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

Suzana de Siqueira Santos[1,2], André Fujita[1] and Catherine Matias[3,4,5]

[1] Universidade de São Paulo, Brasil
[2] Fundação Getulio Vargas, Brasil
[3] Centre National de la Recherche Scientifique, Paris, France
[4] Sorbonne Université, Paris, France
[5] Université de Paris Cité, Paris, France

Mars 2022

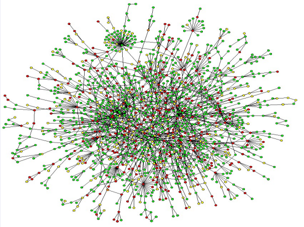
Outline

1. Model selection for statistical analysis of random graphs
2. The empirical spectral density of a graph
3. Model selection procedure
4. Results and some experiments

Outline

Data: networks, their properties and beyond I

Some networks characteristics



- Potentially large number n of interacting entities,
- Potentially **sparse** networks: number of edges $\ll O(n^2)$,
- **Scale-free property** : Degree distribution has a power law $\mathbb{P}(D_i = k) = ck^{-\gamma}, (\gamma > 0)$,
- **Small world property**: shortest path length is small on average (less than 6),
- **Transitivity/clustering property**: is there a large amount of triangles?
- ...

Data: networks, their properties and beyond II

Some challenges

- Go beyond these (local) descriptors and capture **higher-level structures**, such as topological patterns, cliques, nodes groups, etc,
- Propose relevant models that will capture those structures **without any a priori information on which structures we are looking for**,
- Go from static to **dynamic** models,
- Go from pairwise interactions to **higher order** interactions,
- ...

Random graphs models for networks

Random graphs are the mathematical tools that model the networks.

Some existing models, advantages and drawbacks

- Erdös-Rényi, **simple and mathematically well-understood**, too homogeneous;

$$\textit{edges are i.i.d.} \sim \mathcal{B}(p)$$

Random graphs models for networks

Random graphs are the mathematical tools that model the networks.

Some existing models, advantages and drawbacks

- Erdös-Rényi, **simple and mathematically well-understood**, too homogeneous;
- Models based on degree distribution, **scale-free property**, only a partial descriptor of the graph, greedy numerical simulations with fixed-degrees models;
  Nodes degrees are fixed, samples obtained through rewiring algorithm

Random graphs models for networks

Random graphs are the mathematical tools that model the networks.

Some existing models, advantages and drawbacks

- Erdös-Rényi, **simple and mathematically well-understood**, too homogeneous;
- Models based on degree distribution, **scale-free property**, only a partial descriptor of the graph, greedy numerical simulations with fixed-degrees models;
- Generative processes (like preferential attachment), **dynamic model**, depends on parameters that may not be inferred from data (initialisation, stop, exact procedure, . . .),

  Start from a small graph, add nodes and connect them with higher prob. to nodes with large degrees

Random graphs models for networks

Random graphs are the mathematical tools that model the networks.

Some existing models, advantages and drawbacks

- Erdös-Rényi, **simple and mathematically well-understood**, too homogeneous;
- Models based on degree distribution, **scale-free property**, only a partial descriptor of the graph, greedy numerical simulations with fixed-degrees models;
- Generative processes (like preferential attachment), **dynamic model**, depends on parameters that may not be inferred from data (initialisation, stop, exact procedure, . . .),
- Exponential random graph, **natural from a statistical point of view**, big inference issues ;

$$\mathbb{P}_\theta(\mathbf{Y} = \mathbf{y}) = c(\theta)^{-1} \exp(\theta^\intercal S(\mathbf{y}))$$

Random graphs models for networks

Random graphs are the mathematical tools that model the networks.

Some existing models, advantages and drawbacks

- Erdös-Rényi, **simple and mathematically well-understood**, too homogeneous;
- Models based on degree distribution, **scale-free property**, only a partial descriptor of the graph, greedy numerical simulations with fixed-degrees models;
- Generative processes (like preferential attachment), **dynamic model**, depends on parameters that may not be inferred from data (initialisation, stop, exact procedure, ...),
- Exponential random graph, **natural from a statistical point of view**, big inference issues ;
- ...

