DENSE RANDOM GRAPHS: SPECTRA OF GRAPHONS

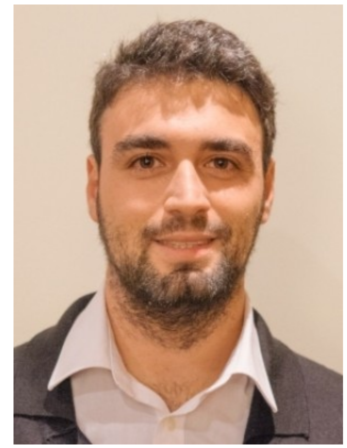
Frank den Hollander
Leiden University, The Netherlands

Spring School in Probability: Complex Networks,
Technische Universität Darmstadt, Germany, 2-6 March 2020.

Spectra of random matrices have been analysed for almost a century. In recent years, many interesting results have been derived for spectra of random matrices associated with networks.

QUESTION ADDRESSED IN THIS TALK:

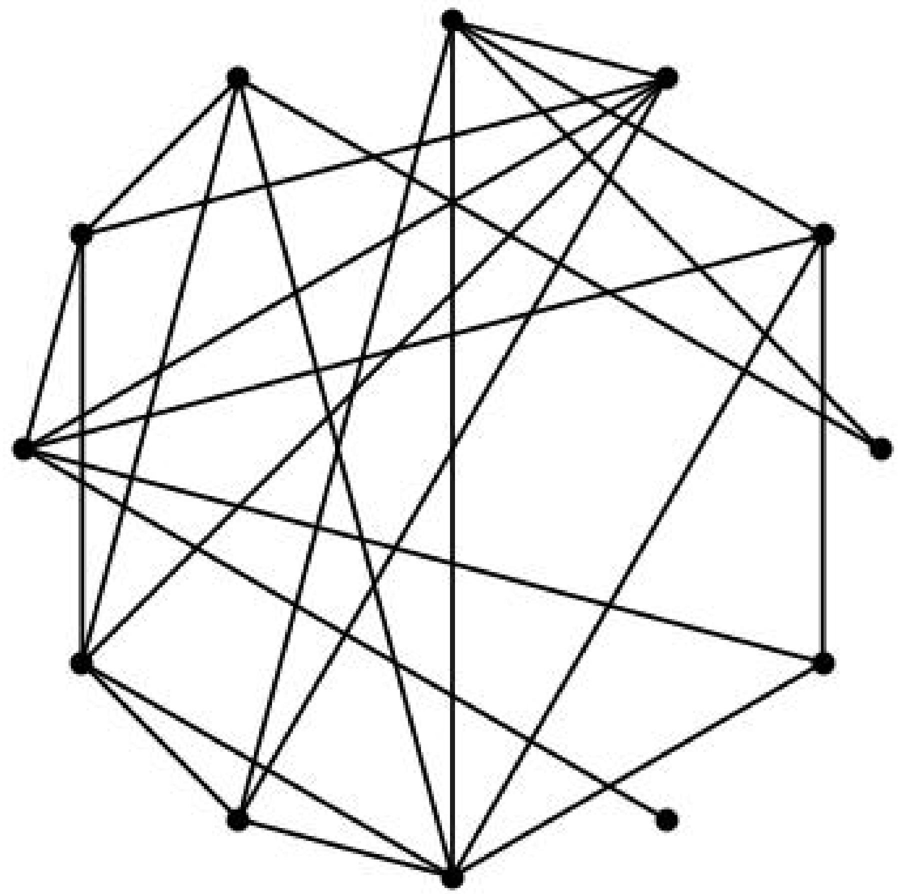
What can be said about the spectrum of the adjacency matrix of a large inhomogeneous Erdős-Rényi random graph?



Arijit Chakrabarty   Rajat Hazra   Matteo Sfragara

Random Matrices: Theory and Applications 2020

+ work in progress



Erdős-Rényi random graph: $n = 12$, $p = 0.3$

§ SETTING

1. Let $(\varepsilon_n)_{n\in\mathbb{N}}$ be a sequence of positive numbers such that

$$\lim_{n\to\infty} \varepsilon_n = 0, \qquad \lim_{n\to\infty} n\varepsilon_n = \infty.$$

Let $r\colon [0,1]\times[0,1] \to [0,\infty)$ be a continuous function such that $r(x,y) = r(y,x)$ for all $x,y \in [0,1]$.

2. Fix $n \in \mathbb{N}$, and consider the inhomogeneous Erdös-Rényi random graph ER_n on n vertices where an edge is placed between the pair of vertices $\{i,j\}$ with probability

$$\varepsilon_n\, r\left(\frac{i}{n},\frac{j}{n}\right), \qquad 1 \leq i,j \leq n,$$

independently for different edges. Write $\mathbb{P}$ for the law of ER_n.

