Basics of Spectral Graph Theory

Konstantin Avrachenkov (Inria, France)

IEEE SPS School at Aalto University, 29 March 2022

Plan

1. Motivations and definitions
   (Consensus, Network sampling, Epidemics, SSL, SC)
2. Main approaches in random matrix theory
   (Method of moments, Stieltjes transform)
3. Spectra of random graphs
   (ER, SBM, RGG, Soft GBM)

Motivations and definitions (Consensus)

Consider n agents that have m communication links.

Each agent i has a local state x_i and the agents would like to synchronize their states.

Namely, the goal is to construct an algorithm such that

$$x_i(t) \to \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i(0), \quad \text{as} \quad t \to \infty.$$

Motivations and definitions (Consensus)

The communication structure can be represented by a graph $G = (V, E)$ with $n = |V|$ and $m = |E|$.

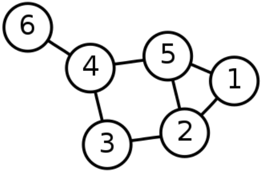

Figure: Graph Example with $n = 6$ and $m = 7$ (from Wikipedia).

A graph can be described by the adjacency matrix $A = (a_{ij})$:

$$a_{ij} = \begin{cases} 1, & \text{if there is a link between } i \text{ and } j, \\ 0, & \text{otherwise.} \end{cases}$$

In the above example $a_{12} = 1$ but $a_{13} = 0$.

Motivations and definitions (Consensus)

One of the simplest consensus algorithms in continuous time is described by

$$\dot{x}_i = \sum_{j=1, j\neq i}^{n} a_{ij}(x_j - x_i).$$

Or in the matrix form

$$\dot{x} = -Lx,$$

where $L = D - A$ is the (combinatorial) Laplacian and $D = diag(d_i) = diag(\sum_{j=1}^{n} a_{ij})$ is the diagonal matrix of nodes' degrees.

The rate of convergence is dominated by the Laplacian spectral gap:

$$\gamma(L) = \min_k \{\lambda_k(L) : \lambda_k(L) > 0\},$$

where λ_k is the kth eigenvalue of L. Note that $L\mathbb{1} = 0$.

Motivations and definitions (Consensus)

Let us state properties of the adjacency matrix A and Laplacian L:

If A is symmetric (here for most of the time we consider undirected graphs), the eigenvalues are real and the spectral theorem applies:

$$A = \sum_{i=1}^{n} \lambda_i(A) v_i v_i^T,$$

where v_i is an eigenvector associated with the ith eigenvalue.

$$L = D - A = diag(A\underline{1}) - A$$

Since $x^T L x = \sum_{i \sim j} (x_i - x_j)^2 \geq 0$, L is positive semidefinite and $\lambda_i(L) \geq 0$. Note that $L\underline{1} = 0$ and hence $\lambda_1(L) = 0$.

Motivations and definitions (Epidemics)

Consider the following model of epidemics on a network:

If $X_i(t) = 0$, the node is healthy at time t and $X_i(t) = 1$, otherwise.

A node recovers with rate 1 and contaminates a neighbour with rate β. Namely,

$$\begin{aligned} X_i &: \quad 1 \to 0, \text{with rate } \; 1; \\ X_i &: \quad 0 \to 1, \text{with rate } \; \beta \sum_{j=1}^{n} a_{ij} X_j. \end{aligned}$$

Motivations and definitions (Epidemics)

If the following condition on the spectral radius holds

$$\rho(A) = \max_k |\lambda_k(A)| < \frac{1}{\beta},$$

the epidemics dies out fast, i.e.

$$P[X(t) \neq 0] \leq \sqrt{n||X(0)||_1} e^{(\beta\rho(A)-1)t},$$

and

$$E[\text{time to extinction}] \leq \frac{\log(n)+1}{1-\beta\rho(A)}.$$

However, if this condition on the Laplacian spectral gap holds

$$\frac{\gamma(L)}{2} > \frac{1}{\beta},$$

the epidemics survives for an exponentially long time.

Motivations and definitions (Network sampling)

Analysing (online) social networks one would like to know:

- How young is given social network?
- How many friends has an average network member?
- What proportion of population supports some political party?
- etc

Motivations and definitions (Network sampling)

All such questions are related to the problem of estimating an average of a function $f(\cdot)$ defined on the network nodes.

Let $G = (V, E)$ be an undirected graph representing a social network.

Then, we are interested to estimate

$$\mu(G) = \frac{1}{n} \sum_{v \in V} f(v). \tag{1}$$

Motivations and definitions (Network sampling)

Uniform node sampling is typically very costly and biased.

One work-around can be achieved by using the random walk based sampling.

Let $\{X_t,\ t = 0, \ldots, T\}$ are nodes sampled by T steps of the random walk.

Motivations and definitions (Network sampling)

Since the standard random walk visits more often large degree nodes, the following estimator

$$\hat{\mu}^{(T)}(G) = \frac{1}{T}\sum_{t=1}^{T} f(X_t)$$

is biased.

One way around to remove the bias is to use Metropolis-Hastings chain with the following transition matrix

$$P_{ij}^{MH} = \begin{cases} \frac{1}{\max(d_i,d_j)} & \text{if } j \neq i \\ 1 - \sum_{k\neq i} \frac{1}{\max(d_i,d_k)} & \text{if } j = i. \end{cases}$$

Motivations and definitions (Network sampling)

By using the CLT for MCs, one can show the following central limit theorem for MH Chain.

