Spectral density of random graphs: convergence properties and application in model fitting

SUZANA DE SIQUEIRA SANTOS*
Universidade de São Paulo, Instituto de Matemática e Estatística, Departamento de Ciência da Computação, São Paulo, Brazil
Fundação Getulio Vargas, Escola de Matemática Aplicada, Rio de Janeiro, Brazil
*Corresponding author: suzana.santos@fgv.br

ANDRÉ FUJITA
Universidade de São Paulo, Instituto de Matemática e Estatística, Departamento de Ciência da Computação, São Paulo, Brazil

AND

CATHERINE MATIAS
Sorbonne Université, Université de Paris, Centre National de la Recherche Scientifique, Laboratoire de Probabilités, Statistique et Modélisation, Paris, France

[Received on 16 August 2021]

Random graph models are used to describe the complex structure of real-world networks in diverse fields of knowledge. Studying their behavior and fitting properties are still critical challenges, that in general, require model-specific techniques. An important line of research is to develop generic methods able to fit and select the best model among a collection. Approaches based on spectral density (i.e., distribution of the graph adjacency matrix eigenvalues) appeal to that purpose: they apply to different random graph models. Also, they can benefit from the theoretical background of random matrix theory. This work investigates the convergence properties of model fitting procedures based on the graph spectral density and the corresponding cumulative distribution function. We also review the convergence of the spectral density for the most widely used random graph models. Moreover, we explore through simulations the limits of these graph spectral density convergence results, particularly in the case of the block model, where only partial results have been established.

Keywords: random graphs, spectral density, model fitting, model selection, convergence.

1. Introduction

The study of real-world networks is fundamental in several areas of knowledge [15, 28, 35, 46]. Analyzing their structure is sometimes challenging, as it may change across time and instances from the same group/phenotype. For example, a functional brain network may change across time and individuals with the same phenotype. Therefore, we can think of a network as a random graph, which is a realization of a random process.

Several models were proposed to describe random processes that generate graphs. For instance, in the Erdős-Rényi (ER) random graph [18], each pair of nodes connects independently at random with probability p. We obtain a generalization when considering nodes i and j connect independently at random with non-homogeneous probability p_{ij}. Later, [22, 27, 42] proposed an intermediate setup

- block models - to generate graphs with non-homogeneous connection probabilities taking a finite number of values. In what follows, we distinguish between the (deterministic) block model (BM) and the stochastic block model (SBM). In the former, pairs of nodes connect independently. In the latter, the connections are only *conditionally* independent (given the nodes' groups).

Models can also characterize different properties of networks. Examples include networks with spatial associations between nodes [geometric random graph - GRG, 36], a regular structure (d-regular graph - DR), small-world structure [Watts-Strogatz - WS, 48], a power-law distribution of the vertex degrees [Barabási-Albert - BA, 6], or occurrence probability depending on pre-defined summary statistics and covariates [exponential random graph model, ERGM, 11, 23].

Given an empirical network, a significant task is to fit a collection of models through parameter estimation and select the best model in the collection. There are well-established procedures for estimating the model's parameters [1, 11, 42] for some models, such as the ER, SBM, and ERGM. However, those estimators present a lack of generality; they are specific for each model. Every time a new complex network model is proposed, it becomes necessary to develop a specific parameter estimator. Besides, the next question is, among the proposed set of models, which one best fits the observed network? The usual statistical answer relies on model selection criteria in some setups (e.g., ER, SBM) but not all, and are again specific to the model.

In this context, Takahashi et al. proposed fitting and selecting random graph models based on the Kullback-Leibler divergence between graph spectral densities [43]. The interest of their approach has been validated on synthetic and real datasets. The rationale behind this idea is that there is a close relationship between the graph structure and the graph spectrum [47].

The empirical spectral density characterizes some random graph/matrix ensembles. For example, Wigner et al. [50] showed analytically (semicircle law) for ER and DR graphs. In contrast, Takahashi et al. [43] showed empirically for GRG, BA, and WS graphs.

The semicircle law states that a large class of random symmetric matrices' spectral density converges weakly almost surely to a specific distribution called the semicircle distribution [2, 24, 50]. We can apply this result to the ER and d-regular random graphs under certain conditions [44]. However, despite the advances in the study of the convergence of graph spectral densities, we know little about the theoretical properties of the procedures proposed by Takahashi et al.

In this work, we prove that if we replace the Kullback-Leibler divergence between spectral densities by the ℓ_1 distance, the parameter estimated by the procedure proposed by Takahashi et al. [43] is consistent under the general assumptions:

1. the empirical spectral density converges weakly to a limiting distribution,
2. the map between the parameter space and the limiting distribution is injective and continuous, and
3. the parameter space is compact (this technical condition could be relaxed).

Furthermore, we show that we can extend the results when considering the eigenvalues' cumulative distribution to fit the models. Note that to obtain theoretical support on the procedures, we need to consider slight modifications of these, e.g., replacing the Kullback-Leibler divergence with the ℓ_1 distance. This is due to technicalities that may not be overcome yet. For example, convergence in total variation norm is necessary for convergence of the Kullback-Leibler divergence (see Corollary 3.1 from [37] and Lemma 2.4 from [38]). This means that theoretical support for the original procedures by [43] would require stronger assumptions on the convergence of the empirical spectral density. Our theoretical results (based on the ℓ_1 distance) are the first to pave the road to explaining the good performance of these procedures.

In Section 2, we present the notation and main definitions used in this paper. Section 3 summarizes the literature on convergence results for the random graph's spectral density under different models. We describe a model-fitting procedure (that slightly differs from [43]) in Section 4. We establish its convergence properties in Section 5; see Theorems 5.1 and 5.2. Then, in Section 6, we evaluate the method performance by simulation experiments. We illustrate applications on real-world data in Section 7. Finally, we discuss the results in Section 8 and show our conclusions in Section 9.

2. Notation and definitions

Let $G = (V,E)$ be an undirected graph in which $V = \{1,2,\ldots,n\}$ is a set of vertices, and E is a set of edges connecting the elements of V. The spectrum of G is the set of eigenvalues of its adjacency matrix $\mathbf{A} = (\mathbf{A}_{ij})_{1\leqslant i,j\leqslant n}$. Since G is undirected, $\mathbf{A}$ is symmetric, and then its eigenvalues are real values $\lambda_1^G \geqslant \lambda_2^G \geqslant \ldots \geqslant \lambda_n^G$.

Let δ_x be the Dirac measure at point x. Then, we define the empirical spectral distribution (ESD) as

$$\mu^G = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i^G/\sqrt{n}}. \tag{2.1}$$

Let $\mathbf{1}\{A\}$ be the indicator function of set A. Then, the corresponding empirical eigenvalues' cumulative distribution function (CDF) is

$$F^G(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\Big\{\frac{\lambda_i^G}{\sqrt{n}} \leqslant x\Big\}. \tag{2.2}$$

Let $f : \mathbb{R} \mapsto \mathbb{R}$ be a continuous and bounded function. If we integrate μ^G over f we obtain

$$\mu^G(f) = \int_{\mathbb{R}} f(\lambda)\mu^G(d\lambda) = \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i^G/\sqrt{n}). \tag{2.3}$$

Let $(G_n)_{n\geqslant 1}$ be a sequence of graphs on the (sequence of) sets of vertices V (in other words, each G_n is a graph over n nodes). We say that $(\mu^{G_n})_{n\geqslant 1}$ *converges weakly* to μ (and denote $\mu^{G_n} \Rightarrow \mu$) if for any real bounded and continuous function $f : \mathbb{R} \mapsto \mathbb{R}$, we have $\mu(f) = \lim_{n\to\infty} \mu^{G_n}(f)$.

In what follows, we will consider a (kernel) function ϕ and introduce for any $x \in \mathbb{R}$ the function $\phi_x(\cdot) = \phi(x - \cdot)$. In this case, the quantity in Eq (2.3) becomes a convolution function:

$$x \mapsto \mu^{G_n}(\phi_x) = \int_{\mathbb{R}} \phi(x-\lambda)\mu^{G_n}(d\lambda) = \frac{1}{n}\sum_{i=1}^{n} \phi(x - \lambda_i^{G_n}/\sqrt{n}). \tag{2.4}$$

A *random graph* is a random variable (r.v.) taking its values in the set of graphs. A *random graph model* is a collection of graphs that are either finite or countable, together with a probability distribution P on this collection. Let $(\mathscr{G}_n)_{n\geqslant 1}$ be a sequence of random graphs on the set of vertices $V = \{1,2,\ldots,n\}$. Then the eigenvalues of $\mathscr{G}_n$ are random variables, and the ESD $\mu^{\mathscr{G}_n}$ is a random probability measure. Now, there are many ways in which the random measure $\mu^{\mathscr{G}_n}$ may converge weakly. We say that $\mu^{\mathscr{G}_n}$ converges weakly in expectation to a probability measure μ if for any real bounded and continuous function $f : \mathbb{R} \mapsto \mathbb{R}$, we have

$$\mathbb{E}\mu^{\mathscr{G}_n}(f) \underset{n\to\infty}{\to} \mathbb{E}\mu(f).$$

Similarly, we say that $\mu^{\mathscr{G}_n}$ converges weakly in P-probability to a probability measure μ if for any real bounded and continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ and any $\varepsilon > 0$, we have that,

$$P\Big(|\mu^{\mathscr{G}_n}(f) - \mu(f)| > \varepsilon\Big) \to 0,$$

as $n \to \infty$. Finally, if

$$P\left(\lim_{n\to\infty} \mu^{\mathscr{G}_n}(f) = \mu(f)\right) = 1$$

for any bounded and continuous $f : \mathbb{R} \mapsto \mathbb{R}$, we say that $\mu^{\mathscr{G}_n}$ converges weakly P-almost surely to μ.

Let f and g be functions of n. Then, as $n \to \infty$, we say that $f = O(g)$ if $|f|/|g|$ is bounded from above; $f = o(g)$ if $f/g \to 0$; and $f = \omega(g)$ if $|f|/|g| \to \infty$.

Finally, we introduce the Stieltjes transform [5], which is an important tool for studying the limiting spectral distribution of random matrices. Given a probability distribution μ, its Stieltjes transform s_μ is a function on the upper-half complex plane defined for $z \in \mathbb{C}^+$ as

$$s_\mu(z) = \int_{\mathbb{R}} \frac{d\mu(x)}{x - z}.$$

3. Random graph models' spectral density

In this section, we summarize results on the spectral density of random graphs. First, we describe Wigner's law (also called semicircle law) for random symmetric matrices. We show examples of random graph models for which the law is valid. Then we summarize analytical and empirical results on the spectral density under different models when Wigner's law does not hold.

3.1 *Wigner's law*

Many results on the limiting ESD of random graphs rely on Wigner's law for random symmetric matrices. Wigner [50] proved that for symmetric matrices whose entries are real-valued independent and identically distributed (i.i.d.) random variables, with mean zero and unit variance, the ESD converges weakly in expectation to the semicircle law μ_{sc} defined as

$$\mu_{sc}(dx) = \frac{1}{2\pi}\sqrt{4 - x^2}\mathbf{1}\{|x| \leqslant 2\}dx. \tag{3.1}$$

Later, Grenander [24] proved that this convergence holds weakly in probability. Finally, Arnold [2] proved that (under the same assumptions) the ESD converges weakly almost surely to the semicircle distribution.

3.2 *Erdős-Rényi random graph model (ER)*

One of the simplest examples of random graph models in terms of construction is the model proposed by [18]. Given a probability p, and the number of vertices n, each pair of vertices is connected with probability p, independently of the other pairs.

