EURANDOM PREPRINT SERIES

2018-010

July, 2018

Spectra of Adjacency and Laplacian Matrices of Inhomogeneous Erdős Rényi Random Graphs

A. Chakrabarty, R. Subhra Hazra, F. den Hollander, M. Sfragara
ISSN 1389-2355

Spectra of Adjacency and Laplacian Matrices of Inhomogeneous Erdős-Rényi Random Graphs

Arijit Chakrabarty, Rajat Subhra Hazra, Frank den Hollander, and Matteo Sfragara

Abstract: We consider inhomogeneous Erdős-Rényi random graphs $\mathbb{G}_N$ on N vertices in the non-dense regime. The edge between the pair of vertices $\{i, j\}$ is retained with probability $\varepsilon_N\, f(\frac{i}{N}, \frac{j}{N})$, $1 \le i \ne j \le N$, independently of other edges, where $f\colon [0,1] \times [0,1] \to [0,\infty)$ is a continuous function such that $f(x,y) = f(y,x)$ for all $x, y \in [0,1]$. We study the empirical distribution of both the adjacency matrix A_N and the Laplacian matrix Δ_N associated with $\mathbb{G}_N$ in the limit as $N \to \infty$ when $\lim_{N\to\infty} \varepsilon_N = 0$ and $\lim_{N\to\infty} N\varepsilon_N = \infty$. In particular, we show that the empirical spectral distributions of $(N\varepsilon_N)^{-1/2} A_N$ and $(N\varepsilon_N)^{-1/2}\Delta_N$ converge to deterministic limits weakly in probability. For the special case where $f(x,y) = r(x)r(y)$ with $r\colon [0,1] \to [0,\infty)$ a continuous function, we give an explicit characterization of the limiting distributions. We further show some applications of our results to constrained random graphs, Chung-Lu random graphs and social networks.

1. Introduction and main results

Spectra of random matrices have been analyzed for close to a century. In recent years, many interesting results have been derived for random matrices associated with *random graphs*, like the adjacency matrix and the Laplacian matrix (Bauer and Golinelli (2001), Bhamidi et al. (2012), Bordenave and Lelarge (2010), Ding et al. (2010), Dumitriu and Pal (2012), Farkas et al. (2001), Jiang (2012a,b), Khorunzhy et al. (2004), Lee and Schnelli (2016), Tran et al. (2013)).

The focus of the present paper is on inhomogeneous Erdős-Rényi random graphs, which are rooted in the theory of complex networks. We consider the regime where the degrees of the vertices diverge sublinearly with the size of the graph. In this regime, we identify the scaling limit of the empirical spectral distribution, both for the adjacency matrix and the Laplacian matrix. For the special case where the connection probabilities have a product structure, we obtain an explicit description of the scaling limit of the empirical spectral distribution in terms of objects that are rooted in free probability. It is known that in the absence of inhomogeneity, i.e., for standard Erdős-Rényi random graphs, in the sparse regime the empirical spectral distributions of the adjacency matrix and the Laplacian matrix converge (after appropriate scaling and centering) to a semicircle law, respectively, a free additive convolution of a Gaussian and a semicircle law. See, for example, Bryc et al. (2006), Ding et al. (2010), Jiang (2012a). Our results extend these results to the inhomogeneous setting.

There are some recent results on the largest eigenvalue of sparse inhomogeneous Erdős-Rényi random graphs (Benaych-Georges et al. (2017)), and also on the empirical spectral distribution of adjacency matrices via the theory of graphons (Zhu (2018)). Inhomogeneous Erdős-Rényi random graphs with a product structure in the connection probabilities arises naturally in different contexts. In Dembo and Lubetzky (2016) the latter have been shown to play a crucial role in the identification of the limiting spectral distribution of the adjacency matrix of the configuration model. Our methodology allows us to look at some important applications. For instance, a Chung-Lu type random graph is used to model sociability distribution in networks. We show how to use the rescaled empirical spectral distribution and free probability to statistically recover the sociability distribution. Another important application is constrained random graphs. Given a sequence of positive integers, among the probability distributions for which the sequence of average degrees matches the given sequence, the one that maximizes the entropy is the canonical Gibbs measure. It is known that, under a sparsity condition, the connection probabilities arising out of the canonical Gibbs measure asymptotically have a product structure (Squartini et al. (2015)). We show that our results on the adjacency matrix can be easily extended to cover such situations. The spectrum of the Laplacian of a random graph is well known to be connected to properties of the random walk on the graph, algebraic connectivity, and Kirchhoff's law, among others. The explicit bearing of the spectral distribution of the Laplacian on the corresponding graph are left for future research, for which our results may serve as a starting point.

The paper is organized as follows. The setting is defined in Section 1, and three scaling theorems are stated, Theorems 1.1–1.4. A number of technical lemmas are stated and proved in Section 2. These serve as preparation for the proof of Theorems 1.1–1.3, which is given in Section 3. Theorem 1.4, which is a randomized version of Theorem 1.1, is proved in Section 4. In Section 5, various applications are discussed. Appendix A collects a few basic facts that are needed along the way.

1.1. Setting

Let $f\colon [0,1]\times[0,1]\to[0,\infty)$ be a continuous function, satisfying

$$f(x,y)=f(y,x) \qquad \forall\, x,y\in[0,1]\,.$$

A sequence of positive real numbers $(\varepsilon_N\colon N\geq 1)$ is fixed that satisfies

$$\lim_{N\to\infty}\varepsilon_N=0\,, \qquad \lim_{N\to\infty}N\varepsilon_N=\infty\,. \tag{1.1}$$

Consider the random graph $\mathbb{G}_N$ on vertices $\{1,\dots,N\}$ where, for each (i,j) with $1\leq i<j\leq N$, an edge is present between vertices i and j with probability $\varepsilon_N f(\frac{i}{N},\frac{j}{N})$, independently of other pairs of vertices. In particular, $\mathbb{G}_N$ is an undirected graph with no self loops and no multiple edges. Boundedness of f ensures that $\varepsilon_N f(\frac{i}{N},\frac{j}{N})\leq 1$ for all $1\leq i<j\leq N$ when N is large enough. If $f\equiv c$ with c a constant, then $\mathbb{G}_N$ is the Erdős-Rényi graph with edge retention probability $\varepsilon_N c$. For general f, $\mathbb{G}_N$ can be thought of as an *inhomogeneous* version of the Erdős-Rényi graph.

The adjacency matrix of $\mathbb{G}_N$ is denoted by A_N. Clearly, A_N is a symmetric random matrix whose diagonal entries are zero and whose upper triangular entries are independent

Bernoulli random variables, i.e.,

$$A_N(i,j) \triangleq \mathrm{BER}\left(\varepsilon_N f\left(\tfrac{i}{N},\tfrac{j}{N}\right)\right), \qquad 1 \le i \ne j \le N\,.$$

Write P to denote the law of A_N.

1.2. Scaling

Our first theorem states the existence of the limiting spectral distribution of A_N after suitable scaling. Here, and elsewhere in the paper, ESD is the abbreviation for *empirical spectral distribution*: the probability measure that puts mass $1/N$ at every eigenvalue, respecting its algebraic multiplicity.

Theorem 1.1. *There exists a compactly supported and symmetric probability measure μ on $\mathbb{R}$ such that*

$$\lim_{N\to\infty} \mathrm{ESD}\left((N\varepsilon_N)^{-1/2} A_N\right) = \mu \text{ weakly in probability}. \tag{1.2}$$

Furthermore, if

$$\min_{0\le x,y\le 1} f(x,y) > 0\,,$$

then μ is absolutely continuous with respect to Lebesgue measure.

The Laplacian of $\mathbb{G}_N$ is the $N\times N$ matrix Δ_N defined as

$$\Delta_N(i,j) = \begin{cases} -\sum_{k=1}^N A_N(i,k), & i=j\,,\\ A_N(i,j), & i\ne j\,.\end{cases}$$

Our second theorem is the analogue of Theorem 1.1 with A_N replaced by Δ_N.

Theorem 1.2. *There exists a symmetric probability measure ν on $\mathbb{R}$ such that*

$$\lim_{N\to\infty} \mathrm{ESD}\left((N\varepsilon_N)^{-1/2}(\Delta_N - D_N)\right) = \nu \text{ weakly in probability},$$

where

$$D_N = \mathrm{Diag}\left(\mathrm{E}(\Delta_N(1,1)),\dots,\mathrm{E}(\Delta_N(N,N))\right). \tag{1.3}$$

Remark 1.1. The ESD of a random matrix is a random probability measure. Note that μ and ν are both deterministic, i.e., a law of large numbers is in force. Theorems 1.1 and 1.2 are existential, in the sense that explicit descriptions of μ and ν are missing. We have some control on the Stieltjes transform of μ (see Remark 3.1 below) and we know that ν has a finite moment generating function (see (3.9) below).

1.3. Multiplicative structure

Our next theorem identifies μ and ν under the additional assumption that f has a *multiplicative structure*, i.e.,

$$f(x,y) = r(x)r(y)\,, \qquad x,y \in [0,1]\,, \tag{1.4}$$

for some continuous function $r\colon\ [0,1] \to [0,\infty)$. The statement is based on the theory of (possibly unbounded) self-adjoint operators affiliated with a W^*-probability space. A few relevant definitions are given below. For details the reader is referred to (Anderson et al., 2010, Section 5.2.3).

Definition. *A C^*-algebra $\mathcal{A} \subset B(\mathcal{H})$, with $\mathcal{H}$ a Hilbert space, is a W^*-*algebra *when $\mathcal{A}$ is closed under the weak operator topology. If, in addition, τ is a state such that there exists a unit vector $\xi \in \mathcal{H}$ satisfying*

$$\tau(a) = \langle a\xi, \xi\rangle \qquad \forall\, a \in \mathcal{H}\,,$$

*then $(\mathcal{A},\tau)$ is a W^*-*probability space*. In that case a densely defined self-adjoint (possibly unbounded) operator T on $\mathcal{H}$ is said to be* affiliated with $\mathcal{A}$ *if $h(T) \in \mathcal{A}$ for any bounded measurable function h defined on the spectrum of T, where $h(T)$ is defined by the spectral theorem. Finally, for an affiliated operator T, its* law $\mathcal{L}(T)$ *is the unique probability measure on $\mathbb{R}$ satisfying*

$$\tau\left(h(T)\right) = \int_{\mathbb{R}} h(x)\big(\mathcal{L}(T)\big)(dx)$$

for every bounded measurable $h\colon\ \mathbb{R} \to \mathbb{R}$.

The distribution of a single self-adjoint operator is defined above. For two or more self-adjoint operators $T_1,\dots,T_n$, a description of their *joint distribution* is a specification of

$$\tau\big(h_1\left(T_{i_1}\right)\cdots h_k\left(T_{i_k}\right)\big)\,,$$

for all $k \geq 1$, all $i_1,\dots,i_k \in \{1,\dots,n\}$, and all bounded measurable functions $h_1,\dots,h_k$ from $\mathbb{R}$ to itself. Once the above is specified, it is immediate that $\mathcal{L}(p(T_1,\dots,T_k))$ can be calculated for any polynomial p in k variables such that $p(T_1,\dots,T_k)$ is self-adjoint.