Proposition

(Central Limit Theorem for MH) For MH Markov chain, it holds that

$$\sqrt{T}\left(\hat{\mu}_{MH}^{(T)}(G) - \mu(G)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_{MH}^2), \quad \text{as} \quad T \to \infty,$$

where $\sigma_{MH}^2 = \frac{2}{n} f^T Z f - \frac{1}{n} f^T f - \left(\frac{1}{n} f^T \underline{1}\right)^2$ *and where* $Z = [I - P^{MH} + \frac{1}{n}\underline{11}^T]^{-1}$ *is the fundamental matrix.*

There are nice expressions and bounds for σ_{MH} in terms of the eigenvalues of P^{MH}.

Motivations and definitions (Network sampling)

An even more efficient estimator is the Respondent Driven Sampling estimator (RDS-estimator):

$$\hat{\mu}_{RDS}^{(T)}(G) = \frac{\sum_{t=1}^{T} f(X_t)/d(X_t)}{\sum_{t=1}^{T} 1/d(X_t)}.$$

Here the bias towards large degree nodes is corrected by the estimator and we can sample nodes with the standard random walk.

Motivations and definitions (Network sampling)

Note that the transition probability matrix of the standard random walk can be expressed as

$$P = D^{-1}A,$$

where $D = diag(d_i)$ is the diagonal matrix of nodes' degrees.

It is often convenient to work with the symmetrized version

$$\tilde{A} = D^{-1/2}AD^{-1/2},$$

which has the same spectrum,

and with the normalized Laplacian:

$$\mathcal{L} = I - \tilde{A} = I - D^{-1/2}AD^{-1/2},$$

with $\lambda_i(\mathcal{L}) \in [0, 2]$ and $\lambda_1(\mathcal{L}) = 0$.

Motivations and definitions (Semi-supervised learning)

Finally, let us consider graph-based semi-supervised learning.

Now $G = (V, E)$ is the similarity graph on data points, e.g., G can be kNN graph or RBF based graph with

$$w_{ij} = \exp(-||X_i - X_j||^2/\gamma),$$

where X_i is the normalized vector of attributes of the ith data point. Then,

$$a_{ij} = 1\{w_{ij} \geq \theta\},$$

where θ regulates the sparsity of the graph.

Motivations and definitions (Semi-supervised learning)

Typically, obtaining labelled data is expensive and time-consuming.

The main idea of the graph-based semi-supervised learning is to propagate information from few available labelled points to unlabelled data.

Suppose we would like to classify n data points into K classes.

Define an $n \times K$ matrix Y as

$$Y_{ik} = \begin{cases} 1, & \text{if } i \in V_k \text{, i.e., point } i \text{ is labelled as a class } k \text{ point,} \\ 0, & \text{otherwise.} \end{cases}$$

We refer to each column Y_{*k} of matrix Y as a labeling function.

Motivations and definitions (SSL)

Then, one fairly general class of SSL methods can be expressed as an optimization problem

$$\min_F \{ \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \| {d_i}^{\sigma-1} F_{i*} - {d_j}^{\sigma-1} F_{j*} \|^2 + \mu \sum_{i=1}^{N} {d_i}^{2\sigma-1} \| F_{i*} - Y_{i*} \|^2 \}$$

$$F_{ik} > F_{ik'}, \quad \forall k' \neq k \Rightarrow \quad \text{Data point} \quad i \quad \text{is classified into class} \quad k.$$

Now we have two parameters μ and σ.

Motivations and definitions (SSL)

The solution can in fact be given in the explicit form:

$$F_{*k} = \frac{\mu}{2+\mu}\left(I - \frac{2}{2+\mu}D^{-\sigma}AD^{\sigma-1}\right)^{-1} Y_{*k},$$

for $k = 1, ..., K$.

A simple and efficient way to compute F_{*k} is by power iterations:

$$F_{*k}^{(t+1)} = \frac{2}{2+\mu}D^{-\sigma}AD^{\sigma-1}F_{*k}^{(t)} + \frac{\mu}{2+\mu}Y_{*k}, \quad t = 1, 2, ...$$

The spectral gap of $D^{-\sigma}AD^{\sigma-1}$ dictates the rate of convergence.

Motivations and definitions (SSL)

In particular cases, we have

- if $\sigma = 1$, the Standard Laplacian method:
  $F_{*k} = \frac{\mu}{2+\mu}(I - \frac{2}{2+\mu}D^{-1}A)^{-1}Y_{*k}$,
- if $\sigma = 1/2$, the Normalized Laplacian method:
  $F_{*k} = \frac{\mu}{2+\mu}(I - \frac{2}{2+\mu}D^{\frac{-1}{2}}AD^{\frac{-1}{2}})^{-1}Y_{*k}$,
- if $\sigma = 0$, PageRank based method:
  $F_{*k} = \frac{\mu}{2+\mu}(I - \frac{2}{2+\mu}AD^{-1})^{-1}Y_{*k}$.

Motivations and definitions (Graph clustering)

Graph clustering (or unsupervised learning) is a very established research area with many applications.