Let $(\mathscr{G}_{p,n})_{n\geqslant 1}$ be a sequence of ER random graphs, and $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ denote the eigenvalues of the adjacency matrix $\mathbf{A}$ (here, we do not stress that $\mathbf{A} = \mathbf{A}_n$ depends on n). Based on Wigner's semicircle law, it can be proved that if $p = \omega(\frac{1}{n})$, then the ESD of the scaled matrix $\mathbf{A}/\sqrt{np(1-p)}$, given by

$$\tilde{\mu}^{\mathscr{G}_{p,n}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{np(1-p)}}, \tag{3.2}$$

converges weakly almost surely (and also in probability and expectation) to the semicircle distribution [44]. Another formulation says the ER graph $\mathscr{G}_{p,n}$ ESD, denoted by $\mu^{\mathscr{G}_{p,n}}$ and defined as in Equation

(2.1), converges to

$$\mu_{p,sc}(dx) = \frac{1}{2\pi p(1-p)}\sqrt{4p(1-p)-x^2}\mathbf{1}\{|x| \leqslant 2\sqrt{p(1-p)}\}dx. \tag{3.3}$$

Note that if $np = O(1)$, the convergence no longer holds.

3.3 *(Deterministic) Block model (BM)*

The (deterministic) block model was introduced in the 70's by [10, 49]; see also [20]. As already mentioned, a natural generalization of the ER model is to consider that any pair of nodes i, j connect independently but with its probability p_{ij}. From a statistical point of view, such a model has too many parameters. We can interpret the block model as a particular case where we restrict these p_{ij} to a finite number of values. More precisely, each node belongs to one of the M groups, determining its probability of connection. That is, pairs of nodes connect with probabilities $p_{m,l}$, which depend on the respective groups m, l of these nodes.

In what follows, we consider, more specifically, the model studied in [4]. Let $\mathscr{G}$ be a random graph with n nodes $V = \{1,2,\ldots,n\}$ belonging to M groups $\{\Omega_1,\Omega_2,\ldots,\Omega_M\}$ with equal size $K = n/M$. Assume that the conditional probabilities of connections are as follows

$$P(\mathbf{A}_{ij} = 1 | i \in m, j \in l) = \begin{cases} p_0 & \text{if } m \neq l\,, \\ p_m & \text{if } m = l\,. \end{cases}$$

Let $\mathbf{A}$ denote the adjacency matrix of $\mathscr{G}$ and $\bar{\mathbf{A}}$ denote its expectation matrix. Let $i \in \Omega_m$ and $j \in \Omega_l$ be two vertices of $\mathscr{G}$, then the $n \times n$ matrix $\bar{\mathbf{A}}$ is such that $\bar{\mathbf{A}}_{ij} = p_m$, if $m = l$, and $\bar{\mathbf{A}}_{ij} = p_0$, otherwise. Denote $p^\star = \max_{1\leqslant m\leqslant M} p_m$, and $\gamma(n) = 1/\sqrt{np^\star(1-p^\star)}$. The centered and normalized adjacency matrix of $\mathscr{G}$ is defined as

$$\tilde{\mathbf{A}} = \frac{(\mathbf{A} - \bar{\mathbf{A}})}{\gamma(n)}. \tag{3.4}$$

An expression for the Stieltjes transform of the spectral density of $\tilde{\mathbf{A}}$ was obtained by [4], as described in Proposition 3.1.

PROPOSITION 3.1 (COROLLARY 3 FROM [4]). If $\lim_{n\to+\infty} np_m = +\infty$ and $p_m/p_0 \leqslant c$ for some constant $c > 0$ and for all $m = 1,2,\ldots,M$, then, almost surely, the ESD associated with $\tilde{\mathbf{A}}$ converges weakly to a distribution function whose Stieltjes transform is

$$s(z) = \sum_{m=1}^{M} c_m(z),$$

where the functions $(c_m)_{1\leqslant m\leqslant M}$ are the unique solutions to the system of equations

$$c_m(z) = \frac{-1/M}{z + \zeta_m c_m(z) + \zeta_0 \sum_{l\neq m} c_l(z)}, \quad 1 \leqslant m \leqslant M,$$

with

$$\zeta_l = \lim_{n\to+\infty} \frac{p_l(1-p_l)}{p^\star(1-p^\star)}, \quad 0 \leqslant l \leqslant M,$$

that satisfy the conditions

$$\Im(c_m(z))\Im(z) > 0, \text{ for } \Im(z) > 0, \quad 1 \leqslant m \leqslant M,$$

where $\Im(z)$ denotes the imaginary part of z.

The recent paper [52] proposes explicit moment formula for the limiting distribution. Notice that the results from [4] were obtained only for the centered matrix $\tilde{\mathbf{A}}$. For the Erdős-Rényi model, Tran et al. [44] proved that the centered and non-centered adjacency matrices respective spectral densities are approximately identical. To investigate whether the BM graphs have a similar property, we did simulation experiments comparing the ESD of $\tilde{\mathbf{A}}$ and $\mathbf{A}/\gamma(n)$. We did simulations to obtain the ESD under various scenarios. We show the results in Section 6.1.

3.4 *Stochastic block model (SBM)*

Similarly to the previous model, the *stochastic block model (SBM)* produces graphs with groups of vertices connected with a particular edge density [21, 27, 42]. The main difference between SBM and the deterministic block model lies in a random assignation of each node to one of the groups with some pre-specified group probability. This results in groups with random sizes, while the latter are fixed in BM. Moreover, the entries of an adjacency matrix in BM are independent random variables, while for SBM, these are conditionally independent random variables. Handling the non-independence of the entries of the adjacency matrix in SBM is more challenging. Note that many authors abusively call SBM what is only a BM. More precisely, given a set of vertices $V = \{1, \ldots, n\}$, each vertex first picks a group at random among M possibilities independently of the others and with probabilities $(\pi_1, \pi_2, \ldots, \pi_M)$. Then conditional on these latent (unobserved) groups, two vertices from groups m, l connect independently of the others with probability p_{ml}. To our knowledge, no theoretical results about the convergence of the ESD have been obtained for the SBM yet.

3.5 *Configuration random graph model*

The configuration model is a particular instance of the inhomogeneous ER model and has at least two different usages in the literature. Let $\underline{d} = (d_1, \ldots, d_n)$ be a degree sequence of a graph.

The fixed-degree model $FD(\underline{d})$ is the collection of all graphs with this degree sequence and uniform probability on this collection. We obtain this model by simulation: start with a real graph from degree sequence $\underline{d}$ and use a rewiring algorithm to shuffle the edges.

The random-degree model $RD(\underline{d})$ considers all graphs over n nodes such that the entries $\mathbf{A}_{ij}$ are independent with (nonidentical) Bernoulli distribution $\mathscr{B}(p_{ij})$ where $p_{ij} = d_i d_j / C$ and $C > 0$ is a normalizing constant chosen such that $0 \leqslant p_{ij} \leqslant 1$ (for instance, take $C = \max_{i \neq j} d_i d_j$). It is a random graph model where graphs have degrees only approximately given by $\underline{d}$. Indeed, if D_i denotes the random degree of node i, then

$$\mathbb{E}(D_i) = \sum_{j \neq i} \mathbb{E}(\mathbf{A}_{ij}) = \sum_{j \neq i} p_{ij} = \frac{d_i}{C} \sum_{j \neq i} d_j = \frac{d_i(2|E| - d_i)}{C}.$$

Whenever d_i is not too large and $C \simeq 2|E|$ we get $\mathbb{E}(D_i) \simeq d_i$.

Let $\mathbf{A}'$ denote the normalized adjacency matrix:

$$\mathbf{A}' = \frac{\mathbf{A}}{\sqrt{n} \sum_{i=1}^{n} d_i}.$$

Preciado and Rahimian [39] proved that for the random-degree model proposed by Chung and Lu [12], under specific conditions regarding the expected degree sequence (sparsity, finite moments, and controlled growth of degrees), the ESD of the normalized adjacency matrix $\mathbf{A}'$ converges weakly almost surely to a deterministic density function. Although the analytical expression of this function is unknown, the authors show its moments explicitly.

3.6 *d-regular random graph (DR)*

A d-regular random graph is a random graph over n nodes with all degrees equal to d. In this case, the parameter d is then directly determined by the observed graph. Although there is no need to estimate d, studying this random graph's spectral density can help obtain a measure of goodness of fit. This can be useful for graphs that "are almost" regular, in which degrees vary due to measurement errors.

If $d \geqslant 3$ is fixed with n, and $(\mathscr{G}_{d,n})_{n\geqslant 1}$ is a sequence of d-regular random graphs, then the ESD defined as

$$\tilde{\mu}^{\mathscr{G}_{d,n}} = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i} \tag{3.5}$$

converges weakly almost surely to

$$\mu_d(dx) = \frac{d\sqrt{4(d-1)-x^2}}{2\pi(d^2-x^2)}\mathbf{1}\{|x| \leqslant 2\sqrt{d-1}\}dx. \tag{3.6}$$

Note that $\tilde{\mu}^{\mathscr{G}_{d,n}}$ is the distribution of the eigenvalues of $\mathbf{A}$, without scaling by $\sqrt{n}$. The result dates back to the work [33]. It is also valid when considering that d grows slowly with n [17, 44]. The distribution μ_d is known as a Kesten-McKay distribution. The result may also be rephrased in the following way [7]: the ESD of the matrix $\mathbf{A}/\sqrt{d-1}$ converges weakly almost surely to

$$\tilde{\mu}_d(dx) = \left(1 + \frac{1}{d-1} - \frac{x^2}{d}\right)^{-1} \frac{\sqrt{4-x^2}}{2\pi}\mathbf{1}\{|x| \leqslant 2\}dx. \tag{3.7}$$

Dumitriu and Pal [17] showed that the ESD of $\mathbf{A}/\sqrt{d-1}$ converges in probability to μ_{sc} whenever d tends to infinity slowly, namely $d = n^{\varepsilon_n}$ with $\varepsilon_n = o(1)$.

Let $\mathbf{J}$ be a $n \times n$ matrix containing only ones, and set

$$\mathbf{M} = \left[\frac{d}{n}\left(1 - \frac{1}{n}\right)\right]^{-1/2}\left(\mathbf{A} - \frac{d}{n}\mathbf{J}\right). \tag{3.8}$$

Tran et al. [Theorem 1.5 in 44] showed that the ESD of $\mathbf{M}/\sqrt{n}$ converges weakly to the semicircle distribution μ_{sc} when d tends to infinity with n.

3.7 *Geometric random graph model*

To construct a geometric random graph (GRG), we randomly draw n vertices on a D-dimensional unit cube (or torus) and connect pairs of vertices whose distance is at most r_n (with $r_n \to 0$). A statistical motivation for this model is that the nodes' coordinates represent (latent) attributes. The metric captures the similarity between individuals and translates into connectivity in the graph.

The following results concern the ESD of (unscaled) adjacency matrix in GRG under different regimes. Bordenave [9] proved its convergence when $nr_n^D \to c \in (0,\infty)$, which corresponds to the thermodynamic regime. He also proposed an approximation of the limit when the constant c becomes large.

Later, Hamidouche et al. [26] proved that in the so-called connectivity regime, when $\left(\frac{n}{\log n}\right) r_n^D \to c \in (0,\infty)$ as $n \to \infty$, the ESD converges to the ESD of a deterministic geometric graph. Finally, Bordenave [9] also characterizes the ESD of the adjacency matrix, when normalized by n, in the dense regime, i.e, when r_n^D scales as a constant (namely $r_n^D = O(1)$ and $r_n^D = \omega(1)$).

3.8 *Preferential attachment models*

Preferential attachment random graph models dynamically generate graphs by successively adding a new node in a network together with a set of edges. Probability of connecting the new vertex to an existing node is proportional to its current degree. One of the best known preferential attachment model is the Barabási-Albert (BA) one [6]. In this model, the probability of connecting a vertex to other existing nodes is proportional to their degrees to power a scaling exponent p_s.

There is no theoretical result about the convergence of the empirical spectral density for the Barabási-Albert model. However, [19] has described the shape of its distribution through simulations.

In the linear preferential attachment tree model [8], the probability of connecting a new vertex is proportional to the out-degree plus $1+a$, where a is a process parameter. Under certain conditions, the properly rescaled ESD converges to a deterministic probability function as $n \to +\infty$ [8].

