Spectral signatures of breaking of ensemble equivalence

Dionigi, P.

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 [19], 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 [2] 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 [15]). 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 [28] for a review). A powerful and versatile network model was developed in the seminal work [20], which was further developed especially in [30] 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 [25] and was, in the context of graph theory, further developed in what we call *Maximum-Entropy Networks* (see [32]). 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, [24]), 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 [26], where thermodynamic properties of certain stellar systems were considered. Since then, many studies have appeared where ensemble equivalence was questioned. In particular, in [16] 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 [37] for a review). In a series of papers [36, 38, 39], 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 [31, 23, 22] 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 [29].

§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 invaluable 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 [41] 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 [27, 3] for references and [1] 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 [13, 32]). 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

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

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

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

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 $\vec{C}$

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

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 [6] 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 ([10]). 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 [28]) and write the partition function as

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

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 [42] 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 [38] 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 [35, 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:

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

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

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

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 $\mathbb{P}_{\mathrm{mic}}$ and $\mathbb{P}_{\mathrm{can}}$, 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 [23, 22, 31] 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

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

• *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 [41]. 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 $\text{Var}[a_{ij}] = 1$. 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 a mean degree $p(n-1) = d > (\log n)^a$ with $a > 3$, after a proper scaling and centering of the matrix elements, the

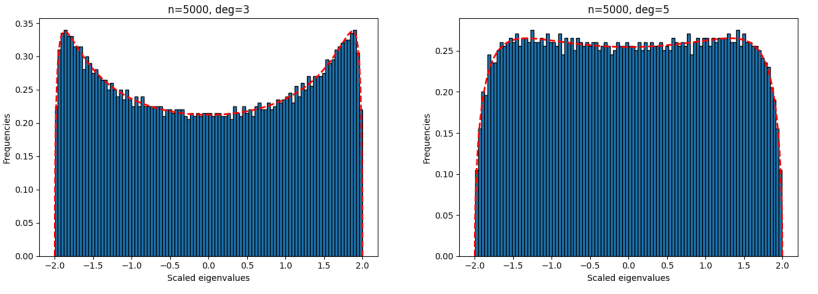
[1] In Wigner matrices the diagonal elements can be chosen independently from a different distribution than the off-diagonal elements without changing the asymptotic behaviour of the ESD.

ESD and many other spectral characteristic were extensively studied in [18, 17]. For the case with fixed d, less is known. This is an active field of research with many open problems. See [12, 4] 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* $d = 5$.

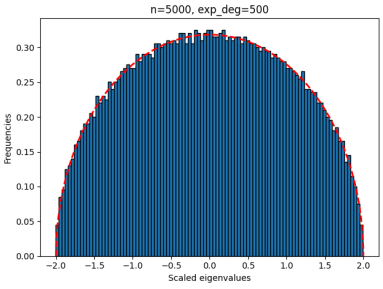
Further properties of spectral statistics of random regular graph with fixed degree were studied in [7, 8]. For a growing $d = d(n) \gg 1$, it was proved in [40] that

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

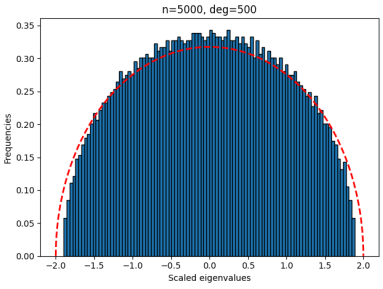
By (1.8.4) and the above observations, ESD cannot be the right quantity to look at Macrostate BEE. Indeed, for sufficient large degrees, the ESDs of the microcanonical and the canonical ensemble (i.e., the random regular graph and the ERRG in the homogenous case) are asymptotically equivalent.

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 [5]. 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:

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



(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).*

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 [40], 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 corresponding 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 $\mathrm{O}(\sqrt{d})$ 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 [33], 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).

Bibliography

[1] G. Akemann, J. Baik, and P. Di Francesco. *The Oxford Handbook of Random Matrix Theory*. Oxford University Press, Sept. 2015.