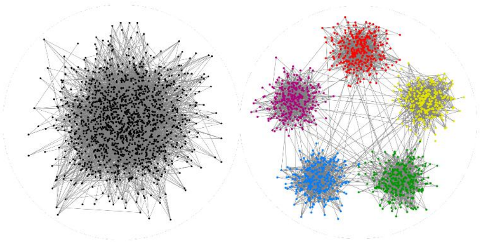

Figure: From [Abbe 2017]

Motivations and definitions (Graph clustering)

Consider the vector $x = (x_i) \in \{-1, 1\}^n$ corresponding to the partition $V = V_1 \sqcup V_2$:

$$x_i = \begin{cases} 1, & \text{if } i \in V_1 \\ -1, & \text{if } i \in V_2 \end{cases}.$$

Take the adjacency matrix $A = (A_{ij})$, the diagonal matrix D, where $D_{ii} = \sum_j A_{ij}$, and the Laplacian $L = D - A$. Then,

$$Cut(V_1, V_2) = \sum_{i \in V_1, j \in V_2} A_{ij} = \frac{1}{4} \sum_{i,j \in [n]} A_{ij}(x_i - x_j)^2 \ \propto \ x^T L x.$$

Motivations and definitions (Graph clustering)

Continuous relaxation:

$$\arg \min_{|V_1|=|V_2|=n/2} Cut(V_1, V_2) = \arg \min_{\substack{x \in \{-1,1\}^n \\ x \perp 1_n}} x^T L x$$

$$\longrightarrow \arg \min_{\substack{x \in \mathcal{R}^n \\ ||x||_2^2 = \sqrt{n} \\ x \perp 1_n}} x^T L x$$

Spectral clustering based on eigenvectors of Laplacian matrix:

- First eigenvector of L is $v^{(1)} = (1, \ldots, 1)^T$ with $\lambda_1 = 0$;
- Second eigenvector or Fiedler vector $v^{(2)}$ provides the solution to the relaxed minimum cut problem;
- Cluster node i according to the sign of $v_i^{(2)}$.

The difficulty depends on how far $\lambda_2(L)$ is from the rest of eigenvalues.

Main approaches in random matrix theory

We demonstrate the main approaches in random matrix theory on the classical Wiegner matrices.

Definition

A Wiegner matrix M_n is a symmetric real valued matrix with upper-triangular independent zero mean and unit variance entries (of course, $M_{n,ij} = M_{n,ji}$).

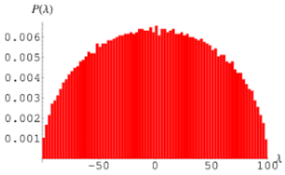

Figure: Histogram of eignevalues, $n = 5000$ (from Wolfram MathWorld).

Main approaches in random matrix theory

Define the Emprirical Spectral Distribution (ESD)

$$\mu_n = \mu(X_n) = \frac{1}{n}\sum_{i=1}^{n}\delta(x - \lambda_i(X_n)).$$

Theorem (Semicircular law)

Let M_n be the Wiegner ensemble. Then the ESDs of $X_n = \frac{1}{\sqrt{n}}M_n$ converge weakly, almost surely (and hence, also in probability and in expectation) to the Wigner semi-circular distribution

$$\mu_{sc}(x)dx = \frac{1}{2\pi}\sqrt{4 - x^2}1\{|x| \leq 2\}dx. \qquad (2)$$

Main approaches in RMT (Method of moments)

Theorem (Carleman's condition)

Let μ be a distribution and denote $m_1, m_2, \dots$ its sequence of moments which are assumed to be all finite. If the condition

$$\sum_{k=1}^{\infty} m_{2k}^{-\frac{1}{2k}} = +\infty,$$

is fulfilled, then μ is uniquely determined by the sequence $m_1, m_2, \dots$

Note that a slightly easier condition is $|m_k| \leq CD^k k!$.

Main approaches in RMT (Method of moments)

Thus, we can prove the convergence by analyzing the moments

$$m_k(\mu_n) = \int_{\mathbb{R}} x^k d\mu_n = \frac{1}{n}\sum_{i=1}^{n} \lambda_i^k(X_n) = \frac{1}{n}\mathrm{tr}X_n^k$$

or due to good concentration of measure, even the expectation

$$\bar{m}_k(\mu_n) = \frac{1}{n}E[\mathrm{tr}X_n^k] = \frac{1}{n}\sum_{i_1,\ldots,i_k=1}^{n} E[x_{i_1 i_2}\cdots x_{i_{k-1} i_k} x_{i_k i_1}]. \qquad (3)$$

Each term in (3), $\mathbf{i} = i_1 i_2 \cdots i_k i_1$ corresponds to a closed path consisting of k edges.

Main approaches in RMT (Method of moments)

Since the entries of X_n have mean zero and are independent (up to the symmetry), the summand $E[x_{i_1 i_2} \cdots x_{i_{k-1} i_k} x_{i_k i_1}]$ will be zero unless every edge in the path is traversed an even number of times.

Thus, we already see that the odd moments should be zero.

Furthermore, there are at most $k/2$ unique edges and at most $k/2+1$ distinct vertices.

Let the weight t of a sequence $\mathbf{i}$ be the number of distinct indices $i_1, \ldots, i_k$. By the above observation, a nonzero term in (3) have a weight $t \leq k/2+1$.

Main approaches in RMT (Method of moments)

Let us show that the terms with $t < k/2 + 1$ are negligible as $n \to \infty$.