3. Let A_n be the adjacency matrix of ER_n. Write

$$\lambda_i(A_n), \qquad 1 \le i \le n,$$

for the real eigenvalues of A_n. Define the empirical spectral distribution of A_n as

$$\mathrm{ESD}(A_n) = \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i(A_n)},$$

which is a random probability distribution on $\mathbb{R}$.

§ SCALING

THEOREM 1:

There exists a compactly supported symmetric probability measure μ *on* $\mathbb{R}$ *such that, weakly in* $\mathbb{P}$*-probability,*

$$\lim_{n\to\infty} \mathrm{ESD}\left(A_n \Big/ \sqrt{n\varepsilon_n}\right) = \mu.$$

Furthermore, if

$$\min_{x,y\in[0,1]} r(x,y) > 0,$$

then μ *is absolutely continuous with respect to Lebesgue measure. The density of* μ *can be characterised implicitly via an integral equation for its Stieltjes transform.*

Zhu 2018

It is possible to identify μ when r is of rank 1, i.e.,

$$r(x,y) = \nu(x)\nu(y), \qquad x,y \in [0,1],$$

for some continuous function $\nu\colon [0,1] \to [0,\infty)$.

THEOREM 2:

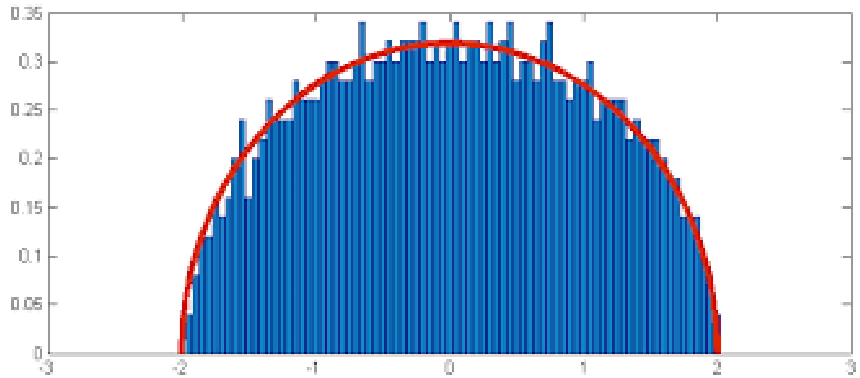
If r is of product form, then

$$\mu = \mu_\nu \boxtimes \mu_*,$$

where

$$\mu_\nu = \text{LAW}[\nu(U)], \quad U = \text{UNIF}[0,1],$$
$$\mu_* = \text{standard Wigner semicircle law},$$

and $\boxtimes$ denotes free multiplicative convolution.



Wigner 1955

In free probability, the Wigner semicircle law takes over the role of the normal law in classical probability. The so-called free cumulants replace the classical cumulants, in the sense that partitions are replaced by non-crossing partitions.

Just as the cumulants of degree ≥ 2 are all zero if and only if the distribution is normal, the free cumulants of degree ≥ 2 are all zero if and only if the distribution is the Wigner semicircle law.

THEOREM 3:

Theorems 1–2 can be generalized to the situation where the function r is random, depends on n and converges to a deterministic limit as $n \to \infty$.

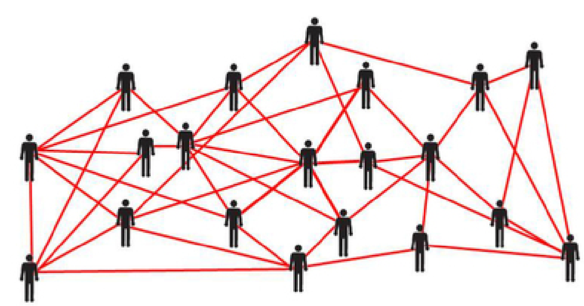
Key ingredients of the proof are:

centering, Gaussianisation, perturbation, decoupling, combinatorics from free probability, ...

§ APPLICATION 1 statistics for social networks

Consider a community of n individuals, represented by the vertices in ER_n. Data is available about which individuals are acquainted. Based on this data, the sociability pattern of the community has to be inferred statistically.

Let ρ denote a probability measure on $[0, \infty)$ with bounded support. 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 individual i.



Pick n so large that

$$0 \leq \varepsilon_n R_i R_j \leq 1 \qquad \forall\, 1 \leq i, j \leq n.$$

Suppose that i, j are acquainted with probability $\varepsilon_n R_i R_j$, which is represented by an edge in ER_n between vertices i, j. The data that is available is the adjacency matrix A_n.

