying parameters.

Let $A_{\mathrm{CM}(\vec{d}_n)}$ be the adjacency matrix of $\mathrm{CM}(\vec{d}_n)$. Let $(\lambda_i)_{i\in[n]}$ be the eigenvalues of $A_{\mathrm{CM}(\vec{d}_n)}$, ordered such that $\lambda_1 \geq \cdots \geq \lambda_n$. We are interested in the behaviour of λ_1 as $n \to \infty$. Our main theorem reads as follows.

Theorem 4.1.2. *Subject to Assumption 4.1.1,*

$$\mathbb{E}[\lambda_1] = \frac{m_2}{m_1} + \frac{m_1 m_3}{m_2^2} - 1 + o(1), \qquad n \to \infty, \tag{4.1.1}$$

and

$$\frac{\lambda_1}{\mathbb{E}[\lambda_1]} \to 1 \text{ in } \mathbb{P}\text{-probability.}$$

In [54] we looked at an alternative version of the configuration model, called the *Chung-Lu model* $\mathrm{CM}_n^*(\vec{d}_n)$, where the *average degrees*, rather than the degrees themselves, are fixed at $\vec{d}_n$. This is an ensemble with soft constraints; in the considered regime for the degrees, it coincides with a maximum-entropy ensemble, also called 'canonical' configuration model [114]. For this model we showed that, subject to Assumption 4.1.1,

$$\mathbb{E}^*[\lambda_1] = \frac{m_2}{m_1} + \frac{m_1 m_3}{m_2^2} + o(1), \qquad n \to \infty, \tag{4.1.2}$$

and $\lambda_1/\mathbb{E}^*[\lambda_1] \to 1$ in $\mathbb{P}^*$-probability, where $\mathbb{P}^*$ and $\mathbb{E}^*$ denote expectation with respect to the law of $\mathrm{CM}_n^*(\vec{d}_n)$ and λ_1 is the largest eigenvalue of the $A_{\mathrm{CM}_n^*(\vec{d}_n)}$. The notable difference between (4.1.1) and (4.1.2) is the shift by -1.

For the special case where all the degrees are equal to d, we have $m_0 = m_\infty = d$ and $m_k = nd^k$, and so $\mathbb{E}[\lambda_1] = d + o(1)$ and $\mathbb{E}^*[\lambda_1] = d + 1 + o(1)$. In fact, $\mathbb{P}(\lambda_1 = d) = 1$. Since in this model the degrees

can fluctuate with the same law (*soft* constraint in the physics literature), this case reduces to the Erdős-Rényi random graph for which results on $\mathbb{E}^*[\lambda_1]$ were already well known in [68, 86] and further analyzed in [58].

Breaking of ensemble equivalence. The shift by -1 was proven in [53] for the homogeneous case with equal degrees and is a *spectral signature* of *breaking of ensemble equivalence* [114]. Indeed, a d-regular graph is the 'micro-canonical' version of a random graph where all degrees are equal and 'hard', and the Erdős-Rényi random graph is the corresponding 'canonical' version where all degrees are equal and 'soft'. More in general, $\mathrm{CM}(\vec{d}_n)$ is the micro-canonical configuration model where the constraint on the degrees is 'hard', while $\mathrm{CM}_n^*(\vec{d}_n)$ is the canonical version where the constraint is 'soft'. We refer the reader to [71] for the precise definition of these two configuration model *ensembles* and for the proof that they are not asymptotically equivalent *in the measure sense* [122]. This means that the *relative entropy per node* of $\mathbb{P}$ with respect to $\mathbb{P}^*$ has a strictly positive limit as $n \to \infty$. This shows that the choice of constraint matters, not only on a microscopic scale but also on a macroscopic scale. Indeed, for non-equivalent ensembles one expects the existence of certain macroscopic properties that have different expectation values in the two ensembles (*macrostate (in)equivalence* [122]). The fact that the largest eigenvalue picks up this discrepancy is interesting. What is remarkable is that the shift by -1, under the hypotheses considered, holds true also in the case of heterogeneous degrees and remains the same *irrespective* of the scale of the degrees and of the distribution of the degrees on this scale.

Outline. The remainder of this paper is organised as follows. In Section 4.2 we look at the issue of simplicity of the graph. In Section 4.3 we bound the spectral norm of the matrix

$$H = A_{\mathrm{CM}(\vec{d}_n)} - \mathbb{E}[A_{\mathrm{CM}(\vec{d}_n)}]. \tag{4.1.3}$$

We use the proof of [34] to show that $\|H\| = o(m_\infty)$ with high probability. Using the latter we show that

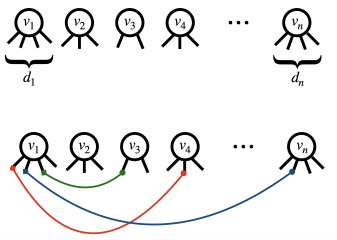
$$\lambda_1 \sim \lambda_1 \left(\mathbb{E}[A_{\mathrm{CM}(\vec{d}_n)}] \right) \sim \frac{m_2}{m_1}$$

with high probability. In Section 4.4 we use the estimates in Section 4.3 to prove Theorem 4.1.2.

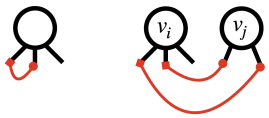
4.2 Configuration model and simple graphs with a given degree sequence

In the context of random graphs with hard constraints on the degrees, we work with both the *configuration model* (a random *multi-graph* with a prescribed degree sequence) and with the *conditional configuration model* (a random *simple* graph with a prescribed degree sequence). We view the second as a special case of the first.

Configuration model. The configuration model with degree sequence $\vec{d} = (d_1, \dots, d_n)$, $\mathrm{CM}(\vec{d}_n)$, generates a graph through a *perfect matching* P of the set of half-edges $\mathcal{E} = \cup_{i \in [n]} \{i\} \times [d_i]$, i.e., $P \colon \mathcal{E} \to \mathcal{E}$ such that P is an isomorphism and an involution, and $P(\alpha) \neq \alpha$ for all $\alpha \in \mathcal{E}$.



(a) *Pairing procedure.*



(b) *The configuration model generates multi-graphs with self-loops (left) and multiple edges (right).*

It turns out that a pairing scheme, such as matching half-edges from left to right or selecting pairs uniformly at random, is considered admissible as long as it yields the correct probability for a perfect matching

(see, for example, [79, Lemma 7.6]). An important property of each admissible pairing scheme is that

$$\mathbb{P}_{\mathrm{CM}(\vec{d}_n)}(\alpha \sim \beta) = \frac{1}{m_1 - 1} \qquad \forall \alpha, \beta \in \mathcal{E}. \tag{4.2.1}$$

We can therefore endow $\mathcal{E}$ with the lexicographic order. An element $\alpha \in \mathcal{E}$, $\alpha = (i, \ell)$, is associated with a vertex $v(\alpha) = i$ through the first component of α. An edge e is an element of $\mathcal{E} \times \mathcal{E}$ given by the pairing $e = \{\alpha, \beta\} = \{\alpha, P(\alpha)\}$. We define the configuration $\mathcal{C}$ as the set of all such e. Given the lexicographic order, we may assume that $\alpha < \beta$ and identify arbitrarily the head and the tail of the edge: $t(e) = v(\alpha)$, $h(e) = v(\beta)$. We can then order the edges in their order of appearance via $t(e)$, forming a list $\{e_i,\, i = 1, \dots, m_1/2\}$. These properties of the configuration model will be used in Subsection 4.3.1. For the configuration model it is easy to check that[1]

$$\mathbb{E}[a_{ii}] = \frac{d_i(d_i - 1)}{m_1 - 1}, \qquad \mathbb{E}[a_{ij}] = \frac{d_i d_j}{m_1 - 1}, \quad i \neq j. \tag{4.2.2}$$

Indeed every matrix element a_{ij}, can be expressed as

$$\sum_{k=1}^{d_i} \sum_{h=1}^{d_j} \mathbb{1}\big(\alpha \sim \beta, \alpha = (i,k), \beta = (j,h)\big),$$

where $\mathbb{1}(E)$ is the indicator function of the event E. Taking expectations and using (4.2.1), we get 4.2.2.

Simple graphs with a given degree sequence. A linked but different model is the one that samples uniformly at random a simple graph with a given degree sequence $\vec{d}$, $\mathcal{G}(\vec{d})$[2]. An immediate question is whether there is any relation between $\mathrm{CM}(\vec{d})$ and $\mathcal{G}(\vec{d})$. It turns out that the following is true (see [79] for reference):

[1] We adopt the convention that a self-loop contributes 2 to the degree, i.e., a_{ii} is twice the number of self-loops attached to vertex i. This convention is useful because it yields $\sum_j a_{ij} = d_i$.

[2] With a little abuse of notation we permit our simple graphs to have self-loops. We maintain the convention explained in the previous footnote.

Lemma 4.2.1. *Let be G_n a graph generated via the configuration model $\mathrm{CM}(\vec{d}_n)$ with degree sequence $\vec{d}$. Then*

$$\mathbb{P}_{\mathrm{CM}(\vec{d}_n)}(G_n \text{ is simple}) = \exp\left[-O\left(\frac{m_1^2}{n^2}\right)\right] \tag{4.2.3}$$

and

$$\mathbb{P}_{\mathrm{CM}(\vec{d}_n)}\left(G_n = \cdot \mid G_n \text{ is simple}\right) \text{ is uniform.} \tag{4.2.4}$$

Thus, we can identify the law of the *uniform random graph with a given degree sequence* with the law of the *configuration model conditional on simplicity*, i.e.,

$$\mathbb{P}_{\mathcal{G}(\vec{d})}(\cdot) = \mathbb{P}_{\mathrm{CM}(\vec{d})}(\cdot \mid \mathrm{G \text{ is simple}}).$$

It follows that the expectation under the uniform random graph with a given degree sequence can be expressed as a the expectation of the configuration model conditional on simplicity. As explained in [78, Remark 1.14], we have

$$\begin{aligned}\mathbb{E}_{\mathcal{G}(\vec{d})}[a_{ij}] = \mathbb{P}_{\mathcal{G}(\vec{d})}(i \sim j) &= (1+o(1))\frac{d_i d_j}{m_1 + d_i d_j} \\ &\leq \frac{d_i d_j}{m_1(1+O(1/m_\infty))} = \frac{d_i d_j}{m_1} + O(1/m_\infty).\end{aligned} \tag{4.2.5}$$

4.3 Bound on $\|H\|$

In this section we derive the following bound on the spectral norm of the matrix H defined in (4.1.3). This bound will play a crucial role in the proof of Theorem 4.1.2 in Section 4.4.

Theorem 4.3.1. *For every $K > 0$, there exists a constant $C > 0$ such that*

$$\|H\| \leq C\sqrt{m_\infty} \text{ with probability } 1 - o(n^{-K}),$$

where the constant C depends on $\frac{m_\infty}{m_0}$ and K only.

The proof is given in Section 4.3.1 and we follow the proof of Lemma 16 of [34]. In Sections 4.3.2–4.3.3 we derive some estimates that are needed in Section 4.3.1.

We have

$$\|H\| = \max\{\lambda_1(H), |\lambda_n(H)|\}.$$

Recall (4.1.3). In Lemma 4.3.3 below we will see that $\mathbb{E}[A_{\mathrm{CM}(\vec{d}_n)}]$ is asymptotically rank 1 (more precisely, the other eigenvalues are of order $O(n^{-1})$. This means that $\lambda_i\left(A_{\mathrm{CM}(\vec{d}_n)}\right) = O(n^{-1})$, $i \neq 1$. Therefore, by Weyl's inequality, it easy to see that $\lambda_1(H)$ and $\lambda_n(H)$ are bounded above and below by $\lambda_2(A_{G_n})$ and $\lambda_n(A_{G_n})$ with error $O(n^{-1})$, while $\lambda_1(A_{G_n})$ is in a window of size $\sqrt{m_\infty}$ around $\frac{m_2}{m_1-1}$. We therefore have the following Corollary, which will be needed in Section 4.4:

Corollary 4.3.2. *For every $K > 0$,*

$$\lambda_1(A_{G_n}) - \lambda_1(\mathbb{E}[A_{G_n}]) = O(\sqrt{m_\infty}) \text{ with probability } 1 - o(n^{-K}).$$

4.3.1 Spectral estimates

Notation. Abbreviate $G_n = \mathrm{CM}(\vec{d}_n)$. Let U be uniformly distributed on $[n]$, let d_U be the degree of a vertex that is picked uniformly at random, and let

$$\omega_n = \mathbb{E}[d_U]$$

be the average of the empirical degree distribution. Under Assumption 4.1.1, $\omega_n \to \infty$ and $\omega_n = o(n)$. We define the *normalised degree sequence* $(\hat{d}_i)_{i\in[n]}$ as

$$\hat{d}_i = \frac{d_i}{\omega_n} \tag{4.3.1}$$

and the *normalised adjacency matrix* $\hat{A}_{G_n}$ as

$$\hat{A}_{G_n} = \frac{A_{G_n}}{\sqrt{\omega_n}}.$$

In the following we will need multiples of the vector

$$\tilde{e}_i = \frac{d_i}{\sqrt{m_1 - 1}},\ 1 \leq i \leq n, \tag{4.3.2}$$

which is the eigenvector corresponding to the rank-1 approximation of $\mathbb{E}[A_{G_n}]$, as it easy to check using (4.2.2).

Two lemmas. The matrix $\mathbb{E}[A_{G_n}]$ is asymptotically rank 1:

Lemma 4.3.3. *Define*

$$\tilde{e} = (\tilde{e}_1, \ldots, \tilde{e}_n)^t$$

and

$$A_{\text{shift}} = \mathbb{E}[A_{G_n}] - |\tilde{e}\rangle\langle\tilde{e}| = -\frac{1}{m_1 - 1} \operatorname{diag}(d_1, \ldots, d_n).$$

Then

$$\|A_{\text{shift}}\| = O(n^{-1}).$$

Proof. Using (4.2.2), we have A_{shift} is a diagonal matrix and hence $\|A_{\text{shift}}\| \leq \frac{m_\infty}{m_1 - 1} = O(n^{-1})$ under Assumption 4.1.1. Note that when all the degrees are equal to d, then this bound is sharp, i.e., $\|A_{\text{shift}}\| = \frac{d}{dn-1}$, and $\mathbb{E}[A_{G_n}]$ is exactly rank 1.

It is shown in [49] that the empirical spectral distribution of A_{G_n} has a deterministic limit given by $\mu = \mu_{sc} \boxtimes \mu_{\hat{D}}$, where μ_{sc} is the standard Wigner semicircle law, $\mu_{\hat{D}}$ is the distribution of $\hat{d}_U$, and $\boxtimes$ is the free product defined in [10, 22, 128].

Lemma 4.3.4. *Let $(\hat{d}_i)_{i \in [n]}$ be the normalised degree sequence, and suppose that Assumption 4.1.1 holds and*

$$\frac{1}{n} \sum_{i=1}^{n} \delta_{\hat{d}_i} \Rightarrow \mu_{\hat{D}}. \tag{4.3.3}$$

Then $\mu_{\hat{D}}$ is compactly supported.

Proof. By Assumption 4.1.1 D(2),

$$0 < c \leq \liminf_{n \to \infty} \frac{\min_i \hat{d}_i}{\max_i \hat{d}_i} \leq \limsup_{n \to \infty} \frac{\max_i \hat{d}_i}{\min_i \hat{d}_i} \leq C < \infty.$$

Hence the support of $\hat{d}_U$ is contained in a multiple of $[c, C]$, and (4.3.3) implies that $\mu_{\hat{D}}$ is compactly supported.

Key estimates. From [49, Theorem 1] we know that, under Assumption 4.1.1, the empirical spectral distribution of $\hat{A}_{G_n}$, written

$$\mu_{\hat{A}_{G_n}},$$

converges weakly to $\mu_{\text{sc}} \boxtimes \mu_{\hat{D}}$. The fact that μ_{sc} and $\mu_{\hat{D}}$ are compactly supported implies that $\mu_{\text{sc}} \boxtimes \mu_{\hat{D}}$ is compactly supported too. However, this still allows that H has $o(n)$ outliers, which possibly control $\|H\|$. It is hard to analyse $\|H\|$ for sparse graphs, because it is related to expansion properties of the graph, mixing times of random walks on the graph, and more. To prove that $\|H\| = O(\sqrt{m_\infty})$, we will use the argument described in [34], which is an adaptation of the argument in [63]. We are only after the order of $\|H\|$, not sharp estimates.