Given $\mathbf{i}$ of weight t, there are $n(n-1)\cdots(n-t+1) \leq n^t$ sequences equivalent to it.

The contribution of each term in this equivalence class is

$$\frac{1}{n} E[x_{i_1 i_2} \cdots x_{i_{k-1} i_k} x_{i_k i_1}] = O\left(\frac{1}{n} \frac{1}{\sqrt{n}^k}\right)$$

Thus, the total contribution to (3) is at most

$$O\left(\frac{n^t}{n^{k/2+1}}\right) \to 0, \quad n \to \infty.$$

Then, the terms with $t = k/2 + 1$ with k even correspond to trees, and a sequence $i_1 i_2 \cdots i_k i_1$ represents a closed path on such trees which traverses each edge exactly twice, once in each direction.

Main approaches in RMT (Method of moments)

Counting such trees gives for $k/2$ even

$$m_k(\mu_n) \to \frac{1}{k/2+1}\binom{k}{k/2}, \quad n \to \infty,$$

which are the even moments of the semicircular distribution.

Main approaches in RMT (Stieltjes transform)

The next method is based on Stieltjes transform of ESD:

$$\begin{aligned} s_n(z) &= \int_{\mathbb{R}} \frac{1}{x-z} d\mu_n(x) \\ &= \frac{1}{n}\mathrm{tr}(X_n - zI_n)^{-1} = \frac{1}{n}\mathrm{tr}(\frac{1}{\sqrt{n}}M_n - zI_n)^{-1}, \end{aligned}$$

for $z \in \mathbb{C}\backslash\mathbb{R}$. Properties of Stieltjes transform:

$$s_n(z) = -\frac{1}{z}\left[1 + \frac{1}{zn}\mathrm{tr}\left(\frac{M_n}{\sqrt{n}}\right) + \frac{1}{z^2 n}\mathrm{tr}\left(\frac{M_n}{\sqrt{n}}\right)^2 + \ldots\right]$$

and hence

$$s_n(z) = -\frac{1}{z} - \frac{1}{z^2 n}O(1).$$

Main approaches in RMT (Stieltjes transform)

Imaginary part of $s_n(z)$ is positive for z in the upper half plane.

$s_n(z)$ is analytic at all points in the upper half plane.

For z such that $Im(z) > 0$,

$$|s_n(z)| \leq \frac{1}{Im(z)}. \tag{4}$$

The density function can be recovered as follows:

$$\mu(x) = \lim_{\epsilon \to 0^+} \frac{s(x + i\epsilon) - s(x - i\epsilon)}{2\pi i}.$$

The Stieltjes transform of the semicircular law:

$$s_{sc}(z) = \frac{-z + \sqrt{z^2 - 4}}{2}.$$

Main approaches in RMT (Stieltjes transform)

First we show that we can work only with the expected Stieltjes transform.

Lemma

For fixed z in the upper half plane,

$$|s_n(z) - E[s_n(z)]| \to 0, \quad \textit{almost surely.} \tag{5}$$

Proof outline: Note that

$$\begin{aligned}
\sqrt{n(n-1)}\, s_{n-1}\left(\frac{\sqrt{n}}{\sqrt{n-1}} z\right) &= \sqrt{n(n-1)}\, \frac{1}{n-1} \operatorname{tr}\left(\frac{M_{n-1}}{\sqrt{n-1}} - \frac{\sqrt{n}}{\sqrt{n-1}} zI\right)^{-1} \\
&= \frac{\sqrt{n}}{\sqrt{n-1}} \left(\frac{\sqrt{n}}{\sqrt{n-1}}\right)^{-1} \operatorname{tr}\left(\frac{M_{n-1}}{\sqrt{n}} - zI\right)^{-1} \\
&= \operatorname{tr}\left(\frac{M_{n-1}}{\sqrt{n}} - zI\right)^{-1}.
\end{aligned}$$

Main approaches in RMT (Stieltjes transform)

Next, consider

$$\sqrt{n(n-1)}s_{n-1}\left(\frac{\sqrt{n}}{\sqrt{n-1}}z\right) - ns_n(z)$$

$$= \operatorname{tr}\left(\frac{M_{n-1}}{\sqrt{n}} - zI\right)^{-1} - \operatorname{tr}\left(\frac{M_n}{\sqrt{n}} - zI\right)^{-1} = \sum_{i=1}^{n-1}\frac{1}{\lambda_i(M_{n-1})/\sqrt{n} - z} - \sum_{i=1}^{n}\frac{1}{\lambda_i(M_n)/\sqrt{n} - z}.$$

Then, by using Cauchy's Interlace Theorem, i.e.

$$\lambda_1(M_n) \le \lambda_1(M_{n-1}) \le \lambda_2(M_n) \le \ldots \le \lambda_{n-1}(M_{n-1}) \le \lambda_n(M_n)$$

and the bound (4), we can conclude that

$$\sum_{i=1}^{n-1}\frac{1}{\lambda_i(M_{n-1})/\sqrt{n} - z} - \sum_{i=1}^{n}\frac{1}{\lambda_i(M_n)/\sqrt{n} - z} = O(1).$$

Next, divide the both sides of the above equation by n.