Definition. *Let $(\mathcal{A},\tau)$ be a W^*-probability space and $a_1, a_2 \in \mathcal{A}$. Then a_1 and a_2 are* freely independent *if*

$$\tau\left(p_1(a_{i_1})\cdots p_n(a_{i_n})\right) = 0\,,$$

for all $n \geq 1$, all $i_1,\dots,i_n \in \{1,2\}$ with $i_j \neq i_{j+1}$, $j = 1,\dots,n-1$, and all polynomials $p_1,\dots,p_n$ in one variable satisfying

$$\tau\left(p_j(a_{i_j})\right) = 0, \qquad j = 1,\dots,n\,.$$

For (possibly unbounded) operators $a_1,\dots,a_k$ and $b_1,\dots,b_m$ affiliated with $\mathcal{A}$, the collections $(a_1,\dots,a_k)$ and $(b_1,\dots,b_m)$ are freely independent *if and only if*

$$p\left(h_1(a_1),\dots,h_k(a_k)\right) \ \ \textit{and} \ \ q\left(g_1(b_1),\dots,g_m(b_m)\right)\,,$$

are freely independent for all bounded measurable $h_1, \ldots, h_k$ and $g_1, \ldots, g_m$, and all polynomials p and q in k and m non-commutative variables, respectively. It is immediate that the two operators in the above display are bounded, and hence belong to $\mathcal{A}$.

We are now in a position to state our third and last theorem.

Theorem 1.3. *If f is as in (1.4), then*

$$\mu = \mathcal{L}\left(r^{1/2}(T_u)T_s r^{1/2}(T_u)\right) , \tag{1.5}$$

and

$$\nu = \mathcal{L}\left(r^{1/2}(T_u)T_s r^{1/2}(T_u) + \alpha r^{1/4}(T_u)T_g r^{1/4}(T_u)\right) , \tag{1.6}$$

where

$$\alpha = \left(\int_0^1 r(x)\,dx\right)^{1/2} .$$

Here, T_g and T_u are commuting self-adjoint operators affiliated with a W^*-probability space $(\mathcal{A}, \tau)$ such that, for bounded measurable functions h_1, h_2 from $\mathbb{R}$ to itself,

$$\tau\left(h_1(T_g)h_2(T_u)\right) = \left(\int_{-\infty}^{\infty} h_1(x)\phi(x)\,dx\right)\left(\int_0^1 h_2(u)\,du\right) , \tag{1.7}$$

with ϕ the standard normal density. Furthermore, T_s has a standard semicircle distribution and is freely independent of (T_g, T_u).

Remark 1.2. The right-hand side of (1.5) is the same as the free multiplicative convolution of the standard semicircle law and the law of $r(U)$, where U is a standard uniform random variable.

Remark 1.3. The fact that T_g and T_u commute, together with (1.7), specifies their joint distribution. In fact, they are standard normal and standard uniform, respectively, independently of each other in the *classical sense*. Free independence of T_s and (T_g, T_u), plus the fact that the former follows the standard semicircle law, specifies the joint distribution of T_s, T_g, T_u.

Remark 1.4. In order to admit the unbounded operator T_g, a W^*-probability space is needed. If all the operators would have been bounded, then a C^*-probability space would have sufficed.

1.4. Randomization

Theorem 1.1 can be generalized to the situation where the function f is random. Such a randomization helps us to address the applications listed in Section 5. Suppose that $(\varepsilon_N\colon\, N \geq 1)$ is a sequence of positive numbers satisfying (1.1). Suppose further that, for every ≥ 1, $(R_{Ni}\colon\, 1 \leq i \leq N)$ is a collection of non-negative random variables such that there is a deterministic $C < \infty$ for which

$$\sup_{N\geq 1} \max_{1\leq i\leq N} R_{Ni} \leq C \text{ a.s.} \tag{1.8}$$

In addition, suppose that there is a probability measure μ_r on $\mathbb{R}$ such that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta_{R_{Ni}} = \mu_r \text{ weakly a.s.} \tag{1.9}$$

The non-negativity of R_{Ni} and (1.8) ensure that μ_r is concentrated on $[0, C]$. Furthermore, the first line of (1.1) ensures that the additional assumption

$$\sup_{N\geq 1} \varepsilon_N \leq \frac{1}{C} \tag{1.10}$$

entails no a loss of generality.

For fixed N and conditional on $(R_{N1}, \ldots, R_{NN})$, the random graph $\mathbb{G}_N$ is constructed as before, except that there is an edge between i and j with probability $\varepsilon_N R_{Ni} R_{Nj}$, which is at most 1 by (1.10) for all $1 \leq i < j \leq N$. In other words, $\mathbb{G}_N$ has two levels of randomness: one in the choice of $(R_{N1}, \ldots, R_{NN})$ and one in the choice of the set of edges. Once again, A_N is the adjacency matrix of $\mathbb{G}_N$. The following is a *randomized* version of Theorem 1.1.

Theorem 1.4. *Under the assumptions* (1.1) *and* (1.8)–(1.9)*,*

$$\lim_{N\to\infty} \mathrm{ESD}\left((N\varepsilon_N)^{-1/2} A_N\right) = \mu_r \boxtimes \mu_s, \ \textit{weakly in probability},$$

where μ_s is the standard semicircle law.

2. Preparatory approximations

The proofs of Theorems 1.1–1.3 in Section 3 rely on several preparatory approximations, which we organize in Lemmas 2.1–2.5 below. Along the way we need several basic facts, which we collect in Appendix A.

2.1. Centering

The first approximation is that the mean of each off-diagonal entry of A_N and Δ_N can be subtracted, with negligible perturbation in the respective empirical spectral distributions.

Lemma 2.1. *Let A_N^0 and Δ_N^0 be $N \times N$ matrices defined by*

$$\begin{aligned} A_N^0(i,j) &= (N\varepsilon_N)^{-1/2} \left[A_N(i,j) - \mathrm{E}\left(A_N(i,j)\right)\right], \\ \Delta_N^0(i,j) &= (N\varepsilon_N)^{-1/2} \left[\Delta_N(i,j) - \mathrm{E}\left(\Delta_N(i,j)\right)\right], \end{aligned}$$

for all $1 \leq i, j \leq N$. Then

$$\lim_{N\to\infty} L\left(\mathrm{ESD}(A_N^0), \mathrm{ESD}((N\varepsilon_N)^{-1/2} A_N)\right) = 0 \quad \textit{in probability},$$
$$\lim_{N\to\infty} L\left(\mathrm{ESD}(\Delta_N^0), \mathrm{ESD}((N\varepsilon_N)^{-1/2} (\Delta_N - D_N))\right) = 0 \quad \textit{in probability},$$

where $L(\eta_1, \eta_2)$ denotes the Lévy distance between the probability measures η_1 and η_2, and D_N is the diagonal matrix defined in (1.3)*.*

Proof. An appeal to Fact A.1 shows that

$$
\begin{aligned}
&L^3\left(\mathrm{ESD}(A_N^0), \mathrm{ESD}((N\varepsilon_N)^{-1/2}A_N)\right)\\
&\le \frac{1}{N^2\varepsilon_N}\sum_{i=1}^N\sum_{j=1}^N \mathrm{E}^2\left(A_N(i,j)\right)\\
&= \frac{1}{N^2\varepsilon_N}\sum_{i=1}^N\sum_{j=1,\neq i}^N \varepsilon_N^2 f^2\left(\frac{i}{N},\frac{j}{N}\right)\\
&= [1+o(1)]\,\varepsilon_N\int_0^1\int_0^1 f^2(x,y)\,dx\,dy\,, \qquad N\to\infty\,.
\end{aligned}
$$

The first claim follows by recalling that $\varepsilon_N \downarrow 0$. The proof the second claim is verbatim the same. $\square$

2.2. Gaussianisation

One of the crucial steps in studying the scaling properties of ESD is to replace each entry by a Gaussian random variable.

Lemma 2.2. *Let $(G_{i,j}\colon\, 1\le i\le j)$ be a family of i.i.d. standard Gaussian random variables. Define $N\times N$ matrices A_N^g and Δ_N^g by*

$$
A_N^g(i,j) \;=\; \begin{cases}\sqrt{\frac{1}{N}f\left(\frac{i}{N},\frac{j}{N}\right)\left(1-\varepsilon_N f\left(\frac{i}{N},\frac{j}{N}\right)\right)}G_{i\wedge j,i\vee j}, & i\neq j\,,\\ 0, & i=j\,,\end{cases} \tag{2.1}
$$

$$
\Delta_N^g(i,j) \;=\; \begin{cases} A_N^g(i,j), & i\neq j\,,\\ -\sum_{\substack{1\le k\le N\\ k\neq i}} A_N^g(i,k), & i=j\,.\end{cases} \tag{2.2}
$$

Fix $z\in\mathbb{C}\setminus\mathbb{R}$ and a three times continuously differentiable function $h\colon\,\mathbb{R}\to\mathbb{R}$ such that

$$
\max_{0\le j\le 3}\sup_{x\in\mathbb{R}}|h^{(j)}(x)|<\infty\,.
$$

For an $N\times N$ real symmetric matrix M, define

$$
H_N(M)=\frac{1}{N}\operatorname{Tr}\left((M-zI_N)^{-1}\right)\,,
$$

where I_N is the identity matrix of order N. Then

$$
\lim_{N\to\infty}\mathrm{E}\left[h\left(\Re H_N(A_N^g)\right)-h\left(\Re H_N(A_N^0)\right)\right] \;=\; 0\,, \tag{2.3}
$$

$$
\lim_{N\to\infty}\mathrm{E}\left[h\left(\Im H_N(A_N^g)\right)-h\left(\Im H_N(A_N^0)\right)\right] \;=\; 0\,, \tag{2.4}
$$

and

$$
\lim_{N\to\infty}\mathrm{E}\left[h\left(\Re H_N(\Delta_N^g)\right)-h\left(\Re H_N(\Delta_N^0)\right)\right] \;=\; 0\,, \tag{2.5}
$$

$$\lim_{N\to\infty} \mathrm{E}\left[h\left(\Im H_N(\Delta_N^g)\right) - h\left(\Im H_N(\Delta_N^0)\right)\right] \;=\; 0\,, \tag{2.6}$$

where $\Re$ and $\Im$ denote the real and the imaginary part of a complex number, respectively.