For random regular graphs with fixed degree, the problem of computing $\|H\|$ was solved in [64]. We are not aware of any proof concerning the second largest eigenvalue of configuration models with bounded degrees, but any sharp bound must come from techniques of the type employed in [26, 64]. In what follows we prove Theorem 4.3.1 based on [63] and its adaptation to the inhomogeneous setting in [34]. In the following, we present the proof as outlined in the latter paper, adapting it to our case to make our paper self-contained. This adjustment is necessary because the result we need is somewhat obscured in the context in which it appears in [34].

Let $D_n = \text{diag}(d_1, \dots, d_n)$. The transition kernel of a random walk on the graph G_n is given by $P_n = D_n^{-1} A_{G_n}$. The *random* matrix P_n has as principal normalised eigenvector $\vec{1} = (1, \dots, 1)/\sqrt{n}$ with eigenvalue 1. Define

$$\lambda^* = \max\{\lambda_2(P_n), |\lambda_n(P_n)|\}.$$

Note that the matrices P_n and $S_n = D_n^{1/2} P_n D_n^{-1/2}$ have the same spectrum by a similarity transformation. Hence we can write the Rayleigh formula

$$\lambda^* = \max_{z:\ \langle z, \vec{1}\rangle = 0} \frac{|\langle P_n z, z\rangle|}{\|z\|^2} = \max_{x:\ \langle x, \sqrt{\vec{d}}\rangle = 0} \frac{|\langle S_n x, x\rangle|}{\|x\|^2},$$

where $\sqrt{\vec{d}} = (\sqrt{d_1}, \ldots, \sqrt{d_n})$. Define $M_n = D_n^{-1} A_n D_n^{-1}$. Let $\tilde{\lambda}$ be the second largest eigenvalue of M_n. Then

$$\langle S_n x, x\rangle = \langle D_n^{1/2} M_n D_n^{1/2} x, x\rangle = \langle M_n D_n^{1/2} x, D_n^{1/2} x\rangle.$$

Since

$$\frac{\langle x, x\rangle}{\|x\|^2} \geq \frac{1}{m_\infty} \frac{\langle D_n^{1/2} x, D_n^{1/2} x\rangle}{\|x\|^2},$$

putting $y = D_n^{1/2} x$ we see that

$$\lambda^* \leq m_\infty \max_{\langle y, \vec{1}\rangle = 0} \frac{|\langle M_n y, y\rangle|}{\|y\|^2} = m_\infty \tilde{\lambda}, \tag{4.3.4}$$

which gives a bound of the type $\|H\| = O(m_\infty^2 \tilde{\lambda})$. Note that the matrix elements of M_n can be expressed as

$$(M_n)_{ij} = \frac{a_{ij}}{d_i d_j}, \tag{4.3.5}$$

where a_{ij} counts the number of edges between vertices i and j, with the convention for the diagonal elements stated earlier. In view of (4.3.4), in order to obtain a bound on λ^* we must focus on $\tilde{\lambda}$. In fact, we must show that

$$\tilde{\lambda} = O(m_\infty^{-3/2}), \tag{4.3.6}$$

in order to obtain the desired bound $\|H\| = O(\sqrt{m_\infty})$. To achieve this we proceed in steps:

- We reduce the computation of $\tilde{\lambda}$ to the analysis of two terms.
- In (4.3.7) below we identify the leading order of $\mathbb{E}[\tilde{\lambda}]$ from these two terms, which turns out to be $O(m_\infty^{-3/2})$.
- We show that the other terms are of higher order and therefore are negligible.
- We prove concentration around the mean through concentration of the leading order term in Lemma 4.3.8 below.

– Recalling the denominator of (4.3.5), we get a bound on $\|H\|$ after multiplying by m_∞^2.

To prove (4.3.6), we reduce the problem to a maximisation problem in a simpler space. Namely, let $\varepsilon \in (0,1)$, and

$$T = \left\{ x \in \left(\frac{\varepsilon}{\sqrt{n}}\mathbb{Z}\right)^n : \sum_{i\in[n]} x_i = 0, \sum_{i\in[n]} x_i^2 \le 1 \right\}.$$

Then, using the formula for the volume of the n-dimensional ball, we have

$$|T| \le \left(\frac{(2+\varepsilon)\sqrt{n}}{2\varepsilon}\right)^n \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)} \le \left(\frac{(2+\varepsilon)\sqrt{2\pi e}}{2\varepsilon}\right)^n.$$

The maximisation over $\mathbb{R}^n$ can be reduced to over T and this leads to an error that depends on ε.

Lemma 4.3.5. *Let*

$$\lambda = \max\{|\langle x, M_n y\rangle| : x, y \in T\}.$$

Then

$$\tilde{\lambda} \le (1-\varepsilon)^{-2}\lambda.$$

Proof. Let be $\mathcal{S} = \{x \in \mathbb{R}^n : \sum_{i\in[n]} x_i = 0, \|x\| \le 1\}$. We want to show that for every $x \in \mathcal{S}$ there is a vector $y \in T$ such that $\|x-y\| \le \varepsilon$ and $\sum_{i\in[n]}(x_i - y_i) = 0$. Let us write the components of x as

$$x_i = \varepsilon\frac{m_i}{\sqrt{n}} + f_i, \qquad i \in [n],$$

where $m_i \in \mathbb{Z}$ and $f_i \in [0, \varepsilon n^{-1/2})$ is an error term. Because $\sum_{i\in[n]} x_i = 0$, we choose m_i such that $\sum_{i\in[n]} f_i v = v\varepsilon f n^{-1/2}$, where f is a non-negative integer smaller than n. We relabel the indices in such a way that $m_i \le m_j$ when $i \le j$. Now consider the vector y given by

$$y_i = \begin{cases} \varepsilon\frac{m_i+1}{\sqrt{n}} & \text{if } i \le f, \\ \varepsilon\frac{m_i}{\sqrt{n}} & \text{if } i > f. \end{cases}$$

It follows that $\sum_{i\in[n]} y_i = 0$ and $\|y\| \le 1$, and therefore $y \in T$. Furthermore, because $|x_i - y_i| \le \varepsilon n^{-1/2}$ by construction, we have the required

property $\|x - y\| \leq 1$. Iterating the previous argument, we can express every vector $x \in \mathcal{S}$ in terms of a sequence of vectors $(y^{(i)})_{i\in[n]}$ in T such that

$$x = \sum_{i\in[n]} \varepsilon^i y^{(i)}.$$

Therefore, because (4.3.4) is maximized on $\mathcal{S}$, we have

$$\langle x, M_n x\rangle = \sum_{i,j\in[n]\times[n]} \varepsilon^{i+j} \langle x^{(i)} M_n x^{(j)}\rangle \leq \frac{1}{(1-\varepsilon)^2} \max\{|\langle y, M_n z\rangle| \colon\ y, z \in T\},$$

from which the claim follows.

Our next goal is to show that $|\langle x, M_n y\rangle| = o(m_\infty^{-3/2})$ for all $x, y \in T$ with a suitably high probability. This can be done in the following way. Fix $x, y \in T$ and define the random variable

$$X = \sum_{i,j\in[n]\times[n]} x_i (M_n)_{ij} y_j.$$

Define the set of indices

$$\mathcal{B} = \left\{(i,j) \in [n]\times[n] \colon\ 0 < |x_i y_j| < \frac{\sqrt{m_\infty}}{n}\right\}.$$

Then we can rewrite $X = X' + X''$ with

$$X' = \sum_{(i,j)\in\mathcal{B}} x_i (M_n)_{ij} y_j, \qquad X'' = \sum_{(i,j)\notin\mathcal{B}} x_i (M_n)_{ij} y_j.$$

In Section 4.3.2 we show that $\mathbb{E}[X']$ is of the correct order and that X' is well concentrated around its mean. In Section 4.3.3 we analyse X'', which is of a different nature and requires that we exclude subgraphs in the configuration model that are too dense.

4.3.2 Estimate of first contribution

1. Use (4.2.2) and (4.3.5) to write out $\mathbb{E}[m_{ij}] = \mathbb{E}[(M_n)_{ij}]$. This gives

$$\mathbb{E}[X'] = \sum_{(i,j)\in\mathcal{B}} \frac{x_i y_j}{m_1 - 1} - \sum_{(i,i)\in\mathcal{B}} \frac{x_i y_i}{d_i^2 (m_1 - 1)}.$$

In view of the bound on $|x_i y_i|$, the last term gives a contribution

$$\left| \sum_{(i,i)\in\mathcal{B}} \frac{x_i y_i}{d_i^2(m_1-1)} \right| = O\left(\frac{1}{m_1 m_\infty^{3/2}} \right).$$

Since $x, y \in T$, we have that $\sum_{i\in[n]} x_i = 0$ and $\sum_{j\in[n]} y_j = 0$, and therefore $\sum_{(i,j)\in[n]\times[n]} x_i y_j = 0$. Hence

$$\left| \sum_{(i,j)\in\mathcal{B}} x_i y_j \right| = \left| \sum_{(i,j)\notin\mathcal{B}} x_i y_j \right|,$$

where we can bound the right-hand side as

$$\left| \sum_{(i,j)\notin\mathcal{B}} x_i y_j \right| \leq \sum_{(i,j)\colon |x_i y_j| \geq \frac{\sqrt{m_\infty}}{n}} |x_i y_j| \leq \sum_{(i,j)\colon |x_i y_j| \geq \frac{\sqrt{m_\infty}}{n}} \frac{x_i^2 y_j^2}{|x_i y_j|} \leq \frac{n}{\sqrt{m_\infty}} \sum_{(i,j)} x_i^2 y_j^2 \leq -$$

We can therefore conclude that

$$|\mathbb{E}[X']| \leq \frac{n}{(m_1-1)\sqrt{m_\infty}} + O\left(\frac{1}{m_1 m_\infty^{3/2}} \right). \tag{4.3.7}$$

2. To prove that X' is concentrated around its mean, we use an argument originally developed in [33, 104] and used for the configuration model in [34, 63]. This argument employs the *martingale structure* of the configuration model *conditional on partial pairings*. Define

$$\chi(x) = \begin{cases} x, & \text{if } |x| < \frac{\sqrt{m_\infty}}{n}, \\ 0, & \text{otherwise.} \end{cases}$$

We can then express X' as

$$X' = \sum_{e\in\mathcal{C}} \frac{\chi(x_{t(e)} y_{h(e)})}{d_{t(e)} d_{h(e)}} + \sum_{e\in\mathcal{C}} \frac{\chi(x_{h(e)} y_{t(e)})}{d_{t(e)} d_{h(e)}} = X'_a + X'_b.$$

We divide the set of pairings $\mathcal{C}$ into three sets

$$\mathcal{C}_1 = \left\{ e \in \mathcal{C} \colon\ |x_{t(e)}| > \frac{1}{\varepsilon\sqrt{n}} \right\},$$
$$\mathcal{C}_2 = \left\{ e \in \mathcal{C} \colon\ |y_{t(e)}| > \frac{1}{\varepsilon\sqrt{n}},\ |x_{t(e)}| \le \frac{1}{\varepsilon\sqrt{n}} \right\},$$
$$\mathcal{C}_3 = \left\{ e \in \mathcal{C} \colon\ |y_{t(e)}| \le \frac{1}{\varepsilon\sqrt{n}},\ |x_{t(e)}| \le \frac{1}{\varepsilon\sqrt{n}} \right\},$$

and write

$$X'_a = X_1 + X_2 + X_3, \qquad X_i = \sum_{e \in \mathcal{C}_i} \frac{\chi(x_{t(e)} y_{h(e)})}{d_{t(e)} d_{h(e)}}.$$

We can do a similar decomposition for X'_b.

3. The following martingale lemma from [34, 63] is the core estimate that we want to apply to each of the X_i's.

Lemma 4.3.6. *Let be G_1 and G_2 two graphs generated via the configuration model through perfect matchings P_1 and P_2. Let $\{e_i^{(1)}\}_{i\ge1}$ be the edges of G_1 and $\{e_i^{(2)}\}_{i\ge1}$ be the edges of G_2, ordered as above. Define an equivalence relation on the probability space Ω by setting $G_1 \equiv_k G_2$ when $\{e_i^{(1)}\}_{i=1}^k = \{e_i^{(2)}\}_{i=1}^k$, i.e., the first k pairings match. Let Ω_k be the set of equivalence classes, and $\mathcal{F}_k$ the corresponding σ-algebra with $\mathcal{F}_0 = \mathcal{F}$. Consider a bounded measurable $f\colon\ \mathcal{G}(\vec{d}) \to \mathbb{R}$, and set $Y_k = \mathbb{E}[f \mid \mathcal{F}_k]$. Note that*

$(Y_k)_{0\le k\le m_1/2}$ is a Doob martingale with $\mathbb{E}[Y_k \mid \mathcal{F}_{k-1}] = Y_{k-1}$ and $Y_0 = \mathbb{E}[f]$, $Y_{m_1/2} = f$.

Define $Z_k = Y_k - Y_{k-1}$, and suppose that there exist functions $(g_k(\zeta))_{1\le k\le \frac{1}{2}m_1}$ such that

$$\mathbb{E}\left[\mathrm{e}^{\zeta^2 Z_k^2} \mid \mathcal{F}_{k-1} \right] \le g_k(\zeta), \qquad 1 \le k \le m_1/2.$$

Then, for all $t \ge 0$ and $\zeta > 0$,

$$\mathbb{P}\left(|f - \mathbb{E}[f]| \ge t \right) \le 2\,\mathrm{e}^{-\zeta t} \prod_{k=1}^{m_1/2} g_k(\zeta). \tag{4.3.8}$$

Remark 4.3.7. The condition on the existence $g_k(\zeta)$ can be rephrased as the existence of a random variable W_k that stochastically dominates Z_k on $(\Omega_{k-1}, \mathcal{F}_{k-1})$ (see [63]). ♠

Lemma 4.3.8. *[34, Lemma 15] There exist constants $B_\ell > 0$, $\ell = 1, 2, 3$, depending only on the ratio $\frac{m_\infty}{m_0}$, such that*

$$\mathbb{P}\left(|X_\ell - \mathbb{E}[X_\ell]| \geq \frac{t}{m_\infty^{3/2}}\right) \leq 2\,\mathrm{e}^{-tn+B_\ell n}, \qquad \ell = 1, 2, 3. \tag{4.3.9}$$

Proof. While the properties in $\mathcal{C}_3$ allow us to apply standard martingale arguments to capture X_3 (see below), the properties in $\mathcal{C}_1$ and $\mathcal{C}_2$ force us to use Lemma 4.3.6 to capture X_1 and X_2.

i. From the definition of $\mathcal{C}_1$ and $\mathcal{C}_2$ it follows that a bound on X_1 implies by symmetry a bound on X_2 (with, possibly, different constants). We will therefore focus on X_1, the result for X_2 carrying through trivially. Without loss of generality we may reorder the indices in such a way that $|x_i| \geq |x_{i+1}|$ and for each $e_i = \{\alpha_i, \beta_i\}$ use the lexicographic order $\alpha_{i+1} > \alpha_i$ and $\beta_i > \alpha_i$ (i.e., we perform the pairing sequentially from left to right; see [79, Lemma 7.6]). It follows that $v(\alpha_i) \leq v(\alpha_{i+1})$ and $|x_{v(\alpha_i)}| \geq |x_{v(\alpha_{i+1})}|$. Define

$$\hat{\chi}(x, y) = \begin{cases} xy, & \text{if } |xy| < \frac{\sqrt{m_\infty}}{n} \text{ and } |x| > \frac{1}{\epsilon\sqrt{n}}, \\ 0, & \text{otherwise.} \end{cases}$$

For each $e = \{\alpha, \beta\}$ in the configuration $\mathcal{C}$ we have

$$X_1 = \sum_{e \in \mathcal{C}} q(e)$$

with $q(e) = \hat{\chi}(x_{v(\alpha)}, y_{v(\beta)})/d_{v(\alpha)}d_{v(\beta)}$.

ii. Next, take $Y_k = \mathbb{E}[X_1 \mid \mathcal{F}_k]$, with $Y_0 = \mathbb{E}[X_1]$ and $Y_m = X_1$. Then $(Y_k)_{k \in \mathbb{N}_0}$ is a Doob martingale and, by the definition of the configuration model, we can write $Z_k = Y_k - Y_{k-1}$ as

$$Z_k(\mathcal{C}) = \frac{2^{\frac{1}{2}m_1 - k}(\frac{1}{2}m_1 - k)!}{(m_1 - 2k)!} \left(\sum_{\mathcal{C}' \equiv_k \mathcal{C}} X_1(\mathcal{C}') - \frac{1}{m_1 - 2k + 1} \sum_{\mathcal{C}'' \equiv_{k-1} \mathcal{C}} X_1(\mathcal{C}'') \right).$$