Main approaches in RMT (Stieltjes transform)

$$\sqrt{\frac{n-1}{n}} s_{n-1}\left(\frac{\sqrt{n}}{\sqrt{n-1}} z\right) - s_n(z) = O\left(\frac{1}{n}\right) \tag{6}$$

And hence, by continuity of Stieltjes transform,

$$s_n(z) = s_{n-1}(z) + O\left(\frac{1}{n}\right).$$

Then, applying McDiarmid's inequality, yields

$$P\left[|s_n(z) - E[s_n(z)]| \geq \frac{\kappa}{\sqrt{n}}\right] \leq Ce^{-c\kappa^2},$$

for some absolute constants c and C. Taking $\kappa = \epsilon n^{1/4}$ and applying Borel-Cantelli Lemma, we prove the statement

$$|s_n(z) - E[s_n(z)]| \to 0, \quad \text{a.s.}$$

Main approaches in RMT (Stieltjes transform)

Next, we can concentrate on $E[s_n(z)]$. The Schur complement plays a crucial role. Let

$$A_n = \begin{bmatrix} A_{n-1} & \frac{1}{\sqrt{n}} Y \\ \frac{1}{\sqrt{n}} Y^T & \frac{1}{\sqrt{n}} M_{n,nn} - z \end{bmatrix},$$

where $A_n = \frac{1}{\sqrt{n}} M_n - zI$, $A_n = \frac{1}{\sqrt{n}} M_{n-1} - zI$ and Y is the rightmost column of M_n with the last entry removed. Then, we can write

$$A_{n,nn}^{-1} = \frac{1}{(\frac{1}{\sqrt{n}} M_{n,nn} - z) - \frac{1}{n} Y^T (\frac{1}{\sqrt{n}} M_{n-1} - zI)^{-1} Y}$$

$$= \frac{1}{-z - \frac{1}{n} Y^T (\frac{1}{\sqrt{n}} M_{n-1} - zI)^{-1} Y + o(1)}.$$

Main approaches in RMT (Stieltjes transform)

We note that by symmetry

$$\begin{aligned} E[s_n(z)] &= E\left[\frac{1}{n}\mathrm{tr}\left(\frac{M_n}{\sqrt{n}} - zI\right)^{-1}\right] \\ &= E\left[\left(\frac{M_n}{\sqrt{n}} - zI\right)^{-1}_{nn}\right] \\ &= E[A^{-1}_{n,nn}]. \end{aligned}$$

Thus,

$$E[s_n(z)] = E\left[\frac{1}{-z - \frac{1}{n}Y^T(\frac{1}{\sqrt{n}}M_{n-1} - zI)^{-1}Y + o(1)}\right].$$

Main approaches in RMT (Stieltjes transform)

Let $R_{n-1} = \left(\frac{1}{\sqrt{n}} M_{n-1} - zI\right)^{-1}$. We would like to show that

$$\frac{1}{n} Y^T R_{n-1} Y = E[s_n(z)] + o(1). \tag{7}$$

Let us use double conditioning

$$E[E[Y^T R_{n-1} Y | R]] = E\left[\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} E[y_i r_{ij} y_j | R]\right] = E\left[\sum_{i=1}^{n-1} r_{ii}\right]$$

Then, with the help of (6), we get

$$\frac{1}{n} E[\mathrm{tr} R_{n-1}] = E\left[\sqrt{\frac{n-1}{n}} s_{n-1}\left(\frac{\sqrt{n}}{\sqrt{n-1}} z\right)\right] = E[s_n(z)] + o(1).$$

Main approaches in RMT (Stieltjes transform)

Thus, we obtain

$$E[s_n(z)] = \frac{1}{-z - E[s_n(z)]} + o(1).$$

Since the imaginary part of $s_n(z)$ should be positive, the fixed point solution of the above equation is

$$s(z) = \frac{-z + \sqrt{z^2 - 4}}{2},$$

which coincides with Stieltjes transform of semicircular law.

Spectra of random graphs (Erdős-Rényi model)

Erdős-Rényi random graph model can be described by the adjacency matrix

$$A_{ij} = A_{ji} \sim Ber(p(n)).$$

Note that A is not Wiegner ensemble.
Let us introduce the centered and normalized version:

$$\hat{A} = \gamma(n)A = \bar{A} + \tilde{A},$$

where

$$\tilde{A}_{ij} \sim Cen(p, \gamma),$$

and where

$$Cen(p, \gamma) = \begin{cases} \gamma(1-p), & \text{w.p.} \quad p; \\ -\gamma p, & \text{w.p.} \quad 1-p; \end{cases}$$

with $\gamma(n) = (np(n)(1-p(n)))^{-1/2}$.

Spectra of random graphs (Erdős-Rényi model)

In the case of ER model we need to check Lindeberg's condition:

$$\lim_{n\to\infty} \max_{i=1,\dots n} \sum_{j=1}^{n} \int_{|x|>\theta} x^2 \mathrm{d}P_{\tilde{A}_{ij}}(x) = 0, \quad \forall\theta.$$

For the $Cen(p(n), \gamma(n))$ the above condition results in the requirement

$$p(n) = \omega(n^{-1}), \quad \text{as} \quad n \to \infty,$$

or equivalently, the average degree $np(n)$ should diverge. Then,

$$\mu(\tilde{A}) \xrightarrow{a.s.} \mu_{sc}.$$

Spectra of random graphs (Erdős-Rényi model)

Note that a single eigenvalue has a negligible contribution to ESD when $n \to \infty$.

Therefore, one needs to study the spectral norm of a random matrix separately.