3.9 *Watts-Strogatz random graph model*

Watts-Strogatz (WS) random graphs [48] present small-world properties, such as short average path lengths (i.e. the majority of vertices can be reached from all other vertices by a small number of steps) and a higher clustering coefficient (i.e. the number of triangles in the graph) than ER random graphs. The process generating WS graphs starts with a regular lattice of size n, with each vertex connected to the $K/2$ nearest vertices on each ring's side. For each vertex i, and each edge connected to i, with probability p_r, it replaces the edge by a new one connecting i to a vertex chosen at random. To the best of our knowledge, all results regarding the limiting ESD of WS graphs are empirical [19].

3.10 *Miscellenaous*

Our list of random graphs models is not exhaustive and focuses on the most common ones. We conclude it by mentioning a very recent result [3] on the soft geometric block model. This model combines the block model approach and geometric random graphs. In this reference, the authors describe the limiting ESD in terms of the Fourier transform [see Theorem 1 in 3].

4. Fitting a model

Consider a parameterized random graph model $\{P_\theta; \theta \in \Theta\}$ where P_θ is a probability distribution on a collection of graphs, described through the parameter $\theta \in \Theta$, $\Theta \subset \mathbb{R}^d$. Let $\mathscr{G}$ be a random graph from some distribution P (that might belong or not to the parametric model $\{P_\theta; \theta \in \Theta\}$). We want to estimate a parameter θ such that P_θ is close to P. In many random graphs models, the ESD of graphs generated under the distribution P_θ converges weakly in some sense (for P_θ) to some distribution that we denote μ_θ whenever this limit exists.

ESDs are discrete distributions. In general, the limits μ_θ are absolutely continuous measures (that is, they have densities with respect to the Lebesgue measure on the spectrum set Λ). It is thus natural to consider a kernel estimator of the spectrum's density as defined in Eq (2.4). We denote by $\phi_{x,\sigma}(\cdot) =$

$\sigma^{-1}\phi(\frac{x-\cdot}{\sigma})$ a kernel function (a function such that $\int \phi(x)dx = 1$) with parameters $x \in \mathbb{R}$ and $\sigma > 0$. The kernel estimator of the spectral density with bandwidth σ is defined as

$$\mu^{\mathscr{G}}(\phi_{x,\sigma}) = \frac{1}{n\sigma}\sum_{i=1}^{n}\phi\left(\frac{x-\lambda_i^{\mathscr{G}}/\sqrt{n}}{\sigma}\right).$$

Suppose the bandwidth $\sigma = \sigma_n$ goes to 0 as n tends to infinity. In that case, this estimator should approximate the limiting spectral density μ_θ of the graph. Note that we choose to focus on the ESD derived from the normalized adjacency matrix, with normalization (or scaling) $\sqrt{n}$, while the literature review from Section 3 contains results with different normalization. This point is further investigated in Section 6.

Now, given a divergence measure D on the set of probability measures over the spectrum of the graphs, we estimate θ by solving the minimization problem

$$\hat{\theta} = \underset{\theta\in\Theta}{\operatorname{argmin}} D(\mu^{\mathscr{G}}(\phi_{\cdot,\sigma}), \mu_\theta(\phi_{\cdot,\sigma})). \tag{4.1}$$

We either use an analytical expression for μ_θ (when available) or rely on Monte Carlo estimates. In the latter case, we sample M graphs $\mathscr{G}_1, \mathscr{G}_2, \cdots, \mathscr{G}_M$ from the distribution P_θ. Then, for each graph $\mathscr{G}_m$, we numerically obtain $\mu^{\mathscr{G}_m}(\phi_{x,\sigma})$ and we estimate $\mu_\theta(\phi_{x,\sigma})$ by

$$\hat{\mu}_\theta(\phi_{x,\sigma}) = \frac{1}{M}\sum_{m=1}^{M}\mu^{\mathscr{G}_m}(\phi_{x,\sigma}) = \frac{1}{nM\sigma}\sum_{m=1}^{M}\sum_{i=1}^{n}\phi\left(\frac{x-\lambda_i^{\mathscr{G}_m}/\sqrt{n}}{\sigma}\right).$$

The parameter estimator problem (4.1) is based on the assumption that the random graph $\mathscr{G}$ is generated by a known family of random graphs $\{P_\theta; \theta \in \Theta\}$ (the model). That is, we assume that $P \in \{P_\theta; \theta \in \Theta\}$. However, usually, the model $\{P_\theta; \theta \in \Theta\}$ is unknown. Then given a list of random graph models, we choose the one with the best fit according to some objective criterion. If we consider random graphs models with the same complexity (number of parameters), we choose the model with the smallest divergence (D) between the spectral densities, as described in Algorithm 4.1. Otherwise, we penalize the divergence D by a function of the number of parameters to prevent from overfitting.

A previous work by [43] uses AIC as a criterion to choose a model. Their approach for problem (4.1) is based on the Kullback-Leibler divergence $KL(\mu^{\mathscr{G}}(\phi_x), \mu_\theta(\phi_x))$, where ϕ is the Gaussian kernel. In this case, we obtain

$$D(\mu^{\mathscr{G}}(\phi_{x,\sigma}), \mu_\theta(\phi_{x,\sigma})) = \frac{1}{n\sigma}\sum_{i=1}^{n}\int \phi(x-\lambda_i^{\mathscr{G}}/\sqrt{n})\log\left(\frac{\sum_{i=1}^{n}\phi(x-\lambda_i^{\mathscr{G}}/\sqrt{n})}{n\sigma\mu_\theta(\phi_{x,\sigma})}\right)dx.$$

Algorithm 4.1 Procedure for fitting and selecting a random graph model.

Input: Graph $\mathscr{G}$, a list of random graph models $\{P_\theta^i; \theta \in \Theta^i \subset \mathbb{R}^d\}$, a finite subset $\tilde{\Theta}^i \subset \Theta^i$, for $i = 1, 2, \ldots, N$.

Output: Return model and parameter with best fit.

Min $\leftarrow +\infty$

Argmin $\leftarrow (0,0)$

Compute the kernel density estimator $\mu^{\mathscr{G}}(\phi_{\cdot,\sigma})$.

for each parameterized random graph model $\{P_\theta^i; \theta \in \Theta^i\}$, $i = 1, 2, \cdots, N$, **do**

  for each $\theta^j \in \tilde{\Theta}^i$ **do**

if the limiting ESD from $P^i_{\theta^j}$, denoted by $\mu^i_{\theta^j}$, is known analytically **then**
$D_{i,j} \leftarrow D(\mu^{\mathscr{G}}(\phi_{\cdot,\sigma}), \mu^i_{\theta^j}(\phi_{\cdot,\sigma}))$.
else
Sample M graphs $\mathscr{G}_1, \mathscr{G}_2, \cdots, \mathscr{G}_M$ from $P^i_{\theta^j}$.
for each graph $\mathscr{G}_m$ **do**
Compute the kernel density estimator $\mu^{\mathscr{G}_m}(\phi_{\cdot,\sigma})$.
end for
$\hat{\mu}^i_{\theta^j}(\phi_{\cdot,\sigma}) \leftarrow \frac{1}{M}\sum_{m=1}^{M} \mu^{\mathscr{G}_m}(\phi_{\cdot,\sigma})$.
$D_{i,j} \leftarrow D(\mu(\phi_{\cdot,\sigma}), \hat{\mu}^i_{\theta^j}(\phi_{\cdot,\sigma}))$
end if
if $D_{i,j} <$ Min **then**
Argmin $\leftarrow (i,j)$
Min $\leftarrow D_{i,j}$
end if
end for
end for
return Argmin.

Now we focus on D being the ℓ_1 distance so that we have

$$D(\mu^{\mathscr{G}}(\phi_{\cdot,\sigma}), \mu_\theta(\phi_{\cdot,\sigma})) = \|\mu^{\mathscr{G}}(\phi_{\cdot,\sigma}) - \mu_\theta(\phi_{\cdot,\sigma})\|_1 = \int |\mu^{\mathscr{G}}(\phi_{x,\sigma}) - \mu_\theta(\phi_{x,\sigma})| dx.$$

We also suppose that the ESD from the parameterized random graph model is known analytically. In that case, the convergence properties of $\hat{\theta}$ are described in the next section.

5. Convergence properties

In what follows, we consider as divergence the ℓ_1 distance and denote it $||\cdot||_1$. We also denote $\mathscr{M}_1(\Lambda)$ the set of probability measures over the spectrum Λ.

THEOREM 5.1 Let $\{P_\theta; \theta \in \Theta\}$ denote a parameterized random graph model and assume that for any $\theta \in \Theta$, there exists μ_θ the limiting ESD with respect to weak, almost sure convergence. We assume that $\mu_\theta \in \mathscr{M}_1(\Lambda)$, where Λ is a bounded set. Consider $(\mathscr{G}_n)_{n\geqslant 1}$ a sequence of random graphs from distribution $P_{\theta^\star}$. Let ϕ be a kernel and $\sigma = \sigma_n$ a bandwidth that converges to 0. If the map $\theta \in \Theta \mapsto \mu_\theta(d\lambda) \in \mathscr{M}_1(\Lambda)$ is injective, continuous and Θ is compact, then the minimizer

$$\hat{\theta}_n = \underset{\theta\in\Theta}{\operatorname{argmin}} ||\mu^{\mathscr{G}_n}(\phi_{\cdot,\sigma}) - \mu_\theta(\phi_{\cdot,\sigma})||_1$$

converges in probability to $\theta^\star$ as $n \to \infty$.

Proof. The proof is based on Theorem 3.2.8 from [14] about minimum contrast estimators. The first step is establishing the convergence of the contrast function $D_n(\theta, \theta^\star) := \|\mu^{\mathscr{G}_n}(\phi_{\cdot,\sigma}) - \mu_\theta(\phi_{\cdot,\sigma})||_1$. Assuming that μ_θ exists for any $\theta \in \Theta$, we get that for any $x \in \mathbb{R}$,

$$\mu^{\mathscr{G}_n}(\phi_{x,\sigma}) = \frac{1}{n}\sum_{i=1}^{n} \phi_{x,\sigma}(\lambda_i^{\mathscr{G}_n}/\sqrt{n}) \underset{n\to\infty}{\longrightarrow} \mu_{\theta^\star}(\phi_{x,\sigma}),$$

$P_{\theta^\star}$-almost surely. Then, from dominated convergence theorem, we have

$$D_n(\theta,\theta^\star) \underset{n\to\infty}{\longrightarrow} \|\mu_{\theta^\star}(\phi_{\cdot,\sigma}) - \mu_\theta(\phi_{\cdot,\sigma})\|_1 := D(\theta,\theta^\star),$$

$P_{\theta^\star}$-almost surely. Now, the second step is to establish that the limiting function $\theta \mapsto D(\theta,\theta^\star)$ has a strict minimum at $\theta^\star$. Notice that the limit $D(\theta,\theta^\star)$ is minimum if and only if $\mu_{\theta^\star}(\phi_{\cdot,\sigma}) = \mu_\theta(\phi_{\cdot,\sigma})$ almost surely (wrt the Lebesgue measure), namely

$$\forall x \in \mathbb{R}, \quad \int \phi[(x-\lambda)/\sigma]\mu_{\theta^\star}(d\lambda) = \int \phi[(x-\lambda)/\sigma]\mu_\theta(d\lambda).$$

Classical properties of kernels [see for e.g. Proposition 1.2 in 45] imply that letting $\sigma \to 0$, we get

$$\mu_\theta(d\lambda) = \mu_{\theta^\star}(d\lambda).$$

Since $\theta \mapsto \mu_\theta$ is injective, this implies $\theta = \theta^\star$. Thus, the limit $D(\theta,\theta^\star)$ is minimum if and only if $\theta = \theta^\star$.