Proof. We only prove (2.5). The proofs of the other claims are similar. We use ideas from Chatterjee (2005). Let $z = u + iv \in \mathbb{C}^+$ and $n = N(N-1)/2$. Define $\phi\colon \mathbb{R}^n \to \mathbb{C}$ as

$$\phi(x) = H_N(\Delta(x)) \tag{2.7}$$

where $\Delta(x)$ is the $N \times N$ symmetric Laplacian matrix given by

$$\Delta(x)(i,j) = \begin{cases} -\sum_{k=1,k\neq i}^{N} x_{i,k} & i = j \\ x_{i\wedge j, i\vee j} & i \neq j. \end{cases}$$

Note that $\partial\Delta(x)/\partial x_{ij}$ is the $N \times N$ matrix that has -1 at the i-th and j-th diagonal and 1 at (i,j)-th and (j,i)-th entry. The following identities were derived in (Chatterjee, 2005, Section 2):

$$\begin{aligned} \frac{\partial\phi}{\partial x_{i,j}} &= -N^{-1}\,\mathrm{Tr}\left(\frac{\partial\Delta}{\partial x_{i,j}}K^2\right), \\ \frac{\partial^2\phi}{\partial x_{i,j}^2} &= 2N^{-1}\,\mathrm{Tr}\left(\frac{\partial\Delta}{\partial x_{i,j}}K\frac{\partial\Delta}{\partial x_{i,j}}K^2\right), \\ \frac{\partial^3\phi}{\partial x_{i,j}^3} &= -6N^{-1}\,\mathrm{Tr}\left(\frac{\partial\Delta}{\partial x_{i,j}}K\frac{\partial\Delta}{\partial x_{i,j}}K\frac{\partial\Delta}{\partial x_{i,j}}K^2\right), \end{aligned} \tag{2.8}$$

where $K(x) = (\Delta(x) - zI)^{-1}$. Now using these identities we get

$$\Big\|\frac{\partial\phi}{\partial x_{ij}}\Big\|_\infty \le \frac{4}{|\Im z|^2}\frac{1}{N},\quad \Big\|\frac{\partial^2\phi}{\partial x_{ij}^2}\Big\|_\infty \le \frac{8}{|\Im z|^3}\frac{1}{N},\quad \Big\|\frac{\partial^3\phi}{\partial x_{ij}^3}\Big\|_\infty \le \frac{48}{|\Im z|^4}\frac{1}{N}\,.$$

If we define

$$\begin{aligned} \lambda_2(\phi) &= \sup\left\{\Big\|\frac{\partial\phi}{\partial x_{i,j}}\Big\|_\infty^2,\ \Big\|\frac{\partial^2\phi}{\partial x_{i,j}^2}\Big\|_\infty\right\}, \\ \lambda_3(\phi) &= \sup\left\{\Big\|\frac{\partial\phi}{\partial x_{i,j}}\Big\|_\infty^3, \Big\|\frac{\partial^2\phi}{\partial x_{i,j}^2}\Big\|_\infty^2, \Big\|\frac{\partial^3\phi}{\partial x_{i,j}^3}\Big\|_\infty\right\}, \end{aligned}$$

then there exists constants C_2 and C_3 depending on $\Im z$ such that $\lambda_2(\phi) \le C_2 N^{-1}$ and $\lambda_3(\phi) \le C_3 N^{-1}$. Hence, using $\lambda_r(\Re\phi) \le \lambda_r(\phi)$ and

$$U = \Re\left(H_N(\Delta_N^0)\right), \qquad V = \Im\left(H_N(\Delta_N^g)\right), \tag{2.9}$$

we have from (Chatterjee, 2005, Theorem 1.1)

$$|\mathrm{E}[h(U)] - \mathrm{E}[h(V)]|$$

$$
\begin{aligned}
&\leq C_1(h)\lambda_2(\phi)\sum_{1\leq i\neq j\leq N}\big(\mathrm{E}[A_N^0(i,j)^2 I(|A_N^0(i,j)|>K)\\
&\quad+\mathrm{E}[A_N^g(i,j)^2 I(|A_N^g(i,j)|>K)\big)\\
&\quad+C_2(h)\frac{\lambda_3(\phi)}{(N\varepsilon_N)^{3/2}}\sum_{i\neq j}\big(\mathrm{E}[A_N^0(i,j)^3 I(|C_N(i,j)|>k)\\
&\quad+\mathrm{E}[(A_N^g(i,j)^3 I(|A_N^g(i,j)|>k)\big)\ .
\end{aligned}
\tag{2.10}
$$

Using the fact that $\varepsilon_N \downarrow 0$, we have that $\mathrm{E}[A_N^0(i,j)^4] = \mathrm{O}(N^{-2}\varepsilon_N^{-1})$. Also

$$
P(|A_N^0(i,j)| > K) \leq \mathrm{O}(N^{-1}).
$$

So, by the Cauchy-Schwartz inequality and the above bounds, we have

$$
\mathrm{E}[A_N^0(i,j)^2 I(|A_N^0(i,j)| > K) \leq \mathrm{O}\left(\varepsilon_N^{-1/2}N^{-3/2}\right).
$$

Since $N\varepsilon_N \to \infty$, we have

$$
\lambda_2(\phi)\sum_{1\leq i\neq j\leq N}\mathrm{E}[A_N^0(i,j)^2 I(|A_N^0(i,j)|>K) \leq CN^{-1/2}\varepsilon_N^{-1/2}\to 0\,,\quad N\to\infty\,.
$$

Similarly, we have

$$
\lambda_3(\phi)\sum_{i\neq j}\mathrm{E}[A_N^0(i,j)^3 I(|A_N^0(i,j)|>K) \leq \frac{C}{N^{5/2}\varepsilon_N^{3/2}}N^2\varepsilon_N \to 0\,,\quad N\to\infty\,.
$$

Using Gaussian tail bounds, we can also show that the other two terms in (2.10) go to 0, which settles (2.5). A similar computation can be done for the imaginary part in (2.9), which proves (2.6). The proofs of (2.3) and (2.4) are analogous (and, in fact, closer to the argument in Chatterjee (2005)). □

2.3. Leading order variance

Next, we show that another minor tweak to the entries of A_N^g and Δ_N^g results in a negligible perturbation.

Lemma 2.3. *Define an $N \times N$ matrix A_N by*

$$
\bar{A}_N(i,j) \quad = \quad \sqrt{\frac{1}{N}f\left(\frac{i}{N},\frac{j}{N}\right)}\,G_{i\wedge j,i\vee j},\qquad 1\leq i,j\leq N\,,
\tag{2.11}
$$

and let

$$
\bar{\Delta}_N = \bar{A}_N - X_N\,,
$$

where X_N is a diagonal matrix of order N, defined by

$$
X_N(i,i) = \sum_{1\leq k\leq N,k\neq i}\bar{A}_N(i,k),\qquad 1\leq i\leq N\,.
$$

Then

$$\lim_{N\to\infty} L\left(\mathrm{ESD}(A_N^g), \mathrm{ESD}(\bar{A}_N)\right) = 0 \quad \textit{in probability}, \tag{2.12}$$

$$\lim_{N\to\infty} L\left(\mathrm{ESD}(\Delta_N^g), \mathrm{ESD}(\bar{\Delta}_N)\right) = 0 \quad \textit{in probability}. \tag{2.13}$$

Proof. To prove (2.13), yet another application of Fact A.1 implies that

$$\begin{aligned}
&\mathrm{E}\left[L^3\left(\mathrm{ESD}(\Delta_N^g), \mathrm{ESD}(\bar{\Delta}_N)\right)\right] \\
&\le \frac{1}{N}\mathrm{E}\left(\mathrm{Tr}\left[\left(\Delta_N^g - \bar{\Delta}_N\right)^2\right]\right) \\
&= \frac{1}{N}\sum\sum_{1\le i\ne j\le N} \mathrm{Var}\left(\bar{A}_N(i,j) - A_N^g(i,j)\right) \\
&\quad + \frac{1}{N}\sum_{i=1}^{N} \mathrm{Var}\left(\sum_{\substack{j=1\\ j\ne i}}^{N}\left(\bar{A}_N(i,j) - A_N^g(i,j)\right)\right) + \frac{1}{N^2}\sum_{i=1}^{N} f\left(\frac{i}{N},\frac{i}{N}\right) \\
&= \frac{4}{N^2}\sum\sum_{1\le i<j\le N} f\left(\frac{i}{N},\frac{j}{N}\right)\left(1-\sqrt{1-\varepsilon_N f\left(\frac{i}{N},\frac{j}{N}\right)}\right)^2 \\
&\qquad + \frac{1}{N^2}\sum_{i=1}^{N} f\left(\frac{i}{N},\frac{i}{N}\right) \\
&\to 0, \qquad N\to\infty,
\end{aligned}$$

because f is bounded. Thus, (2.13) follows. The proof of (2.12) is similar. □

2.4. Decoupling

The (diagonal) entries of X_N are nothing but the row sums of $\bar{A}_N$. However, the correlation between an entry of $\bar{A}_N$ and that of X_N is small. The following decoupling lemma shows that it does not hurt when the entries of X_N are replaced by a mean-zero Gaussian random variable of the same variance that is independent of $\bar{A}_N$.

Lemma 2.4. *Let $(Z_i\colon\, i\ge 1)$ be a family of i.i.d. standard normal random variables, independent of $(G_{i,j}\colon\, 1\le i\le j)$. Define a diagonal matrix Y_N of order N by*

$$Y_N(i,i) = Z_i\sqrt{\frac{1}{N}\sum_{1\le j\le N,\, j\ne i} f\left(\frac{i}{N},\frac{j}{N}\right)}, \qquad 1\le i\le N,$$

and let

$$\tilde{\Delta}_N = \bar{A}_N + Y_N.$$

Then, for every $k\in\mathbb{N}$,

$$\lim_{N\to\infty}\frac{1}{N}\mathrm{E}\left(\mathrm{Tr}\left[(\tilde{\Delta}_N)^{2k} - (\bar{\Delta}_N)^{2k}\right]\right) = 0, \tag{2.14}$$

and

$$\lim_{N\to\infty}\frac{1}{N^2}\mathrm{E}\left(\mathrm{Tr}^2\left[(\tilde{\Delta}_N)^{k}\right] - \mathrm{Tr}^2\left[(\bar{\Delta}_N)^{k}\right]\right) = 0. \tag{2.15}$$

Proof. Without loss of generality we may assume that $f \le 1$. For $N \ge 1$, define the $N \times N$ matrices $\bar{M}_N$ and $\tilde{M}_N$ by

$$\bar{M}_N(i,j) = \begin{cases} N^{-1/2} G_{i\wedge j, i\vee j}, & i \neq j\,, \\ N^{-1/2} G_{i,i} - \sum_{k=1,\, k\neq i}^N \bar{M}_N(i,k), & i = j\,, \end{cases}$$

and

$$\tilde{M}_N(i,j) = \begin{cases} \bar{M}_N(i,j), & i \neq j\,, \\ N^{-1/2} G_{i,i} + Z_i \sqrt{\frac{N-1}{N}}, & i = j\,. \end{cases}$$

Note that, in the special case where f is identically 1, $\bar{M}_N$ and $\tilde{M}_N$ are identical to $\bar{\Delta}_N$ and $\tilde{\Delta}_N$, respectively. For $k \in \mathbb{N}$ and Π a partition of $\{1,\dots,2k\}$, let

$$\begin{aligned} \Psi(\Pi,N) \quad = \quad & \Big\{ i \in \{1,\dots,N\}^{2k} \colon\ i_u = i_v \\ & \qquad \iff u,v \text{ belong to the same block of } \Pi \Big\}\,. \end{aligned} \tag{2.16}$$

For fixed Π and N, an immediate application of Wick's formula shows that, for all $i,j \in \Psi(\Pi,N)$,

$$\mathrm{E}\left(\prod_{u=1}^{2k} \bar{M}_N(i_u, i_{u+1}) \right) = \mathrm{E}\left(\prod_{u=1}^{2k} \bar{M}_N(j_u, j_{u+1}) \right)\,,$$

with the convention that $i_{2k+1} \equiv i_1$, and

$$\mathrm{E}\left(\prod_{u=1}^{2k} \tilde{M}_N(i_u, i_{u+1}) \right) = \mathrm{E}\left(\prod_{u=1}^{2k} \tilde{M}_N(j_u, j_{u+1}) \right)\,,$$

Therefore, for any $i \in \Psi(\Pi,N)$, we can unambiguously define

$$\psi(\Pi,N) = \mathrm{E}\left(\prod_{u=1}^{2k} \bar{M}_N(i_u, i_{u+1}) \right) - \mathrm{E}\left(\prod_{u=1}^{2k} \tilde{M}_N(i_u, i_{u+1}) \right)\,.$$

As shown in (Bryc et al., 2006, Lemma 4.12), for a fixed Π,

$$\lim_{N\to\infty} N^{-1} |\psi(\Pi,N)|\, \#\Psi(\Pi,N) = 0\,, \tag{2.17}$$

where $\#$ denotes cardinality of a set.