Now that we have an expression for Z_k, we can use the method of switching (see, for example, [132]). Indeed, given a $\mathcal{C}' \equiv_k \mathcal{C}$, we can define a quantity $\mathcal{C}'_\eta$ as follows. Given the first k pairings of $\mathcal{C}$, let I be the set of points already paired, and let $\{\alpha, \beta\}$ be the k-th pair. Put $\eta \notin I - \{\beta\}$ and $\{\eta, \gamma\} \in \mathcal{C}'$. Then $\mathcal{C}'_\eta$ is the pairing obtained from $\mathcal{C}'$ by mapping

$$\{\alpha, \beta\}, \{\eta, \gamma\} \to \{\alpha, \eta\}, \{\gamma, \beta\}.$$

Is easy to see that $\mathcal{C}'_\eta \equiv_{k-1} \mathcal{C}$ and that $\{\{\mathcal{C}'_\eta \colon \eta \notin I - \{\beta\}\} \mid \mathcal{C}' \equiv_k \mathcal{C}\}$ is a partition of $\{\mathcal{C}'' \mid \mathcal{C}'' \equiv_{k-1} \mathcal{C}\}$. We can therefore rewrite

$$\begin{aligned} Z_k(\mathcal{C}) &= \frac{2^{\frac{1}{2}m_1-k}(\frac{1}{2}m_1-k)!}{(m_1-2k)!} \sum_{\mathcal{C}' \equiv_k \mathcal{C}} \sum_{\eta \notin I} (X_1(\mathcal{C}') - X_1(\mathcal{C}'_\eta)) \\ &= \sum_{\eta \notin I} \sum_{\gamma \notin I, \gamma \neq \eta} \frac{q(\{\alpha,\beta\}) + q(\{\eta,\gamma\}) - q(\{\alpha,\eta\}) - q(\{\gamma,\beta\})}{(2m-2k+1)(2m-2k-1)}. \end{aligned}$$

Because $\sum_i x_u^2 \leq 1$, there are at most $\varepsilon^2 n$ indices of x such that $|x_i| > 1/(\varepsilon\sqrt{n})$. By the definition of X_1, the lexicographic ordering of $\{\alpha_i, \beta_i\}$ and the ordering of $|x_j| > |x_{j+1}|$, there exists a $\tilde{k}$ such that $Z_{\tilde{k}} = 0$. Take $\tilde{k} = \varepsilon^2 m_\infty n$. For $k > \tilde{k}$, we have that

$$m_1 - 2k - 1 \geq m_1 - 2\varepsilon^2 m_\infty n - 1 \geq m_0 n,$$

where the free parameter ε has to be fixed such that the last inequality holds. (Note that, because there exists a constant θ such that $\frac{m_\infty}{m_0} < \theta$, we can always choose an ε small enough so that this holds). Hence we can bound

$$|Z_k(\mathcal{C})| \leq \frac{1}{(m_0 n)^2} \sum_{\eta \notin I} \sum_{\gamma \notin I, \gamma \neq \eta} \left(|q(\{\alpha,\beta\})| + |q(\{\eta,\gamma\})| + |q(\{\alpha,\eta\})| + |q(\{\gamma,\beta\})|\right).$$

iii. Define

$$y^\alpha = \frac{1}{|x_{v(\alpha)}|} \min\left\{|y x_{v(\alpha)}|, \frac{\sqrt{m_\infty}}{n}\right\}.$$

Because $\alpha < \beta$, we can bound

$$q(\{\alpha,\beta\}) = \frac{\hat{\chi}(x_{v(\alpha)}, y_{v(\beta)})}{d_{v(\alpha)} d_{v(\beta)}} \leq \frac{y^\alpha_{v_\beta} |x_{v(\alpha)}|}{m_0^2}.$$

A similar bound holds for $\{\alpha,\eta\}$ for the same reason. For the other two edges, $\{\eta,\gamma\}$ and $\{\gamma,\beta\}$, we need to upper bound with a symmetric term, because we do not know whether $\gamma > \eta$ or $\gamma < \eta$. Thus, we have the upper bound

$$q(\{\eta,\gamma\}) \leq \frac{1}{m_0^2}\left(y^\gamma_{v_\eta} |x_{v(\gamma)}| + y^\eta_{v_\gamma} |x_{v(\eta)}|\right)$$

(the same bound holds for $\{\gamma, \beta\}$). Moreover, by the lexicographic order, $x_{v(\alpha)}$ bounds all the other components, and therefore $y^{\alpha}_{v_\beta}|x_{v(\alpha)} \leq y^{\alpha}_{v_\beta}|x_{v(\alpha)}$. Now note that $\sum_i |y_i| \leq \sqrt{n}$ (because $\sum_i y_i^2 \leq 1$), and so

$$\sum_{\eta \notin I} y^{\alpha}_{v(\eta)} \leq \sum_{\eta \notin I} |y_{v(\eta)}| \leq m_\infty \sum_i |y_i| \leq m_\infty \sqrt{n}.$$

By the previous considerations, substituting into the expression for $Z(\mathcal{C})$, we have

$$\begin{aligned} |Z_k(\mathcal{C})| &\leq \frac{1}{m_0^2}\left(y^{\alpha}_{v_\beta}|x_{v(\alpha)}| + \left(y^{\alpha}_{v_\eta}|x_{v(\alpha)}| + y^{\alpha}_{v_\gamma}|x_{v(\alpha)}|\right) + y^{\alpha}_{v_\eta}|x_{v(\alpha)}| + \left(y^{\alpha}v_\beta|x_{v(\alpha)}| + y^{\alpha}_{v}\right.\right. \\ &\leq \frac{4m_\infty^2}{m_0^4}|x_{v(\alpha)}|\left(y^{\alpha}_{v(\beta)} + \frac{1}{\sqrt{n}}\right). \end{aligned}$$

iv. From the above bounds we are able to obtain an upper bound for $\mathbb{E}[\exp(\zeta^2 Z_k^2) \mid \mathcal{F}_{k-1}]$ and then use lemma 4.3.6. Indeed,

$$\mathbb{E}\left[\mathrm{e}^{\zeta^2 Z_k^2} \mid \mathcal{F}_{k-1}\right] \leq \frac{1}{m_1 - 2k - 1} \sum_{\omega \notin I \setminus \beta} \exp\left[16\zeta^2 m_\infty^4 m_0^{-8}(x_{v(\alpha)})^2\right]\left(y^{\alpha}_{v(\omega)} + \frac{1}{\sqrt{n}}\right)^2.$$

Looking at (4.3.8), we see that we have to fix $\zeta = m_\infty^{3/2} n$ in order to achieve the required bound. Using that $x_{v(\alpha)} \leq \sqrt{m_\infty}/(\epsilon\sqrt{n})$ (because $x_{v(\alpha)} y^{\alpha}_{v(\beta)} \leq \sqrt{m_\infty} n$ and $y^{\alpha}_{v(\beta)} \geq \varepsilon/\sqrt{n}$), the exponent in the previous display is bounded by $64\theta^8\varepsilon^{-2}$. Using that $\mathrm{e}^x \leq 1 + x\mathrm{e}^x$, $x \geq 0$, and putting $B = 64\theta^8\epsilon^{-2}$, we have

$$\begin{aligned} \mathbb{E}\left[e^{\zeta^2 Z_k^2} | \mathcal{F}_{k-1}\right] &\leq 1 + \frac{B}{m_1 - 2k - 1} \sum_{\omega \notin I \setminus \beta} \zeta^2 m_\infty^4 m_0^{-8}(x_{v(\alpha)})^2 \left(y^{\alpha}_{v(\omega)} + \frac{1}{\sqrt{n}}\right)^2 \\ &\leq 1 + B\zeta^2 m_\infty^4 m_0^{-9}(x_{v(\alpha)})^2 \sum_\omega \left(y^2_{v(\omega)} + 2y_{v(\omega)}\frac{1}{\sqrt{n}} + \frac{1}{n}\right) \\ &\leq \exp\left[4B\frac{\zeta^2 m_\infty^5}{m_0^9 n}(x_{v(\alpha)})^2\right]. \end{aligned}$$

Next, let us pick an index $i(k)$ such that, for all $\mathcal{C} \in \Omega$,

$$\mathbb{E}\left[e^{\zeta^2 Z_k^2} \mid \mathcal{F}_{k-1}\right] \leq \exp\left[4B\frac{\zeta^2 m_\infty^5}{m_0^9 n}(x_{i(k)})^2\right].$$

One possible choice is to take $i(k) = \lceil k/m_\infty \rceil$. We finally get

$$\mathbb{P}\left(|X_1 - \mathbb{E}[X_1]| \geq \frac{t}{m_\infty^{3/2}} \right) \leq 2\,\mathrm{e}^{-\frac{\zeta t}{m_\infty^{3/2}}} \mathrm{e}^{\sum_{k=1}^{\frac{1}{2}m_1} 4B \frac{\zeta^2 m_\infty^5}{m_0^9 n} (x_{i(k)})^2} \leq 2\,\mathrm{e}^{-tn + 4B\frac{m_\infty^9}{m_0^9} n},$$

which proves what was said at the beginning of the proof (4.3.9) for $\ell = 1, 2$.

v. Finally, consider X_3. In view of the bounds in $\mathcal{C}_3$, this case can be dealt with via classical martingale arguments (see, for example, the McDiarmid inequality and its generalizations in [31]). Considering the variables $Y_k = \mathbb{E}[X_3 \mid \mathcal{C}_3]$, we have that $|Y_k - Y_{k-1}| \leq 4/(\varepsilon^2 n m_0^2)$. Thus, given our choice of $\varepsilon = \varepsilon(\theta)$ being constant, we have

$$\mathbb{P}\left(|X_3 - \mathbb{E}[X_3]| \geq t m_\infty^{-3/2} \right) \leq 2\,\mathrm{e}^{-\frac{t^2 \varepsilon^4 n^2 m_0^4}{16 m_\infty^3 m_1}} \leq 2\,\mathrm{e}^{-C(\varepsilon,\theta) t^2 n},$$

where $C(\varepsilon, \theta) > 0$.

4. Combining the results for $\ell = 1, 2, 3$ in the above lemma, we get an exponential bound on X'_a of the type

$$\mathbb{P}\left(|X'_a - \mathbb{E}[X'_a]| \geq \frac{t}{m_\infty^{3/2}} \right) \leq 2\mathrm{e}^{-tn + B_a n},$$

where B_a is a suitable constant, possibly different from any of the B_ℓ. By symmetry, the same holds for X'_b, which proves the concentration around $m_\infty^{-3/2}$ of X'.

Remark 4.3.9. Up to now we have worked with multi-graphs, so we have to pass to simple graphs with a prescribed degree sequence. Is easy to see that, because of (4.2.3), we can express the final result as saying that, conditional on the event that the graph is simple, there exist two constants $\hat{\xi} < \xi$ in $(0, 1)$ and a constant $K(\theta, \xi) > 0$ such that

$$\mathbb{P}\left(|X' - \mathbb{E}[X']| \geq K m_\infty^{-3/2} \right) \leq \mathrm{e}^{O(\hat{d}^2)} \hat{\xi}^n \leq \xi^n.$$

♠

4.3.3 Estimate of second contribution

It remains to show that the pairs with $x, y \notin \mathcal{B}$ give a bounded contribution of the order $O(m_\infty^{-3/2})$ with sufficiently high probability. For this it suffices to show that a simple random graph G with a prescribed degree sequence $\vec{d}$ cannot have too dense subgraphs (see [63, Lemma 2.5] and [34, Lemma 16]):

Lemma 4.3.10. *Let G be a simple random graph of size n drawn uniformly at random with a given degree sequence $\vec{d}$. Let $A, B \subseteq [n]$ be two subsets of the vertex set, and let $e(A,B)$ be the set of edges $e = \{\alpha, \beta\}$ such that either $\alpha \in A, \beta \in B$ or $\alpha \in B, \beta \in A$. Since $\mu(A,B) = \theta|A||B|\frac{m_\infty}{n}$ with $\theta > m_\infty/m_0$ a sufficiently large constant, for any $K > 0$ there exist a constant $C = C(\theta, K)$ such that with probability $1 - o(n^{-K})$ any pair A, B with $|A| \le |B|$ satisfies at least one of the following:*

$$e(A,B) \le C\mu(A,B) \tag{4.3.10}$$

$$e(A,B)\log\left(\frac{e(A,B)}{\mu(A,B)}\right) \le C|B|\log\left(\frac{n}{|B|}\right). \tag{4.3.11}$$

The above lemma has the following corollary.

Corollary 4.3.11. *For all $x, y \in T$,*

$$X'' = O\left(\frac{1}{m_\infty^{3/2}}\right) \quad \text{with probability at least } 1 - O(n^{-K}),$$

where K is the constant in Lemma 4.3.10.

Proof. Fix $x, y \in T$ and define

$$S_i(x) = \left\{\ell\colon \frac{\varepsilon^{2-i}}{\sqrt{n}} \le |x_\ell| < \frac{\varepsilon^{1-i}}{\sqrt{n}}\right\}, \quad i \in I,$$

$$S_j(y) = \left\{\ell\colon \frac{\varepsilon^{2-i}}{\sqrt{n}} \le |y_\ell| < \frac{\varepsilon^{1-i}}{\sqrt{n}}\right\}, \quad j \in J,$$

where $I = \{i | S_i(x) \neq \emptyset\}$ (J is defined similarly). Then, for $x \in T$, write

$$x_u|_S = \begin{cases} x_u, & \text{if } u \in S, \\ 0, & \text{otherwise.} \end{cases}$$

In order to apply Lemma 4.3.10, define $A_i = S_i(x)$ and $B_j = S_j(y)$, and let a_i and b_i be their cardinality, respectively. Divide the set of indices into two groups:

$$\mathcal{E} = \big\{(i,j)\colon\ i,j > 0, \varepsilon^{2-i-j} > \sqrt{m_\infty}, a_i \le b_j\big\},$$
$$\mathcal{E}' = \big\{(i,j)\colon\ i,j > 0, \varepsilon^{2-i-j} > \sqrt{m_\infty}, a_i > b_j\big\}.$$

By the definition of X'' and the set $\mathcal{B}$, we have

$$X'' = \sum_{x_i y_j > \sqrt{m_\infty}/n} x_i A_{ij} y_j = \sum_{(i,j)\in\mathcal{E}} (x|_{A_i})^t A y|_{B_j} + \sum_{(i,j)\in\mathcal{E}'} (x|_{A_i})^t A y|_{B_j}.$$

It suffices to show that either of the contributions coming from $\mathcal{E}$ or $\mathcal{E}'$ are $O(m_\infty^{-3/2})$ (the other will follow by symmetry). Focus on $\mathcal{E}$. Putting $e_{i,j} = e(A_i, B_j)$ and $\mu_{i,j} = \mu(A_i, B_j)$, we see that the bound can be rewritten as

$$\frac{1}{n} \sum_{(i,j)\in\mathcal{E}} \frac{e_{i,j}}{\varepsilon^{i+j}} = O(\sqrt{m_\infty}).$$

Divide $\mathcal{E}$ into the union of $\mathcal{E}_a$ and $\mathcal{E}_b$, where $\mathcal{E}_a$ satisfies (4.3.10) and $\mathcal{E}_b$ satisfies (4.3.11). Clearly $\mathcal{E} = \mathcal{E}_a \cup \mathcal{E}_b$ and

$$\frac{1}{n} \sum_{(i,j)\in\mathcal{E}} \frac{e_{i,j}}{\varepsilon^{i+j}} \le \frac{1}{n} \sum_{(i,j)\in\mathcal{E}_a} \frac{e_{i,j}}{\varepsilon^{i+j}} + \frac{1}{n} \sum_{(i,j)\in\mathcal{E}_b} \frac{e_{i,j}}{\varepsilon^{i+j}}.$$

If we are able to show that both contributions from $\mathcal{E}_a$ and $\mathcal{E}_b$ are $O(\sqrt{m_\infty})$, then the theorem follows. It is easy to see that $\mathcal{E}_a$ gives a bounded contribution. Indeed,

$$\frac{1}{n} \sum_{(i,j)\in\mathcal{E}_a} \frac{e_{i,j}}{\varepsilon^{i+j}} \le \frac{1}{n^2} \sum_{(i,j)\in\mathcal{E}_a} \frac{C a_i b_j \theta m_\infty}{\varepsilon^{i+j}} \le \frac{C' m_\infty}{n^2} \sum_{(i,j)\in\mathcal{E}_a} \frac{a_i b_j}{\varepsilon^{2(i+j)}\sqrt{m_\infty}} = O(\sqrt{m_\infty}),$$

where in the last step we use that, because $\sum_i x_i^2 \le 1$,

$$\sum_{i\in I} \frac{a_i}{\varepsilon^{2(i-2)}} \le n, \qquad \sum_{j\in J} \frac{b_j}{\varepsilon^{2(i-2)}} \le n.$$

It remains to show that

$$\frac{1}{n} \sum_{(i,j)\in\mathcal{E}_b} \frac{e_{i,j}}{\varepsilon^{i+j}} = O(\sqrt{m_\infty}).$$

In order to do so, we divide $\mathcal{E}_b$ into subsets $\mathcal{E}_b^{(\ell)}$, $\ell = 1, \ldots, 5$, having the following properties:

$$
\begin{aligned}
&(1) \quad \varepsilon^j \geq \varepsilon^i \sqrt{m_\infty}. \\
&(2) \quad e_{i,j} \leq \frac{\mu_{i,j}}{\varepsilon^{i+j}\sqrt{m_\infty}}. \\
&(3) \quad \log\left(\frac{e_{i,j}}{\mu_{i,j}}\right) \geq \frac{1}{4}\log\left(\frac{n}{b_j}\right). \\
&(4) \quad \frac{n}{b_j} \leq \mathrm{e}^{-4j}. \\
&(5) \quad \frac{n}{b_j} > \mathrm{e}^{-4j}.
\end{aligned}
$$

For $j > i$ we have that $\mathcal{E}_b^{(\ell)} \nsubseteq \mathcal{E}_b^{(\ell)}$ and $\mathcal{E}_b = \cup_\ell \mathcal{E}_b^{(\ell)}$. Thus it suffices to show a bound of $O(\sqrt{n})$ for each of the quantities $S_\ell = 1/n \sum_{\mathcal{E}_b^{(\ell)}} e_{i,j}/\varepsilon^{i+j}$.

