Spectral Clustering of Sparse Graphs and the Non-Backtracking Matrix

Marianna Bolla
BME Math. Inst. marib@math.bme.hu
Probability Seminar at the Rényi Institute

Application: Hannu Reittu, VTT, Finland
Simulation: Daniel Zhou, BSM
Thanks to: László Lovász, Tamás Móri, Katalin Friedl, Dániel Keliger

December 6, 2023.

Outline

- Non-backtracking matrix of simple graphs.
- Sparse stochastic block model.
- Belief propagation.
- Inflation–deflation.
- Non-backtracking matrix of edge-weighted graphs.
- k-means clustering with node representatives.
- Bond percolation for simulated data.
- Application for real-world data.

Non-Backtracking (Hashimoto) matrix of simple graphs

$G = (V, E)$ simple graph, $|V| = n$, $|E| = m$;
the entries of the non-backtracking matrix $\mathbf{N} = (n_{ef})$ are indexed by the directed edges (bidirected edges of E), $|E^{\rightarrow}| = 2m$:

$$n_{ef} = \delta_{e \to f} \delta_{f \neq e^{-1}}, \qquad n_{i \to j,\, s \to l} = \delta_{js}(1 - \delta_{il}),$$

where $e = \{i \to j\}$ and $f = \{s \to l\}$ are directed edges, and $e \to f$ with $e = (e_1, e_2)$ and $f = (f_1, f_2)$ means that $e_2 = f_1$;
$e^{-1} = \{j \to i\}$.

Relation to line-graphs

Proposition

If $\mathbf{N} = \begin{pmatrix} \mathbf{N}_{11} & \mathbf{N}_{12} \\ \mathbf{N}_{21} & \mathbf{N}_{22} \end{pmatrix}$, *where the two (row/column) blocks correspond to the edges and their inverses (in the same order), then*

$$\mathbf{N}_{11}^* = \mathbf{N}_{22}, \quad \mathbf{N}_{22}^* = \mathbf{N}_{11}, \quad \mathbf{N}_{12}^* = \mathbf{N}_{12}, \quad \mathbf{N}_{21}^* = \mathbf{N}_{21}.$$

Further, $\mathbf{N}_{11} + \mathbf{N}_{12} + \mathbf{N}_{21} + \mathbf{N}_{22}$ *is equal to the* $m \times m$ *adjacency matrix of the line-graph of* G.

In Lovász, Combinatorial Exercises it is proved that if the line-graphs of two simple graphs, provided they both have node-degrees at least 4, are isomorphic, then they are isomorphic too. However, if the degree condition does not hold, it can happen that two not isomorphic graphs have isomorphic line-graphs. For example, a triangle and a star on 4 vertices.

Transpose (*), involution, and swapping

Though $\mathbf{N}$ is not a normal matrix, even not always diagonalizable (the algebraic and geometric multiplicity of some of its eigenvalues may not be the same), it exhibits some symmetry: $n^*_{ef} = n_{e^{-1}\,f^{-1}}$.

With the notation $\check{x}_e := x_{e^{-1}}$ for the coordinates of $\mathbf{x}, \check{\mathbf{x}} \in \mathbb{R}^{2m}$: if $\mathbf{x} = (\mathbf{x}_1^*, \mathbf{x}_2^*)^*$, then $\check{\mathbf{x}} = (\mathbf{x}_2^*, \mathbf{x}_1^*)^*$ (swapping).
Let $\mathbf{V}$ denote the following involution on $\mathbb{R}^{2m}$ ($\mathbf{V} = \mathbf{V}^{-1}$, $\mathbf{V}^2 = \mathbf{I}$):
$\mathbf{V} = \begin{pmatrix} \mathbf{O} & \mathbf{I}_m \\ \mathbf{I}_m & \mathbf{O} \end{pmatrix}$. Then $\mathbf{V}\mathbf{x} = \check{\mathbf{x}}$ and $\mathbf{V}\check{\mathbf{x}} = \mathbf{x}$;
$\mathbf{N}^* = \mathbf{V}\mathbf{N}\mathbf{V}$ and $\mathbf{N}^*\check{\mathbf{x}} = (\check{\mathbf{N}\mathbf{x}})$.

Consequently: if $\mathbf{x}$ is a right eigenvector of $\mathbf{N}$, then $\check{\mathbf{x}}$ is a left eigenvector of $\mathbf{N}$ (and right eigenvector of $\mathbf{N}^*$) with the same eigenvalue.

Eigenvalues of N (Ihara formula)

N is a Frobenius-type matrix, its largest absolute value eigenvalue $\lambda(\mathbf{N})$ is positive real, and it can also have some other „structural" real eigenvalues. Since the characteristic polynomial of **N** has real coefficients, its complex eigenvalues occur in conjugate pairs in the bulk of its spectrum.