An immediate observation is that, for all $1 \le i,j,i',j' \le N$,

$$\mathrm{Cov}\left(\tilde{M}_N(i,j), \tilde{M}_N(i',j') \right) = 0 \quad \text{if} \quad (i\wedge j, i\vee j) \neq \left(i'\wedge j', i'\vee j'\right)\,,$$

and likewise for $\tilde{\Delta}_N$. Furthermore,

$$\mathrm{Var}\left(\tilde{M}_N(i,j) \right) = \mathrm{Var}\left(\bar{M}_N(i,j) \right)\,, \qquad 1 \le i,j \le N\,,$$

and likewise for $\tilde{\Delta}_N$ and $\bar{M}_N$. For $N \geq 1$ and $1 \leq i, j, i', j' \leq N$, define

$$\eta_N(i,j,i',j') = \begin{cases} \frac{\mathrm{Cov}\left(\bar{\Delta}_N(i,j),\bar{\Delta}_N(i',j')\right)}{\mathrm{Cov}\left(\bar{M}_N(i,j),\bar{M}_N(i',j')\right)}\,, & \text{if the denominator is non-zero}\,, \\ 0\,, & \text{otherwise}\,. \end{cases}$$

It is easy to check that the assumption $f \leq 1$ ensures that $|\eta_N(i,j,i',j')| \leq 1$. Therefore, for all N and $1 \leq i, j, i', j' \leq N$,

$$\begin{aligned} \mathrm{Cov}\left(\bar{\Delta}_N(i,j), \bar{\Delta}_N(i',j')\right) &= \eta_N(i,j,i',j')\mathrm{Cov}\left(\bar{M}_N(i,j), \bar{M}_N(i',j')\right)\,, \\ \mathrm{Cov}\left(\tilde{\Delta}_N(i,j), \tilde{\Delta}_N(i',j')\right) &= \eta_N(i,j,i',j')\mathrm{Cov}\left(\tilde{M}_N(i,j), \tilde{M}_N(i',j')\right)\,. \end{aligned}$$

For fixed Π, N and $i \in \Psi(\Pi, N)$, by an appeal to Wick's formula the above implies that there exists a $\xi(i, N) \in [-1, 1]$ such that

$$\mathrm{E}\left(\prod_{u=1}^{2k} \bar{\Delta}_N(i_u, i_{u+1})\right) - \mathrm{E}\left(\prod_{u=1}^{2k} \tilde{\Delta}_N(i_u, i_{u+1})\right) = \xi(i,N)\psi(\Pi,N)\,,$$

and therefore, by (2.17),

$$\begin{aligned} &\sum_{i\in\Psi(\Pi,N)} \left|\mathrm{E}\left(\prod_{u=1}^{2k} \bar{\Delta}_N(i_u, i_{u+1})\right) - \mathrm{E}\left(\prod_{u=1}^{2k} \tilde{\Delta}_N(i_u, i_{u+1})\right)\right| \\ &= \sum_{i\in\Psi(\Pi,N)} |\xi(i,N)|\,|\psi(\Pi,N)| \leq |\psi(\Pi,N)|\,\#\Psi(\Pi,N) = o(N)\,, \quad N \to \infty\,. \end{aligned}$$

Since this holds for every partition Π of $\{1, \ldots, 2k\}$, (2.14) follows. The proof of (2.15) follows along similar lines. □

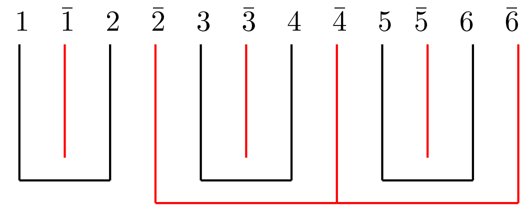
2.5. Combinatorics from free probability

The final preparation is a general result from random matrix theory. To state this, the following notions from the theory of free probability are borrowed, the details of which can be found in Nica and Speicher (2006).

Definition. *For an even positive integer k, $NC_2(k)$ is the set of non-crossing pair partitions of $\{1, \ldots, k\}$. For $\sigma \in NC_2(k)$, its* Kreweras complement *$K(\sigma)$ is the maximal non-crossing partition $\bar{\sigma}$ of $\{\bar{1}, \ldots, \bar{k}\}$, such that $\sigma \cup \bar{\sigma}$ is a non-crossing partition of $\{1, \bar{1}, \ldots, k, \bar{k}\}$. For example,*

$$\begin{aligned} K\left(\{(1,4),(2,3)\}\right) &= \{(1,3),(2),(4)\}\,, \\ K\left(\{(1,2),(3,4),(5,6)\}\right) &= \{(1),(2,4,6),(3),(5)\}\,. \end{aligned}$$

The second example is illustrated as:



For $\sigma \in NC_2(k)$ and $N \geq 1$, define

$$
\begin{aligned}
&S(\sigma, N) \\
&= \left\{ i \in \{1, \ldots, N\}^k \colon \ i_u = i_v \iff u, v \text{ belong to the same block of } K(\sigma) \right\}
\end{aligned}
$$

and

$$
C(k, N) = \{1, \ldots, N\}^k \setminus \left(\bigcup_{\sigma \in NC_2(k)} S(\sigma, N) \right) .
$$

In other words, $S(\sigma, N)$ is the same as $\Psi(K(\sigma), N)$ defined in (2.16).

Lemma 2.5. *Suppose that, for each $N \geq 1$, $W_{N,1}, \ldots, W_{N,k}$ are $N \times N$ real (and possibly asymmetric) random matrices, where k is a positive even number. Suppose further that, for each $u = 1, \ldots, k$,*

$$
\max_{1 \leq i,j \leq N} \mathrm{E}\left[W_{N,u}(i,j)^k \right] = O\left(N^{-k/2} \right) \tag{2.18}
$$

and

$$
\lim_{N \to \infty} \mathrm{E}\left[\left(\frac{1}{N} \sum_{i \in C(k,N)} P_i \right)^2 \right] = 0 \, , \tag{2.19}
$$

and that, for every $\sigma \in NC_2(k)$, there exists a deterministic and finite $\beta(\sigma)$ such that

$$
\lim_{N \to \infty} \mathrm{E}\left(\frac{1}{N} \sum_{i \in S(\sigma,N)} P_i \right) = \beta(\sigma) \, , \tag{2.20}
$$

$$
\lim_{N \to \infty} \mathrm{E}\left[\left(\frac{1}{N} \sum_{i \in S(\sigma,N)} P_i \right)^2 \right] = \beta(\sigma)^2 \, , \tag{2.21}
$$

where

$$
P_i = W_{N,1}(i_1, i_2) \ldots W_{N,k-1}(i_{k-1}, i_k) W_{N,k}(i_k, i_1) \, , \qquad i \in \{1, \ldots, N\}^k \, .
$$

Furthermore, let $V_1, V_2, \ldots$ be i.i.d. random variables drawn from some distribution with all moments finite, independent of $(W_{N,j} \colon \ N \geq 1, \, 1 \leq j \leq k)$, and let

$$
U_N = \mathrm{Diag}(V_1, \ldots, V_N), \qquad N \geq 1 \, .
$$

Then, for all choices of $n_1, \dots, n_k \geq 0$,

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\left(U_N^{n_1} W_{N,1} \dots U_N^{n_k} W_{N,k}\right) = c \quad \textit{in } L^2$$

for some deterministic $c \in \mathbb{R}$.

Proof. The fact that the sets $S(\sigma, N)$ are disjoint for different $\sigma \in NC_2(k)$ allows us to write

$$\operatorname{Tr}\left(U_N^{n_1} W_{N,1} \dots U_N^{n_k} W_{N,k}\right) = \sum_{\sigma\in NC_2(k)} \sum_{i\in S(\sigma,N)} \tilde{P}_i + \sum_{i\in C(k,N)} \tilde{P}_i \,,$$

where

$$\tilde{P}_i = \prod_{j=1}^{k} \left(V_{i_j}^{n_j} W_{N,j}(i_j, i_{j+1})\right), \qquad i \in \{1, \dots, N\}^k \,.$$

In order to show that the second sum in the right-hand side is negligible after scaling by N, the independence of $(V_1, V_2, \dots)$ and $(W_{N,j}\colon\ N \geq 1,\, 1 \leq j \leq k)$, together with the fact that the common distribution of the former has finite moments, implies that

$$\mathrm{E}\left[\left(\frac{1}{N} \sum_{i\in C(k,N)} \tilde{P}_i\right)^2\right] \leq K N^{-2} \sum_{i,j\in C(k,N)} \mathrm{E}(P_i P_j) \,,$$

where K is a finite constant. Assumption (2.19) shows that

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i\in C(k,N)} \tilde{P}_i = 0 \quad \text{in } L^2 \,.$$

In order to complete the proof, it suffices to show that for every $\sigma \in NC_2(k)$ there exists a $\theta(\sigma) \in \mathbb{R}$ with

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i\in S(\sigma,N)} \tilde{P}_i = \theta(\sigma) \qquad \text{in } L^2 \,. \tag{2.22}$$

To that end, fix $\sigma \in NC_2(k)$ and note that, for $i \in S(\sigma, N)$,

$$\mathrm{E}(\tilde{P}_i) = \mathrm{E}\left(P_i\right) \mathrm{E}\left(\prod_{j=1}^{k} V_{i_j}^{n_j}\right) = \mathrm{E}\left(P_i\right) \prod_{u\in K(\sigma)} \mathrm{E}\left(V_1^{\sum_{j\in u} n_j}\right) \,, \tag{2.23}$$

the product in the last line being taken over every block u of $K(\sigma)$. Putting

$$\theta(\sigma) = \beta(\sigma) \prod_{u\in K(\sigma)} \mathrm{E}\left(V_1^{\sum_{j\in u} n_j}\right) \,,$$

we see that (2.20) gives

$$\lim_{N\to\infty} \mathrm{E}\left[\frac{1}{N} \sum_{i\in S(\sigma,N)} \tilde{P}_i\right] = \theta(\sigma) \,. \tag{2.24}$$