[2] N. Alon and J. H. Spencer. *The Probabilistic Method*. Wiley-Interscience series in discrete mathematics and optimization. Wiley, Hoboken, N.J, 3rd ed edition, 2008. OCLC: ocn173809124.

[3] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge studies in advanced mathematics. Cambridge University Press, Cambridge, 2009.

[4] L. Avena, R. S. Hazra, and N. Malhotra. Limiting Spectra of inhomogeneous random graphs, Dec. 2023. arXiv:2312.02805 [math].

[5] J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5), Sept. 2005.

[6] A. Barvinok and J. Hartigan. The number of graphs and a random graph with a given degree sequence: Random Graph with a Given Degree Sequence. *Random Structures & Algorithms*, 42(3):301–348, May 2013.

[7] R. Bauerschmidt, J. Huang, A. Knowles, and H.-T. Yau. Bulk eigenvalue statistics for random regular graphs. *The Annals of Probability*, 45(6A), Nov. 2017.

[8] R. Bauerschmidt, J. Huang, and H.-T. Yau. Local Kesten–McKay Law for Random Regular Graphs. *Communications in Mathematical Physics*, 369(2):523–636, July 2019.

[9] S. Chakraborty, R. van der Hofstad, and F. d. Hollander. Sparse random graphs with many triangles, Feb. 2022. arXiv:2112.06526 [math].

[10] S. Choudum. A simple proof of the Erdos-Gallai theorem on graph sequences. *Bulletin of the Australian Mathematical Society*, 33(1):67–70, Feb. 1986.

[11] F. R. K. Chung. *Spectral graph theory*. Number no. 92 in Regional conference series in mathematics. Published for the Conference Board of the mathematical sciences by the American Mathematical Society, Providence, R.I, 1997.

[12] S. Coste and J. Salez. Emergence of extended states at zero in the spectrum of sparse random graphs. *The Annals of Probability*, 49(4), May 2021.

[13] D. Cvetkovic, P. Rowlinson, and S. Simic. *Eigenspaces of Graphs*. Cambridge University Press, 1st edition, Jan. 1997.

[14] D. M. Cvetković, M. Doob, and H. Sachs. *Spectra of graphs: theory and application*. Dt. Verl. der Wiss, Berlin, 1980.

[15] J. A. Davis. Clustering and Hierarchy in Interpersonal Relations: Testing Two Graph Theoretical Models on 742 Sociomatrices. *American Sociological Review*, 35(5):843, Oct. 1970.

[16] R. S. Ellis, K. Haven, and B. Turkington. Large Deviation Principles and Complete Equivalence and Nonequivalence Results for Pure and Mixed Ensembles. *Journal of Statistical Physics*, 101(5):999–1064, Dec. 2000.

[17] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral Statistics of Erdős-Rényi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues. *Communications in Mathematical Physics*, 314(3):587–640, Sept. 2012.

[18] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős–Rényi graphs I: Local semicircle law. *The Annals of Probability*, 41(3B), May 2013.

[19] P. Erdős and A. Rényi. On Random Graphs I. *Publicationes Mathematicae Debrecen*, 6:290–297, 1959.

[20] O. Frank and D. Strauss. Markov graphs. *Journal of the American Statistical Association*, 81(395):832–842, 1986. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

[21] S. Ganguly, E. Hiesmayr, and K. Nam. Upper tail behavior of the number of triangles in random graphs with constant average degree, Feb. 2022. arXiv:2202.06916 [cs, math].

[22] D. Garlaschelli, F. den Hollander, and A. Roccaverde. Ensemble nonequivalence in random graphs with modular structure. *Journal of Physics A: Mathematical and Theoretical*, 50(1):015001, Jan. 2017.

[23] D. Garlaschelli, F. den Hollander, and A. Roccaverde. Covariance Structure Behind Breaking of Ensemble Equivalence in Random Graphs. *Journal of Statistical Physics*, 173(3-4):644–662, Nov. 2018.