Finally, the map $\theta \mapsto \mu_\theta(d\lambda)$ is continuous on the compact set Θ and thus uniformly continuous. This implies that

$$\forall \varepsilon > 0, \exists \eta > 0 \text{ such that if } \|\theta - \theta'\| \leqslant \eta \text{ then } |\mu_\theta(d\lambda) - \mu_{\theta'}(d\lambda)| \leqslant \varepsilon.$$

Thus we get that

$$\begin{aligned}
|D_n(\theta,\theta^\star) - D_n(\theta',\theta^\star)| &\leqslant \|\mu_\theta(\phi_{\cdot,\sigma}) - \mu_{\theta'}(\phi_{\cdot,\sigma})\|_1 \\
&\leqslant \sigma^{-1} \int\int_\Lambda \phi[(x-\lambda)/\sigma] \times |\mu_\theta(d\lambda) - \mu_{\theta'}(d\lambda)| dx \\
&\leqslant \int\int \phi(y) \times |\mu_\theta(x-\sigma y) - \mu_{\theta'}(x-\sigma y)| dy dx \\
&\leqslant \varepsilon |\Lambda|,
\end{aligned}$$

where $|\Lambda|$ denotes the length of the interval Λ. For clarity, we assumed in the last line that μ_θ is a measure with a density, but this has no consequence as soon as $\|\theta - \theta'\| \leqslant \eta$. This establishes a uniform convergence (wrt θ) of the contrast function D_n as n tends to infinity. Relying on Theorem 3.2.8 from [14], we obtain the desired result. □

We can establish a similar result whether we consider cumulative distribution functions (CDF) of eigenvalues instead of spectral densities.

THEOREM 5.2 Let $\{P_\theta; \theta \in \Theta\}$ denote a parameterized random graph model and assume that for any $\theta \in \Theta$, there exists μ_θ the limiting ESD with respect to weak almost sure convergence. We assume that $\mu_\theta \in \mathscr{M}_1(\Lambda)$, where Λ is a bounded set. Consider $(\mathscr{G}_n)_{n\geqslant 1}$ a sequence of random graphs from distribution $P_{\theta^\star}$. If the map $\theta \in \Theta \mapsto \mu_\theta(d\lambda) \in \mathscr{M}_1(\Lambda)$ is injective, continuous and Θ is compact, then the minimizer

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}} \|F^{\mathscr{G}_n} - F_\theta\|_1$$

converges in probability to $\theta^\star$ as $n \to \infty$.

Proof. The proof follows the same lines as the previous one, by first noting that for any x, the function $\lambda \mapsto \mathbf{1}\{\lambda \leqslant x\}$ is continuous almost everywhere. Thus

$$F^{\mathscr{G}_n}(x) \underset{n\to\infty}{\longrightarrow} F_\theta(x),$$

$P_{\theta^\star}$-almost surely. For the second step, notice that F_θ entirely characterizes the distribution μ_θ. Finally, we similarly prove that

$$\begin{aligned}|D_n(\theta,\theta^\star) - D_n(\theta',\theta^\star)| &\leqslant \|F_\theta - F_{\theta'}\|_1 \\ &\leqslant \int_\Lambda \int_\Lambda \mathbf{1}\{\lambda \leqslant x\}|\mu_\theta(d\lambda) - \mu_{\theta'}(d\lambda)|dx \\ &\leqslant \varepsilon|\Lambda|^2\end{aligned}$$

as soon as $\|\theta - \theta'\| \leqslant \eta$. □

6. Simulations

In what follows, we describe experiments to study the block model spectral density, illustrate Theorems 5.1 and 5.2, and evaluate the model selection procedures' performance. We used the Gaussian kernel to estimate the ESD and the Silverman's criterion [41] to choose the bandwidth σ in all experiments. To generate random graphs, we used package *igraph* for *R* [13].

First, in Section 6.1, we analyze the ESD of BM random graphs. Indeed, the results in [4] are limited to a specific BM and we empirically explore the validity of the convergence of the ESD in more general BM. Then, in Section 6.2, we analyze the parameter estimator's empirical behavior to confirm results given by Theorems 5.1 and 5.2 in scenarios in which a limiting spectral density exists and has a known analytical form. In Section 6.3, we further explore scenarios in which the limiting spectral density may not exist. The distribution of the random graph model's eigenvalues is estimated empirically, as described in Algorithm 4.1. Finally, in Section 6.4, we show experiments for the model selection procedure.

6.1 *Block Model spectral density*

We generated BM graphs randomly and compared the ESD of the adjacency matrix $\mathbf{A}/\gamma(n)$ with the centered matrix $\tilde{\mathbf{A}}$ given in (3.4) and the theoretical density derived by [4] that are recalled in Proposition 3.1. We considered five different scenarios, described as follows.

1. **Scenario 1.** Graph generated by the same model described in [4] with a small number of blocks ($M = 3$). We set each block size to $K = 300$. The probability of connecting vertices from different groups is set to $p_0 = 0.2$ and the probabilities within the groups (p_1, p_2, p_3) to $(0.8, 0.5, 0.6)$, respectively.

2. **Scenario 2.** Graph generated by the same model described in [4] with a larger number of groups ($M = 10$). We set each block size to $K = 300$. We set the probability to connect vertices from different groups to $p_0 = 0.2$ and the probabilities within the groups $(p_1, p_2, p_3, p_4, p_5, p_6, p_7, p_8, p_9, p_{10})$ to $(0.8, 0.5, 0.6, 0.7, 0.4, 0.9, 0.55, 0.42, 0.38, 0.86)$, respectively.

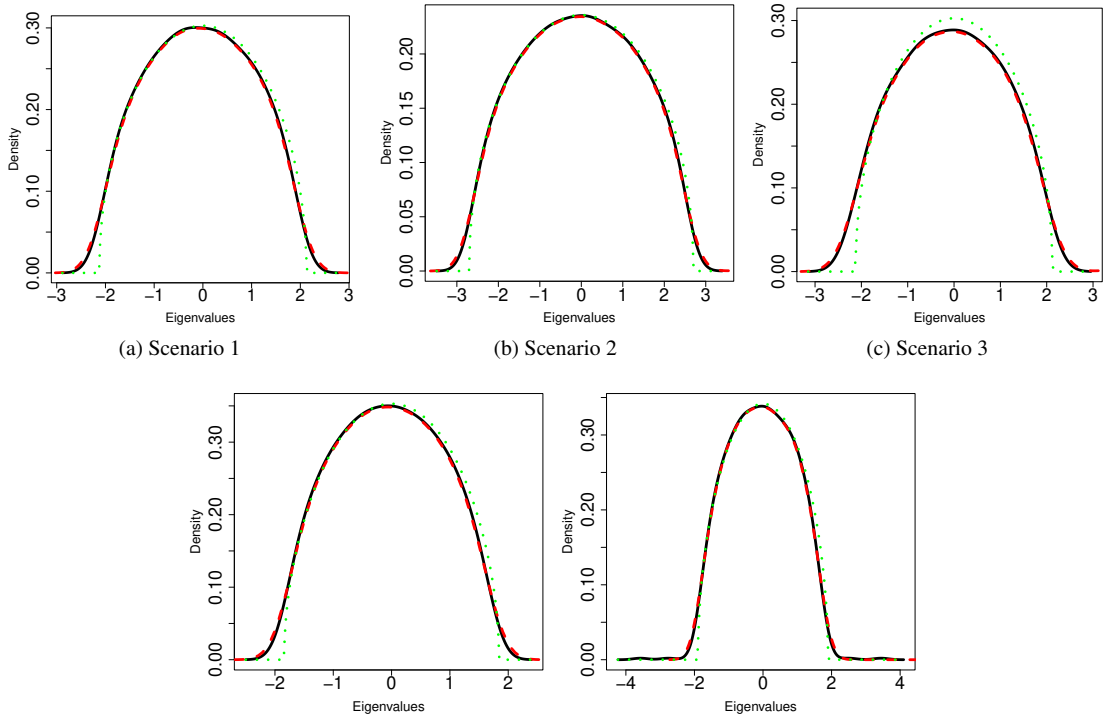
3. **Scenario 3.** Generalization of the model described by [4] for different block sizes. We considered $M = 3$ blocks of sizes 100, 80, and 300, respectively. We set the probability of connecting vertices from different groups to $p_0 = 0.2$, and the probabilities within the groups (p_1, p_2, p_3) to $(0.8, 0.5, 0.6)$, respectively.

4. **Scenario 4.** Generalization of the model described by [4] for p_0 larger than the probabilities within the groups. We considered $M = 3$ blocks of size $K = 300$. We set the probability of

connecting vertices from different groups to $p_0 = 0.9$ and the probabilities within the groups (p_1, p_2, p_3) to $(0.8, 0.5, 0.6)$, respectively.

5. **Scenario 5.** Generalization of the model described by [4] with different probabilities of connecting vertices from different groups. We considered $M = 3$ blocks of size $K = 300$. We set the probabilities of connecting vertices from different groups (previously p_0) to $p_{12} = 0.1$, $p_{13} = 0.2$, and $p_{23} = 0.05$. The probabilities within the groups (p_1, p_2, p_3) are set to $(0.8, 0.5, 0.6)$, respectively.

We compared the ESD of both the centered matrix $\tilde{\mathbf{A}}$ and the non-centered matrix $\mathbf{A}/\gamma(n)$ to [4]'s theoretical density in Figure 1. Note that the normalizing term $\gamma(n)$ depends on the unknown parameter value $p^\star$. The density is computed numerically by inverting the Stieltjes transform.



(a) Scenario 1 (b) Scenario 2 (c) Scenario 3

(d) Scenario 4 (e) Scenario 5

FIG. 1: Comparison between the empirical spectral distribution (ESD) of the block model for matrices $\tilde{\mathbf{A}}$ (solid black line), $\mathbf{A}/\gamma(n)$ (red dashed line), and the theoretical distribution derived by [4] (green dotted line). We considered different scenarios: (a) a small number of blocks, (b) a larger number of blocks, (c) blocks of different sizes, (d) probability between groups larger than the probabilities within the groups, and (e) different probabilities of connecting vertices from different groups.

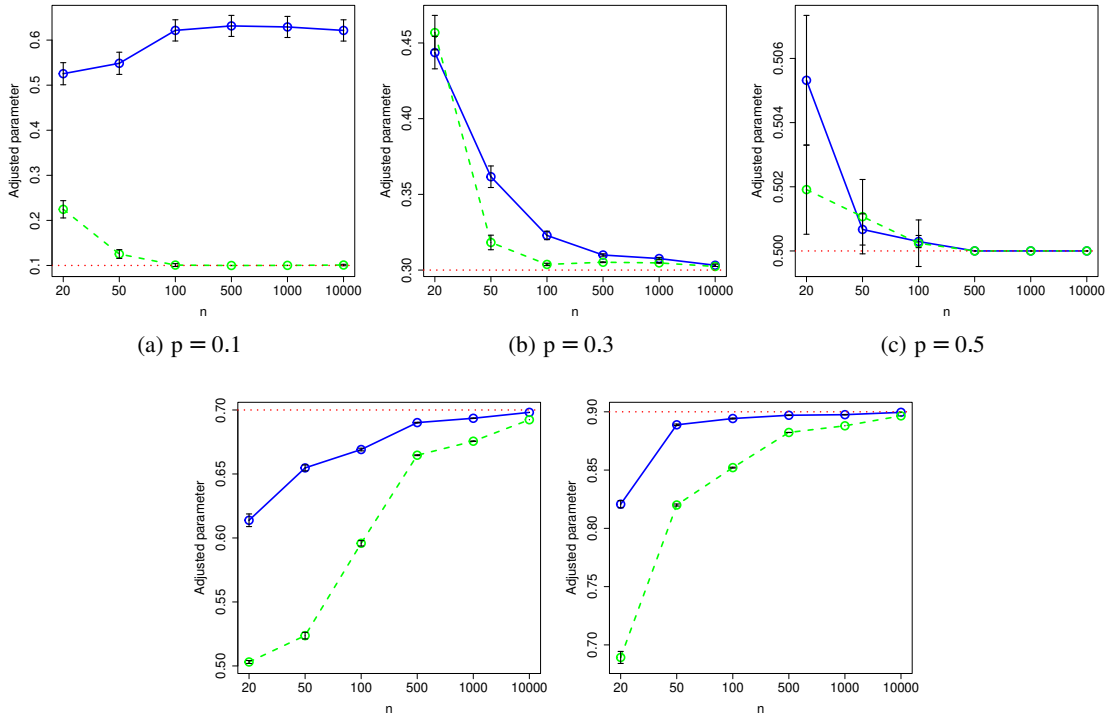
For all the scenarios considered here, the ESDs of $\mathbf{A}/\gamma(n)$ and $\tilde{\mathbf{A}}$ are very close, suggesting that their ESD converges to the same distribution. For scenarios 1, 2, 4, and 5, the ESDs of $\tilde{\mathbf{A}}$ and $\mathbf{A}/\gamma(n)$ are close to the theoretical distribution derived by [4], suggesting that Avrachenkov *et al.*'s results could

be further investigated for models with different probabilities to connect vertices from different groups, and large p_0. For scenario 3, the empirical densities are farther from the theoretical density. This might be due to numerical errors during ESD estimation or might indicate that Avrachenkov *et al.*'s results do not apply to graphs with blocks of different sizes.