Let us call $i, j \in \mathbb{N}^k$ "disjoint" if no coordinate of i matches any coordinate of j, i.e.,

$$\min_{1\le u,v\le k} |i_u - j_v| \ge 1\,.$$

Since $K(\sigma)$ has exactly $\frac{1}{2}k+1$ blocks, (2.18) implies that

$$\lim_{N\to\infty} N^{-2} \sum_{\substack{i,j\in S(\sigma,N)\\ i,j \text{ not disjoint}}} \mathrm{E}(\tilde{P}_i\tilde{P}_j) = 0\,.$$

If $i, j \in S(\sigma, N)$ are disjoint, then it is immediate that

$$\mathrm{E}(\tilde{P}_i\tilde{P}_j) = \left[\prod_{u\in K(\sigma)} \mathrm{E}\left(V_1^{\sum_{j\in u} n_j}\right)\right]^2 \mathrm{E}(P_iP_j)\,.$$

The above two displays, in conjunction with (2.21), show that

$$\lim_{N\to\infty} \mathrm{E}\left[\left(\frac{1}{N}\sum_{i\in S(\sigma,N)} \tilde{P}_i\right)^2\right] = \theta(\sigma)^2\,.$$

This, along with (2.24), establishes (2.22), from which the proof follows. □

3. Proof of Theorems 1.1–1.3

Proof of Theorem 1.1. Theorem 2.1 of Chakrabarty (2017) implies that as $N\to\infty$,

$$\lim_{N\to\infty} \mathrm{ESD}\left(\bar{A}_N\right) = \mu \quad \text{weakly in probability}\,,$$

for a compactly supported symmetric probability measure μ. Lemma 2.3 immediately tells us that

$$\lim_{N\to\infty} \mathrm{ESD}\left(A_N^g\right) = \mu \quad \text{weakly in probability}\,,$$

and hence for h and H_N as in Lemma 2.2,

$$\lim_{N\to\infty} \mathrm{E}\left[h\left(\Re H_N(A_N^g)\right)\right] = h\left(\Re\int_{\mathbb{R}} \frac{1}{x-z}\,\mu(dx)\right)\,.$$

The claim in (2.3) shows that A_N^g can be replaced by A_N^0 in the above display. Since the right-hand side is deterministic and the above holds for any h satisfying the hypothesis of Lemma 2.2, it follows that

$$\lim_{N\to\infty} \Re H_N(A_N^0) = \Re\int_{\mathbb{R}} \frac{1}{x-z}\,\mu(dx) \quad \text{in probability}\,.$$

A similar argument works for the imaginary part, which shows that

$$\lim_{N\to\infty} \mathrm{ESD}(A_N^0) = \mu \quad \text{weakly in probability}\,.$$

Lemma 2.1 completes the proof of (1.2).

Finally, if f is bounded away from 0, then the combination of (Chakrabarty, 2017, Lemma 3.1) and (Biane, 1997, Corollary 2) implies that μ is absolutely continuous with respect to the Lebesgue measure (see also Chakrabarty and Hazra (2016)). Thus, the proof of Theorem 1.1 follows. □

Remark 3.1. The moments of μ are specified in Chakrabarty (2017). It turns out that the same limiting spectral distribution arises in a different random matrix model (see Chakrabarty et al. (2016)).

Remark 3.2. A close inspection of the proof reveals that it suffices to assume that f is Riemann integrable instead of continuous. In other words, if f is symmetric and bounded, and its set of discontinuities has Lebesgue measure zero, then the result holds. However, continuity will be used later in (3.3) in the proof of Theorem 1.2.

Proof of Theorem 1.2. The proof comes in 3 Steps.

1. Riemann approximation. For $N \geq 1$, define the $N \times N$ diagonal matrix Q_N by

$$Q_N(i,i) = F(i/N)Z_i, \qquad 1 \leq i \leq N\,, \tag{3.1}$$

where

$$F(x) = \left(\int_0^1 f\,(x,y)\;dy\right)^{1/2}, \qquad 0 \leq x \leq 1\,, \tag{3.2}$$

and $(Z_i\colon\ i \geq 1)$ is as in Lemma 2.4. Fact A.2 in Appendix A implies that

$$\begin{aligned}&\left|\left(\frac{1}{N}\operatorname{Tr}\left((\tilde{\Delta}_N)^k\right)\right)^{1/k} - \left(\frac{1}{N}\operatorname{Tr}\left((\bar{A}_N + Q_N)^k\right)\right)^{1/k}\right| \\ &\qquad \leq \left(\frac{1}{N}\operatorname{Tr}\left[(Y_N - Q_N)^k\right]\right)^{1/k}.\end{aligned}$$

Since, f being continuous,

$$\begin{aligned}&\mathrm{E}\left[N^{-2}\operatorname{Tr}^2\left[(Y_N - Q_N)^k\right]\right] = O(1) \\ &\times \sup_{0\leq x\leq 1}\left[F(x) - \left(\frac{1}{N}\sum_{j=1, j\neq[Nx]/N}^{N} f\left(x, \frac{j}{N}\right)\right)^{1/2}\right]^{2k} \to 0\,, \quad N \to \infty\,,\end{aligned} \tag{3.3}$$

we get, for every even k,

$$\left(\frac{1}{N}\operatorname{Tr}\left((\tilde{\Delta}_N)^k\right)\right)^{1/k} - \left(\frac{1}{N}\operatorname{Tr}\left((\bar{A}_N + Q_N)^k\right)\right)^{1/k} \to 0 \text{ in } L^{2k}\,, \quad N \to \infty\,. \tag{3.4}$$

Our next step is to show that, for every even integer k,

$$\lim_{N\to\infty}\frac{1}{N}\operatorname{Tr}\left((\bar{A}_N + Q_N)^k\right) = \gamma_k \quad \text{in } L^2 \tag{3.5}$$

for some $\gamma_k \in \mathbb{R}$. The above will follow once we show that, for all $m \geq 1$ and $n_1, \dots, n_m \geq 0$,

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\left(Q_N^{n_1} \bar{A}_N \dots Q_N^{n_m} \bar{A}_N\right) = \theta \quad \text{in } L^2 \tag{3.6}$$

for some $\theta \in \mathbb{R}$ (depending on $m, n_1, \dots, n_m$). To that end, define the diagonal matrices U_N and B_N by

$$\begin{aligned} U_N(i,i) &= Z_i\,, \\ B_N(i,i) &= F(i/N)\,, \qquad i = 1, \dots, N\,. \end{aligned}$$

Observe that

$$Q_N = B_N U_N = U_N B_N\,,$$

and hence the left-hand side of (3.6) is the same as

$$\frac{1}{N} \operatorname{Tr}\left(U_N^{n_1} W_{N1} \dots U_N^{n_m} W_{Nm}\right)\,, \tag{3.7}$$

where

$$W_{Nj} = B_N^{n_j} \bar{A}_N, \qquad j = 1, \dots, m\,.$$

In order to apply Lemma 2.5 we need to verify its hypotheses.

2. Verification of the hypotheses. Our next claim is that $W_{N1}, \dots, W_{Nm}$ satisfy (2.18)–(2.21). To that end, observe that for $N \geq 1$ and $j = 1, \dots, m$,

$$W_{Nj}(u,v) = F^{n_j}\left(\frac{u}{N}\right) f^{1/2}\left(\frac{u}{N}, \frac{v}{N}\right) N^{-1/2} G_{u\wedge v,\, u\vee v}, \qquad 1 \leq u, v \leq N\,.$$

Let

$$H_j(x,y) = F^{n_j}(x) f^{1/2}(x,y),\ (x,y) \in [0,1]^2\,.$$

Fix a partition Π of $\{1, \dots, m\}$. Recall the notation $\Psi(\Pi, N)$ in the proof of Lemma 2.4. Clearly, for every $i \in \Psi(\Pi, N)$,

$$\mathrm{E}\left(\prod_{j=1}^m W_{Nj}(i_j, i_{j+1})\right) = N^{-m/2} \psi(\Pi) \left(\prod_{j=1}^m H_j\left(\frac{i_j}{N}, \frac{i_{j+1}}{N}\right)\right)\,,$$

where

$$\psi(\Pi) = \mathrm{E}\left(\prod_{j=1}^m G_{i_j \wedge i_{j+1}, i_j \vee i_{j+1}}\right)\,,$$

which does not depend on $i \in \Psi(\Pi, N)$. The standard arguments leading to a proof via the method of moments of the Wigner semicircle law show that

$$\begin{aligned} &\lim_{N\to\infty} N^{-m/2+1}\, \psi(\Pi)\, \#\Psi(\Pi, N) \\ &= \begin{cases} 1, & \text{if } m \text{ is even, and } \Pi = K(\sigma) \text{ for some } \sigma \in NC_2(m)\,, \\ 0, & \text{otherwise}\,. \end{cases} \end{aligned}$$

Assume for the moment that m is even, and let $\sigma \in NC_2(m)$. It is known that $K(\sigma)$ has $m/2+1$ blocks. Define a function $\mathcal{L}_\sigma : \{1,\dots,m\} \to \{1,\dots,\frac{1}{2}m+1\}$ such that

$$\mathcal{L}_\sigma(j) = \mathcal{L}_\sigma(k) \text{ if and only if } j,k \text{ are in the same block of } K(\sigma)\,.$$

It follows that for $\Pi = K(\sigma)$,

$$\begin{aligned}\lim_{N\to\infty} \frac{1}{N} \sum_{i\in\Psi(\Pi,N)} \mathrm{E}\left(\prod_{j=1}^{m} W_{Nj}(i_j,i_{j+1})\right) \\ = \int_{[0,1]^{(m/2)+1}} \prod_{(u,v)\in\sigma, u<v} H_u\left(x_{\mathcal{L}_\sigma(u)}, x_{\mathcal{L}_\sigma(v)}\right) dx_1 \dots dx_{(m/2)+1}\,.\end{aligned}$$

This shows that hypothesis (2.20) holds. The hypotheses (2.19) and (2.21) follow similarly by an analogue of the standard arguments, while (2.18) is trivial.

Thus, $W_{N1},\dots,W_{Nm}$ and U_N satisfy the hypotheses of Lemma 2.5. The claim of that lemma shows that the random variable in (3.7) converges in L^2 to a finite deterministic constant as $N\to\infty$, i.e., (3.6) holds. This in turn proves (3.5), which in conjunction with (3.4) shows that

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{Tr}\left((\tilde{\Delta}_N)^k\right) = \gamma_k \quad \text{in } L^2\,.$$

Lemma 2.4 asserts that

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{Tr}\left((\bar{\Delta}_N)^k\right) = \gamma_k \quad \text{in } L^2\,, \tag{3.8}$$

and hence also in probability.