The statistical inference problem is to estimate ρ from A_n. To standardise ρ, we assume that w.o.l.g.

$$\int_0^\infty x\rho(dx) = 1.$$

Since, weakly $\mathbb{P}$-a.s.,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{R_i} = \rho,$$

Theorem 3 gives that, weakly in $\mathbb{P}$-probability,

$$\lim_{n\to\infty} \mathrm{ESD}\left(A_n \Big/ \sqrt{n\varepsilon_n}\right) = \rho \boxtimes \mu_*.$$

In practice, ε_n is unknown, which can be worked around by arguing that, weakly in $\mathbb{P}$-probability,

$$\lim_{n\to\infty} \mathrm{ESD}\left(\sqrt{\frac{n}{\mathrm{Tr}(A_n^2)}}\, A_n\right) = \rho \boxtimes \mu_*,$$

because, in $\mathbb{P}$-probability,

$$\mathrm{Tr}(A_n^2) = \sum_{1\le i,j\le n} A_n(i,j)A_n(j,i) = \sum_{1\le i,j\le n} A_n(i,j)^2 \sim n^2\varepsilon_n.$$

The procedure is that $\rho \boxtimes \mu_*$ can be statistically estimated from A_n. Subsequently, ρ can be estimated because the moments of $\rho \boxtimes \mu_*$ are functions of the moments of ρ and μ_*.

Indeed, since the moments of μ_* are known, the moments of ρ can be recursively computed from the moments of $\rho \boxtimes \mu_*$. Since ρ is compactly supported, it can in turn be computed via its moments.

§ APPLICATION 2 configuration model

Let $\mathcal{G}_n$ be the set of simple graphs on n vertices. We fix the degrees of all the vertices, namely, vertex i has degree d_i^*, where

$$\vec{d}_n^* = \{d_i^*\}_{1\le i\le n}$$

is a sequence of positive integers of which we only require that it is graphical, i.e., there is at least one simple graph matching these degrees.

hard configuration model

The soft ensemble P_n is the unique probability distribution on $\mathcal{G}_n$ with the following two properties:

(I) The average degree of vertex i, defined by

$$\sum_{G \in \mathcal{G}_n} d_i(G) P_n(G),$$

equals d_i^* for all $i \leq i \leq n$.

(II) The entropy of P_n, defined by

$$-\sum_{G \in \mathcal{G}_n} P_n(G) \log P_n(G),$$

is maximal.

P_n models a random graph of which no prior information is available other than the average degrees.

soft configuration model

Property (II) forces P_n to take the form Jaynes 1957

$$P_n(G) = \frac{1}{Z_n(\vec{\theta}^*)} \exp\left[-\sum_{i=1}^{n} \theta_i^* d_i(G)\right], \qquad G \in \mathcal{G}_n,$$

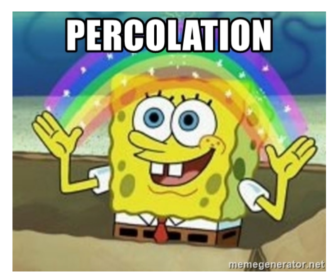
where $\vec{\theta}_n^* = \{\theta_i^*\}_{1 \le i \le n}$ is the unique sequence of Lagrange multipliers such that property (I) is satisfied.

Reparametrisation yields

$$P_n(G) = \prod_{1 \le i < j \le n} (p_{ij}^*)^{A_n[G](i,j)} \, (1 - p_{ij}^*)^{1 - A_n[G](i,j)}, \; G \in \mathcal{G}_n,$$

where $A_n[G]$ is the adjacency matrix of G, and

$$p_{ij}^* = \frac{x_i^* x_j^*}{1 + x_i^* x_j^*}, \quad x_i^* = e^{-\theta_i^*}, \qquad 1 \le i \ne j \le n.$$



Property (I) requires that

$$d_i^* = \sum_{\substack{1 \le j \le n \\ j \ne i}} p_{ij}^*, \qquad 1 \le i \le n,$$

which constitutes a set of n equations for n unknowns.