Theorem (Vu)

Let M be a Wigner matrix with independent random elements M_{ij}, $i, j = 1, \ldots n$ having zero mean and variance at most $\sigma^2(n)$. If the entries are bounded by $K(n)$ and there exist a constant C' such that $\sigma(n) \geq C' n^{-1/2} K(n) \log^2(n)$, then there exists a constant C such that with high probability (w.h.p.)

$$\|M\|_2 \leq 2\sigma(n)\sqrt{n} + C(K(n)\sigma(n))^{1/2} n^{1/4} log(n). \quad (8)$$

Spectra of random graphs (Erdős-Rényi model)

The above result applies to ER model with $K = \sqrt{\frac{1-p(n)}{np(n)}}$.

Namely, if the link probability $p(n)$ satisfies an additional condition

$$p(n) \geq C'' \frac{log^4(n)}{n},$$

we have w.h.p.

$$\|\widetilde{A}^{\mathrm{ER}}\|_2 \leq 2 + C \sqrt[4]{\frac{1-p(n)}{np(n)}} \log n.$$

This means that the edge of the semicircular law indeed sharply defines the edge of the limiting spectral distribution.

Spectra of random graphs (Erdős-Rényi model)

It is also interesting to investigate the spectral norm of $\hat{A}^{ER}$.

First note that by the inequality

$$|F^A(x) - F^B(x)| \leq \frac{\mathrm{rank}(A-B)}{n},$$

and the fact that $\overline{A}^{\mathrm{ER}} = \hat{A}^{\mathrm{ER}} - \tilde{A}^{\mathrm{ER}}$ has unit rank for any n, the limiting spectral distribution of $\hat{A}^{ER}$ is also the semicircular law.

The spectral norm of the two matrices $\hat{A}^{\mathrm{ER}}$ and $\tilde{A}^{\mathrm{ER}}$ are different, because the largest eigenvalue changes when a unit rank matrix is added.

Spectra of random graphs (Erdős-Rényi model)

From Bauer-Fike Theorem, we have

$$|\lambda_i(\hat{A}^{\mathrm{ER}}) - \lambda_i(\bar{A}^{\mathrm{ER}})| \leq ||\tilde{A}^{\mathrm{ER}}||_2,$$

and in particular

$$\left|\lambda_n(\hat{A}^{\mathrm{ER}}) - \gamma(n)np(n)\right| \leq 2.$$

Note that, for dense and sparse networks, $\gamma(n)np(n) \gg 2$.
Hence the above result implies that

$$\lambda_n(\widehat{A}^{\mathrm{ER}}) \to n\gamma(n)p_n \quad \text{a.s.}$$

Spectra of random graphs (SBM)

Consider a random graph with n nodes and M communities Ω_m, for $m = 1, \ldots, M$, of equal sizes $K = n/M$, which is assumed to be an integer.

$$\begin{cases} A_{ij} = A_{ji} \; \sim \; Ber(p_m), & \text{if } i, j \in \Omega_m \\ A_{ij} = A_{ji} \; \sim \; Ber(p_0), & \text{if } i \in \Omega_\ell \text{ and } j \in \Omega_m, \ell \neq m. \end{cases} \tag{9}$$

This random graph is called Stochastic Block Model (SBM).

Spectra of random graphs (SBM)

We shall again need to consider the normalized and centered adjacency matrix:

$$\begin{cases} \widetilde{A}_{ij} = \widetilde{A}_{ji} \sim \mathcal{C}(p_m, \gamma) & \text{if } i, j \in \Omega_m \\ \widetilde{A}_{ij} = \widetilde{A}_{ji} \sim \mathcal{C}(p_0, \gamma) & \text{if } i \in \Omega_\ell \text{ and } m \in \Omega_m \\ & \quad \text{with } \ell \neq m, \end{cases} \tag{10}$$

with $\gamma(n) = (np^*(1-p^*))^{-1}$ where $p^* = \max_{m=1,\ldots M} p_m$. Additionally, we assume that all the probabilities p_m scales at the same rate, i.e. $\lim_{n\to+\infty} \frac{p_i}{p_j} = c_{ij}$ for some $c_{ij} > 0$.

Spectra of random graphs (SBM)

Let us first present the following general result:

Theorem (Girko)

Let the symmetric matrix M satisfy Lindeberg's condition. Additionally, the variances σ_{ij}^2 of its entries satisfy the conditions

$$\sup_n \max_{i=1,2,..n} \sum_j \sigma_{ij}^2 < \infty$$

and $\inf_{i,j} n\sigma_{ij}^2 = c > 0$. Then, as $n \to +\infty$, almost surely $F^M(x,n)$, the spectral distribution function of M converges for any x to a deterministic distribution function $S(x)$ whose Stieltjes transform $s(z)$ is given by

$$s(z) = \int \frac{\mathrm{d}S(x)}{x-z} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} c_i(z,n)$$

Spectra of random graphs (SBM)

where $c_i(z,n)$ is the unique solution to a (possibly infinite) system of equations

$$c_i(z,n) = \left\{ \left[-zI - \left(\delta_{pl} \sum_s c_s(z,n) \sigma_{sl}^2 \right)_{p,l=1}^{\infty} \right]^{-1} \right\}_{ii}.$$

We can specify the above general result to SBM.