• For $\ell = 1$ we note that, since $e_{i,j} \leq a_i m_\infty$,

$$
S_1 \leq \frac{1}{n} \sum_i \sum_{j|\varepsilon^j \geq \varepsilon^i \sqrt{m_\infty}} \frac{a_i m_\infty}{\varepsilon^{i+j}} = bO\left(\frac{1}{n}\sum_i \frac{a_i\sqrt{m_\infty}}{\varepsilon^{2i}}\right) = O(\sqrt{m_\infty}).
$$

• For $\ell = 2$ we obtain

$$
S_2 \leq \frac{1}{n}\sum_{ij} \frac{\mu_{ij}}{\varepsilon^{2(i+j)}\sqrt{m_\infty}} = O(\frac{\sqrt{m_\infty}}{n^2}\sum_{ij}\frac{a_i b_j}{\varepsilon^{2(i+j)}}) = O(\sqrt{n}).
$$

• For $\ell = 3$, because the pairs $(i,j) \in \mathcal{E}_b$ have property (4.3.11), it follows easily that $e_{ij} = O(b_j)$. Furthermore, because $\mathcal{E}_b^{(3)} \nsubseteq \mathcal{E}_b^{(1)}$, we have that $\forall (i,j) \in \mathcal{E}_b^{(3)}$, $e^j < e_i\sqrt{m_\infty}$. It follows that

$$
S_3 = O\left(\frac{1}{n}\sum_j \sum_{i|\varepsilon^i > e^j/\sqrt{m_\infty}} \frac{b_j}{\varepsilon^{i+j}}\right) = O\left(\frac{1}{n}\sum_j \frac{\sqrt{m_\infty}\, b_j}{\varepsilon^{2j}}\right) = O(\sqrt{m_\infty}).
$$

• For $\ell = 4$ we take advantage of the fact that $(i,j) \in \mathcal{E}_b^{(4)}$ do not belong to $\mathcal{E}_b^{(3)}$ and $\mathcal{E}_b^{(2)}$. This implies that

$$
\frac{e_{ij}}{m_{ij}} \leq \frac{1}{e^j} \qquad \frac{e_{ij}}{\mu_{ij}} \geq \frac{1}{\varepsilon^{i+j}\sqrt{m_\infty}}.
$$

Hence we have $\varepsilon^{-i} \le \sqrt{m_\infty}$ and, by (4.3.11), also $e_{ij} = O(jb_j)$. We can therefore conclude that

$$S_4 = O\left(\frac{1}{n}\sum_j \sum_{i|\varepsilon^{-i}\le\sqrt{m_\infty}} \frac{jb_j}{\varepsilon^{i+j}}\right) = O\left(\frac{\sqrt{m_\infty}}{n}\sum_j \frac{jb_j}{\varepsilon^j}\right) = O(\sqrt{m_\infty}),$$

where in the last equality we use that $\sum_{j\in J} b_j/(n\varepsilon^2) = O(1)$.

• For $\ell = 5$, using the property in (5) and (4.3.11), we have

$$e_{ij} \le Cn\varepsilon^{4j}\log\varepsilon^{-4j} = O(nj\varepsilon^{4j}).$$

Also, using that $\mathcal{E}_b^{(4)} \not\subseteq \mathcal{E}_b^{(4)}$, we have $\varepsilon^j < \varepsilon^i\sqrt{m_\infty}$, from which we conclude that

$$S_5 = O\left(\sum_j \sum_{i|\varepsilon^i>\varepsilon^j/\sqrt{m_\infty}} j\varepsilon^{3j-i}\right) = O\left(\sqrt{m_\infty}\sum_j j\varepsilon^{2j}\right) = O(\sqrt{m_\infty}).$$

This completes the proof.

4.4 Proof of the main theorem

Expansion. Throughout this section we abbreviate $A = A_{G_n}$ and condition on the event that G_n is simple. Recall Lemma 4.3.3. Compute

$$\begin{aligned}
&Av_1 = \lambda_1 v_1,\\
&(H + |\tilde{e}\rangle\langle\tilde{e}| + (\mathbb{E}[A] - |\tilde{e}\rangle\langle\tilde{e}|))\, v_1 = \lambda_1 v_1,\\
&\left(H + |\tilde{e}\rangle\langle\tilde{e}| - \mathrm{diag}\left(\frac{d_1}{m_1-1},\dots,\frac{d_n}{m_1-1}\right)\right) v_1 = \lambda_1 v_1.
\end{aligned}$$

Rewriting the equation we have,

$$\langle\tilde{e}, v_1\rangle\tilde{e} = \left(\lambda_1\mathbb{I} + \mathrm{diag}\left(\frac{d_1}{m_1-1},\dots,\frac{d_n}{m_1-1}\right) - H\right) v_1.$$

Therefore componentwise we have the following inequality,

$$\left(\lambda_1 + \frac{m_0}{m_1-1}\right)\left(\mathbb{I} - \frac{H}{\lambda_1 + \frac{m_0}{m_1-1}}\right) v_1 \le \langle\tilde{e}, v_1\rangle\tilde{e} \le \left(\lambda_1 + \frac{m_\infty}{m_1-1}\right)\left(\mathbb{I} - \frac{H}{\lambda_1 + \frac{m_\infty}{m_1-1}}\right)$$

If x, y and e are non-negative vector (the non-negativity of v_1 follows from Perron-Frobenius theory) with $x \geq y$, then $\langle e, x\rangle \geq \langle e, y\rangle$, we can use Corollary 4.3.2 and invert the matrix multiplying v_1. Indeed, given that $\|H\| = O(\sqrt{m_\infty})$ and $\lambda_1 \sim m_2/m_1$ on the event with probability $1 - o(n^{-K})$ of Corollary 4.3.2 and Theorem 4.3.1, we can invert and expand

$$\left(\mathbb{I} - \frac{H}{\lambda_1 + \frac{m_0}{m_1 - 1}}\right)^{-1} = \sum_{k \in \mathbb{N}_0} \frac{H^k}{(\lambda_1 + \frac{m_0}{m_1 - 1})^k},$$

and similarly for m_∞. Thus,

$$\lambda_1 \leq \frac{m_2}{m_1 - 1} - \frac{m_0}{m_1 - 1} + \sum_{k \in \mathbb{N}} \frac{\langle \tilde{e}, H^k \tilde{e}\rangle}{(\lambda_1 + \frac{m_0}{m_1 - 1})^k}$$
$$\lambda_1 \geq \frac{m_2}{m_1 - 1} - \frac{m_\infty}{m_1 - 1} + \sum_{k \in \mathbb{N}} \frac{\langle \tilde{e}, H^k \tilde{e}\rangle}{(\lambda_1 + \frac{m_\infty}{m_1 - 1})^k}.$$

Our final goal is to determine the expectation $\mathbb{E}[\lambda_1]$, which splits as

$$\mathbb{E}[\lambda_1] = \mathbb{E}[\lambda_1 | \mathcal{E}]\, \mathbb{P}(\mathcal{E}) + \mathbb{E}[\lambda_1 | \mathcal{E}^c]\, \mathbb{P}(\mathcal{E}^c). \tag{4.4.1}$$

The event $\mathcal{E}^c$ has probability at most n^{-K}, where K is a large arbitrary constant. Thus, given the deterministic bound $\lambda_1 \leq n$, we may focus on $\mathbb{E}[\lambda_1 | \mathcal{E}]\, \mathbb{P}(\mathcal{E})$. In order to do this, we need to be able to handle terms of the type

$$\frac{m_2}{m_1 - 1} + \frac{\mathbb{E}[\langle \tilde{e}, H\tilde{e}\rangle]}{\frac{m_2}{m_1 - 1}(1 + o(1))} + \frac{\mathbb{E}[\langle \tilde{e}, H^2\tilde{e}\rangle]}{\left(\frac{m_2}{m_1 - 1}\right)^2 (1 + o(1))} + \sum_{k \in \mathbb{N} \setminus \{1,2\}} \frac{\mathbb{E}[\langle \tilde{e}, H^k \tilde{e}\rangle]}{\left(\frac{m_2}{m_1 - 1}\right)^k (1 + o(1))}$$

Since

$$\frac{\mathbb{E}[\langle \tilde{e}, H^k \tilde{e}\rangle]}{\left(\frac{m_2}{m_1 - 1}\right)^k (1 + o(1))} \leq \frac{m_\infty^{k/2}}{\left(\frac{m_2}{m_1 - 1}\right)^{k-1}} = o\left(\frac{1}{m_0^{k/2 - 1}}\right),$$

the last sum is $o(1/\sqrt{m_\infty})$, which is an error term. It therefore remains to study $\langle \tilde{e}, H^k \tilde{e}\rangle$, $k = 1, 2$. The study of these moments for the configuration model is more involved than for random regular graphs.

Case $k = 1$**.** Compute

$$\begin{aligned}\langle \tilde{e}, H\tilde{e}\rangle &= \langle \tilde{e}, (A - \mathbb{E}[A])\,\tilde{e}\rangle \\ &= \frac{1}{m_1 - 1}\left(\sum_{ij} d_i d_j a_{ij} - \frac{1}{m_1 - 1}\sum_{ij} d_i^2 d_j^2\right) = \frac{\sum_j d_j(\sum_{i\sim j} d_i)}{m_1 - 1} - \frac{m_2^2}{(m_1-1)^2}.\end{aligned}$$

Since

$$\mathbb{E}\left[\sum_j d_j \sum_{i\sim j} d_i\right] = \sum_{ij} d_j d_i \mathbb{E} a_{ij} = \frac{m_2^2}{m_1 - 1} - \frac{m_3}{m_1 - 1},$$

where the last term comes from the presence of selfloops. It follows that $\mathbb{E}[\langle \tilde{e}, H^k \tilde{e}\rangle] = O(1/\sqrt{n})$.

Case $k = 2$**.** Compute

$$\begin{aligned}\langle \tilde{e}, H^2\tilde{e}\rangle &= \langle \tilde{e}, (A - \mathbb{E}[A])^2\,\tilde{e}\rangle \\ &= \frac{1}{m_1 - 1}\Bigg(\sum_{ijk} d_i d_k a_{ij} a_{jk} - \frac{1}{m_1 - 1}\sum_{ijk} d_i d_k^2 a_{ij} d_j \\ &\quad - \frac{1}{m_1 - 1}\sum_{ijk} d_i^2 d_j d_k a_{jk} + \frac{1}{(m_1-1)^2}\sum_{ijk} d_i^2 d_j^2 d_k^2\Bigg).\end{aligned}$$

Write

$$\frac{1}{m_1 - 1}\sum_{ijk} d_i d_k^2 a_{ij} d_j = \frac{m_2}{m_1 - 1}\sum_{ijk} d_i a_{ij} d_j = \frac{m_2}{m_1 - 1}\sum_i d_i \left(\sum_{i\sim j} d_j\right)$$

(by symmetry the third term is equal) and

$$\sum_{ijk} d_i d_k a_{ij} a_{jk} = \sum_j \left(\sum_{i\sim j} d_i\right)^2 = \sum_j \sum_{i\sim j} d_i^2 + \sum_k \sum_{\substack{i,j\sim k \\ i\neq j}} d_i d_j = m_3 + \sum_k \sum_{\substack{i,j\sim k \\ i\neq j}} d_i d_j.$$

Indeed in $\sum_j \sum_{i\sim j} d_i^2$, the summand d_i^2 appears exactly d_i times, because the node i has exactly d_i neighbours, and so $\sum_j \sum_{i\sim j} d_i^2 = m_3$. Putting

the terms together, we get

$$\langle \tilde{e}, H^2 \tilde{e} \rangle = \frac{1}{m_1 - 1} \left(m_3 + \sum_k \sum_{\substack{i,j \sim k \\ i \neq j}} d_i d_j - 2 \frac{m_2}{m_1 - 1} \sum_i d_i \left(\sum_{i \sim j} d_j \right) + \frac{m_2^3}{(m_1 - 1)^2} \right).$$

Taking expectations, we get

$$\mathbb{E}[\langle \tilde{e}, H^2 \tilde{e} \rangle] = \frac{1}{m_1 - 1} \left(m_3 + \mathbb{E} \left[\sum_k \sum_{\substack{i,j \sim k \\ i \neq j}} d_i d_j - \frac{m_2^3}{(m_1 - 1)^2} \right] \right).$$

Note that $\mathbb{E}[\sum_k \sum_{i,j \sim k,\, i \neq j} d_i d_j]$ is a sum over the wedges centered at vertex k, summed all k. We can swap the summation over pairs of vertices, and choose a third neighbour to form a wedge, which gives

$$\sum_k \sum_{\substack{i,j \sim k \\ i \neq j}} d_i d_j = \sum_{\substack{i,j \\ i \neq j}} d_i d_j \sum_k \mathbb{1}_{(k \sim i, k \sim j)}.$$

Compute

$$\mathbb{E}\left[\sum_{\substack{i,j\\ i\neq j}} d_i d_j \sum_k \mathbb{1}_{(k\sim i, k\sim j)}\right] = \sum_{\substack{i,j\\ i\neq j}} d_i d_j \sum_k \mathbb{E}\left[\mathbb{1}_{(k\sim i, k\sim j)}\right]$$

$$= \sum_{\substack{i,j\\ i\neq j}} d_i d_j \sum_k \left(\frac{d_i d_j d_k (d_k-1)}{(m_1-1)(m_1-2)} \mathbb{1}_{k\neq i, k\neq j} + \frac{d_i d_j (d_k-1)(d_k-2)}{(m_1-1)(m_1-2)} \mathbb{1}_{k=i \text{ or } k=j}\right)$$

$$= \sum_{\substack{i,j\\ i\neq j}} d_i d_j \sum_k \left(\frac{d_i d_j d_k (d_k-1)}{(m_1-1)(m_1-2)} - \frac{2 d_i d_j (d_k-1)}{(m_1-1)(m_1-2)} \mathbb{1}_{k=i \text{ or } k=j}\right)$$

$$= \frac{1}{(m_1-1)(m_1-2)} \sum_{\substack{i,j\\ i\neq j}} d_i^2 d_j^2 \sum_k \left(d_k(d_k-1) - 2(d_k-1)\mathbb{1}_{k=i \text{ or } k=j}\right)$$

$$= \frac{1}{(m_1-1)(m_1-2)} \left((m_2-m_1)\left(\sum_{i,j} d_i^2 d_j^2 - \sum_i d_i^4\right) - 2\sum_{\substack{i,j\\ i\neq j}} d_i^2 d_j^3 - 2\sum_{\substack{i,j\\ i\neq j}} d_i^3 d_j^2 + 4\sum_{\substack{i,j\\ i\neq j}} d_i^2 d_j^2\right.$$

$$= \frac{1}{(m_1-1)(m_1-2)} \left((m_2-m_1)\left(m_2^2-m_4\right) - 4\left(\sum_{i,j} d_i^2 d_j^3 - \sum_i d_i^5\right) + 4\left(\sum_{i,j} d_i^2 d_j^2 - \sum_i d_i^4\right)\right)$$

$$= \frac{1}{(m_1-1)(m_1-2)} \left((m_2-m_1)\left(m_2^2-m_4\right) - 4\left(m_3 m_2 - m_5\right) + 4\left(m_2^2-m_4\right)\right)$$

$$= \frac{m_2^3 - m_2 m_4 - m_1 m_2^2 + m_1 m_4 - 4m_3 m_2 + 4m_5 + 4m_2^2 - 4m_4}{(m_1-1)(m_1-2)}.$$