Model selection issue

- For a fixed model, specific statistical procedures fit the model to the data through parameter inference algorithms ;
- **Issue**: how do you compare 2 different models?
  - Model selection procedures are required.
  - Criteria such as estimated likelihood can't be used (at least without penalization terms)

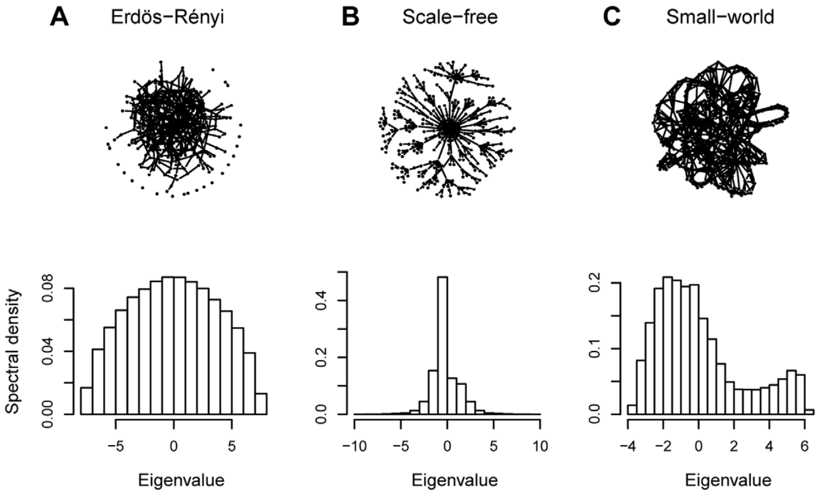
Outline

Empirical spectral density (ESD) I

Spectrum of a matrix

- The spectrum of a matrix is the set of its eigenvalues;
- The adjacency matrix of an undirected graph is a real symmetric matrix, hence its spectrum is real valued.
- The empirical spectral density (ESD) is the empirical distribution over the set of (normalized) eigenvalues.
- **Characteristic:** a lot of the graph's information is contained in its ESD.

Empirical spectral density (ESD) II



Picture from [TSFF12].

Empirical spectral density (ESD) III

Adjacency or Laplacian?

- Laplacian matrices L are 'normalized' versions of adjacency matrices A, e.g $L = I - D^{-1/2}AD^{-1/2}$ where D diagonal with degrees.
- In the following, we focus on adjacency matrices.

Beware: Random graphs induce random ESD

- Once we fix a random graph model, the entries of the adjacency matrix are random variables. Thus the eigenvalues are also random and the ESD becomes a random measure.
- Behaviour of the random ESD, as the number of nodes increases, relates to **random matrix theory**.

Empirical spectral density (ESD) IV

Notation

- A adjacency matrix of (undirected) graph G over n nodes
- $\lambda_1^G \geq \lambda_2^G \geq \cdots \geq \lambda_n^G$ the eigenvalues of A
- ESD and corresponding cumulative distr function (CDF):

$$\mu^G(\cdot) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i^G/\sqrt{n}}(\cdot); \quad F^G(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left\{ \frac{\lambda_i^G}{\sqrt{n}} \leq x \right\},$$

In other words, for any function f, we have
$\mu^G(f) = \int_{\mathbb{R}} f(\lambda)\mu^G(d\lambda) = \frac{1}{n}\sum_{i=1}^{n} f(\lambda_i^G/\sqrt{n})$.

ESD as a tool to estimate a random graph model

Principle and Questions

- It is **easy to compute** the ESD of a random graph from the observation of its adjacency matrix
- AND a lot of the **graph's information** is contained in its spectral density.
- Thus, a natural idea is to use this ESD to discriminate between 2 random graphs models.

Our goals

Takahashi's *et al.* [TSFF12] have proposed such a procedure.

- Better understand the theoretical properties of such a procedure
- For this we need to understand more about the limiting behaviour of the ESD.