3. Uniqueness of the limiting measure. Equation (3.5) ensures that there exists a symmetric probability measure on $\mathbb{R}$ whose k-th moment is γ_k for every even integer k. Our next claim is that such a measure is unique, i.e., $(\gamma_k\colon\ k\geq 1)$ determines the measure. It is not obvious how to check Carleman's condition, and therefore we argue as follows. It suffices to exhibit a probability measure ν whose odd moments are zero and whose k-th moment is γ_k for even k such that

$$\int_{-\infty}^{\infty} e^{tx}\nu(dx) < \infty \quad \forall\, t\in\mathbb{R}\,. \tag{3.9}$$

To do so we bring in the notion of a non-commutative probability space (NCP), which is defined in Appendix A. For $K>0$ and $N\geq 1$, define

$$U_{NK} = \mathrm{Diag}\left(Z_1 \mathbf{1}(|Z_1|\leq K),\dots, Z_1\mathbf{1}(|Z_N|\leq K)\right)\,,$$

and

$$Q_{NK} = B_N U_{NK}\,.$$

The arguments leading to (3.6) can be easily tweaked to show that, for fixed $K>0$ and a fixed polynomial p in two non-commuting variables,

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{E}\,\mathrm{Tr}\left[p\left(\bar{A}_N, Q_{NK}\right)\right] \tag{3.10}$$

exists. Fact A.3 in Appendix A implies that there exist self-adjoint elements q and a in a tracial NCP $(\mathcal{A}, \phi)$ such that the above limit equals $\phi\left[p\left(a, q\right)\right]$ for every polynomial p in two non-commuting variables. Hence

$$\lim_{N\to\infty} \mathrm{EESD}\left[p\left(\bar{A}_N, Q_{NK}\right)\right] = \mathcal{L}\left[p\left(a, q\right)\right] \quad \text{in distibution}\,, \tag{3.11}$$

for any symmetric polynomial p, where EESD denotes the expectation of ESD. Theorem 1.1 implies that the LSD of $\bar{A}_N$, which is $\mathcal{L}(a)$ by (3.11), is compactly supported, and hence a is a bounded element. The spectrum of q is clearly a subset of $[-K, K]$. The second claim in Fact A.3 in Appendix A allows us to assume that $(\mathcal{A}, \phi)$ is a W^*-probability space.

Let

$$\nu_K = \mathcal{L}(a+q)\,.$$

If C is a finite constant such that

$$a \le C\mathbf{1}\,,$$

then clearly

$$a + q \le C\mathbf{1} + q\,. \tag{3.12}$$

Applying the method of moments to Q_{NK}, we find by an appeal to (3.11) that the law of q is same as the law of

$$F(V)Z_1\mathbf{1}(|Z_1| \le K)\,,$$

where V is standard uniform independently of Z_1, and F is as in (3.2). Under the assumption that $f \le 1$, which represents no loss of generality,

$$\int_{-\infty}^{\infty} e^{tx}\left(\mathcal{L}(q)\right)(dx) \le e^{t^2/2},\, t\in\mathbb{R}\,.$$

By (Bercovici and Voiculescu, 1993, Corollary 3.3) applied to (3.12), it follows that

$$\int_{\mathbb{R}} e^{tx}\nu_K(dx) \le \int_{\mathbb{R}} e^{tx}\left(\mathcal{L}(C\mathbf{1}+q)\right)(dx) \le \exp\left(\tfrac{1}{2}t^2 + tC\right)\,, \quad t > 0\,. \tag{3.13}$$

Fact A.1 applied to $\bar{A}_N + Q_{NK_1}$ and $\bar{A}_N + Q_{NK_1}$ shows that

$$\sup_{N\ge 1} L\left(\mathrm{EESD}\left(\bar{A}_N + Q_{NK_1}\right), \mathrm{EESD}\left(\bar{A}_N + Q_{NK_2}\right)\right)$$

is small for large K_1 and K_2. Thus, $(\nu_K\colon\ K>0)$ is Cauchy in the Lévy metric, and hence there exists a probability measure ν such that

$$\lim_{K\to\infty} \nu_K = \nu\,.$$

This, along with (3.13), establishes that

$$\int_{\mathbb{R}} e^{tx}\nu(dx) \le \exp\left(\tfrac{1}{2}t^2 + tC\right)\,, \quad t > 0\,, \tag{3.14}$$

and

$$\lim_{K\to\infty}\int_{\mathbb{R}} x^k\nu_K(dx) = \int_{\mathbb{R}} x^k\nu(dx), \quad k \ge 1\,.$$

Clearly,

$$\int_{-\infty}^{\infty} x^k \nu_K(dx) = \lim_{N\to\infty} N^{-1}\mathrm{E}\,\mathrm{Tr}\left[\left(\bar{A}_N + Q_{NK}\right)^k\right].$$

Therefore, by keeping track of the limit in (3.10), we can show (with some effort) that

$$\lim_{K\to\infty}\int_{\mathbb{R}} x^k \nu_K(dx) = \begin{cases}\gamma_k, & k \text{ even},\\ 0, & k \text{ odd}.\end{cases}$$

Thus, ν has the desired moments. By extending (3.14) to the case $t<0$, we see that (3.9) follows. Thus, ν is the only symmetric probability measure whose even moments are (γ_k).

Equation (3.8) and the claim proved above show that

$$\lim_{N\to\infty} \mathrm{ESD}\left(\bar{\Delta}_N\right) = \nu \quad \text{weakly in probability}.$$

Hence Lemmas 2.1–2.3 imply that

$$\lim_{N\to\infty} \mathrm{ESD}\big((N\varepsilon_N)^{-1/2}(\Delta_N - D_N)\big) = \nu \quad \text{weakly in probability}.$$

as in the proof of Theorem 1.1. □

Proof of Theorem 1.3. Let $(G_{i,j}\colon 1\le i\le j)$ and $(Z_i\colon i\ge 1)$ be as in Lemma 2.4. For $N\ge 1$, define the $N\times N$ matrices

$$\begin{aligned} G_N(i,j) &= N^{-1/2}G_{i\wedge j, i\vee j}, \quad 1\le i,j\le N,\\ R_N &= \mathrm{Diag}\left(\sqrt{r(1/N)},\ldots,\sqrt{r(1)}\right),\\ U_N &= \mathrm{Diag}(Z_1,\ldots,Z_N).\end{aligned}$$

The notation U_N is exactly as in the proof of Theorem 1.2. Let $\bar{A}_N$ and Q_N be as in (2.11) and (3.1), respectively. Observe that, under the assumption (1.4),

$$\bar{A}_N = R_N G_N R_N,$$

and

$$Q_N = \alpha R_N^{1/2} U_N R_N^{1/2},$$

where α is as defined in the statement of Theorem 1.3. Proceeding as in the proofs of Theorems 1.1 and 1.2, wee see that it suffices to show that

$$\lim_{N\to\infty} \mathrm{ESD}\left(R_N G_N R_N\right) = \mathcal{L}\left(r^{1/2}(T_u)T_s T^{1/2}(T_u)\right) \quad \text{weakly in probability} \tag{3.15}$$

and

$$\begin{aligned}&\lim_{N\to\infty} \mathrm{ESD}\left(R_N G_N R_N + \alpha R_N^{1/2} U_N R_N^{1/2}\right)\\ &= \mathcal{L}\left(r^{1/2}(T_u)T_s T^{1/2}(T_u) + \alpha r^{1/4}(T_u)T_g r^{1/4}(T_u)\right) \text{ weakly in probability},\end{aligned} \tag{3.16}$$

where T_s, T_g, T_u are as in the statement. Define U_{NK} to be the "truncated" version of U_N, for a fixed $K > 0$, as in the proof of Theorem 1.2. Both (3.15) and (3.16) will follow once we show that

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\left(p\left(R_N^{1/2}, U_{NK}, G_N\right)\right) = \tau\left(p\left(T_r, T_g', T_s\right)\right) \quad \text{in probability}, \tag{3.17}$$

where $T_r = r^{1/4}(T_u)$ and $T_g' = T_g \mathbf{1}(|T_g| \le K)$, for any symmetric polynomial p in three non-commuting variables. It is a well known fact that, for all $k \ge 1$,

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}(G_N^k) = \tau(T_s^k) \quad \text{in probability}. \tag{3.18}$$

Since R_N and U_{NK} are diagonal matrices, they commute. This, in conjunction with the strong law of large numbers, implies that, for any $k \ge 1$, $m_1, \ldots, m_k$ and $n_1, \ldots, n_k \ge 0$,

$$\begin{aligned}&\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\left(R_N^{m_1} U_{NK}^{n_1} \ldots R_N^{m_k} U_{NK}^{n_k}\right)\\ &= \int_0^1 du\, r^{(m_1+\ldots+m_k)/4}(u) \int_{-K}^{K} (2\pi)^{-1/2} dx\, x^{n_1+\ldots+n_k} e^{-x^2/2} \quad a.s.\end{aligned}$$

The above, in conjunction with (1.7) and the fact that T_g and T_r commute, implies that

$$\lim_{N\to\infty} \frac{1}{N} \operatorname{Tr}\left(p\left(R_N^{1/2}, U_{NK}\right)\right) = \tau\left(p\left(T_r, T_g'\right)\right) \quad a.s. \tag{3.19}$$

for any polynomial p in two variables.

Thus, all that remains to show is the asymptotic free independence of T_s and (T_r, T_g'), which is precisely the claim of Fact A.4 in Appendix A, i.e., (3.18) and (3.19) imply (3.17). Applying (3.17) to $p(x, y, z) = x^2 z x^2$ and $p(x, y, z) = x^2 z x^2 + \alpha x y x$, we get the truncated versions of (3.15) and (3.16), respectively. Yet another application of Fact A.1 in Appendix A allows us to let $K \to \infty$, obtaining (3.15) and (3.16). This completes the proof of (1.5) and (1.6). □

4. Proof of Theorem 1.4

Proof of Theorem 1.4. Lemma 2.1 and the assumption (1.1) imply that the mean of the entries of A_N can be subtracted at the cost of a negligible perturbation of the ESD. The inequalities (1.1) and (1.8) ensure that the Gaussianization as in Lemma 2.2 goes through by conditioning on $R_{N1}, \ldots, R_{NN}$. That is, if $(G_{ij}\colon 1 \le i \le j)$ is a collection of i.i.d. standard normal random variables that are independent of $(R_{Ni}\colon 1 \le i \le N, N \ge 1)$, and A_N^g is an $N \times N$ matrix defined by

$$A_N^g(i,j) = \sqrt{R_{Ni} R_{Nj}}\, G_{i\wedge j, i\vee j}, \qquad 1 \le i, j \le N,$$

then the ESD of $A_N/\sqrt{N\varepsilon_N}$ is close to that of $A_N^g/\sqrt{N}$. Finally, (1.9) by an appeal to Fact A.4 shows that

$$\lim_{N\to\infty} \operatorname{ESD}\left(N^{-1/2} A_n^g\right) = \mu_r \boxtimes \mu_s \text{ weakly in probability},$$

which uses (1.8) yet once again. □

5. Applications

In this section we discuss a few applications.