Hence

$$\mathbb{E}[\langle \tilde{e}, H^2 \tilde{e}\rangle]$$

$$= \frac{1}{m_1-1}\left(m_3 + \frac{m_2^3 - m_2 m_4 - m_1 m_2^2 + m_1 m_4 - 4m_3 m_2 + 4m_5 + 4m_2^2 - 4m_4}{(m_1-1)(m_1-2)} - \frac{m_2^3}{(m_1-1)}\right.$$

Given the event $\mathcal{E}$, by (4.4.1) and Corollary 4.3.2, we have that $\mathbb{E}[\lambda_1]$ concentrates around m_2/m_1 with an $O(\sqrt{m_\infty})$ error, and so we can write

$$\mathbb{E}\left[\frac{\langle \tilde{e}, H^2 \tilde{e}\rangle}{\lambda_1^2}\right] = \frac{\mathbb{E}\left[\langle \tilde{e}, H^2 \tilde{e}\rangle\right]}{(\frac{m_2}{m_1-1})^2}(1+o(1)) + o(1).$$

We run through the various contributions separately (using Assumption 4.1.1). Noting that $nm_0^k \leq m_k \leq nm_\infty^k$ and that. there are positive

constants c, C such that $c \leq \frac{m_\infty}{m_0} \leq C$, we have

$$\frac{m_2^3}{m_2^2(m_1-1)(m_1-2)} = \Theta\left(\frac{1}{n}\right),$$
$$\frac{m_2 m_4}{m_2^2(m_1-2)} = o\left(\frac{1}{\sqrt{n}}\right), \quad \frac{m_1 m_4}{m_2^2(m_1-2)} = \Theta\left(\frac{1}{n}\right), \quad \frac{4m_3 m_2}{m_2^2(m_1-2)} = o\left(\frac{1}{n}\right),$$
$$\frac{4m_5}{m_2^2(m_1-2)} = o\left(\frac{1}{n^{3/2}}\right), \quad \frac{4m_2^2}{m_2^2(m_1-2)} = o\left(\frac{1}{n}\right), \quad \frac{4m_4}{m_2^2(m_1-2)} = o\left(\frac{1}{n^2}\right).$$

Therefore

$$\frac{\mathbb{E}\left[\langle \tilde{e}, H^2 \tilde{e}\rangle\right]}{(m_2/(m_1-1))^2} = \frac{m_3 m_1}{m_2^2} - 1 + o(\frac{1}{\sqrt{n}}),$$

which settles (4.1.1).

Weak law of large numbers. We want to show that

$$\frac{\lambda_1}{\mathbb{E}\lambda_1} \to 1$$

in $\mathbb{P}$-probability. Using Corollary 4.3.2 and the Weyl interlacing inequality, we have that, with probability $1 - n^{-K}$ for $K > 0$,

$$\frac{m_2}{m_1} - O\left(\sqrt{m_\infty}\right) \leq \lambda_1 \leq \frac{m_2}{m_1} + O(\sqrt{m_\infty}).$$

By (4.1.1),

$$\frac{\frac{m_2}{m_1} - O(\sqrt{m_\infty})}{\frac{m_2}{m_1}(1+o(1))} \leq \frac{\lambda_1}{\mathbb{E}[\lambda_1]} \leq \frac{\frac{m_2}{m_1} + O(\sqrt{m_\infty})}{\frac{m_2}{m_1}(1+o(1))}.$$

It follows that

$$1 - O\left(\frac{1}{\sqrt{m_\infty}}\right)) \leq \frac{\lambda_1}{\mathbb{E}[\lambda_1]} \leq 1 + O\left(\frac{1}{\sqrt{m_\infty}}\right)$$

with probability $1 - n^{-K}$, and hence the claim follows.

Chapter 5

Sampling random graph models

Abstract

In this Chapter we give a brief introduction to the problem of random graph sampling and we will show simulations that support our findings of the previous Chapters. Simulations were performed using the computational resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University.

In Chapter 1 we spoke about the strong influence the abundance of real-world data had on the flourishing of Network Science. The interplay between models and data validation caused Network Science to emerge as a powerful interdisciplinary field that studies the structure, dynamics and behavior of complex systems represented as networks. As we pointed out, these networks can range from social interactions and biological systems to technological infrastructures. Each network has peculiar features that need to be captured by mathematical models that aim to emulate reality.Typically, the size of the networks of interest is very large, and as a consequence there is no hope to fully reconstruct real-world network structures from the data. Indeed, with the large size of

the networks come many problems, such as the impossibility to gather all the data needed, the amount of time and costs that this would take, as well as accuracy and storage problems. Therefore, in the realm of Network Science, sampling random graphs from given distributions plays a crucial role, offering researchers a practical and efficient way to gain insight into large-scale networks after the main features (i.e. the ones that are easily accessible) have been incorporated. Once the model is chosen and is found to recreate the observed datas, it becomes an efficient *null model* that can be used to test whether new gathered data are consistent with it or require more sophisticated models.

As we already discussed, Network Science provides many versatile models that are able to capture many different features. Sampling-wise a difference needs to be made between sampling from distributions with soft constraints and from distributions with hard constraint. Ultimately, we will specialize our discussion to the type of constraints that we analyzed in this thesis, i.e., constraints on the degree sequence.

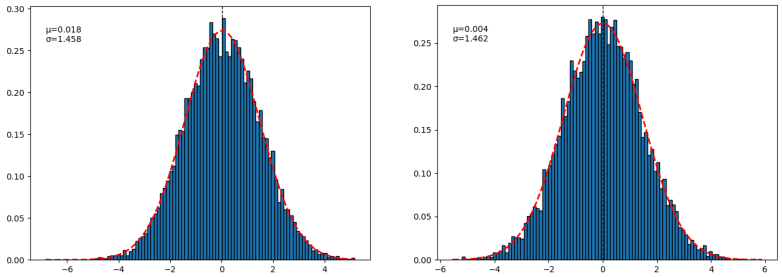
5.1 Sampling from the Canonical Ensemble

Following [112, 127], fast sampling of the canonical model with constraint $\vec{C}(g)$ can be obtained once the Shannon entropy maximization problem has been solved. Indeed, once the functional form of the $p_{ij}\left(\vec{\theta}\right)$ as in (1.6.1) is obtained, it is easy to calculate the value of the Lagrange multipliers $\vec{\theta}$ through maximum likelihood. The precise value of $\vec{\theta}$ needed to express (1.6.1) must be chosen in order to match with what has been measured from data. This is obtained by requiring that the logarithm of the probability of observing $\vec{C}^*$ given $\vec{\theta}$ is maximal, i.e.,

$$\max_{\vec{\theta}} \ln \mathbb{P}\left(\vec{C}(g) = \vec{C}^* \middle| \vec{\theta}\right).$$

This is possible only when the dyadic probabilities p_{ij} can be expressed in closed form from the entropy maximization. When this is not the case, other sampling procedures should be taken into account, most of them based on Monte Carlo approaches (for example, *Hamiltonian Monte Carlo* [24, 25]), or mean-field approaches (for example, the solution for the Strauss model in [98]). See [45] for more examples. Constraints on the degree sequence, i.e., the ones used in this thesis, allow for an explicit

form of (1.6.1), even in the directed case with reciprocity or weights. We will therefore use the methodology and the packages developed in [112].



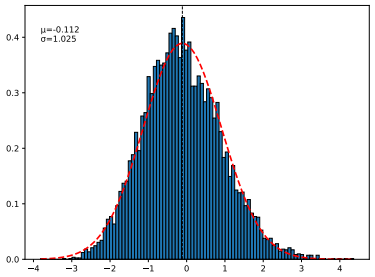
(a) $n = 500$, $\vec{d} = (16, 18, 19, 20, 21)$, $\sqrt{n} \approx 22.4$, $\log n \approx 6.2$.

(b) $n = 1000$, $\vec{d} = (20, 23, 25, 28, 30)$, $\sqrt{n} \approx 31.6$, $\log n \approx 6.9$.

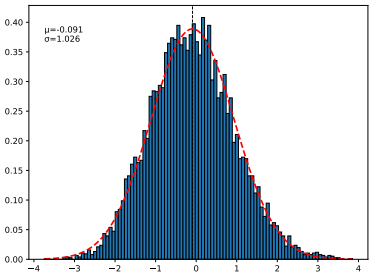
(c) $n = 2000$, $\vec{d} = (28, 33, 35, 38, 42)$, $\sqrt{n} \approx 44.7$, $\log n \approx 6.21$.

(d) $n = 5000$, $\vec{d} = (48, 53, 57, 62, 64)$, $\sqrt{n} \approx 70.7$, $\log n \approx 8.5$.

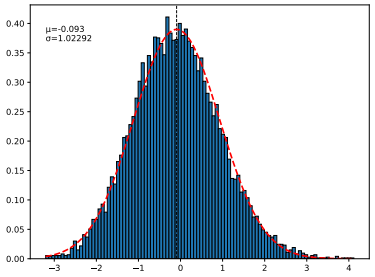
Figure 1: Histograms of $\bar{\lambda}_1$ for different graph sizes n and degree sequences $\vec{d}$. The sample size for each regime is 10^4. Each element specified in the degree sequence appears $\frac{n}{5}$ times. In red is plotted the Gaussian fit; μ is the sample mean (represented by a dashed line in the histogram), σ is the sample standard deviation. We expect $\mu \approx 0$ and $\sigma \approx \sqrt{2}$.



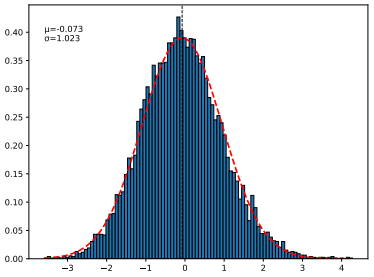
(a) $n = 500$,
$\vec{d} = (16, 18, 19, 20, 21)$,
$\sqrt{n} \approx 22.4$, $\log n \approx 6.2$.



(b) $n = 1000$,
$\vec{d} = (20, 23, 25, 28, 30)$,
$\sqrt{n} \approx 31.6$, $\log n \approx 6.9$.



(c) $n = 2000$,
$\vec{d} = (28, 33, 35, 38, 42)$,
$\sqrt{n} \approx 44.7$, $\log n \approx 6.21$.



(d) $n = 5000$,
$\vec{d} = (48, 53, 57, 62, 64)$,
$\sqrt{n} \approx 70.7$, $\log n \approx 8.5$.

Figure 2: Histograms of $\bar{v}_1(i)$ for different graph sizes n and degree sequences $\vec{d}$. For each of the images, i is chosen to be the last i such that d_i is equal to the 4$^{\text{th}}$ element of the corresponding degree sequence (e.g. for $n = 500$, $v_1(400)$ was analysed with $d_{400} = 20$). The sample size for each regime is 10^4. Each element in the degree sequence appears $\frac{n}{5}$ times. In red is plotted the Gaussian fit; μ is the sample mean (represented by a dashed line in the histogram), σ is the sample standard deviation. We expect $\mu \approx 0$ and $\sigma \approx 1$.

.

5.1.1 Simulations of results of Chapter 3: Largest eigenvalue.

Theorems 3.1.6–3.1.7 show that, after proper scaling and under certain conditions of sparsity and homogeneity, the largest eigenvalue and the components of the largest eigenvector exhibit Gaussian behaviour in the limit as $n \to \infty$. A natural question is how these quantities behave for finite n. Indeed, real-world networks have sizes that range from $n = 10^2$ to $n = 10^9$. Another question is computational feasibility. Indeed, our CLTs require the degrees to lie between $(\log n)^4$ (respectively, $(\log n)^8$) and $\sqrt{n}$. In order to make this possible, n must be at least 10^{11} (respectively, 10^{29}), which is unrealistic. Let us therefore see what simulations have to say[1].

In Figure 1 we show histograms for the quantity

$$\bar{\lambda}_1 = \frac{m_2}{m_1 \sigma_1}(\lambda_1 - \mathbb{E}[\lambda_1]),$$

which should be close to normal with mean 0 and variance 2 (for $\mathbb{E}[\lambda_1]$ the correction term $o(1)$ is neglected). The convergence is fast: already for $n = 500$ the Gaussian shape emerges and represents an excellent fit: the sample mean μ is close to 0 and the sample standard deviation σ is close to $\sqrt{2}$.

5.1.2 Largest eigenvector.

In Figure 2 we show histograms for the quantity

$$\bar{v}_1(i) = \frac{m_2^{3/2}}{m_1 s_1(i)} \left(v_1(i) - d_i/\sqrt{m_2} \right),$$

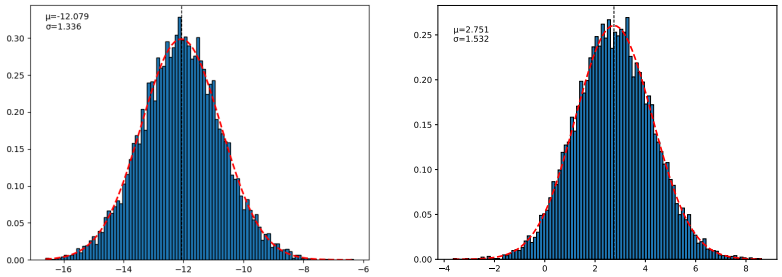
which should be close to normal with mean 0 and variance 1. The fit is again excellent.

5.1.3 Degrees of order $\log n$ and $\sqrt{n}$.

What happens when the degrees are of order $\log n$? As can be seen in Figure 3, in that range the Gaussian approximation for the largest eigen-

[1]Simulations were performed using the computational resources from the Academic Leiden Interdisciplinary Cluster Environment (ALICE) provided by Leiden University.

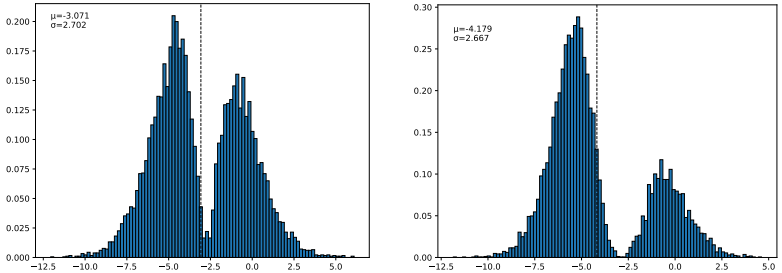
value is visibly worse, especially for the centering. The same happens for the components of the largest eigenvector, as can be seen in Figure 4, where the Gaussian shape is lost and two peaks appear.



(a) $n = 500$,
$\vec{d} = (4, 5, 8, 9)$,
$\sqrt{n} \approx 22.4$, $\log n \approx 6.2$.

(b) $n = 5000$,
$\vec{d} = (6, 7, 9, 10)$,
$\sqrt{n} \approx 70.7$, $\log n \approx 8.5$.

Figure 3: Histograms of $\bar{\lambda}_1$ for different graph sizes n and degree sequences $\vec{d}$ of order $\log n$. The sample size for each regime is 10^4. Each element specified in the degree sequence appears $\frac{n}{4}$ times. In red is plotted the Gaussian fit; μ is the sample mean (represented by a dashed line in the histogram), σ is the sample standard deviation. If Theorem 3.1.6 would hold, then we would expect $\mu \approx 0$ and $\sigma \approx \sqrt{2}$.



(a) $n = 500$, $\vec{d} = (4, 5, 8, 9)$, $\sqrt{n} \approx 22.4$, $\log n \approx 6.2$.

(b) $n = 5000$, $\vec{d} = (6, 7, 9, 10)$, $\sqrt{n} \approx 70.7$, $\log n \approx 8.5$.

Figure 4: Histograms of $\bar{v}_1(i)$ for different graph sizes n and degree sequences $\vec{d}$ of order $\log n$. For each of the images, i has been chosen to be the last i such that d_i is equal to the 3rd element of the specified degree sequence (e.g. for $n = 500$, $v_1(375)$ was analysed with $d_{375} = 8$). The sample size for each regime is 10^4. Each element specified in the degree sequence appears $\frac{n}{4}$; μ is the sample mean (represented by a dashed line in the histogram), σ is the sample standard deviation. If Theorem 3.1.7 would hold, then we would expect $\mu \approx 0$ and $\sigma \approx 1$.