Convergence properties of the ESD

Wigner's law

If the entries of A are **iid with zero mean and unit variance**, then the ESD converges (weakly in expectation) to the semicircle law

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

- Important issues in random matrix theory: in more general contexts,
  - does the ESD still converge to some limit?
  - If yes, what is this limit?
- **Question here**: can we use these results to prove that the model selection procedure based on ESD is consistant?

Spectral densities of random graph models I

Spectral density

- A (parametric) random graph model is a collection of graphs $\mathcal{G}$ together with a family of distributions $\{\mathbb{P}_\theta, \theta \in \Theta\}$ over $\mathcal{G}$.
- Consider a sequence of graphs $(G_n)_{n\geq 1}$ with distribution $\mathbb{P}_\theta$, whenever the limit of the sequence of ESDs $(\mu^{G_n})_{n\geq 1}$ exists, we denote it μ_θ. It's the **spectral density associated to the probability measure** $\mathbb{P}_\theta$.

Example

The Erdös-Rényi (ER) model $\mathcal{G}(n,p)$ is a family of distributions over the set of graphs with n nodes. Here $\theta = p$ is the parameter of the model and it is known that the ER spectral density is

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

Spectral densities of random graph models II

How do you compute the spectral density?

- Recall that the spectral density is defined as the **limit** of ESDs.
- Either you have an analytic formula for the spectral density
- Or you can sample many different graphs from $\mathbb{P}_\theta$ with large size and approximate the spectral density through the ESDs of these graphs.

State of the art about existing spectral densities

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

Table: Convergences of ESD in different models.

DR : d-regular graphs ; BM: blockmodel graphs (+ some constraints).
Remark: Very few results compared to variety of models.

Outline

Model selection by Takahashi's *et al.* [TSFF12]

General idea

- For a fixed parametric model $\{\mathbb{P}_\theta, \theta \in \Theta\}$ and a grid of values $\theta \in \tilde{\Theta}$, compute (the kernel estimator of) the ESD μ_θ under parameter θ.
- Compare each μ_θ with (the kernel estimator of) the ESD μ^G of the observed graph G, through some distance D.
- Select the model and corresponding parameter value with closest distance.

Advantages

- No need for a model specific inference procedure;
- Compares models with completely different number of parameters, with no need to penalize for these numbers.

Algorithm's description

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

Output: Return model and parameter with best fit.

Compute the kernel density estimator $\mu^{\mathcal{G}}$.

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

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

    if the limit of ESD from $P^i_{\theta^j}$, denoted by $\mu^i_{\theta^j}$, is known analytically **then**

      $D_{i,j} \leftarrow D(\mu^{\mathcal{G}}, \mu^i_{\theta^j})$.

    else

      Sample M graphs $\mathcal{G}_1, \mathcal{G}_2, \cdots, \mathcal{G}_M$ from $P^i_{\theta^j}$.

      for each graph $\mathcal{G}_m$ **do**

        Compute the kernel density estimator $\mu^{\mathcal{G}_m}$.

      end for

      $\hat{\mu}^i_{\theta^j} \leftarrow \frac{1}{M} \sum_{m=1}^{M} \mu^{\mathcal{G}_m}$.

      $D_{i,j} \leftarrow D(\mu^{\mathcal{G}}, \hat{\mu}^i_{\theta^j})$

    end if

  end for

end for

return Model and parameter (i, j) with smallest $D_{i,j}$ value

Outline

Consistency results

Theorem

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

$$\hat{\theta}_n = \underset{\theta \in \Theta}{Argmin} \|\mu^{\mathcal{G}_n} - \mu_\theta\|_1$$

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

Consequence: Using D the $\mathbb{L}_1$-norm, [TSFF12]'s procedure will recover the true parameter value.