5.1. Constrained random graphs

Let $\mathcal{S}_N$ be the set of all simple graphs on N vertices. Suppose that we fix the degrees of the vertices, namely, vertex i has degree k_i^*. Here, $k^* = (k_i^*\colon 1 \le i \le N)$ is a sequence of *positive integers* of which we only require that they are *graphical*, i.e., there is at least one simple graph with these degrees. The so-called *canonical ensemble* P_N is the unique probability distribution on S_N with the following two properties:

(I) The *average degree* of vertex i, defined by $\sum_{G\in\mathcal{S}_N} k_i(G)P_N(G)$, equals k_i^* for all $i \le i \le N$.
(II) The *entropy* of P_N, defined by $-\sum_{G\in\mathcal{S}_N} P_N(G)\log P_N(G)$, is maximal.

The name canonical ensemble comes from Gibbs theory in equilibrium statistical physics. The probability distribution P_N describes a random graph of which we have *no prior information* other than the average degrees. It is known that, because of property (II), P_N takes the form (Jaynes (1957))

$$P_N(G) = \frac{1}{Z_N(\theta)} \exp\left[-\sum_{i=1}^N \theta_i k_i(G)\right], \qquad G \in \mathcal{S}_N\,,$$

where $\theta = (\theta_i\colon 1 \le i \le N)$ is a sequence of real-valued Lagrange multipliers that must be chosen in such a way that property (I) is satisfied. The normalization constant $Z_N(\theta)$, which depends on θ, is called the partition function in Gibbs theory.

The matching of property (I) uniquely fixes θ, namely, it turns out that (Squartini et al. (2015))

$$P_N(G) = \prod_{1\le i<j\le N}^{N} (p_{ij}^*)^{A_N[G](i,j)}\,(1-p_{ij}^*)^{1-A_N[G](i,j)}\,, \qquad G \in \mathcal{S}_N\,,$$

where $A_N[G]$ is the adjacency matrix of G, and p_{ij}^* represent a reparameterisation of the Lagrange multipliers, namely,

$$p_{ij}^* = \frac{x_i^* x_j^*}{1+x_i^* x_j^*}\,, \qquad 1 \le i \ne j \le N\,, \tag{5.1}$$

with $x_i^* = e^{-\theta_i^*}$. Thus, we see that P_N is nothing other than an inhomogeneous Erdős-Rényi random graph where the probability that vertices i and j are connected by an edge equals p_{ij}^*. In order to match property (I), these probabilities must satisfy

$$k_i^* = \sum_{\substack{1\le j\le N\\ j\ne i}} p_{ij}^*\,, \qquad 1 \le i \le N\,, \tag{5.2}$$

which constitutes a set of N equations for the N unknowns $x_1^*,\ldots,x_N^*$.

In order to proceed, we need to make assumptions on the sequence $(k^*_{Ni}\colon\ 1\le i\le N)$. For the sake of notational simplification, the dependence on N will be suppressed in the notation. Let $(k^*_i\colon\ 1\le i\le N)$ be a graphical sequence of positive integers in the so-called *sparse regime*, i.e.,

$$m_N := \max_{1\le\ell\le N} k^*_\ell = o(\sqrt{N}\,)\,, \qquad N\to\infty\,. \tag{5.3}$$

Furthermore, assume that

$$\lim_{N\to\infty} m_N = \infty\,, \tag{5.4}$$

and that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N \delta_{k^*_i/m_N} = \mu_r \text{ weakly}\,, \tag{5.5}$$

as $N\to\infty$, for some probability measure μ_r. Let x^*_i and p^*_{ij} be determined by (5.1) and (5.2). Let A_N be the adjacency matrix of an inhomogeneous Erdős-Rényi random graph on N vertices, with p^*_{ij} being the probability of an edge being present between i and j for all $1\le i\ne j\le N$.

It is known that (Squartini et al. (2015))

$$\max_{1\le\ell\le N} x^*_\ell = o(1)\,,$$

in which case (5.1) and (5.2) give

$$\begin{aligned} x^*_i &= [1+o(1)]\,\frac{k^*_i}{\sqrt{\sum_{1\le\ell\le N} k^*_\ell}}\,,\\ p^*_{ij} &= [1+o(1)]\,\frac{k^*_i k^*_j}{\sum_{1\le\ell\le N} k^*_\ell}\,, \end{aligned} \tag{5.6}$$

as $N\to\infty$ with the error term *uniform* in $1\le i\ne j\le N$.

Abbreviate

$$\sigma_N := \sum_{1\le\ell\le N} k^*_\ell\,,$$

and pick

$$\varepsilon_N = \frac{m_N^2}{\sigma_N}\,.$$

It follows from (5.3) and (5.4) that, respectively,

$$\lim_{N\to\infty} N\varepsilon_N = \infty\,,$$

and

$$\lim_{N\to\infty} \varepsilon_N = 0\,.$$

As in the proof of Theorem 1.4, Lemmas 2.1–2.2 imply that the upper triangular entries of A_N can be replaced by independent mean zero normal random variables. In other words, if $(G_{ij}\colon\ 1\le i\le j)$ are i.i.d. standard normal, and A^g_N is the random matrix defined by

$$A^g_N(i,j) = \sqrt{p^*_{ij}}\,G_{i\wedge j,i\vee j},\ 1\le i,j\le N\,,$$

with $p^*_{ii} = 0$ for all i, then $\mathrm{ESD}\big((N\varepsilon_N)^{-1/2} A_N\big)$ and $\mathrm{ESD}\big((N\varepsilon_N)^{-1/2} A^g_N\big)$ are asymptotically close.

The second line of (5.6) implies that

$$\sqrt{p^*_{ij}} = \big[1 + o(1)\big] \sqrt{\varepsilon_N \, \frac{k^*_i k^*_j}{m_N^2}}\,,$$

uniformly in $1 \leq i \neq j \leq N$, and hence

$$\sum_{i,j=1}^N \left[\sqrt{p^*_{ij}} - \sqrt{\varepsilon_N \, \frac{k^*_i k^*_j}{m_N^2}} \right]^2 = o\left(N^2 \varepsilon_N\right) \,.$$

In other words, if $\tilde{A}_N$ is defined by

$$\tilde{A}_N(i,j) = \sqrt{\frac{k^*_i k^*_j}{m_N^2}}\, G_{i\wedge j, i\vee j}\,, \qquad 1 \leq i,j \leq N\,,$$

then

$$\lim_{N\to\infty} \frac{1}{N} \mathrm{E}\left[\mathrm{Tr}\big((N\varepsilon_N)^{-1/2} A^g_N - N^{-1/2} \tilde{A}_N\big)^2 \right] = 0\,.$$

Fact A.1 implies that

$$\lim_{N\to\infty} L\left(\mathrm{ESD}\big((N\varepsilon_N)^{-1/2} A^g_N\big), \mathrm{ESD}\big(N^{-1/2}\tilde{A}_N\big) \right) = 0 \text{ in probability}\,.$$

Finally, by an appeal to Fact A.4, (5.5) implies that

$$\lim_{N\to\infty} \mathrm{ESD}\big(N^{-1/2}\tilde{A}_N\big) = \mu_r \boxtimes \mu_s, \text{ weakly in probability}\,,$$

where μ_s is the standard semicircle law. Hence

$$\lim_{N\to\infty} \mathrm{ESD}\big((N\varepsilon_N)^{-1/2} A_N\big) = \mu_r \boxtimes \mu_s, \text{ weakly in probability}\,.$$

We close by looking at a concrete example of a graphical sequence $(k^*_i \colon 1 \leq i \leq N)$ satisfying (5.3)–(5.5). For $N \geq 1$, let

$$k^*_i = \lfloor i^{1/3} \rfloor\,, \qquad 1 \leq i \leq N\,.$$

Then it is immediate that

$$m_N = \lfloor N^{1/3} \rfloor = o(\sqrt{N})\,,$$

and

$$\lim_{N\to\infty} m_N = \infty\,,$$

and

$$\lim_{N\to\infty} \left(\frac{1}{N} \sum_{i=1}^N \delta_{k^*_i/m_N} \right)(\cdot) = P(U^{1/3} \in \cdot\,) \text{ weakly}\,,$$

where U is a standard uniform random variable. Finally, (van der Hofstad, 2017, Theorem 7.12) implies that $(k^*_i \colon 1 \leq i \leq N)$ is graphical for N large enough.

5.2. Chung-Lu graphs

The following random graph introduced by Chung and Lu (2002) is similar to the one discussed in Section 5.1. For $N \geq 1$, let $(d_{Ni}\colon 1 \leq i \leq N)$ be positive real numbers satisfying the following. For fixed N, let

$$m_N := \max_{1\leq i\leq N} d_{Ni}\,, \qquad \sigma_N := \sum_{i=1}^{N} d_{Ni}\,,$$

and assume that

$$\lim_{N\to\infty} \frac{m_N^2}{\sigma_N} = 0\,, \qquad \lim_{N\to\infty} N\frac{m_N^2}{\sigma_N} = \infty\,,$$

and

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta_{d_{Ni}/m_N} = \mu_r \text{ weakly}\,,$$

for some measure μ_r on $\mathbb{R}$. Consider an inhomogeneous Erdős-Rényi graph on N vertices where an edge exists between i and j with probability $d_{Ni}d_{Nj}/\sigma_N$, for $1 \leq i \neq j \leq N$. This is the so-called Chung-Lu graph. If A_N denotes its adjacency matrix, then Theorem 1.4 implies that

$$\lim_{N\to\infty} \mathrm{ESD}\left((N\varepsilon_N)^{-1/2}A_N\right) = \mu_r \boxtimes \mu_s\,, \text{ weakly in probability}\,,$$

where

$$\varepsilon_N = \frac{m_N^2}{\sigma_N}\,, \qquad N \geq 1\,,$$

and μ_s is the standard semicircle law.

5.3. Social networks

Consider a community consisting of N individuals. Data is available on whether the i-th individual and the j-th individual are acquainted, for every pair $\{i,j\}$ with $1 \leq i,j \leq N$. Based on this data, the *sociability pattern* of the community has to be inferred statistically. Examples arise in social networks and collaboration networks.