5.2 Sampling from the Microcanonical Ensemble

Sampling from uniform distributions is known to be an hard problem. The main reason for this, in graph theory, is the difficulty in estimating the cardinality of the support (1.5.3) of the uniform distribution. Many approximate procedures have been developed over time to overcome this obstacle. In general, there is an interplay between biased sampling and complexity of the algorithm. While most of the procedures to sample with accuracy from $\Gamma_{\vec{C}^*}$ require an exponential complexity time, faster procedures rely on Monte Carlo approaches that suffer from two related types of problems: bias and ergodicity. The latter refers to the fact that, depending on the constraints on the dynamics of the Markov Chain, there can be configurations that are never visited by the MCMC. Biased sampling refers to the fact that our MCMC might sample certain graphs with higher probability, e.g. because of a lack of ergodicity or a high mix-

ing time for the Markov chain. This is usually solved, when possible, by introducing importance sampling. For the case when the constraint is on the degree sequence fast algorithms are available. These algorithm usually are divided into two steps: the first generates an *unbiased* seed which then is fed to the MCMC for the second step. For the case of constraints on the degree sequence the MCMC is shown to mix fast enough to make this method efficient. Usually the shortcoming of these approaches is the limitations on the density and inhomogeneity of our random graphs.

Because of its importance in graph theory, sampling graphs with a given degree sequence, i.e., when the constraint is on the degree sequence, has been studied since the 60s. A good reference is [73]. Three main approaches are used. One is the use of configuration model, which was shown in [83] to be efficient to generate simple graphs only when $m_2 = \mathrm{O}(\frac{m_1}{2}\sqrt{\log n})$ and $\max_i d_i = \mathrm{o}(m_1/2)$ (for example, when $\max_i d_i = \mathrm{O}(\sqrt{\log n})$). A rejection sampling when the degree sequence is above these thresholds will lead to exponential complexity. To overcome this difficulty in [94] Wormald and McKay showed a way to sample from the set of simple graphs with a given degree sequence by implementing switching-based algorithms. In the case of an homogeneous degree sequence, i.e., $d_i = d$ for all i, the microcanonical ensemble coincides with the random d-regular graph model. The problem was solved by implementing the switching algorithm of [94] and was perfectioned in [69, 87, 115]. The algorithm is efficient when $d = \mathrm{O}(n^{-1/3})$. For inhomogeneous degree sequences, it was proved in [82] that when

$$\max_i d_i = \mathrm{o}(\sqrt{n}), \quad m_1 = \Theta(n), \quad m_2 = \mathrm{O}(n),$$

the switching algorithm asymptotically provides a uniform sampling. For the directed case a sequential stub-matching procedure was shown in [81] to lead to asymptotic uniform sampling when $\max_i d_i = \mathrm{O}(m_1^{1/4-\varepsilon})$, provided $m_1 = \sum_i d_i^{\text{in}} = \sum_i d_i^{\text{out}}$.

MCMC methods usually rely on switching chain dynamics performed on a seed graph generated via the *Havel-Hakimi* algorithm [74, 76]. The details of this method and its variations can be found in [73, Chapter 6]. In particular, it was shown that the mixing properties of the Markov chain are linked to *P-stability* of the degree sequence [85].

In our simulations, given the relatively small size of the graphs and the low density and inhomogeneity of the degree sequences taken in account, we opted for a rejection sampling using the configuration model.

In Table 1 we report the details of the simulated graphs and the rejection rate.

Configuration Model					
Size n	Degree Sequence	Mean Degree	$\sqrt{n}$	$\log n$	Rejection Rate
1000	$\vec{d} = (20, 23, 25, 28, 30)$	25.2	≈ 31.6	≈ 6.9	1.67%
2000	$\vec{d} = (28, 33, 35, 38, 42)$	35.2	≈ 44.7	≈ 7.6	1.35%
5000	$\vec{d} = (48, 53, 57, 62, 64)$	56.8	≈ 70.7	≈ 8.5	0.95%
10000	$\vec{d} = (78, 80, 83, 87, 90)$	83.6	100	≈ 9.2	0.73%

Table 1: Configuration model that have been sampled. Each element specified in the degree sequence appears $\frac{n}{5}$ times. The rejection rate has been obtained by sampling 10000 graphs for each different size and degree sequence and counting the non simple realizations.

5.2.1 Simulations of results of Chapter 4: Largest eigenvalue.

In Theorem 4.1.2 we proved that the expectation of λ_1 in the configuration model conditioned on simplicity satisfies

$$\mathbb{E}[\lambda_1] = \frac{m_2}{m_1} + \frac{m_1 m_3}{m_2^2} - 1 + o(1), \qquad n \to \infty.$$

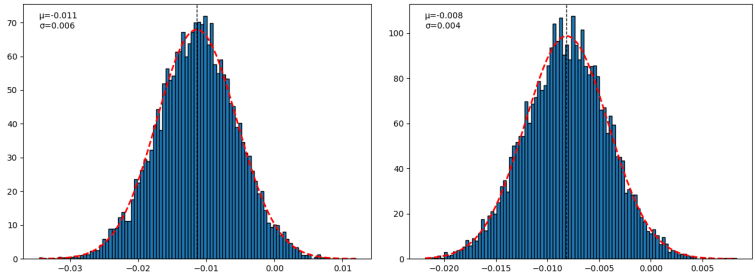
In Figure 5 and Figure 6 we plot

$$\bar{\lambda}_1 = \lambda_1 - \mathbb{E}[\lambda_1]$$

for some degree sequences compatible with the ones studied in Section 5.1.1. It can be seen that, with an increasing size of the graph, the error in the above formula becomes smaller and smaller. Furthermore it can be seen that the empirical standard deviation of λ_1 is much smaller than the one calculated for the Chung-Lu model from the formula for σ^2 in Theorem 3.1.6.

To capture the difference between the largest eigenvalues of the models in Chapter 3 and Chapter 4 we can define the following quantity on the probability space formed by the product measure of the two models

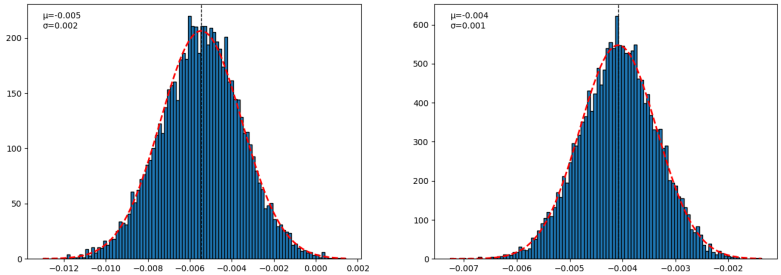
$$\hat{\lambda_1} = \lambda_1^{\text{CL}} - \lambda_1^{\text{CM}}.$$



(a) $n = 1000$,
$\vec{d} = (20, 23, 25, 28, 30)$,
$\sqrt{n} \approx 31.6$, $\log n \approx 6.9$.

(b) $n = 2000$,
$\vec{d} = (28, 33, 35, 38, 42)$,
$\sqrt{n} \approx 44.7$, $\log n \approx 7.6$.

Figure 5: Histograms of $\bar{\lambda}_1$ for different graph sizes n and degree sequences $\vec{d}$. The sample size for each regime is 10^4. Each element specified in the degree sequence appears $\frac{n}{5}$ times. In red is plotted the Gaussian fit; μ is the sample mean (represented by a dashed line in the histogram), σ is the sample standard deviation. We expect $\mu \approx 0$.



(a) $n = 5000$,
$\vec{d} = (48, 53, 57, 62, 64)$,
$\sqrt{n} \approx 70.7$, $\log n \approx 8.5$.

(b) $n = 10000$,
$\vec{d} = (78, 80, 83, 87, 90)$,
$\sqrt{n} = 100$, $\log n \approx 9.2$.

Figure 6: Histograms of $\bar{\lambda}_1$ for different graph sizes n and degree sequences $\vec{d}$. The sample size for each regime is 10^4. Each element specified in the degree sequence appears $\frac{n}{5}$ times. In red is plotted the Gaussian fit; μ is the sample mean (represented by a dashed line in the histogram), σ is the sample standard deviation. We expect $\mu \approx 0$.

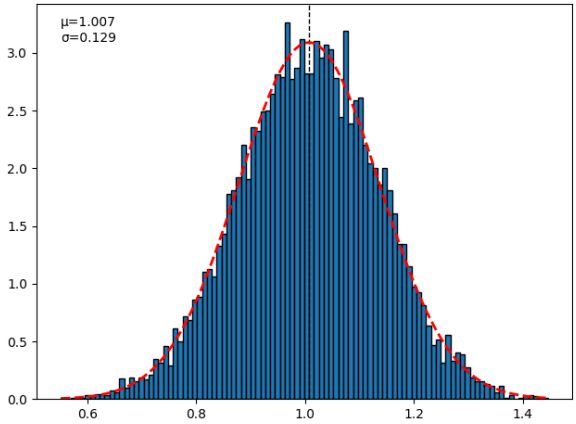


Figure 7: Histograms of $\hat{\lambda}_1$ for $n = 10000$ and degree sequences $\vec{d} = (78, 80, 83, 87, 90)$. The sample size is 10^4. Each element specified in the degree sequence appears $\frac{n}{5}$ times. In red is plotted the Gaussian fit; μ is the sample mean (represented by a dashed line in the histogram), σ is the sample standard deviation. We expect $\mu \approx 1$.

In Figure 7 we plot $\hat{\lambda_1}$. The choice of the product measure corresponds to an independent sampling of λ_1^{CL} and λ_1^{CM}. The histogram supports our conjecture formulated in Chapter 2. Remarkably, the difference is 1, like in the homogenous case, irrespective of the degrees.

Chapter 6

Bibliography

[1] Emmanuel Abbe. "Community detection and stochastic block models: recent developments". In: *The Journal of Machine Learning Research* 18.1 (2017), pp. 6446–6531.

[2] R. Adamczak and P. Wolff. "Concentration Inequalities for Non-Lipschitz Functions with Bounded Derivatives of Higher Order". In: *Probability Theory and Related Fields* 162.3-4 (2015), pp. 531–586.

[3] Gernot Akemann, Jinho Baik, and Philippe Di Francesco. *The Oxford Handbook of Random Matrix Theory*. Oxford University Press, Sept. 2015. ISBN: 978-0-19-874419-1. DOI: 10.1093/oxfordhb/9780198744191.001.0001. URL: https://doi.org/10.1093/oxfordhb/9780198744191.001.0001.

[4] N. Alon, M. Krivelevich, and Van H. Vu. "On the Concentration of Eigenvalues of Random Symmetric Matrices". en. In: *Israel Journal of Mathematics* 131.1 (Dec. 2002), pp. 259–267. ISSN: 0021-2172, 1565-8511. DOI: 10.1007/BF02785860.

[5] Noga Alon. "Eigenvalues and expanders". en. In: *Combinatorica* 6.2 (June 1986), pp. 83–96. ISSN: 0209-9683, 1439-6912. DOI: 10.1007/BF02579166. URL: http://link.springer.com/10.1007/BF02579166 (visited on 12/09/2023).

[6] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. en. 3rd ed. Wiley-Interscience series in discrete mathematics and optimization. OCLC: ocn173809124. Hoboken, N.J: Wiley, 2008. ISBN: 978-0-470-17020-5.

[7] J. Alt, R. Ducatez, and A. Knowles. “Extremal Eigenvalues of Critical Erdős-Rényi Graphs”. In: *arXiv:1905.03243 [math-ph]* (2020).

[8] Johannes Alt, Raphaël Ducatez, and Antti Knowles. “Extremal eigenvalues of critical Erdős–Rényi graphs”. In: *The Annals of Probability* 49.3 (2021), pp. 1347–1401.

[9] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An Introduction to Random Matrices*. Cambridge studies in advanced mathematics. Cambridge: Cambridge University Press, 2009. DOI: 10.1017/CBO9780511801334.

[10] Octavio Arizmendi E. and Victor Pérez-Abreu. “The S-transform of symmetric probability measures with unbounded supports”. en. In: *Proceedings of the American Mathematical Society* 137.09 (Sept. 2009), pp. 3057–3057. ISSN: 0002-9939. DOI: 10.1090/S0002-9939-09-09841-4. URL: http://www.ams.org/jourcgi/jour-getitem?pii=S0002-9939-09-09841-4 (visited on 03/06/2023).

[11] Luca Avena, Rajat Subhra Hazra, and Nandan Malhotra. *Limiting Spectra of inhomogeneous random graphs*. en. arXiv:2312.02805 [math]. Dec. 2023. URL: http://arxiv.org/abs/2312.02805 (visited on 12/23/2023).

[12] Zhidong Bai and Jian-Feng Yao. “Central limit theorems for eigenvalues in a spiked population model”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 44.3 (2008), pp. 447–474.

[13] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. “Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices”. en. In: *The Annals of Probability* 33.5 (Sept. 2005). ISSN: 0091-1798. DOI: 10.1214/009117905000000233. URL: https://projecteuclid.org/journals/annals-of-probability/volume-33/issue-5/Phase-transition-of-the-largest-eigenvalue-for-nonnull-complex-sample/10.1214/009117905000000233.full (visited on 03/06/2023).

[14] Alexander Barvinok and J.A. Hartigan. “The number of graphs and a random graph with a given degree sequence: Random Graph with a Given Degree Sequence”. en. In: *Random Structures & Algorithms* 42.3 (May 2013), pp. 301–348. ISSN: 10429832. DOI: 10.1002/rsa.20409. URL: https://onlinelibrary.wiley.com/doi/10.1002/rsa.20409 (visited on 03/06/2023).

[15] Roland Bauerschmidt, Jiaoyang Huang, and Horng-Tzer Yau. "Local Kesten–McKay Law for Random Regular Graphs". en. In: *Communications in Mathematical Physics* 369.2 (July 2019), pp. 523–636. ISSN: 0010-3616, 1432-0916. DOI: 10.1007/s00220-019-03345-3. URL: http://link.springer.com/10.1007/s00220-019-03345-3 (visited on 03/06/2023).

[16] Roland Bauerschmidt et al. "Bulk eigenvalue statistics for random regular graphs". en. In: *The Annals of Probability* 45.6A (Nov. 2017). ISSN: 0091-1798. DOI: 10.1214/16-AOP1145. URL: https://projecteuclid.org/journals/annals-of-probability/volume-45/issue-6A/Bulk-eigenvalue-statistics-for-random-regular-graphs/10.1214/16-AOP1145.full (visited on 03/06/2023).

[17] Roland Bauerschmidt et al. "Edge rigidity and universality of random regular graphs of intermediate degree". In: *Geom. Funct. Anal.* 30.3 (2020), pp. 693–769. ISSN: 1016-443X,1420-8970. DOI: 10.1007/s00039-020-00538-0. URL: https://doi.org/10.1007/s00039-020-00538-0.

[18] F. Benaych-Georges, C. Bordenave, and A. Knowles. "Largest Eigenvalues of Sparse Inhomogeneous Erdős-Rényi Graphs". In: *arXiv:1704.02953 [math]* (2017).

[19] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. "Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs". In: *The Annals of Probability* 47.3 (2019), pp. 1653–1676. DOI: 10.1214/18-AOP1293. URL: https://doi.org/10.1214/18-AOP1293.

[20] Florent Benaych-Georges, Charles Bordenave, and Antti Knowles. "Spectral radii of sparse random matrices". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 56.3 (2020), pp. 2141–2161. DOI: 10.1214/19-AIHP1033. URL: https://doi.org/10.1214/19-AIHP1033.

[21] Florent Benaych-Georges, Alice Guionnet, and Mylène Maida. "Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices". en. In: *Electronic Journal of Probability* 16.0 (2011), pp. 1621–1662. ISSN: 1083-6489. DOI: 10.1214/EJP.v16-929. URL: https://projecteuclid.org/euclid.ejp/1464820229 (visited on 03/22/2021).

[22] Hari Bercovici and Dan Voiculescu. "Free Convolution of Measures with Unbounded Support". en. In: *Indiana University Mathematics Journal* 42.3 (1993).

[23] B. Bercu, B. Delyon, and E. Rio. *Concentration Inequalities for Sums and Martingales*. SpringerBriefs in Mathematics. Cham: Springer International Publishing, 2015.

[24] M. J. Betancourt et al. *The Geometric Foundations of Hamiltonian Monte Carlo*. en. arXiv:1410.5110 [stat]. Oct. 2014. URL: http://arxiv.org/abs/1410.5110 (visited on 07/18/2023).