Abbreviate

$$m_n = \max_{1 \le i \le n} d_i^*.$$

We focus on the regime

$$\lim_{n \to \infty} m_n = \infty, \quad \lim_{n \to \infty} m_n/\sqrt{n} = 0.$$

Squartini, de Mol, den Hollander, Garlaschelli 2015

It turns out that in this regime

$$p^*_{ij} = [1 + o(1)]\,\frac{d^*_i d^*_j}{\sigma_n}, \qquad n \to \infty,$$

with

$$\sigma_n = \sum_{1 \le i \le n} d^*_i.$$

Pick

$$\varepsilon_n = m_n^2/\sigma_n.$$

Then

$$\lim_{n\to\infty} \varepsilon_n = 0, \qquad \lim_{n\to\infty} n\varepsilon_n = \infty,$$

and

$$p^*_{ij} = [1 + o(1)]\,\varepsilon_n\,(d^*_i/m_n)(d^*_j/m_n).$$

Under the assumption that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{d_i^*/m_n} = \rho$$

for some probability measure ρ, Theorem 3 gives that, weakly in $\mathbb{P}$-probability,

$$\lim_{n\to\infty} \mathrm{ESD}\Big(A_n \Big/ \sqrt{n\varepsilon_n}\Big) = \rho \boxtimes \mu_*.$$

This identifies the scaling of the ESD for the network that is modeled by the soft configuration model as a function of the imposed average degrees.

§ LARGEST EIGENVALUE

Suppose that $\epsilon_n \equiv 1$ and that the reference graphon r satisfies

$$r \in (0,1) \text{ is continuous a.e.,}$$
$$\log r,\ \log(1-r) \in L^1([0,1]^2).$$

Then we can use the LDP for the inhomogeneous Erdős-Rényi random graph.

Let $\lambda_1 = \lambda_1(\widetilde{h}^G)$ denote the maximal eigenvalue of the adjacency matrix $A_n[G]$ of G.

LDP

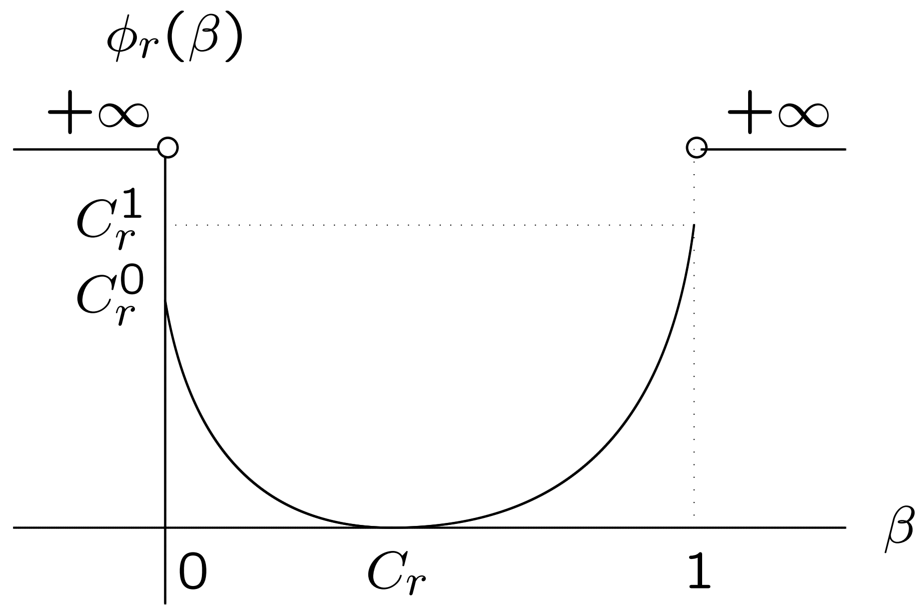
THEOREM 4: Chakrabarty, Hazra, dH & Sfragara in progress

The family $(\mathbb{P}_{n,r}(\lambda_1/n \in \cdot\,))_{n\in\mathbb{N}}$ *satisfies the LDP on* $\mathbb{R}$ *with rate* $\binom{n}{2}$ *and with rate function* ϕ_r *given by*

$$\phi_r(\beta) = \inf_{\substack{h\in\mathcal{W} \\ \|h\|=\beta}} I_r(h)$$

with $\|h\|$ *the operator norm of* h.

The claim is an immediate consequence of the LDP and the contraction principle, because $\tilde{h}^G \mapsto \lambda_1(\tilde{h}^G)/n$ is bounded and continuous with respect to the $\delta_\square$-metric.



$$C_r = \|r\|$$

$$C_r^0 = \int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y\,\log\frac{1}{1-r(x,y)}$$

$$C_r^1 = \int_{[0,1]^2} \mathrm{d}x\,\mathrm{d}y\,\log\frac{1}{r(x,y)}$$

HEURISTICS:

When $\beta = C_r$, the graphon h that minimizes $I_r(h)$ such that $\|h\| = C_r$ is the reference graphon $h = r$. When $\beta > C_r$, we need a graphon h with a larger operator norm. The large deviation cannot go above 1, which is represented by the constant graphon $h \equiv 1$, for which $I_r(1) = C_r^1$. A similar observation holds when $\beta < C_r$.