The above situation can be modeled in several ways, one being the following. Denote by ρ the *sociability distribution* of the community, which is a compactly supported probability measure on $[0,\infty)$. Let $(R_i)_{1\leq i\leq N}$ be i.i.d. random variables drawn from ρ. Think of R_i as the *sociability index* of the i-th individual. Fix $\varepsilon_N > 0$ such that $\varepsilon_N m^2 \leq 1$, where m is the supremum of the support of ρ, so that

$$0 \leq \varepsilon_N R_i R_j \leq 1\,, \qquad 1 \leq i \neq j \leq N\,. \tag{5.7}$$

Suppose that, conditional on $(R_i)_{1\leq i\leq N}$, the i-th and the j-th individual are acquainted with probability $\varepsilon_N R_i R_j$. In other words, the graph in which the vertices are individuals and the edges are mutual acquaintances is an inhomogeneous Erdős-Rényi random graph

$\mathbb{G}_N$ with random connection parameters that are controlled by ν. The data that are available is the adjacency matrix A_N of this graph. The goal is to draw information about ρ from this data. This statistical inference problem boils down to estimating ρ from A_N. In order to make the model identifiable, we assume that

$$\int_0^\infty x\rho(dx) = 1\,. \tag{5.8}$$

It is immediate that

$$\lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N \delta_{R_i} = \rho\,, \text{ weakly, almost surely}\,.$$

Theorem 1.4 implies that if $N^{-1} \ll \varepsilon_N \ll 1$, then

$$\lim_{N\to\infty} \mathrm{ESD}\left((N\varepsilon_N)^{-1/2}A_N\right) = \rho \boxtimes \mu_s, \text{ weakly, in probability}\,,$$

where μ_s is the standard semicircle law. In practice, ε_N will be unknown, which can worked around by using (5.8) to argue that

$$\lim_{N\to\infty} \mathrm{ESD}\left(\sqrt{\frac{N}{\mathrm{Tr}(A_N^2)}}\,A_N\right) = \rho \boxtimes \mu_s, \text{ weakly in probability}\,.$$

Thus, $\rho \boxtimes \mu_s$ can be statistically estimated, in principle, from A_N. Subsequently, ρ can be derived because the moments of $\rho \boxtimes \mu_s$ are a function of those of ρ and μ_s. The moments of the latter being known, the former can be recursively calculated from the estimated moments of $\rho \boxtimes \mu_s$. Since ρ is compactly supported, estimating its moments amounts to estimating the measure.

Appendix A: Basic facts

The following is (Bai and Silverstein, 2010, Corollary A.41), and is also a corollary of the Hoffman-Wielandt inequality.

Fact A.1. *If L denotes the Lévy distance between two probability measures, then for $N \times N$ symmetric matrices A and B,*

$$L^3\left(\mathrm{ESD}(A), \mathrm{ESD}(B)\right) \le \frac{1}{N}\,\mathrm{Tr}\left[(A-B)^2\right]\,.$$

The following is a consequence of the Minkowski and k-Hoffman-Wielandt inequalities. The latter can be found in Exercise 1.3.6 of Tao (2012).

Fact A.2. *For real symmetric matrices A and B of the same order, and an even positive integer k,*

$$\left|\mathrm{Tr}^{1/k}(A^k) - \mathrm{Tr}^{1/k}(B^k)\right| \le \mathrm{Tr}^{1/k}\left[(A-B)^k\right]\,.$$

Definition. *A* non-commutative probability space (NCP) $(\mathcal{A}, \phi)$ *is a unital* $*$*-algebra* $\mathcal{A}$ *equipped with a linear functional* $\phi\colon \mathcal{A} \to \mathbb{C}$ *that is unital, i.e.,*

$$\phi(\mathbf{1}) = 1\,,$$

and positive, i.e.,

$$\phi(a^*a) \geq 0 \textit{ for all } a \in \mathcal{A}\,.$$

An NCP $(\mathcal{A}, \phi)$ *is* tracial *if*

$$\phi(ab) = \phi(ba),\ a, b \in \mathcal{A}\,.$$

Fact A.3. *Suppose that, for every* $n \in \mathbb{N}$, $(\mathcal{A}_n, \phi_n)$ *is a tracial NCP, and there exist self-adjoint* $a_{n1}, \ldots, a_{nk} \in \mathcal{A}_n$ *such that, for every polynomial* p *in* k *non-commuting variables,*

$$\lim_{n\to\infty} \phi_n\left(p\big(a_{n1}, \ldots, a_{nk}\big)\right) = \alpha_p \in \mathbb{C}\,. \tag{A.1}$$

Then there exists a tracial NCP $(\mathcal{A}_\infty, \phi_\infty)$ *and self-adjoint* $a_{\infty 1}, \ldots, a_{\infty k} \in \mathcal{A}_\infty$ *such that, for every polynomial* p *in* k *non-commuting variables,*

$$\phi_\infty\left(p\big(a_{\infty 1}, \ldots, a_{\infty k}\big)\right) = \alpha_p\,.$$

Furthermore, if

$$\sup_{1\leq i\leq k,\, j\geq 1} \left[\phi_\infty\left(a_{\infty i}^{2j}\right)\right]^{1/2j} < \infty\,, \tag{A.2}$$

then $(\mathcal{A}_\infty, \phi_\infty)$ *can be embedded into a* W^**-probability space.*

Proof. Let

$$\mathcal{A}_\infty = \mathbb{C}[X_1, \ldots, X_k]\,,$$

the set of all polynomials in k non-commuting variables. For a monomial

$$p = \alpha X_{i_1} \ldots X_{i_m}\,,$$

define

$$p^* = \overline{\alpha} X_{i_m} \ldots X_{i_1}\,.$$

This defines the $*$-operation on the whole of $\mathcal{A}$. Let

$$\phi_\infty(p) = \alpha_p \text{ for all } p \in \mathcal{A}_\infty\,.$$

It is immediate from (A.1) that ϕ_∞ is positive and unital, i.e., $(\mathcal{A}_\infty, \phi_\infty)$ is an NCP. The desired conclusions are ensured by defining

$$a_{\infty 1} = X_1,\ \ldots,\ a_{\infty k} = X_k\,.$$

Finally, (A.2) implies that $a_{\infty,1}, \ldots, a_{\infty,k}$ are bounded. Hence, by going from polynomials to continuous functions with the help of the Bolzano-Weierstrass theorem, we can embed $(\mathcal{A}_\infty, \phi_\infty)$ into a W^*-probability space. □

The next fact follows from (Mingo and Speicher, 2017, Theorem 4.20) (which is due to Voiculescu) and the discussion immediately following it.

Fact A.4. *Suppose that W_N is an $N \times N$ scaled standard Gaussian Wigner matrix, i.e., a symmetric matrix whose upper triangular entries are i.i.d. normal with mean zero and variance $1/N$. Let D_N^1 and D_N^2 be (possibly random) $N \times N$ symmetric matrices such that there exists a deterministic C satisfying*

$$\sup_{N\geq 1, i=1,2} \|D_N^i\| \leq C < \infty\,,$$

where $\|\cdot\|$ denotes the usual matrix norm (which is same as the largest singular value for a symmetric matrix). Furthermore, assume that there is a W^*-probability space $(\mathcal{A},\tau)$ in which there are self-adjoint elements d_1 and d_2 such that, for any polynomial p in two variables, it

$$\lim_{N\to\infty} \frac{1}{N}\operatorname{Tr}\left(p\left(D_N^1, D_N^2\right)\right) = \tau\left(p(d_1,d_2)\right) \ \ a.s.$$

Finally, suppose that (D_N^1, D_N^2) is independent of W_N. Then there exists a self-adjoint element s in $\mathcal{A}$ (possibly after expansion) that has the standard semicircle distribution and is freely independent of (d_1, d_2), and is such that

$$\lim_{N\to\infty} \frac{1}{N}\operatorname{Tr}\left(p\left(W_N, D_N^1, D_N^2\right)\right) = \tau\left(p(s, d_1,d_2)\right) \ \ a.s.$$

for any polynomial p in three variables.

References

G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices.* Cambridge University Press, 2010.

Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer Series in Statistics, New York, second edition, 2010.

M. Bauer and O. Golinelli. Random incidence matrices: moments of the spectral density. *Journal of Statistical Physics*, 103(1-2):301–337, 2001.

F. Benaych-Georges, C. Bordenave, and A. Knowles. Largest eigenvalues of sparse inhomogeneous Erdos-Renyi graphs. *arXiv:1704.02953*, 2017.

H. Bercovici and D. Voiculescu. Free convolution of measures with unbounded support. *Indiana University Mathematics Journal*, 42:733–773, 1993.

S. Bhamidi, S. N. Evans, and A. Sen. Spectra of large random trees. *Journal of Theoretical Probability*, 25(3):613–654, 2012.

P. Biane. On the free convolution with a semi-circular distribution. *Indiana University Mathematics Journal*, 46(3):705–718, 1997.

C. Bordenave and M. Lelarge. Resolvent of large random graphs. *Random Structures & Algorithms*, 37(3):332–352, 2010.

W. Bryc, A. Dembo, and T. Jiang. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34(1):138, 2006.

A. Chakrabarty. The Hadamard product and the free convolutions. *Statistics and Probability Letters*, 127:150–157, 2017.

A. Chakrabarty and R. S. Hazra. Remarks on absolute continuity in the context of free probability and random matrices. *Proceedings of the American Mathematical Society*, 144:1335–1441, 2016.

A. Chakrabarty, R. S. Hazra, and D. Sarkar. From random matrices to long range dependence. *Random Matrices: Theory and Applications*, 5:1650008, 2016.

S. Chatterjee. A simple invariance theorem. Available at http://arxiv.org/pdf/math/0508213v1.pdf, 2005.

F. Chung and L. Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences of the United States of America*, 99 (25):15879–15882, 2002.

A. Dembo and E. Lubetzky. Empirical spectral distributions of sparse random graphs. *arXiv:1610.05186*, 2016.

X. Ding, T. Jiang, et al. Spectral distributions of adjacency and laplacian matrices of random graphs. *Annals of Applied Probability*, 20(6):2086–2117, 2010.

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

I. J. Farkas, I. Derényi, A.-L. Barabási, and T. Vicsek. Spectra of ?real-world? graphs: Beyond the semicircle law. *Physical Review E*, 64(2):026704, 2001.

E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630, 1957.

T. Jiang. Empirical distributions of laplacian matrices of large dilute random graphs. *Random Matrices: Theory and Applications*, 1(03):1250004, 2012a.

T. Jiang. Low eigenvalues of laplacian matrices of large random graphs. *Probability Theory and Related Fields*, 153(3-4):671–690, 2012b.

O. Khorunzhy, M. Shcherbina, and V. Vengerovsky. Eigenvalue distribution of large weighted random graphs. *Journal of Mathematical Physics*, 45(4):1648–1672, 2004.

J. O. Lee and K. Schnelli. Local law and tracy-widom limit for sparse random matrices. *arXiv:1605.08767*, 2016.

J. A. Mingo and R. Speicher. *Free Probability and Random Matrices*. Springer, 2017.

A. Nica and R. Speicher. *Lectures on the Combinatorics of Free Probability*. Cambridge University Press, New York, 2006.

T. Squartini, J. de Mol, F. den Hollander, and G. Garlaschelli. Breaking of ensemble equivalence in networks. *Physical Review Letters*, 115:268701, 2015.

T. Tao. *Topics in random matrix theory*, volume 132. American Mathematical Soc., 2012.

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

R. van der Hofstad. *Random Graphs and Complex Networks*, volume 1. Cambridge University Press, 2017.

Y. Zhu. Graphon approach to limiting spectral distributions of Wigner-type matrices. *arXiv.1806.11246*, 2018.