6.2 *Illustration of our convergence theorems for ER and DR models*

To illustrate Theorems 5.1 and 5.2, we considered scenarios in which we know the ESD analytic form, such as the ER and d-regular random graph models. Note however that in the latter case, the convergence is known for the adjacency matrix with no scaling, or scaled with $\sqrt{d}$ while our theorems rely on a $\sqrt{n}$ normalization. All experiments described in this section rely on the *optimize* function from the *R* package *stats* to solve (4.1), with precision $\varepsilon = 10^{-8}$. It uses a combination of golden section search and successive parabolic interpolation.

For the ER model, we expect the ESD of $\mathbf{A}/\sqrt{n}$ to converge to the semicircle law when $p = \omega(n^{-1})$. In that case, the assumptions of Theorems 5.1 and 5.2 hold when $\Theta = [0, 0.5]$ or $\Theta = [0.5, 1]$, and then the ℓ_1 distance minimizer for problem (4.1) converges in probability to the true parameter. We generated 1 000 graphs of sizes $n = 20, 50, 100, 500, 1\,000, 10\,000$ with parameters $p = 0.1, 0.3, 0.5, 0.7, 0.9$. For $p < 0.5$, $p = 0.5$, and $p > 0.5$, we considered the intervals $\Theta = [0, 0.5]$, $\Theta = [0, 1]$, and $\Theta = [0.5, 1]$, respectively. In Figure 2, we show the average estimated parameter with a 95% confidence interval for the ER model by using the spectral density (blue solid line) and the cumulative distribution (green dashed line). The red dotted line indicates the true parameter.

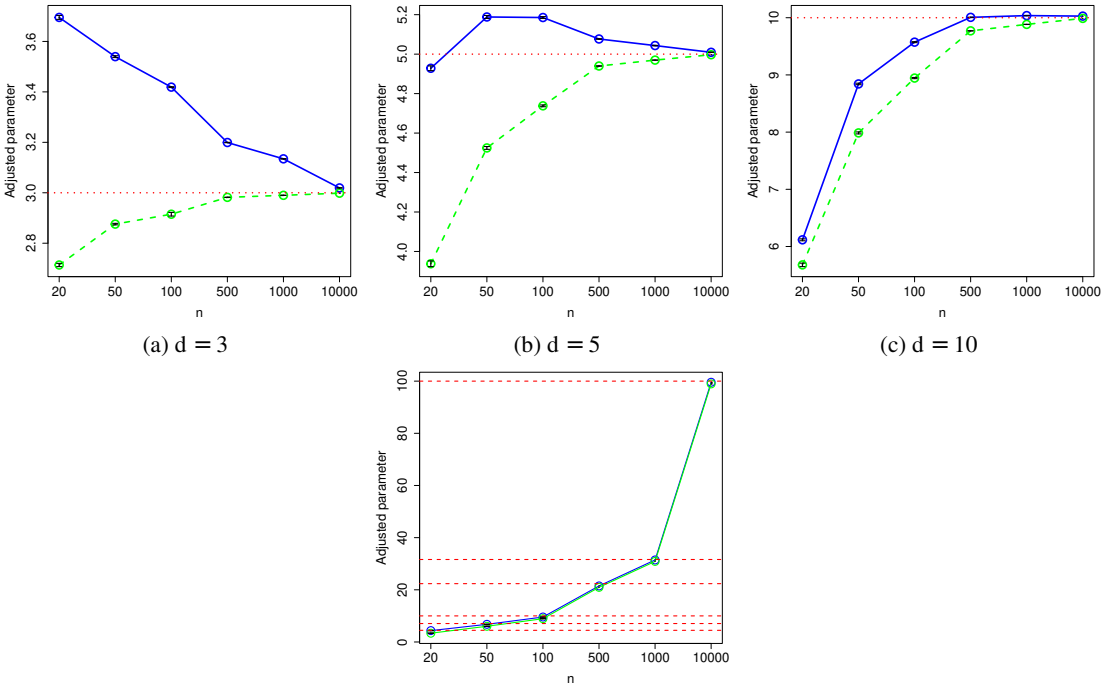


(a) p = 0.1 (b) p = 0.3 (c) p = 0.5

(d) p = 0.7 (e) p = 0.9

FIG. 2: Average estimated parameter for the ER model. We generated 1 000 Erdős-Rényi random graphs of sizes $n = 20, 50, 100, 500, 1\,000, 10\,000$, and probability of connecting two vertices $p = 0.1$ (a), $p = 0.3$ (b), $p = 0.5$ (c), $p = 0.7$ (d), and $p = 0.9$ (e). In each plot, the x-axis and the y-axis correspond, respectively, to the graph size (n) and estimated parameter. The points and error bars indicate the average estimated parameters and the corresponding 95% confidence intervals. The blue solid lines correspond to the results based on the ℓ_1 distance between the ESD and the limiting theoretical spectral density of the ER model (semicircle law). The green dashed lines correspond to the results based on the eigenvalues' cumulative distribution. The red dotted lines indicate the true value.

Similarly, we performed simulation experiments with d-regular random graphs in scenarios in which the ESD of $\mathbf{A}$ converges weakly almost surely to the Kesten-McKay law. In what follows, we slightly modified the procedure from Algorithm 4.1 to rely on the spectrum of unnormalized adjacency matrices $\mathbf{A}$ (rather than $\mathbf{A}/\sqrt{n}$). We generated 1 000 graphs of sizes $n = 20, 50, 100, 500, 1\,000, 10\,000$ with parameters $d = 3, 5, 10, \sqrt{n}$, namely fixed or slowly growing values d. Finally, we set $\Theta = [0, n]$. Figure 3 shows the average estimated parameter with a 95% confidence interval for the d-regular model by using the spectral density (blue solid line) and the cumulative distribution (green dashed line).



(a) d = 3 (b) d = 5 (c) d = 10

(d) d = $\sqrt{n}$

FIG. 3: Average estimated parameter for the DR model. We generated 1 000 d-regular random graphs of sizes $n = 20, 50, 100, 500, 1\,000, 10\,000$, and degree $d = 3$ (a), $d = 5$ (b), $d = 10$ (c), and $d = \sqrt{n}$ (d). In each plot, the x-axis and the y-axis correspond, respectively, to the graph size (n) and estimated parameter. The points and error bars indicate the average estimated parameters and the corresponding 95% confidence intervals. The blue solid lines correspond to the results based on the ℓ_1 distance between the ESD and the limiting theoretical spectral density of the DR model (Kesten-McKay law). The green dashed lines correspond to the results based on the eigenvalues' cumulative distribution. The red dotted lines indicate the true value.

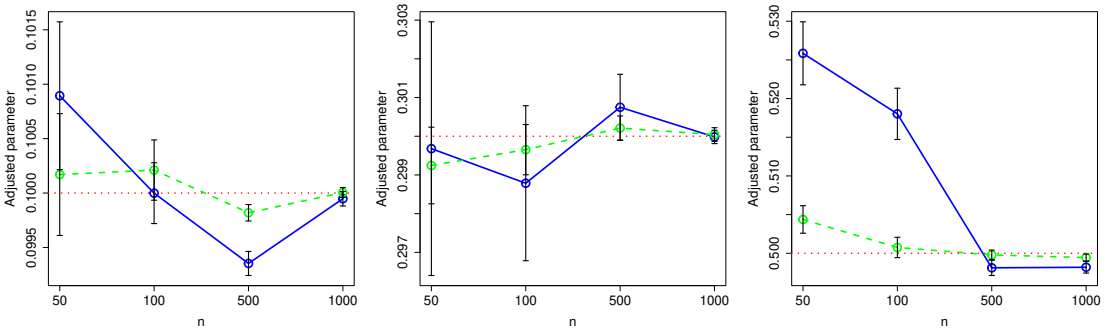
As expected by Theorems 5.1 and 5.2, we observe in Figures 2 and 3 that the estimated parameter approximates the true parameter as the number of vertices increases.

6.3 *Convergence of the parameter estimates in other scenarios*

In this section, we explore the performance of the parameter estimation procedures based on ℓ_1 distance (either through kernel or CDF). Our experiments include scenarios in which the assumptions of Theorems 5.1 and 5.2 may not hold, and a limiting spectral density does not exist, or it is unknown, such as the GRG, WS, and BA random graphs. In that case, to evaluate the model fitting procedure's performance based on the ℓ_1 distance, we estimate the model's spectral density empirically, as described in Algorithm 4.1. Our experiments also include BM graphs that [43] did not consider. For that model, we could compute the limiting spectral density numerically by inverting the Stieltjes transform (as is done in the previous section). However, as it is computationally slow, we empirically estimate the BM spectral density for this experiment.

We generated 1 000 graphs of sizes $n = 50, 100, 500, 1\,000$ with parameters $r = 0.1, 0.3, 0.5, 0.7, 0.9$ for the GRG model, $p_r = 0.1, 0.3, 0.5, 0.7, 0.9$ for WS, and $p_s = 1.1, 1.3, 1.5, 1.7, 1.9$ for BA. For the two first models, in the search grid, we considered values between 0 and 1, varying with a step of size 0.001. For the BA model, we considered the interval $[1, 4]$, and inspected values using a step size of 0.01. We considered two blocks of the same sizes and equal probabilities of connection within the groups for the BM graphs. We generated 1 000 graphs of sizes $n = 100, 600$. We set the probability of connecting vertices from different groups to $p_0 = 0.2$ and the probabilities within the groups to 0.7. Notice that for this model, the search space of the grid is bi-dimensional. We considered the space $[0.1, 0.3] \times [0.6, 0.8]$, with a step of 0.001. For computing the eigenvalues, we used the $\sqrt{n}$ normalization of the adjacency matrix.

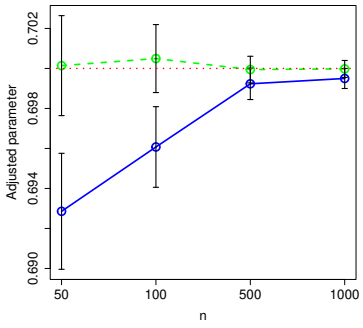
Figure 4 shows the average estimated parameter and a 95% confidence interval for the GRG model. Results obtained for the WS, BA, and BM models are shown in Figures 5, 6, and 7 respectively. The blue solid lines correspond to the results based on the ℓ_1 distance between the ESD of the observed graph and the average ESD of the model, estimated as described by Algorithm 4.1. Similarly, the green dashed lines correspond to the results based on the eigenvalues' empirical cumulative distribution. The red dotted lines indicate the true parameter.