Corollary

Let $\tilde{A}$ be the normalized centered SBM adjacency matrix. If $p_m(n) \in \omega(n^{-1})$, then almost surely the eigenvalue distribution function converge weakly to a distribution function with Stieltjes transform

$$s(z) = \sum_{m=1}^{M} c_m(z) \tag{11}$$

being $c_m(z)$ the unique solution to the system of equation

$$c_m(z) = \frac{-1/M}{z + \varsigma_m c_m(z) + \varsigma_0 \sum_{\ell \neq m} c_\ell(z)}, \quad m = 1, \ldots, M, \tag{12}$$

with $\varsigma_\ell = \lim_{n \to +\infty} \dfrac{p_\ell(1-p_\ell)}{p^*(1-p^*)}$.

Spectra of random graphs (SBM)

The above result implies that in general the limiting spectral distribution of an SBM is not a semicircular law any longer.

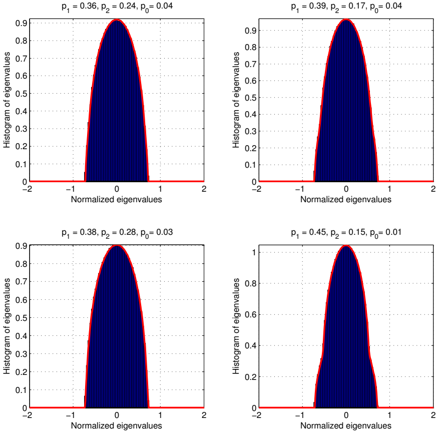

Figure: Comparison between empirically obtained spectrum (histogram) and explicit solution(line) of 2-community SBM

Spectra of random graphs (SBM)

Similarly to ER model, we can also investigate the edge of the limiting distribution and the isolated eigenvalues of $\hat{A}^{SBM}$.

Specifically, if $p_0(n)$ satisfies the inequality $p_0(n) \geq C' \frac{\log^4(n)}{n}$ for some constant $C' > 0$.
Then, there exists a constant $C > 0$ such that w.h.p.

$$\|\widetilde{A}\|_2 \leq 2\sqrt{M^{-1}\left(1 + (M-1)\varsigma_0\right)} + C\sqrt[4]{\frac{1 - p_0(n)}{np_0(n)}}\log(n).$$

Spectra of random graphs (SBM)

We also observe that $\bar{A}^{SBM} = \gamma(n) P \otimes J_K$, where

$$P = \begin{pmatrix} p_1 & p_0 & \dots & p_0 \\ p_0 & p_2 & \ddots & p_0 \\ \vdots & & \ddots & \vdots \\ p_0 & \dots & \dots & p_M \end{pmatrix}, \quad J_K = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \ddots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \dots & \dots & 1 \end{pmatrix},$$

where the size of J_K is $K \times K$.

Spectra of random graphs (SBM)

Next, note that

$$\lambda_{i,j}(A \otimes B) = \lambda_i(A)\lambda_j(B)$$

and if the SBM is homogeneous $(p_1 = p_2 = ... = p_M)$, $\lambda_M(P) = p_1 + (M-1)p_0$, $\lambda_i(P) = p_1 - p_0$ for $i \leq M-1$, and $\lambda_j(J_K) = K = n/M$, which leads to

$$\lambda_n(\hat{A}^{SBM}) = \gamma(n)\frac{n}{M}(p_1 + (M-1)p_0),$$

$$\lambda_i(\hat{A}^{SBM}) = \gamma(n)\frac{n}{M}(p_1 - p_0), \quad i = n-M+1, ..., n-1.$$

We also conclude that in this case the spectral gap has a simple expression: $\gamma(n)np_0$.

Spectra of random graphs (SBM)

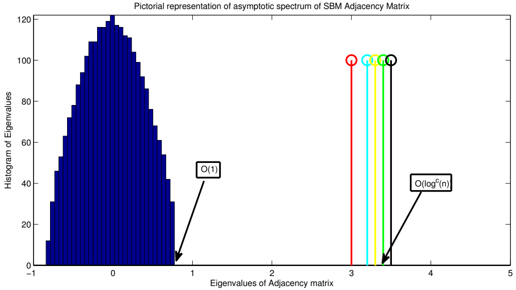

Figure: Extremal eigenvalues and LSD of SBM normalized adjacency matrix.

Spectra of random graphs (SBM)

Let us mention that the limiting spectral distribution of the normalized Laplacian

$$\mathcal{L} = I - D^{-1/2} A D^{-1/2}$$

can be easily obtained from the limiting spectral distribution of the normalized adjacency matrix.

Namely, let $P' = D^{-1/2} A D^{-1/2}$ and consider the case of two communities. We can show that the ESD of the matrix $\frac{1}{2}\sqrt{n}P'$ is asymptotically equivalent to the ESD of the matrix $\frac{1}{\sqrt{n}}A''$, defined as

$$A''_{ij} = \begin{cases} A_{ij}/(p_1 + p_0), & \text{if } i,j \in \Omega_1 \\ A_{ij}/(p_2 + p_0), & \text{if } i,j \in \Omega_2 \\ A_{ij}/\sqrt{(p_1 + p_0)(p_2 + p_0)}, & \text{otherwise.} \end{cases}$$

Spectra of random graphs (Soft GBM)

Model parameters

number of nodes n, geometric dimension d and two measurables functions $F_{\rm in}, F_{\rm out} : \mathbb{T}^d \to [0,1]$.