[25] Michael Betancourt. *A Conceptual Introduction to Hamiltonian Monte Carlo*. en. arXiv:1701.02434 [stat]. July 2018. URL: http://arxiv.org/abs/1701.02434 (visited on 07/18/2023).

[26] Charles Bordenave. "A new proof of Friedman's second eigenvalue Theorem and its extension to random lifts". en. In: *arXiv:1502.04482 [math]* (Mar. 2019). arXiv: 1502.04482. URL: http://arxiv.org/abs/1502.04482 (visited on 06/10/2021).

[27] Charles Bordenave. "A new proof of Friedman's second eigenvalue theorem and its extension to random lifts". In: *Ann. Sci. Éc. Norm. Supér. (4)* 53.6 (2020), pp. 1393–1439. ISSN: 0012-9593,1873-2151. DOI: 10.24033/asens.2450. URL: https://doi.org/10.24033/asens.2450.

[28] Charles Bordenave and Marc Lelarge. "Resolvent of large random graphs". In: *Random Structures Algorithms* 37.3 (2010), pp. 332–352. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.20313. URL: https://doi.org/10.1002/rsa.20313.

[29] S. Boucheron, G. Lugosi, and P. Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. ISBN: 978-0-19-953525-5.

[30] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. en. 1st ed. Oxford: Oxford University Press, 2013. ISBN: 978-0-19-953525-5.

[31] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: a nonasymptotic theory of independence*. en. 1st ed. OCLC: ocn818449985. Oxford: Oxford University Press, 2013. ISBN: 978-0-19-953525-5.

[32] Paul Bourgade and H-T Yau. “The eigenvector moment flow and local quantum unique ergodicity”. In: *Communications in Mathematical Physics* 350.1 (2017), pp. 231–278.

[33] Andrei Broder and Eli Shamir. “On the second eigenvalue of random regular graphs”. en. In: *28th Annual Symposium on Foundations of Computer Science (sfcs 1987)*. Los Angeles, CA, USA: IEEE, Oct. 1987, pp. 286–294. ISBN: 978-0-8186-0807-0. DOI: 10.1109/SFCS.1987.45. URL: http://ieeexplore.ieee.org/document/4568282/ (visited on 05/15/2023).

[34] Andrei Z. Broder et al. “Optimal Construction of Edge-Disjoint Paths in Random Graphs”. en. In: *SIAM Journal on Computing* 28.2 (Jan. 1998), pp. 541–573. ISSN: 0097-5397, 1095-7111. DOI: 10.1137/S0097539795290805. URL: https://epubs.siam.org/doi/10.1137/S0097539795290805 (visited on 10/12/2023).

[35] M. Capitaine, C. Donati-Martin, and D. Féral. “Central limit theorems for eigenvalues of deformations of Wigner matrices”. In: *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques* 48.1 (2012), pp. 107–133. DOI: 10.1214/10-AIHP410. URL: https://doi.org/10.1214/10-AIHP410.

[36] Mireille Capitaine, Catherine Donati-Martin, and Delphine Féral. “The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations”. In: *The Annals of Probability* 37.1 (2009), pp. 1–47. DOI: 10.1214/08-AOP394. URL: https://doi.org/10.1214/08-AOP394.

[37] Claudio Castellano and Romualdo Pastor-Satorras. “Relating topological determinants of complex networks to their spectral properties: structural and dynamical effects”. In: *Physical Review X* 7.4 (2017), p. 041024.

[38] A. Chakrabarty et al. “Spectra of Adjacency and Laplacian Matrices of Inhomogeneous Erdős-Rényi Random Graphs”. In: *Random Matrices: Theory and Applications (to appear)* (2019).

[39] Arijit Chakrabarty, Sukrit Chakraborty, and Rajat Subhra Hazra. “Eigenvalues Outside the Bulk of Inhomogeneous Erdős–Rényi Random Graphs”. In: *Journal of Statistical Physics* 181.5 (2020), pp. 1746–1780. DOI: 10.1007/s10955-020-02644-7. URL: https://doi.org/10.1007/s10955-020-02644-7.

[40] Arijit Chakrabarty et al. “Spectra of adjacency and Laplacian matrices of inhomogeneous Erdos-Renyi random graphs”. In: *Random Matrices Theory Appl.* 10.1 (2021), Paper No. 2150009, 34. ISSN: 2010-3263,2010-3271. DOI: 10.1142/S201032632150009X. URL: https://doi.org/10.1142/S201032632150009X.

[41] Arijit Chakrabarty et al. “Spectra of adjacency and Laplacian matrices of inhomogeneous Erdős-Rényi random graphs”. In: *Random Matrices Theory and Applications* 10 (2021), p. 2150009.

[42] S.A. Choudum. “A simple proof of the Erdos-Gallai theorem on graph sequences”. en. In: *Bulletin of the Australian Mathematical Society* 33.1 (Feb. 1986), pp. 67–70. ISSN: 0004-9727, 1755-1633. DOI: 10.1017/S0004972700002872. URL: https://www.cambridge.org/core/product/identifier/S0004972700002872/type/journal_article (visited on 12/12/2023).

[43] Fan Chung and Linyuan Lu. “Connected components in random graphs with given expected degree sequences”. en. In: *Annals of Combinatorics* 6.2 (2002), pp. 125–145. ISSN: 0218-0006. DOI: 10.1007/PL00012580. URL: http://link.springer.com/10.1007/PL00012580 (visited on 09/20/2021).

[44] Fan Chung, Linyuan Lu, and Van Vu. “Eigenvalues of random power law graphs”. en. In: *Annals of Combinatorics* 7.1 (2003), pp. 21–33. ISSN: 0218-0006, 0219-3094. DOI: 10.1007/s000260300002. URL: http://link.springer.com/10.1007/s000260300002 (visited on 09/14/2021).

[45] Ton Coolen, Alessia Annibale, and Ekaterina Roberts. *Generating Random Networks and Graphs*. Oxford University Press, Mar. 2017. ISBN: 978-0-19-870989-3. DOI: 10.1093/oso/9780198709893.001.0001. URL: https://doi.org/10.1093/oso/9780198709893.001.0001.

[46] Simon Coste and Justin Salez. “Emergence of extended states at zero in the spectrum of sparse random graphs”. en. In: *The Annals of Probability* 49.4 (May 2021). ISSN: 0091-1798. DOI: 10.1214/20-AOP1499. URL: https://projecteuclid.org/journals/annals-of-probability/volume-49/issue-4/Emergence-of-extended-states-at-zero-in-the-spectrum-of/10.1214/20-AOP1499.full (visited on 12/16/2023).

[47] Dragos Cvetkovic, Peter Rowlinson, and Slobodan Simic. *Eigenspaces of Graphs*. 1st. Cambridge University Press, Jan. 1997. ISBN: 978-0-521-57352-8 978-0-521-05718-9 978-1-139-08654-7. DOI: 10.1017/CBO9781139086547. URL: https://www.cambridge.org/core/product/identifier/9781139086547/type/book (visited on 03/23/2023).

[48] James A. Davis. "Clustering and Hierarchy in Interpersonal Relations: Testing Two Graph Theoretical Models on 742 Sociomatrices". en. In: *American Sociological Review* 35.5 (Oct. 1970), p. 843. ISSN: 00031224. DOI: 10.2307/2093295. URL: http://www.jstor.org/stable/2093295?origin=crossref (visited on 12/10/2023).

[49] Amir Dembo, Eyal Lubetzky, and Yumeng Zhang. "Empirical Spectral Distributions of Sparse Random Graphs". en. In: *In and Out of Equilibrium 3: Celebrating Vladas Sidoravicius*. Ed. by Maria Eulália Vares et al. Vol. 77. Series Title: Progress in Probability. Cham: Springer International Publishing, 2021, pp. 319–345. ISBN: 978-3-030-60753-1 978-3-030-60754-8. DOI: 10.1007/978-3-030-60754-8_15. URL: http://link.springer.com/10.1007/978-3-030-60754-8_15 (visited on 03/06/2023).

[50] F. den Hollander et al. "Ensemble Equivalence for Dense Graphs". In: *Electronic Journal of Probability* 23 (2018), Paper no. 12, 1–26.

[51] S. Dhara and S. Sen. "Large Deviation for Uniform Graphs with given Degrees". In: *arXiv:1904.07666 [math]* (2020).

[52] T. Ding X.and Jiang. "Spectral Distributions of Adjacency and Laplacian Matrices of Random Graphs". In: *The Annals of Applied Probability* 20.6 (2010), pp. 2086–2117.

[53] Pierfrancesco Dionigi et al. "A spectral signature of breaking of ensemble equivalence for constrained random graphs". In: *Electronic Communications in Probability* 26.none (Jan. 2021). ISSN: 1083-589X. DOI: 10.1214/21-ECP432. URL: https://projecteuclid.org/journals/electronic-communications-in-probability/volume-26/issue-none/A-spectral-signature-of-breaking-of-ensemble-equivalence-for-constrained/10.1214/21-ECP432.full (visited on 12/27/2021).

[54] Pierfrancesco Dionigi et al. “Central limit theorem for the principal eigenvalue and eigenvector of Chung–Lu random graphs”. en. In: *Journal of Physics: Complexity* 4.1 (Mar. 2023), p. 015008. ISSN: 2632-072X. DOI: 10.1088/2632-072X/acb8f7. URL: https://iopscience.iop.org/article/10.1088/2632-072X/acb8f7 (visited on 03/06/2023).

[55] Richard S. Ellis, Kyle Haven, and Bruce Turkington. “Large Deviation Principles and Complete Equivalence and Nonequivalence Results for Pure and Mixed Ensembles”. In: *Journal of Statistical Physics* 101.5 (Dec. 2000), pp. 999–1064. ISSN: 1572-9613. DOI: 10.1023/A:1026446225804. URL: https://doi.org/10.1023/A:1026446225804.

[56] László Erdős et al. “Spectral Statistics of Erdős-Rényi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues”. en. In: *Communications in Mathematical Physics* 314.3 (Sept. 2012), pp. 587–640. ISSN: 0010-3616, 1432-0916. DOI: 10.1007/s00220-012-1527-7. URL: http://link.springer.com/10.1007/s00220-012-1527-7 (visited on 03/06/2023).

[57] László Erdős et al. “Spectral statistics of Erdős–Rényi graphs I: Local semicircle law”. en. In: *The Annals of Probability* 41.3B (May 2013). ISSN: 0091-1798. DOI: 10.1214/11-AOP734. URL: https://projecteuclid.org/journals/annals-of-probability/volume-41/issue-3B/Spectral-statistics-of-Erd%c5%91sR%c3%a9nyi-graphs-I-Local-semicircle-law/10.1214/11-AOP734.full (visited on 03/06/2023).

[58] László Erdős et al. “Spectral statistics of Erdős–Rényi graphs I: Local semicircle law”. en. In: *The Annals of Probability* 41.3B (May 2013). ISSN: 0091-1798. DOI: 10.1214/11-AOP734. URL: https://projecteuclid.org/journals/annals-of-probability/volume-41/issue-3B/Spectral-statistics-of-Erd%c5%91sR%c3%a9nyi-graphs-I-Local-semicircle-law/10.1214/11-AOP734.full (visited on 03/06/2023).

[59] Paul Erdős and Alfréd Rényi. “On Random Graphs I”. In: *Publicationes Mathematicae Debrecen* 6 (1959), pp. 290–297.

[60] Delphine Féral and Sandrine Péché. “The largest eigenvalue of rank one deformation of large Wigner matrices”. en. In: *Communications in Mathematical Physics* 272.1 (2007), pp. 185–228. ISSN: 0010-3616, 1432-0916. DOI: 10.1007/s00220-007-0209-3.

URL: http://link.springer.com/10.1007/s00220-007-0209-3 (visited on 03/26/2021).

[61] Delphine Féral and Sandrine Péché. "The largest eigenvalues of sample covariance matrices for a spiked population: Diagonal case". en. In: *Journal of Mathematical Physics* 50.7 (2009), p. 073302. ISSN: 0022-2488, 1089-7658. DOI: 10.1063/1.3155785. URL: http://aip.scitation.org/doi/10.1063/1.3155785 (visited on 06/09/2021).

[62] Ove Frank and David Strauss. "Markov graphs". In: *Journal of the American Statistical Association* 81.395 (1986). Publisher: [American Statistical Association, Taylor & Francis, Ltd.], pp. 832–842. ISSN: 01621459. URL: http://www.jstor.org/stable/2289017 (visited on 12/10/2023).

[63] J. Friedman, J. Kahn, and E. Szemerédi. "On the second eigenvalue of random regular graphs". In: *Proceedings of the twenty-first annual ACM symposium on theory of computing*. STOC '89. Number of pages: 12 Place: Seattle, Washington, USA. New York, NY, USA: Association for Computing Machinery, 1989, pp. 587–598. ISBN: 0-89791-307-8. DOI: 10.1145/73007.73063. URL: https://doi.org/10.1145/73007.73063.

[64] Joel Friedman. *A proof of Alon's second eigenvalue conjecture and related problems*. en. arXiv:cs/0405020. May 2004. URL: http://arxiv.org/abs/cs/0405020 (visited on 05/15/2023).

[65] Joel Friedman. "A proof of Alon's second eigenvalue conjecture and related problems". In: *Mem. Amer. Math. Soc.* 195.910 (2008), pp. viii+100. ISSN: 0065-9266,1947-6221. DOI: 10.1090/memo/0910. URL: https://doi.org/10.1090/memo/0910.

[66] Z. Füredi and J. Komlós. "The Eigenvalues of Random Symmetric Matrices". In: *Combinatorica* 1.3 (1981), pp. 233–241.

[67] Z. Füredi and J. Komlós. "The eigenvalues of random symmetric matrices". In: *Combinatorica* 1.3 (1981), pp. 233–241. DOI: 10.1007/BF02579329. URL: https://doi.org/10.1007/BF02579329.

[68] Zoltán Füredi and János Komlós. "The eigenvalues of random symmetric matrices". In: *Combinatorica* 1 (1981), pp. 233–241.

[69] Pu Gao and Nicholas Wormald. “Uniform Generation of Random Regular Graphs”. en. In: *SIAM Journal on Computing* 46.4 (Jan. 2017), pp. 1395–1427. ISSN: 0097-5397, 1095-7111. DOI: 10.1137/15M1052779. URL: https://epubs.siam.org/doi/10.1137/15M1052779 (visited on 12/31/2023).

[70] D. Garlaschelli, F. den Hollander, and A. Roccaverde. “Covariance Structure behind Breaking of Ensemble Equivalence in Random Graphs”. In: *Journal of Statistical Physics* 173.3-4 (2018), pp. 644–662.

[71] Diego Garlaschelli, Frank den Hollander, and Andrea Roccaverde. “Covariance Structure Behind Breaking of Ensemble Equivalence in Random Graphs”. en. In: *Journal of Statistical Physics* 173.3-4 (Nov. 2018), pp. 644–662. ISSN: 0022-4715, 1572-9613. DOI: 10.1007/s10955-018-2114-x. URL: http://link.springer.com/10.1007/s10955-018-2114-x (visited on 03/06/2023).

[72] Diego Garlaschelli, Frank den Hollander, and Andrea Roccaverde. “Ensemble nonequivalence in random graphs with modular structure”. en. In: *Journal of Physics A: Mathematical and Theoretical* 50.1 (Jan. 2017), p. 015001. ISSN: 1751-8113, 1751-8121. DOI: 10.1088/1751-8113/50/1/015001. URL: https://iopscience.iop.org/article/10.1088/1751-8113/50/1/015001 (visited on 03/06/2023).

[73] Catherine Greenhill. “Generating graphs randomly”. In: *Surveys in combinatorics 2021*. Ed. by Konrad K. Dabrowski et al. London mathematical society lecture note series. Cambridge: Cambridge University Press, 2021, pp. 133–186. DOI: 10.1017/9781009036214.005.

[74] S. L. Hakimi. “On Realizability of a Set of Integers as Degrees of the Vertices of a Linear Graph. I”. en. In: *Journal of the Society for Industrial and Applied Mathematics* 10.3 (Sept. 1962), pp. 496–506. ISSN: 0368-4245, 2168-3484. DOI: 10.1137/0110037. URL: http://epubs.siam.org/doi/10.1137/0110037 (visited on 07/19/2021).

[75] D.L. Hanson and F.T. Wright. “A Bound on Tail Probabilities for Quadratic Forms in Independent Random Variables”. In: *The Annals of Mathematical Statistics* 42.3 (1971), pp. 1079–1083.