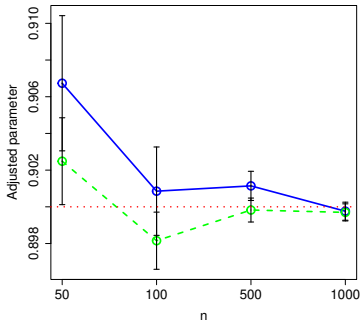
(a) r = 0.1

(b) r = 0.3

(c) r = 0.5

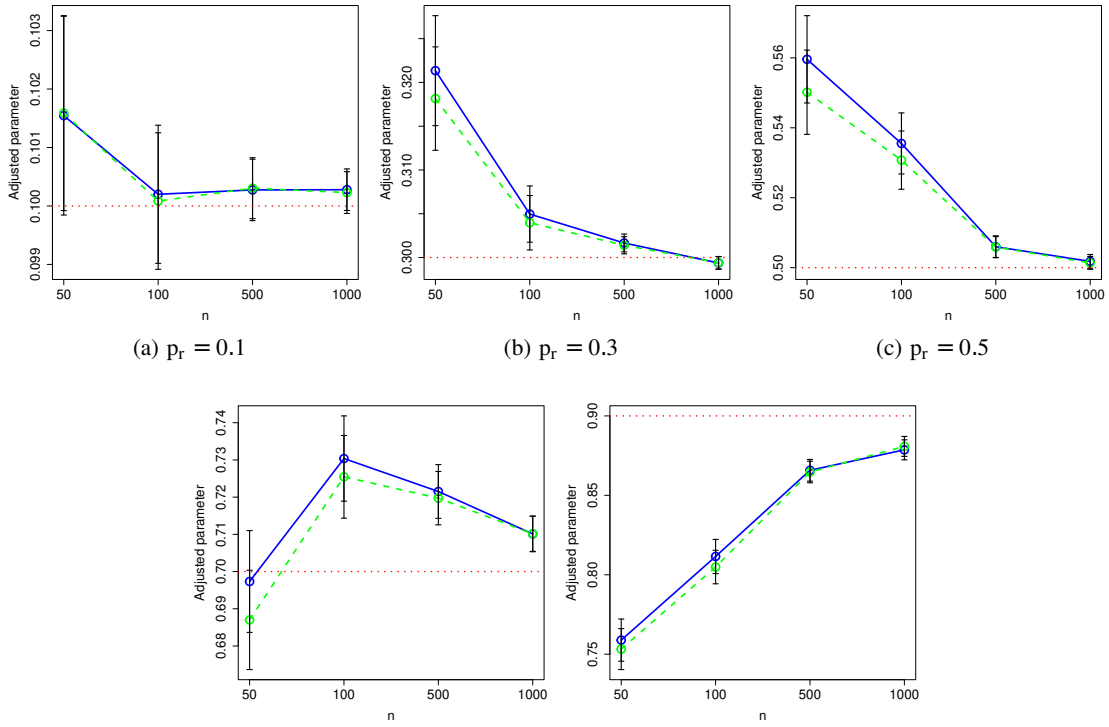


(d) r = 0.7



(e) r = 0.9

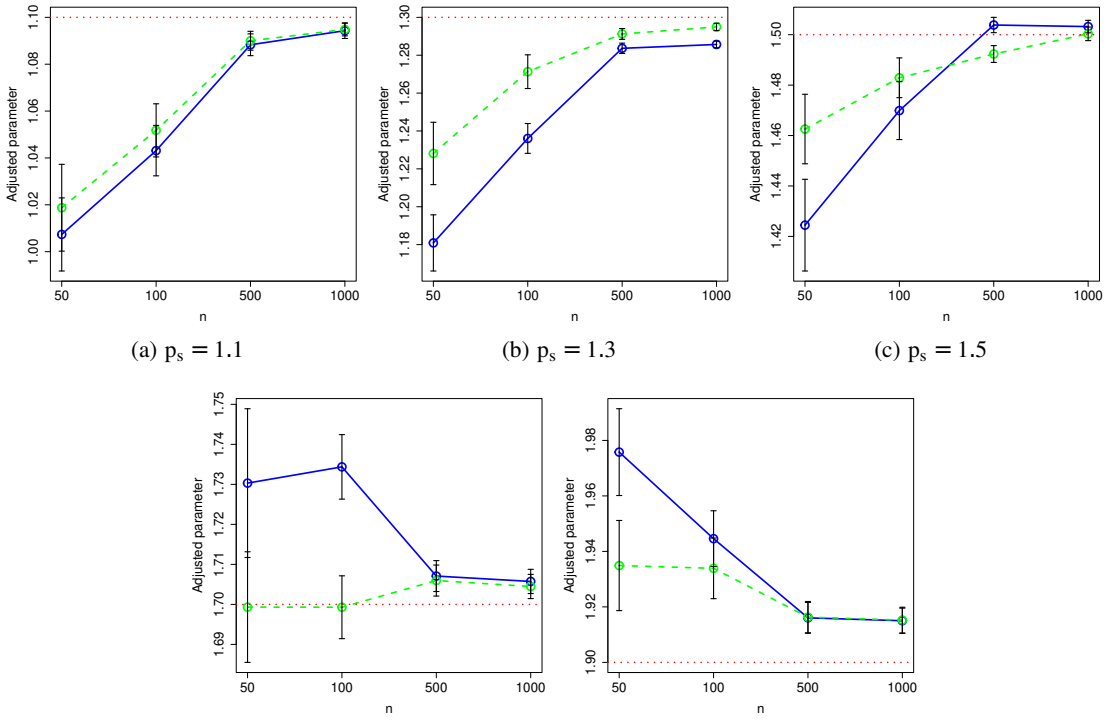
FIG. 4: Average estimated parameter for the GRG model. We generated 1 000 GRG random graphs of sizes $n = 50, 100, 500, 1\,000$, and radius $r = 0.1$ (a), $r = 0.3$ (b), $r = 0.5$ (c), $r = 0.7$ (d), and $r = 0.9$ (e). In each plot, the x-axis and the y-axis correspond, respectively, to the graph size (n) and estimated parameter. The points and error bars indicate the average estimated parameters and the corresponding 95% confidence intervals. The blue solid lines correspond to the results based on the ℓ_1 distance between the ESD of the observed graph and the average ESD of the GRG model. The green dashed lines correspond to the results based on the eigenvalues' cumulative distribution. The red dotted lines indicate the true value.



(a) $p_r = 0.1$ (b) $p_r = 0.3$ (c) $p_r = 0.5$

(d) $p_r = 0.7$ (e) $p_r = 0.9$

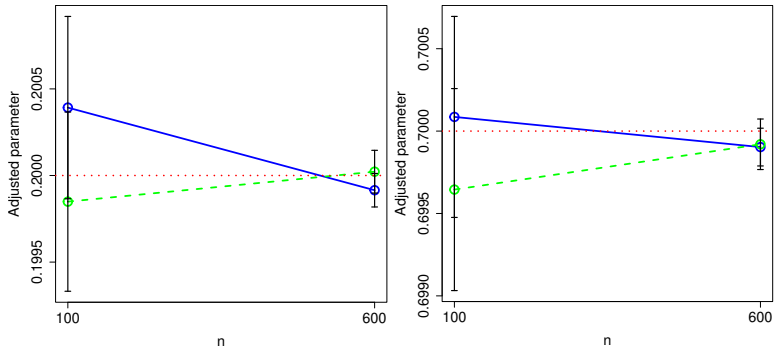
FIG. 5: Average estimated parameter for the WS model. We generated 1 000 WS random graphs of sizes $n = 50, 100, 500, 1\,000$, and probability of reconnecting edges $p_r = 0.1$ (a), $p_r = 0.3$ (b), $p_r = 0.5$ (c), $p_r = 0.7$ (d), and $p_r = 0.9$ (e). In each plot, the x-axis and the y-axis correspond, respectively, to the graph size (n) and estimated parameter. The points and error bars indicate the average estimated parameters and the corresponding 95% confidence intervals. The blue solid correspond to the results based on the ℓ_1 distance between the ESD of the observed graph and the average ESD of the WS model. The green dashed lines correspond to the results based on the eigenvalues' cumulative distribution. The red dotted lines indicate the true value.



(a) $p_s = 1.1$ (b) $p_s = 1.3$ (c) $p_s = 1.5$

(d) $p_s = 1.7$ (e) $p_s = 1.9$

FIG. 6: Average estimated parameter for the BA model. We generated 1 000 BA random graphs of sizes $n = 50, 100, 500, 1\,000$, and scale exponent $p_s = 1.1$ (a), $p_s = 1.3$ (b), $p_s = 1.5$ (c), $p_s = 1.7$ (d), and $p_s = 1.9$ (e). In each plot, the x-axis and the y-axis correspond, respectively, to the graph size (n) and estimated parameter. The points and error bars indicate the average estimated parameters and the corresponding 95% confidence intervals. The blue solid lines correspond to the results based on the ℓ_1 distance between the ESD of the observed graph and the average ESD of the BA model. The green dashed lines correspond to the results based on the eigenvalues' cumulative distribution. The red dotted lines indicate the true value.



(a) $p_0 = 0.2$ (b) $p_1 = 0.7$

FIG. 7: Average estimated parameter for the BM model with two blocks of the same sizes. We generated 1 000 BM random graphs of sizes $n = 100, 600$, with probability to connect vertices from different groups $p_0 = 0.2$ (a), and probability within the clusters $p_1 = 0.7$ (b). In each plot, the x-axis and the y-axis correspond, respectively, to the graph size (n) and estimated parameter. The points and error bars indicate the average estimated parameters and the corresponding 95% confidence intervals. The solid blue lines correspond to the results based on the ℓ_1 distance between the ESD of the observed graph and the average ESD of the BM model. The green dashed lines correspond to the results based on the eigenvalues' cumulative distribution. The red dotted lines indicate the true value.

In most scenarios, we observe that the estimated parameter approximates the actual value as the graph size increases. However, in some cases, results for $n = 500$ and $n = 1\,000$ are similar (see, e.g., BA model, when $p_s = 1.7, 1.9$). This limited convergence might be related to the fact that we only consider parameter values on the search grid (as described in Algorithm 4.1).

6.4 *Model selection*

To evaluate the performance of the model selection approach based on the ℓ_1 distance between the ESD and its theoretical limit, we performed simulation experiments with 1 000 graphs of sizes $n = 50, 100, 500$ generated by the following models: Erdős-Rényi (ER) ($p = \beta$), Geometric (GRG) ($r = \beta$), d-regular (DR) ($d = 10\beta$), Watts-Strogatz (WS) ($p_r = \beta$), and Barabási-Albert (BA) ($p_s = 1 + \beta$), where $\beta = 0.1, 0.5, 0.9$. We also contrasted the performance against the original method by Takahashi et al. [43], based on the Kullback-Leibler divergence between spectral densities. Tables 1, 2, 3 show the confusion matrices obtained by the methods based on the spectral density and cumulative distribution of the eigenvalues for $\beta = 0.1, 0.5$, and 0.9, respectively.