Model definition

- Set of nodes $V = \{1, \dots, n\}$;
- Each node i has random position X_i on the torus $\mathbb{T}^d$;
- Each node i gets randomly community label $\sigma_i \in \{-1, 1\}$;
- Each pair of nodes (i, j) is connected with probability

$$p_{ij} = \begin{cases} F_{\rm in}\left(X_i - X_j\right) & \text{if } \sigma_i = \sigma_j \\ F_{\rm out}\left(X_i - X_j\right) & \text{if } \sigma_i \neq \sigma_j \end{cases}$$

Spectra of random graphs (Soft GBM)

SGBM important particular cases:

- An SGBM where $F_{\text{in}}(x) = p_{\text{in}}$ and $F_{\text{out}}(x) = p_{\text{out}}$ is an instance of Stochastic Block Model (SBM).
- An SGBM where $F_{\text{in}}(x) = 1(|x| \leq r_{\text{in}})$, $F_{\text{out}}(x) = 1(|x| \leq r_{\text{out}})$ with $r_{\text{in}} > r_{\text{out}}$ is an instance of Geometric Block Model (GBM) introduced in
- Euclidean random graphes with known node locations are used in many ML applications.

Spectra of random graphs (Soft GBM)

For $k \in \mathbb{Z}^d$ and $F : \mathbb{T}^d \to \mathbb{R}$ we define the Fourier transform

$$\widehat{F}(k) = \int_{\mathbb{T}^d} F(x) e^{-2i\pi \langle k,x \rangle} \, dx$$

and assume that $F_{\mathrm{in}}(0), F_{\mathrm{out}}(0)$ are equal to the Fourier series of $F_{\mathrm{in}}(\cdot), F_{\mathrm{out}}(\cdot)$ evaluated at 0.

Spectra of random graphs (Soft GBM)

Theorem

Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of A, and

$$\mu_n(x) = \sum_{i=1}^{n} \delta(x - \lambda_i/n)$$

the ESD of the matrix $\frac{1}{n}A$. Then, almost surely $\mu_n(x)$ converges weakly to

$$\mu(x) = \sum_{k \in \mathbb{Z}^d} \delta\left(x - \frac{\widehat{F}_{\mathrm{in}}(k) + \widehat{F}_{\mathrm{out}}(k)}{2}\right) + \delta\left(x - \frac{\widehat{F}_{\mathrm{in}}(k) - \widehat{F}_{\mathrm{out}}(k)}{2}\right).$$

Note that the above result is for fairly dense networks.

Basic references

Consensus algorithms:

DeGroot, M. H. (1974). Reaching a consensus. Journal of the American Statistical Association, 69(345), 118-121.

Network sampling:

Goodman, L. A. (1961). Snowball sampling. The annals of mathematical statistics, 148-170.

Heckathorn, D. D. (1997). Respondent-driven sampling: a new approach to the study of hidden populations. Social problems, 44(2), 174-199.

Avrachenkov, K., Borkar, V. S., Kadavankandy, A., & Sreedharan, J. K. (2018). Revisiting random walk based sampling in networks: evasion of burn-in period and frequent regenerations. Computational social networks, 5(1), 1-19.

Basic references

Epidemics:

Ganesh, A., Massoulié, L., & Towsley, D. (2005). The effect of network topology on the spread of epidemics. In Proceedings 24th IEEE INFOCOM.

Graph-based semi-supervised learning:

Chapelle, O., Scholkopf, B., & Zien, A. (2006). *Semi-supervised learning*. MIT Press.

Avrachenkov, K., Mishenin, A., Gonalves, P., & Sokol, M. (2012). Generalized optimization framework for graph-based semi-supervised learning. In Proceedings of the 2012 SDM.

Spectral clustering:

Von Luxburg, U. (2007). A tutorial on spectral clustering. Statistics and Computing, 17(4), 395-416.

Abbe, E. (2017). Community detection and stochastic block models: Recent developments. JMLR, 18(1), 6446-6531.

Basic references

Random matrix theory:

Wigner, E.P. (1958). On the distribution of the roots of certain symmetric matrices. Annals of Mathematics, 325-327.

Girko, V.L. (2001). *Theory of stochastic canonical equations.* Kluwer.

Vu, V.H. (2005). Spectral norm of random matrices. In Proceedings of ACM STOC.

Bai, Z., & Silverstein, J. W. (2010). *Spectral analysis of large dimensional random matrices*. Springer.

Tao, T. (2012). *Topics in random matrix theory*. AMS.

Spectra of random graphs:

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

Basic references

Ding, X., & Jiang, T. (2010). Spectral distributions of adjacency and Laplacian matrices of random graphs. The Annals of Applied Probability, 2086-2117.
Nadakuditi, R. R., & Newman, M. E. (2012). Graph spectra and the detectability of community structure in networks. Physical review letters, 108(18).
Avrachenkov, K., Cottatellucci, L., & Kadavankandy, A. (2015). Spectral properties of random matrices for stochastic block model. In Proceedings of 13th WiOpt.
Avrachenkov, K., Bobu, A., & Dreveton, M. (2021). Higher-order spectral clustering for geometric graphs. Journal of Fourier Analysis and Applications, 27(2), 1-29.

Thank you

Any questions?

K.Avrachenkov@inria.fr