[24] C. Greenhill. Generating graphs randomly. In K. K. Dabrowski, M. Gadouleau, N. Georgiou, M. Johnson, G. B. Mertzios, and D. Paulusma, editors, *Surveys in combinatorics 2021*, London mathematical society lecture note series, pages 133–186. Cambridge University Press, Cambridge, 2021.

[25] E. T. Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, May 1957. Number of pages: 0 Publisher: American Physical Society.

[26] D. Lynden-Bell, R. Wood, and A. Royal. The gravo-thermal catastrophe in isothermal spheres and the onset of red-giant structure for stellar systems. *Monthly Notices of the Royal Astronomical Society*, 138(4):495–525, Feb. 1968. tex.eprint: https://academic.oup.com/mnras/article-pdf/138/4/495/8078821/mnras138-0495.pdf.

[27] M. L. Mehta. *Random Matrices*. Number v. 142 in Pure and applied mathematics. Elsevier/Academic Press, Amsterdam San Diego, CA, 3rd edition, 2004.

[28] J. Park and M. E. J. Newman. Statistical mechanics of networks. *Physical Review E*, 70(6):066117, Dec. 2004.

[29] A. Roccaverde. *Breaking of Ensemble Equivalence for Complex Networks*. Doctoral Thesis, Leiden University, 2018.

[30] T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock. New Specifications for Exponential Random Graph Models. *Sociological Methodology*, 36(1):99–153, Aug. 2006.

[31] T. Squartini, J. de Mol, F. den Hollander, and D. Garlaschelli. Breaking of Ensemble Equivalence in Networks. *Physical Review Letters*, 115(26):268701, Dec. 2015.

[32] T. Squartini and D. Garlaschelli. *Maximum-Entropy Networks: Pattern Detection, Network Reconstruction and Graph Combinatorics*. SpringerBriefs in Complexity. Springer International Publishing, Cham, 2017.

[33] T. Squartini and D. Garlaschelli. Reconnecting statistical physics and combinatorics beyond ensemble equivalence, Oct. 2017. arXiv:1710.11422 [cond-mat].

[34] H. Touchette. *Equivalence and Nonequivalence of the Microcanonical and Canonical Ensembles: A Large Deviations Study.* Doctoral Thesis, McGill University, Montréal, 2003.

[35] H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1-3):1–69, July 2009. arXiv:0804.0327 [cond-mat].

[36] H. Touchette. Ensemble equivalence for general many-body systems. *EPL (Europhysics Letters)*, 96(5):50010, Dec. 2011. arXiv:1106.2979 [cond-mat, physics:math-ph].

[37] H. Touchette. A basic introduction to large deviations: Theory, applications, simulations, Feb. 2012. arXiv:1106.4146 [cond-mat, physics:math-ph].

[38] H. Touchette. Equivalence and nonequivalence of ensembles: Thermodynamic, macrostate, and measure levels. *Journal of Statistical Physics*, 159(5):987–1016, June 2015. arXiv:1403.6608 [cond-mat].

[39] H. Touchette, R. S. Ellis, and B. Turkington. An Introduction to the Thermodynamic and Macrostate Levels of Nonequivalent Ensembles. *Physica A: Statistical Mechanics and its Applications*, 340(1-3):138–146, Sept. 2004. arXiv:cond-mat/0404655.

[40] L. V. Tran, V. H. Vu, and K. Wang. Sparse random graphs: Eigenvalues and eigenvectors. *Random Structures & Algorithms*, 42(1):110–134, Jan. 2013.

[41] E. P. Wigner. On the Distribution of the Roots of Certain Symmetric Matrices. *The Annals of Mathematics*, 67(2):325, Mar. 1958.

[42] N. C. Wormald. Models of Random Regular Graphs. In J. D. Lamb and D. A. Preece, editors, *Surveys in Combinatorics, 1999*, pages 239–298. Cambridge University Press, 1 edition, July 1999.