Table 1: **Model selection for graphs generated with $\beta = 0.1$.** The number of graphs correctly classified by the model selection approach based on the Kullback-Leibler divergence between empirical spectral densities (left)

, the ℓ_1 distance between empirical spectral densities (middle), and the ℓ_1 distance between empirical cumulative distributions (right). For each model, we generated 1 000 graphs of sizes $n = 50, 100, 500$. The set-up of the models was: the probability of connecting two vertices $p = \beta$ for the Erdős-Rényi (ER), the radius $r = \beta$ for the geometric (GRG), the degree $d = 10\beta$ for the d-regular (DR), the probability of reconnecting edges $p_r = \beta$ for the Watts-Strogatz (WS), and the scale exponent $p_s = 1 + \beta$ for the Barabási-Albert (BA).

n	True model	Selected model				
		ER	DR	GRG	WS	BA
50	ER	905|881|929	6|7|5	0|0|0	89|112|66	0|0|0
	DR	0|0|0	1000|1000|1000	0|0|0	0|0|0	0|0|0
	GRG	1|4|20	0|0|0	999|996|980	0|0|0	0|0|0
	WS	0|0|0	0|0|0	0|0|0	1000|1000|1000	0|0|0
	BA	39|108|16	0|0|0	0|3|0	0|0|0	961|889|984
100	ER	1000|993|998	0|7|2	0|0|0	0|0|0	0|0|0
	DR	0|0|0	1000|1000|1000	0|0|0	0|0|0	0|0|0
	GRG	0|0|0	0|0|0	1000|1000|1000	0|0|0	0|0|0
	WS	0|0|0	0|0|0	0|0|0	1000|1000|1000	0|0|0
	BA	9|94|3	0|0|0	2|2|0	0|0|0	989|904|997
500	ER	1000|994|998	0|6|2	0|0|0	0|0|0	0|0|0
	DR	0|0|0	1000|1000|1000	0|0|0	0|0|0	0|0|0
	GRG	0|0|0	0|0|0	1000|1000|1000	0|0|0	0|0|0
	WS	0|0|0	0|0|0	0|0|0	1000|1000|1000	0|0|0
	BA	0|2|0	0|0|0	0|0|0	0|0|0	998|1000

Table 2: **Model selection for graphs generated with $\beta = 0.5$.** The number of graphs correctly classified by the model selection approach based on the Kullback-Leibler divergence between empirical spectral densities (left)

, the ℓ_1 distance between empirical spectral densities (middle), and the ℓ_1 distance between empirical cumulative distributions (right). For each model, we generated 1 000 graphs of sizes $n = 50, 100, 500$. The set-up of the models was: the probability of connecting two vertices $p = \beta$ for the Erdős-Rényi (ER), the radius $r = \beta$ for the geometric (GRG), the degree $d = 10\beta$ for the d-regular (DR), the probability of reconnecting edges $p_r = \beta$ for the Watts-Strogatz (WS), and the scale exponent $p_s = 1 + \beta$ for the Barabási-Albert (BA).

n	True model	Selected model				
		ER	DR	GRG	WS	BA
50	ER	947|948|911	53|52|89	0|0|0	0|0|0	0|0|0
	DR	0|1|0	1000|999|1000	0|0|0	0|0|0	0|0|0
	GRG	0|2|0	0|0|0	1000|998|1000	0|0|0	0|0|0
	WS	0|21|0	0|0|0	0|0|0	1000|979|1000	0|0|0
	BA	2|64|3	0|0|0	3|103|3	0|0|0	995|833|994
100	ER	966|949|951	53|51|49	0|0|0	0|0|0	0|0|0
	DR	0|0|0	1000|1000|1000	0|0|0	0|0|0	0|0|0
	GRG	0|0|0	0|0|0	1000|1000|1000	0|0|0	0|0|0
	WS	3|6|0	0|0|0	0|0|0	997|994|1000	0|0|0
	BA	1|121|1	0|0|0	95|125|1	0|0|0	904|754|998
500	ER	899|908|922	101|92|78	0|0|0	0|0|0	0|0|0
	DR	0|0|0	1000|1000|1000	0|0|0	0|0|0	0|0|0
	GRG	0|0|0	0|0|0	1000|1000|1000	0|0|0	0|0|0
	WS	0|0|0	0|0|0	0|0|0	1000|1000|1000	0|0|0
	BA	0|31|0	0|0|0	0|11|0	0|0|0	1000|958|1000

Table 3: **Model selection for graphs generated with $\beta = 0.9$.** The number of graphs correctly classified by the model selection approach based on the Kullback-Leibler divergence between empirical spectral densities (left), the ℓ_1 distance between empirical spectral densities (middle), and the ℓ_1 distance between empirical cumulative distributions (right).For each model, we generated 1 000 graphs of sizes $n = 50, 100, 500$. The set-up of the models was: the probability of connecting two vertices $p = \beta$ for the Erdős-Rényi (ER), the radius $r = \beta$ for the geometric (GRG), the degree $d = 10\beta$ for the d-regular (DR), the probability of reconnecting edges $p_r = \beta$ for the Watts-Strogatz (WS), and the scale exponent $p_s = 1 + \beta$ for the Barabási-Albert (BA).

n	True model	Selected model ER	DR	GRG	WS	BA
50	ER	1000|1000|1000	0|0|0	0|0|0	0|0|0	0|0|0
	DR	0|56|4	1000|944|996	0|0|0	0|0|0	0|0|0
	GRG	0|71|0	0|0|0	1000|929|1000	0|0|0	0|0|0
	WS	19|133|22	0|0|0	0|0|0	981|867|978	0|0|0
	BA	0|9|0	0|0|0	3|22|0	0|0|0	997|969|1000
100	ER	1000|1000|1000	0|0|0	0|0|0	0|0|0	0|0|0
	DR	0|1|0	1000|999|1000	0|0|0	0|0|0	0|0|0
	GRG	0|0|0	0|0|0	1000|1000|1000	0|0|0	0|0|0
	WS	5|124|59	0|0|0	0|0|0	995|876|941	0|0|0
	BA	0|0|0	0|0|0	2|2|0	0|0|0	998|998|1000
500	ER	1000|1000|1000	0|0|0	0|0|0	0|0|0	0|0|0
	DR	0|0|0	1000|1000|1000	0|0|0	0|0|0	0|0|0
	GRG	0|0|0	0|0|0	1000|1000|1000	0|0|0	0|0|0
	WS	0|0|0	0|0|0	0|0|0	1000|1000|1000	0|0|0
	BA	0|0|0	0|0|0	0|0|0	0|0|0	1000|1000|1000

Tables 1, 2, and 3 show that the performance is similar across the methods. The number of hits increases as the graphs become larger. The only exception is when $\beta = 0.5$ for ER graphs. In that case, the method confounds some ER graphs with DR graphs. Note that the limiting spectral densities have the form of semicircle laws (though not obtained through the same scaling of the adjacency matrix).

7. Real world networks

To illustrate applications of the model selection approach based on the ℓ_1 distance between spectral densities, we applied the procedure to three real-world networks, described as follows.

1. Yeast protein-protein interaction (PPI) network. It contains 1 966 vertices corresponding to proteins and 2 705 edges indicating whether two proteins physically interact within the cell. We downloaded data from `http://networksciencebook.com/` [51].

2. E. coli bacteria metabolic network. It contains 1 039 vertices corresponding to metabolites and 5 802 edges indicating reactions between pairs of metabolites. We downloaded data from `http://networksciencebook.com/` [40].

3. Facebook ego network. From 10 Facebook ego networks (each one containing friends of a starting user), we selected the one with more than 1 000 vertices. The selected network has 1 034 vertices corresponding to users and 53 498 edges indicating friendships on Facebook. We downloaded the data from `https://snap.stanford.edu/data/ego-Facebook.html` [31].

We considered a list of four candidate models: ER (an utterly random procedure), GRG, WS, and BA (models with few parameters that share characteristics of complex networks). Table 4 shows the ℓ_1 distance between the ESD of the actual network and the average ESD of each model.

Table 4: Model selection for real-world networks. We applied the model selection procedure for three different real networks (yeast protein-protein interaction network, E. coli metabolic network, and Facebook ego network). We considered four different candidate models: Erdős-Rényi (ER), geometric random graph (GRG), Watts-Strogatz (WS), and Barabási-Albert (BA). The values indicate the ℓ_1 distance between the empirical spectral density of the network and the model. We highlighted in bold the smallest value for each network.

Network	ER	GRG	WS	BA
Yeast PPI	0.941	0.433	0.486	**0.108**
E. coli metabolic network	0.636	0.367	0.455	**0.17**
Facebook network	0.296	**0.207**	0.238	0.221

Notice that BA is the only model from our list that builds scale-free graphs, a prevalent pattern in biological networks. For example, many protein-protein interaction networks and metabolic networks are known to share properties with scale-free graphs [29, 30]. Indeed, for both the Yeast PPI and the E. coli metabolic networks, BA is the model with the most similar spectral density. Similarly, the original procedure by [43], based on the Kullback-Leibler divergence between spectral densities, also selects BA (among ER, WS, and BA) for PPI networks of different organisms.

Among the three networks, the Facebook ego network is the least similar to scale-free graphs. It contains social circles corresponding to densely connected groups of vertices [31]. When clusters are well represented spatially, the graph may share characteristics of a geometric graph. In our analysis, the model selection approach selected GRG for this Facebook ego graph.

8. Discussion

By reviewing results on the ESD of random graphs' convergence properties, we observed that the Wigner semicircle law could be applied to the ER and DR models when the adjacency matrices are adequately normalized, as shown in Table 5. For the BM random graph, we can express the limiting ESD of the centered and normalized matrix through the Stieltjes transform under certain conditions (blocks with the same size, same probabilities of connecting vertices from different groups, and inter-groups probability smaller than the intra-groups probabilities). Besides, our simulation experiments suggested that convergence of the ESD of non-centered adjacency matrices and centered BM adjacency matrices are similar. The ESD may also converge to the limiting ESD obtained through the Stieltjes transform when the probabilities to connect vertices from different groups are different, and the inter-groups probability p_0 is larger than the intra-groups probabilities p_m.

Table 5: Convergences of ESD in different models.

Model	Matrix	Limit
ER	$\mathbf{A}/\sqrt{n}$	$\mu_{p,sc}(x) = \frac{1}{2\pi p(1-p)}\sqrt{[4p(1-p)-x^2]_+}$
	$\mathbf{A}/\sqrt{np(1-p)}$	$\mu_{sc}(x) = \frac{1}{2\pi}\sqrt{[4-x^2]_+}$
DR	$\mathbf{A}$	$\mu_d(x) = \frac{d}{2\pi(d^2-x^2)}\sqrt{[4(d-1)-x^2]_+}$
	$\mathbf{A}/\sqrt{d-1}$	$\tilde{\mu}_d(x) = \left(1+\frac{1}{d-1}-\frac{x^2}{d}\right)^{-1}\frac{1}{2\pi}\sqrt{[4-x^2]_+}$
BM	$(\mathbf{A}-\mathbb{E}(\mathbf{A}))/\sqrt{np^\star(1-p^\star)}$	expression through Stieltjes transform

The ESDs of the geometric random graph and the linear preferential attachment models converge weakly almost surely to a limiting density function as $n \to \infty$. However, an analytical expression for this limiting function is unknown.

There is still no theoretical result on the ESD convergence for other models, such as the WS and BA. However, simulation experiments suggest that the ESD of those models approximate to a fixed distribution as the number of vertices increases [19].

Therefore, there are scenarios in which we may assume that the ESD converges weakly to a limiting distribution, as required by Theorems 5.1 and 5.2. If all assumptions hold, then the minimizer of the ℓ_1 distance between the observed graph's eigenvalue distribution and the limiting ESD of the model converges in probability to the true parameter as $n \to \infty$.

When assumptions of Theorems 5.1 and 5.2 are satisfied, our simulations in Section 6.2 show that the estimated parameters approximate the true values as the graph size increases. When we can not obtain a limiting spectral distribution, we estimate it empirically by randomly sampling graphs from the random graph model (as described in Algorithm 4.1). In this case, the estimated parameter also approximates the true parameter as the number of vertices increases, as shown by simulation experiments in Section 6.3.

The model-fitting procedure's performance depends on the choice of the ESD or the CDF. The convergence of the estimated parameter based on the ESD, in general, is faster for the ER model. The model-fitting procedures' performance based on the ESD and CDF was similar for the WS model. For the GRG and BA random graph models, we observed a better performance by using the CDF. The reason is that sometimes ESD estimation/bandwidth choice is difficult. Thus, in that case, using the cumulative distribution of the eigenvalues instead of the spectral density is more convenient.

One of the main limitations of the procedure in Algorithm 4.1 is its high computational cost. In the following complexity analysis, we neglect the number of models N that is supposed to be fixed (and small in general).