THEOREM 5: Chakrabarty, Hazra, dH & Sfragara in progress

(i) ϕ_r *is continuous on* $[0,1]$.

(ii) ϕ_r *is strictly unimodal on* $[0,1]$.

If the reference graphon r is of rank 1, i.e.,

$$r(x,y) = \nu(x)\,\nu(y) \qquad \text{for some } \nu\colon [0,1] \to [0,1],$$

then more can be said. Define

$$m_k = \int_{[0,1]} \mathrm{d}x\,\nu(x)^k, \qquad k \in \mathbb{N}.$$

The following theorem relies on a closer analysis of the variational formula for ϕ_r.

THEOREM 6: Chakrabarty, Hazra, dH & Sfragara in progress

(i) For $\beta \to C_r$,

$$\phi_r(\beta) \sim K_r\,(\beta - C_r)^2$$

with

$$C_r = m_1^2, \qquad K_r = \frac{1}{2}\frac{m_1^4}{m_3^2 - m_4^2}.$$

(ii) For $\beta \uparrow 1$,

$$C_r^1 - \phi_r(\beta) \sim (1-\beta)\log\frac{1}{1-\beta}.$$

(iiii) For $\beta \downarrow 0$,

$$C_r^0 - \phi_r(\beta) \sim \beta\log\frac{1}{\beta}.$$

§ EXPANSION

A key tool in the proof of the last theorems is a series expansion for the operator norm of a graphon h around any graphon of rank 1.

THEOREM 7: Chakrabarty, Hazra, dH & Sfragara in progress

Let $\bar{h} \in \mathcal{W}$ *be of the form* $\bar{h}(x,y) = \bar{\nu}(x)\bar{\nu}(y)$, $x, y \in [0,1]$. *For any* $h \in W$ *such that* $\|h - \bar{h}\| < \|h\|$, *the operator norm* $\|h\|$ *is the unique solution of the equation*

$$\|h\| = \sum_{n \in \mathbb{N}_0} \frac{1}{\|h\|^n} \mathcal{F}_n(h, \bar{h}),$$

where

$$\mathcal{F}_n(h, \bar{h}) = \int_{[0,1]^2} \mathrm{d}x \, \mathrm{d}y \, \bar{\nu}(x)(h - \bar{h})^{\circledast n}(x,y)\bar{\nu}(y)$$

with $\circledast$ *denoting convolution.*

PROOF

We have

$$hu = \lambda u,$$

where λ is the norm and the maximal eigenvalue of h, and u is an eigenfunction of h. Put $g = h - \bar{h}$, and write

$$(g + \bar{h})u = \lambda u \quad \longrightarrow \quad (\lambda - g)u = \bar{h}u$$

$$\longrightarrow \quad u = (\lambda - g)^{-1}\bar{h}u \quad \longrightarrow \quad u = (\lambda - g)^{-1}\bar{\nu}\langle\bar{\nu}, u\rangle.$$

Hence

$$\langle\bar{\nu}, u\rangle = \langle\bar{\nu}, u\rangle\langle\bar{\nu}, (\lambda - g)^{-1}\bar{\nu}\rangle \quad \longrightarrow \quad 1 = \langle\bar{\nu}, (\lambda - g)^{-1}\bar{\nu}\rangle$$

$$\longrightarrow \quad \lambda = \langle\bar{\nu}, (1 - g/\lambda)^{-1}\bar{\nu}\rangle,$$

where $\lambda - g$ is invertible because $\|g\| = \|h - \bar{h}\| < \|h\| = \lambda$, and we take $u > 0$ so that $\langle \bar{\nu}, u \rangle \neq 0$. We can expand

$$\lambda = \left\langle \bar{\nu}, \sum_{n \in \mathbb{N}_0} (g/\lambda)^n \bar{\nu} \right\rangle$$

$$= \sum_{n \in \mathbb{N}_0} \frac{1}{\lambda^n} \int_{[0,1]^{n+1}} \mathrm{d}x_0 \, \mathrm{d}x_1 \cdots \mathrm{d}x_n$$

$$\bar{\nu}(x_0) g(x_0, x_1) \times \cdots \times g(x_{n-1}, x_n) \bar{\nu}(x_n)$$

$$= \sum_{n \in \mathbb{N}_0} \frac{1}{\lambda^n} \mathcal{F}_n(h, \bar{h}).$$

Since $\|h\| = \lambda$, we get the claim. □


AAAHHHH!