[76] Václav Havel. “Poznámka o existenci konečných grafů”. cze. In: *Časopis pro pěstování matematiky* 080.4 (1955). Publisher: Mathematical Institute of the Czechoslovak Academy of Sciences, pp. 477–480. URL: http://eudml.org/doc/19050.

[77] Remco van der Hofstad. *Random Graphs and Complex Networks Vol 1*. Cambridge: Cambridge University Press, 2017. ISBN: 978-1-316-77942-2. DOI: 10.1017/9781316779422. URL: http://ebooks.cambridge.org/ref/id/CBO9781316779422 (visited on 02/25/2020).

[78] Remco van der Hofstad. *Random Graphs and Complex Networks vol 2*. Vol. 2. Cambridge University Press, 2024.

[79] Remco van der van der Hofstad. *Random Graphs and Complex Networks*. en. 1st ed. Cambridge University Press, Nov. 2016. ISBN: 978-1-107-17287-6 978-1-316-77942-2 978-1-316-62506-4. DOI: 10.1017/9781316779422. URL: https://www.cambridge.org/core/product/identifier/9781316779422/type/book (visited on 04/19/2023).

[80] Jiaoyang Huang and Horng-Tzer Yau. “Spectrum of random d-regular graphs up to the edge”. en. In: *Communications on Pure and Applied Mathematics* 77.3 (Mar. 2024), pp. 1635–1723. ISSN: 0010-3640, 1097-0312. DOI: 10.1002/cpa.22176. URL: https://onlinelibrary.wiley.com/doi/10.1002/cpa.22176 (visited on 05/10/2024).

[81] Femke van Ieperen and Ivan Kryven. *Sequential stub matching for uniform generation of directed graphs with a given degree sequence*. en. arXiv:2103.15958 [math]. June 2022. URL: http://arxiv.org/abs/2103.15958 (visited on 12/11/2023).

[82] Svante Janson. “Random graphs with given vertex degrees and switchings”. en. In: *Random Structures & Algorithms* 57.1 (Aug. 2020), pp. 3–31. ISSN: 1042-9832, 1098-2418. DOI: 10.1002/rsa.20911. URL: https://onlinelibrary.wiley.com/doi/10.1002/rsa.20911 (visited on 12/31/2023).

[83] Svante Janson. “The Probability That a Random Multigraph is Simple”. en. In: *Combinatorics, Probability and Computing* 18.1-2 (Mar. 2009), pp. 205–225. ISSN: 0963-5483, 1469-2163. DOI: 10.1017/S0963548308009644. URL: https://www.cambridge.org/core/product/identifier/S0963548308009644/type/journal_article (visited on 04/03/2023).

[84] E. T. Jaynes. “Information theory and statistical mechanics”. In: *Physical Review* 106.4 (May 1957). Number of pages: 0 Publisher: American Physical Society, pp. 620–630. DOI: 10.1103/PhysRev.106.620. URL: https://link.aps.org/doi/10.1103/PhysRev.106.620.

[85] Mark Jerrum and Alistair Sinclair. “Fast uniform generation of regular graphs”. In: *Theoretical Computer Science* 73.1 (1990), pp. 91–100. ISSN: 0304-3975. DOI: https://doi.org/10.1016/0304-3975(90)90164-D. URL: https://www.sciencedirect.com/science/article/pii/030439759090164D.

[86] Ferenc Juhász. “On the spectrum of a random graph”. In: *Algebraic methods in graph theory*. Colloquia Mathematica Societatis János Bolyai 1 (25 1981).

[87] J H Kim and V H Vu. “Generating random regular graphs”. en. In: *Combinatorica* 26 (2006), pp. 683–708.

[88] M. Krivelevich and B. Sudakov. “The Largest Eigenvalue of Sparse Random Graphs”. In: *Combinatorics, Probability and Computing* 12.01 (2003), pp. 61–72.

[89] Michael Krivelevich and Benny Sudakov. “The largest eigenvalue of sparse random graphs”. In: *Combinatorics, Probability and Computing* 12.1 (2003), pp. 61–72. DOI: 10.1017/S0963548302005424.

[90] Can M Le, Elizaveta Levina, and Roman Vershynin. “Concentration of random graphs and application to community detection”. In: *Proceedings of the International Congress of Mathematicians: Rio de Janeiro 2018*. World Scientific. 2018, pp. 2925–2943.

[91] D. Lynden-Bell, Roger Wood, and Astronomer Royal. “The gravo-thermal catastrophe in isothermal spheres and the onset of red-giant structure for stellar systems”. In: *Monthly Notices of the Royal Astronomical Society* 138.4 (Feb. 1968). tex.eprint: https://academic.oup.com/mnras/article-pdf/138/4/495/8078821/mnras138-0495.pdf, pp. 495–525. ISSN: 0035-8711. DOI: 10.1093/mnras/138.4.495. URL: https://doi.org/10.1093/mnras/138.4.495.

[92] Travis Martin, Xiao Zhang, and Mark EJ Newman. “Localization and centrality in networks”. In: *Physical review E* 90.5 (2014), p. 052808.

[93] Brendan D McKay. “The expected eigenvalue distribution of a large regular graph”. In: *Linear Algebra and its applications* 40 (1981), pp. 203–216.

[94] Brendan D McKay and Nicholas C Wormald. "Uniform generation of random regular graphs of moderate degree". en. In: *Journal of Algorithms* 11.1 (Mar. 1990), pp. 52–67. ISSN: 01966774. DOI: 10.1016/0196-6774(90)90029-E. URL: https://linkinghub.elsevier.com/retrieve/pii/019667749090029E (visited on 12/31/2023).

[95] M. L. Mehta. *Random Matrices*. eng. 3rd. Pure and applied mathematics v. 142. Amsterdam San Diego, CA: Elsevier/Academic Press, 2004. ISBN: 978-0-12-088409-4.

[96] C.D. Meyer. *Matrix Analysis and Applied Linear Algebra*. Philadelphia: Society for Industrial and Applied Mathematics, 2000.

[97] Mark EJ Newman. "Finding community structure in networks using the eigenvectors of matrices". In: *Physical review E* 74.3 (2006), p. 036104.

[98] Juyong Park and M. E. J. Newman. "Solution for the properties of a clustered network". en. In: *Physical Review E* 72.2 (Aug. 2005). arXiv:cond-mat/0412579, p. 026136. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.72.026136. URL: http://arxiv.org/abs/cond-mat/0412579 (visited on 12/30/2023).

[99] Juyong Park and M. E. J. Newman. "Statistical mechanics of networks". en. In: *Physical Review E* 70.6 (Dec. 2004), p. 066117. ISSN: 1539-3755, 1550-2376. DOI: 10.1103/PhysRevE.70.066117. URL: https://link.aps.org/doi/10.1103/PhysRevE.70.066117 (visited on 03/06/2023).

[100] Romualdo Pastor-Satorras and Claudio Castellano. "Eigenvector localization in real networks and its implications for epidemic spreading". In: *Journal of Statistical Physics* 173.3 (2018), pp. 1110–1123.

[101] S. Péché. "The largest eigenvalue of small rank perturbations of Hermitian random matrices". en. In: *Probability Theory and Related Fields* 134.1 (Jan. 2006), pp. 127–173. ISSN: 0178-8051, 1432-2064. DOI: 10.1007/s00440-005-0466-z. URL: http://link.springer.com/10.1007/s00440-005-0466-z (visited on 06/09/2021).

[102] A. Roccaverde. "Breaking of Ensemble Equivalence for Complex Networks". PhD Thesis. Leiden University, 2018.

[103] Andrea Roccaverde. “Breaking of Ensemble Equivalence for Complex Networks”. Doctoral Thesis. Leiden University, 2018.

[104] Eli Shamir and Joel Spencer. “Sharp concentration of the chromatic number on random graphs G(n, p)”. In: *Combinatorica* 7.1 (Mar. 1987), pp. 121–129. ISSN: 1439-6912. DOI: 10.1007/BF02579208. URL: https://doi.org/10.1007/BF02579208.

[105] Tom A. B. Snijders et al. “New Specifications for Exponential Random Graph Models”. en. In: *Sociological Methodology* 36.1 (Aug. 2006), pp. 99–153. ISSN: 0081-1750, 1467-9531. DOI: 10.1111/j.1467-9531.2006.00176.x. URL: http://journals.sagepub.com/doi/10.1111/j.1467-9531.2006.00176.x (visited on 03/06/2023).

[106] Alexander Soshnikov. “Universality at the Edge of the Spectrum in Wigner Random Matrices”. In: *Communications in Mathematical Physics* 207.3 (Nov. 1999), pp. 697–733. ISSN: 0010-3616, 1432-0916. DOI: 10.1007/s002200050743.

[107] T. Squartini and D. Garlaschelli. “Reconnecting Statistical Physics and Combinatorics beyond Ensemble Equivalence”. In: *arXiv:1710.11422 [cond-mat.stat-mech]* (2020).

[108] T. Squartini et al. “Breaking of Ensemble Equivalence in Networks”. In: *Physical Review Letters* 115.26 (2015), p. 268701.

[109] Tiziano Squartini and Diego Garlaschelli. “Analytical maximum-likelihood method to detect patterns in real networks”. en. In: *New Journal of Physics* 13.8 (2011), p. 083001. ISSN: 1367-2630. DOI: 10.1088/1367-2630/13/8/083001. URL: https://iopscience.iop.org/article/10.1088/1367-2630/13/8/083001 (visited on 05/12/2021).

[110] Tiziano Squartini and Diego Garlaschelli. *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics*. en. SpringerBriefs in Complexity. Cham: Springer International Publishing, 2017. ISBN: 978-3-319-69436-8 978-3-319-69438-2. DOI: 10.1007/978-3-319-69438-2. URL: http://link.springer.com/10.1007/978-3-319-69438-2 (visited on 03/06/2023).

[111] Tiziano Squartini and Diego Garlaschelli. *Reconnecting statistical physics and combinatorics beyond ensemble equivalence*. en. arXiv:1710.11422 [cond-mat]. Oct. 2017. URL: http://arxiv.org/abs/1710.11422 (visited on 03/06/2023).

[112] Tiziano Squartini, Rossana Mastrandrea, and Diego Garlaschelli. "Unbiased sampling of network ensembles". en. In: *New Journal of Physics* 17.2 (Feb. 2015), p. 023052. ISSN: 1367-2630. DOI: 10.1088/1367-2630/17/2/023052. URL: https://iopscience.iop.org/article/10.1088/1367-2630/17/2/023052 (visited on 12/24/2023).

[113] Tiziano Squartini et al. "Breaking of Ensemble Equivalence in Networks". en. In: *Physical Review Letters* 115.26 (Dec. 2015), p. 268701. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.115.268701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.268701 (visited on 03/06/2023).

[114] Tiziano Squartini et al. "Breaking of ensemble equivalence in networks". In: *Physical review letters* 115.26 (2015), p. 268701.

[115] A. Steger and N. C. Wormald. "Generating Random Regular Graphs Quickly". en. In: *Combinatorics, Probability and Computing* 8.4 (July 1999), pp. 377–396. ISSN: 0963-5483, 1469-2163. DOI: 10.1017/S0963548399003867. URL: https://www.cambridge.org/core/product/identifier/S0963548399003867/type/journal_article (visited on 12/30/2023).

[116] T. Tao. *Topics in Random Matrix Theory*. Vol. 132. Graduate Studies in Mathematics. Providence, Rhode Island: American Mathematical Society, 2012. ISBN: 978-0-8218-7430-1 978-0-8218-8506-2.

[117] H. Touchette. "Equivalence and Nonequivalence of Ensembles: Thermodynamic, Macrostate, and Measure Levels". In: *Journal of Statistical Physics* 159.5 (2015), pp. 987–1016.

[118] H. Touchette, R. S. Ellis, and B. Turkington. "An Introduction to the Thermodynamic and Macrostate Levels of Nonequivalent Ensembles". en. In: *Physica A: Statistical Mechanics and its Applications* 340.1-3 (Sept. 2004). arXiv:cond-mat/0404655, pp. 138–146. ISSN: 03784371. DOI: 10.1016/j.physa.2004.03.088. URL: http://arxiv.org/abs/cond-mat/0404655 (visited on 03/06/2023).

[119] Hugo Touchette. *A basic introduction to large deviations: Theory, applications, simulations*. en. arXiv:1106.4146 [cond-mat, physics:math-ph]. Feb. 2012. URL: http://arxiv.org/abs/1106.4146 (visited on 03/06/2023).

[120] Hugo Touchette. “Ensemble equivalence for general many-body systems”. en. In: *EPL (Europhysics Letters)* 96.5 (Dec. 2011). arXiv:1106.2979 [cond-mat, physics:math-ph], p. 50010. ISSN: 0295-5075, 1286-4854. DOI: 10.1209/0295-5075/96/50010. URL: http://arxiv.org/abs/1106.2979 (visited on 03/06/2023).

[121] Hugo Touchette. “Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels”. en. In: *Journal of Statistical Physics* 159.5 (June 2015). arXiv:1403.6608 [cond-mat], pp. 987–1016. ISSN: 0022-4715, 1572-9613. DOI: 10.1007/s10955-015-1212-2. URL: http://arxiv.org/abs/1403.6608 (visited on 03/06/2023).

[122] Hugo Touchette. “Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels”. In: *Journal of Statistical Physics* 159.5 (2015), pp. 987–1016.

[123] Hugo Touchette. “The large deviation approach to statistical mechanics”. en. In: *Physics Reports* 478.1-3 (July 2009). arXiv:0804.0327 [cond-mat], pp. 1–69. ISSN: 03701573. DOI: 10.1016/j.physrep.2009.05.002. URL: http://arxiv.org/abs/0804.0327 (visited on 03/06/2023).

[124] Linh V. Tran, Van H. Vu, and Ke Wang. “Sparse random graphs: Eigenvalues and eigenvectors”. en. In: *Random Structures & Algorithms* 42.1 (Jan. 2013), pp. 110–134. ISSN: 1042-9832, 1098-2418. DOI: 10.1002/rsa.20406. URL: https://onlinelibrary.wiley.com/doi/10.1002/rsa.20406 (visited on 12/16/2023).

[125] Linh V. Tran, Van H. Vu, and Ke Wang. “Sparse random graphs: eigenvalues and eigenvectors”. In: *Random Structures Algorithms* 42.1 (2013), pp. 110–134. ISSN: 1042-9832,1098-2418. DOI: 10.1002/rsa.20406. URL: https://doi.org/10.1002/rsa.20406.

[126] Antonia M Tulino and Sergio Verdú. “Random matrix theory and wireless communications”. In: *Foundations and Trends in Communications and Information Theory* 1.1 (2004), pp. 1–182.

[127] Nicolò Vallarano et al. “Fast and scalable likelihood maximization for Exponential Random Graph Models”. en. In: *arXiv:2101.12625 [cond-mat, physics:physics]* (June 2021). arXiv: 2101.12625. URL: http://arxiv.org/abs/2101.12625 (visited on 07/19/2021).

[128] Dan Voiculescu. “Multiplication of Certain Non-Commuting Random Variables”. en. In: *Journal of Operator Theory* 18.2 (1987), pp. 223–235. URL: https://www.jstor.org/stable/24714784.

[129] Van H. Vu. “Recent progress in combinatorial random matrix theory”. In: *Probab. Surv.* 18 (2021), pp. 179–200. ISSN: 1549-5787. DOI: 10.1214/20-ps346. URL: https://doi.org/10.1214/20-ps346.

[130] Van H. Vu. “Spectral Norm of Random Matrices”. en. In: *Combinatorica* 27.6 (Nov. 2007), pp. 721–736. ISSN: 0209-9683, 1439-6912. DOI: 10.1007/s00493-007-2190-z.

[131] Eugene P. Wigner. “On the Distribution of the Roots of Certain Symmetric Matrices”. en. In: *The Annals of Mathematics* 67.2 (Mar. 1958), p. 325. ISSN: 0003486X. DOI: 10.2307/1970008. URL: https://www.jstor.org/stable/1970008?origin=crossref (visited on 12/09/2023).

[132] N. C. Wormald. “Models of Random Regular Graphs”. en. In: *Surveys in Combinatorics, 1999*. Ed. by J. D. Lamb and D. A. Preece. 1st ed. Cambridge University Press, July 1999, pp. 239–298. ISBN: 978-0-521-65376-3 978-0-511-72133-5. DOI: 10.1017/CBO9780511721335.010. URL: https://www.cambridge.org/core/product/identifier/CBO9780511721335A084/type/book_part (visited on 09/11/2023).

[133] Y. Zhu. “A Graphon Approach to Limiting Spectral Distributions of Wigner-Type Matrices”. In: *Random Structures & Algorithms* 56.1 (2020), pp. 251–279.