Block model spectral density - beyond [ACK15] I

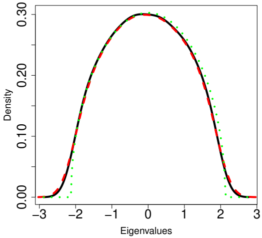
ESD convergence for BM is established in [ACK15] under the particular case of equal size groups ; equal probability p_0 of connecting vertices from different groups ; and intra group probability $p_m > p_0$ for each $1 \leq m \leq M$.

Settings - 5 scenarios

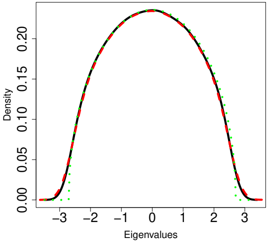
$K=$ bloc sizes ; $M=$ nb groups ; p_0 inter-group probability ; p_m intra-groups probabilities

	K	M	p_0	p_m
S1	300	3	0.2	(0.8,0.5,0.6)
S2	300	10	0.2	$(0.8, 0.5, 0.6, 0.7, 0.4, 0.9, 0.55, 0.42, 0.38, 0.86)$
S3	100,80,300	3	0.2	(0.8,0.5,0.6)
S4	300	3	0.9	(0.8,0.5,0.6)
S5	300	3		$\begin{pmatrix} 0.8 & 0.1 & 0.2 \\ \star & 0.5 & 0.05 \\ \star & \star & 0.6 \end{pmatrix}$

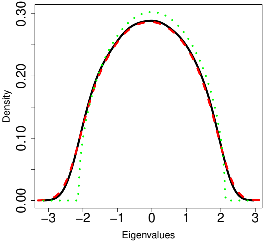
Block model spectral density - beyond [ACK15] II



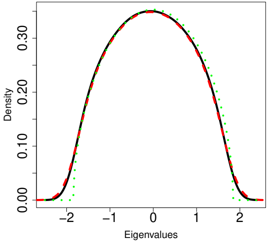
(a) Scenario 1



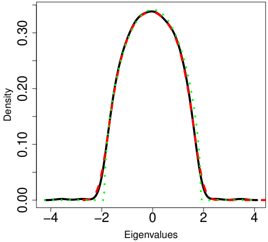
(b) Scenario 2



(c) Scenario 3



(d) Scenario 4



(e) Scenario 5

Block model spectral density - beyond [ACK15] III

Notation $\gamma(n) = \sqrt{np_\star(1-p_\star)}$; $\tilde{A} = (A - E(A))/\gamma(n)$ and $p_\star = \max_{m \geq 1} p_m$.

Comments

- For all the scenarios, the ESDs of $\mathbf{A}/\gamma(n)$ and $\tilde{\mathbf{A}}$ were very close
- For scenarios 1, 2, 4, and 5, the ESD of $\tilde{\mathbf{A}}$ and $\mathbf{A}/\gamma(n)$ are close to the theoretical distribution from [ACK15] ;
- For scenario 3 the ESD are farther from the theoretical density. This might be due to numerical errors during ESD estimation or might indicate that Avrachenkov *et al.*'s results do not apply to graphs with blocks of different sizes.

Conclusions

- We reviewed known results about the limiting behaviour of ESD in different random graphs models
  - Very few results exist compared to the variety of models (difficult problems)
  - For BlockModels, only partial results are known and in some cases (eg different block sizes), the limiting behaviour seems to differ
- We wanted to establish the convergence of the procedures from [TSFF12]
  - Our result is only partial and establishes consistency of parameter estimation within a parametric model ;
  - but not the convergence of the model selection procedure.

Thank you for your attention!

Bibliography

[ACK15] K. Avrachenkov, L. Cottatellucci, and A. Kadavankandy. Spectral properties of random matrices for stochastic block model. In *The 2015 4th International Workshop on Physics-Inspired Paradigms in Wireless Communications and Networks*, pages 537–544, 2015.

[TSFF12] Daniel Yasumasa Takahashi, João Ricardo Sato, Carlos Eduardo Ferreira, and André Fujita. Discriminating Different Classes of Biological Networks by Analyzing the Graphs Spectra Distribution. *PLoS ONE*, 7(12):e49949, 2012.