The spectral decomposition has a complexity $O(n^3)$. Thus, when the analytical form of the limiting spectral distribution is known, the complexity of Algorithm 4.1 is $O(n^3)$, while when we use a Monte Carlo approximation for the limiting ESD, it increases to $O(n^3|\tilde{\Theta}|M)$. Future research for reducing computational cost of Algorithm 4.1 could explore either a faster approximation of the spectrum of each graph and thus its ESD, or an approximation of this ESD. Indeed, a faster approximation of the ESD does not require the direct computation of all eigenvalues [32, 34], and has complexity of the order of $O(|E|C)$, where C is associated to the number of random vectors and number of moments used by the approximation algorithm. For large graphs, we have $C << |E|$. Then, if this approximation is used in our model fitting procedure, the final computational cost could be reduced to $O(|E|)$ (when analytic form of limiting ESD is known) or $O(|E||\tilde{\Theta}|M)$ (when MC approximation is needed). Notice, however, that more experiments are necessary to validate these approximations.

Furthermore, the procedure in Algorithm 4.1 might not be the most recommended approach for classical models such as the ER and BM. The reason is that there are already accurate and efficient parameter estimators [1, 42] for them. However, Algorithm 4.1 gains in generality. Every time we fix a parametric model, we can build a procedure to estimate the parameter and a divergence between the model and the observed graph. This generality is particularly useful when considering a list of random graph models, and we want to choose the one that best approximates the observed graph.

Indeed, the experiments in Section 6.4 suggest that choosing the model that minimizes the ℓ_1 distance between spectral densities may present a hit rate of 100% when the graph size is large enough. However, it is essential to notice that certain combinations of parameters may generate similar graphs from different models, such as the ER and DR graphs.

The illustration of the model selection approach on real-world networks (Section 7) suggested that results are consistent with known properties of biological and social networks. However, it is essential to notice that this procedure does not correspond to a goodness-of-fit test. It might happen that the best model does not fit that well the observed network and is only the least bad. It would be interesting to test whether a given model is likely to generate the observed spectral density in future work.

We focused on studying the ESD of the graph adjacency matrix. However, assumptions of Theorems 5.1 and 5.2 may also hold for other graph matrices, such as the Laplacian and normalized Laplacian, as discussed by [16, 25]. Therefore, results are general in terms of different random graph models and different graph matrices whose ESDs satisfy the assumptions of Theorems 5.1 and 5.2.

9. Conclusions

We reviewed the convergence properties of random graph's ESDs. Our review and simulation study suggest that different random graph models' ESD converge to a limiting density function. Based on the ESD's convergence properties, we prove that the minimizer of the ℓ_1 distance of the ESD and the limiting density function is a consistent estimator of the random graph model parameter. This result also holds for the CDF of the eigenvalues. The main advantage of this model-fitting approach is its generality.

Furthermore, our simulation experiments suggest that the procedures proposed by Takahashi et al. [43] may be modified to rely on ℓ_1 distance between eigenvalue distribution functions to select a random graph model.

In our computational experiments we used code from the *statGraph* package available at CRAN (`https://cran.r-project.org/web/packages/statGraph/`). A zip file containing the R codes of all the experiments are available at `https://www.ime.usp.br/~suzana/code_spectral_density.zip`.

Funding

This work was supported by São Paulo Research Foundation [grant numbers 2015/21162-4, 2017/12074-0, 2018/21934-5, 2019/22845-9, 2020/08343-8]; National Council for Scientific and Technological Development [grant number 303855/2019-3]; Coordination for the Improvement of Higher Education Personnel (Finance code 001); Alexander von Humboldt Foundation; and the Academy of Medical Sciences - Newton Fund.

REFERENCES

1. Ambroise, C. & Matias, C. (2012) New consistent and asymptotically normal parameter estimates for random-graph mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **74**(1), 3–35.

2. Arnold, L. (1967) On the asymptotic distribution of the eigenvalues of random matrices. *Journal of Mathematical Analysis and Applications*, **20**(2), 262–268.

3. Avrachenkov, K., Bobu, A. & Dreveton, M. (2021) Higher-Order Spectral Clustering for Geometric Graphs. *Journal of Fourier Analysis and Applications*, **27**(2), 1–29.

4. Avrachenkov, K., Cottatellucci, L. & Kadavankandy, A. (2015) Spectral properties of random matrices for stochastic block model. In *The 2015 4th International Workshop on Physics-Inspired Paradigms in Wireless Communications and Networks*, pages 537–544.

5. Bai, Z. D. (1999) Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, **9**(3), 611–662.

6. Barabási, A.-L. & Albert, R. (1999) Emergence of Scaling in Random Networks. *Science*, **286**(5439), 509–512.

7. Bauerschmidt, R., Knowles, A. & Yau, H.-T. (2017) Local Semicircle Law for Random Regular Graphs. *Comm. Pure Appl. Math.*, **70**(10), 1898–1960.

8. Bhamidi, S., Evans, S. N. & Sen, A. (2012) Spectra of Large Random Trees. *Journal of Theoretical Probability*, **25**(3), 613–654.

9. Bordenave, C. (2008) Eigenvalues of Euclidean random matrices. *Random Structures & Algorithms*, **33**(4), 515–532.

10. Breiger, R. L., Boorman, S. A. & Arabie, P. (1975) An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, **12**(3), 328 – 383.

11. Chatterjee, S. & Diaconis, P. (2013) Estimating and understanding exponential random graph models. *Ann. Statist.*, **41**(5), 2428–2461.

12. Chung, F. & Lu, L. (2002) Connected Components in Random Graphs with Given Expected Degree Sequences. *Annals of Combinatorics*, **6**(2), 125–145.

13. Csárdi, G. & Nepusz, T. (2006) *The igraph software package for complex network research*. R package version 1.2.6.

14. Dacunha-Castelle, D. & Duflo, M. (1986) *Probability and statistics. Vol. II*. Springer-Verlag, New York.

15. Davidson, E. & Levin, M. (2005) Gene regulatory networks. *PNAS*, **102**(14), 4935–4935.

16. Ding, X. & Jiang, T. (2010) Spectral distributions of adjacency and Laplacian matrices of random graphs. *The Annals of Applied Probability*, **20**(6), 2086–2117.

17. Dumitriu, I. & Pal, S. (2012) Sparse regular random graphs: Spectral density and eigenvectors. *The Annals of Probability*, **40**(5), 2197–2235.

18. Erdős, P. & Rényi, A. (1959) On random graphs. *Publ. Math. Debrecen*, **6**, 290–297.

19. Farkas, I. J., Derényi, I., Barabási, A.-L. & Vicsek, T. (2001) Spectra of "real-world" graphs: Beyond the semicircle law. *Physical Review E*, **64**(2).

20. Faust, K. & Wasserman, S. (1992) Blockmodels: Interpretation and evaluation. *Social Networks*, **14**(1), 5 – 61. Special Issue on Blockmodels.

21. Fienberg, S. E. & Wasserman, S. S. (1981) Categorical Data Analysis of Single Sociometric Relations. *Sociological Methodology*, **12**, 156–192.

22. Frank, O. & Harary, F. (1982) Cluster inference by using transitivity indices in empirical graphs. *J. Amer. Statist. Assoc.*, **77**(380), 835–840.

23. Frank, O. & Strauss, D. (1986) Markov graphs. *J. Amer. Statist. Assoc.*, **81**(395), 832–842.

24. Grenander, U. (2008) *Probabilities on algebraic structures. Reprint of the 1963 original ed.* Mineola, NY: Dover Publications, reprint of the 1963 original ed. edition.

25. Gu, J., Jost, J., Liu, S. & Stadler, P. F. (2016) Spectral classes of regular, random, and empirical graphs. *Linear Algebra and its Applications*, **489**, 30–49.

26. Hamidouche, M., Cottatellucci, L. & Avrachenkov, K. (2019) Spectral Analysis of the Adjacency Matrix

of Random Geometric Graphs. In *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, page 208–214. IEEE Press.

27. Holland, P. W., Laskey, K. B. & Leinhardt, S. (1983) Stochastic blockmodels: First steps. *Social networks*, **5**(2), 109–137.

28. Hückel, E. (1931) Quantentheoretische Beiträge zum Benzolproblem. *Zeitschrift für Physik*, **70**(3), 204–286.

29. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. (2001) Lethality and centrality in protein networks. *Nature*, **411**(6833), 41–42.

30. Jeong, H., Tombor, B., Albert, R. & Oltvai, Z. N. (2000) The large-scale organization of metabolic networks. *Nature*, **407**, 4.

31. Leskovec, J. & Mcauley, J. (2012) Learning to Discover Social Circles in Ego Networks. In Pereira, F., Burges, C. J. C., Bottou, L. & Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems*, volume 25, pages 548–56. Curran Associates, Inc.

32. Lin, L., Saad, Y. & Yang, C. (2016) Approximating Spectral Densities of Large Matrices. *SIAM Review*, **58**(1), 34–65.

33. McKay, B. D. (1981) The expected eigenvalue distribution of a large regular graph. *Linear Algebra and its Applications*, **40**, 203–216.

34. Newman, M. E. J., Zhang, X. & Nadakuditi, R. R. (2019) Spectra of random networks with arbitrary degrees. *Phys. Rev. E*, **99**(4), 042309.

35. Pellegrini, M., Haynor, D. & Johnson, J. M. (2004) Protein interaction networks. *Expert Rev Proteomics*, **1**(2), 239–249.

36. Penrose, M. (2003) *Random geometric graphs*, volume 5 of *Oxford Studies in Probability*. Oxford: Oxford University Press.

37. Piera, F. J. & Parada, P. (2009) On convergence properties of Shannon entropy. *Problems of Information Transmission*, **45**(2), 75–94.

38. Pinski, F., Simpson, G., Stuart, A. & Weber, H. (2015) Kullback–Leibler Approximation for Probability Measures on Infinite Dimensional Spaces. *SIAM J. Math. Anal.*, **47**(6), 4091–4122.

39. Preciado, V. M. & Rahimian, M. A. (2017) Moment-Based Spectral Analysis of Random Graphs with Given Expected Degrees. *IEEE Transactions on Network Science and Engineering*, **4**(4), 215–228.

40. Schellenberger, J., Park, J. O., Conrad, T. M. & Palsson, B. Ø. (2010) BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, **11**(1), 213.

41. Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, Boca Raton.

42. Snijders, T. A. B. & Nowicki, K. (1997) Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure. *J. of Classification*, **14**(1), 75–100.

43. Takahashi, D. Y., Sato, J. R., Ferreira, C. E. & Fujita, A. (2012) Discriminating Different Classes of Biological Networks by Analyzing the Graphs Spectra Distribution. *PLoS ONE*, **7**(12), e49949.

44. Tran, L. V., Vu, V. H. & Wang, K. (2013) Sparse Random Graphs: Eigenvalues and Eigenvectors. *Random Struct. Algorithms*, **42**(1), 110–134.

45. Tsybakov, A. B. (2009) *Introduction to nonparametric estimation*. New York, NY: Springer.

46. van den Heuvel, M. P. & Hulshoff Pol, H. E. (2010) Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, **20**(8), 519–534.

47. Von Collatz, L. & Sinogowitz, U. (1957) Spektren endlicher grafen. *Abh. Math. Semin. Univ. Hambg.*, **21**(1), 63–77.

48. Watts, D. J. & Strogatz, S. H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**(6684), 440–442.

49. White, H. C., Boorman, S. A. & Breiger, R. L. (1976) Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions. *American Journal of Sociology*, **81**(4), 730–780.

50. Wigner, E. P. (1958) On the Distribution of the Roots of Certain Symmetric Matrices. *Annals of Mathematics*, **67**(2), 325–327.

51. Yu, H., Braun, P., Yıldırım, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J.-F., Dricot, A., Vazquez, A., Murray, R. R., Simon, C., Tardivo, L.,

Tam, S., Svrzikapa, N., Fan, C., de Smet, A.-S., Motyl, A., Hudson, M. E., Park, J., Xin, X., Cusick, M. E., Moore, T., Boone, C., Snyder, M., Roth, F. P., Barabási, A-L Tavernier, J., Hill, D. E. & Vidal, M. (2008) High-quality binary protein interaction map of the yeast interactome network. *Science*, **322**(5898), 104–110.

52. Zhu, Y. (2020) A graphon approach to limiting spectral distributions of Wigner-type matrices. *Random Structures & Algorithms*, **56**(1), 251–279.