Ihara formula: **N** has $m - n$ eigenvalues equal to 1 and $m - n$ eigenvalues equal to -1, whereas its further eigenvalues are those of the $2n \times 2n$ matrix

$$\mathbf{K} = \begin{pmatrix} \mathbf{O} & \mathbf{D_A} - \mathbf{I}_n \\ -\mathbf{I}_n & \mathbf{A} \end{pmatrix},$$

where **A** is the adjacency- and $\mathbf{D_A}$ is the degree-matrix of the graph (diagonal, contains the degrees=row-sums of **A**).

K always has at least one additional eigenvalue 1, the geometric multiplicity of which is equal to the number of the connected components of G and $\lambda_{max}(\mathbf{N}) = \lambda_{max}(\mathbf{K}) \leq \lambda_{max}(\mathbf{A})$.

Real eigenvalues and eigenvectors of N (beyond the Ihara formula)

Two auxiliary matrices are introduced: the $2m \times n$ matrix **End** has entries $end_{ei} = 1$ if i is the end-node of the (directed) edge e and 0, otherwise; the $2m \times n$ matrix **Start** has entries $start_{ei} = 1$ if i is the start-node of the (directed) edge e and 0, otherwise. Then for any vector $\mathbf{u} \in \mathbb{R}^n$ and for any edge $e = \{i \to j\}$ the following holds:

$$(\mathbf{End}\,\mathbf{u})(e) = u_j \quad \text{and} \quad (\mathbf{Start}\,\mathbf{u})(e) = u_i.$$

Consequently, $\mathbf{End}\,\mathbf{u}$ is the $2m$-dimensional inflated version of the n-dimensional vector $\mathbf{u}$, where the coordinate u_j of $\mathbf{u}$ is repeated as many times, as many edges have end-node j; likewise, in the $2m$-dimensional inflated vector $\mathbf{Start}\,\mathbf{u}$, the coordinate u_i of $\mathbf{u}$ is repeated as many times, as many edges have start-node i.
As each edge is considered in both possible directions, these multiplicities are the node-degrees d_j and d_i, respectively.

$$\mathbf{End}^* \mathbf{End} = \mathbf{Start}^* \mathbf{Start} = \mathrm{diag}(d_1, \ldots, d_n) = \mathbf{D_A}$$

For any vector $\mathbf{x} \in \mathbb{R}^{2m}$, define

$$x_i^{out} := \sum_{j:\, j \sim i} x_{i \to j} \quad \text{and} \quad x_i^{in} := \sum_{j:\, j \sim i} x_{j \to i} \quad (i = 1, \ldots, n).$$

These become the coordinates of the n-dimensional (column) vectors $\mathbf{x}^{in}$ and $\mathbf{x}^{out}$. Trivially,

$$\mathbf{x}^{out} = \mathbf{Start}^* \mathbf{x} \quad \text{and} \quad \mathbf{x}^{in} = \mathbf{End}^* \mathbf{x} \quad (i = 1, \ldots, n).$$

Calculation

$$
\begin{aligned}
(\mathbf{N}^*\mathbf{x})_i^{out} &= \sum_{e:\, e_1=i} (\mathbf{N}^*\mathbf{x})_e = \sum_{e:\, e_1=i} \sum_{f\to e,\, f\neq e^{-1}} x_f \\
&= \sum_{e:\, e_1=i} [\sum_{f\to e} x_f - x_{e^{-1}}] \\
&= \sum_{f:\, f_2=i} x_f \sum_{e:\, e_1=i} 1 - \sum_{e:\, e_1=i} x_{e^{-1}} \\
&= x_i^{in} d_i - \sum_{e:\, e_2^{-1}=i} x_{e^{-1}} = d_i x_i^{in} - x_i^{in} = (d_i - 1) x_i^{in}
\end{aligned}
$$

Calculation

$$
\begin{aligned}
(\mathbf{N}^*\mathbf{x})_i^{in} &= \sum_{e:\, e_2=i} (\mathbf{N}^*\mathbf{x})_e = \sum_{e:\, e_2=i} \sum_{f\to e,\, f\neq e^{-1}} x_f \\
&= \sum_{j=1}^{n} a_{ji} \sum_{f:\, f_2=j,\, f_1\neq i} x_f \\
&= \sum_{j=1}^{n} a_{ji} \sum_{f:\, f_2=j} x_f - \sum_{j=1}^{n} a_{ji} x_{i\to j} \\
&= \sum_{j=1}^{n} a_{ij} x_j^{in} - \sum_{j:\, j\sim i} x_{i\to j} = (\mathbf{A}\mathbf{x}^{in})_i - x_i^{out},
\end{aligned}
$$

where we used that the (0-1) adjacency matrix $\mathbf{A}$ of the graph is symmetric with entries $a_{ij} = a_{ji} = \delta_{i\sim j}$.

Summarizing, if $\mathbf{x}$ is a (right) eigenvector of $\mathbf{N}^*$ with (real) eigenvalue μ, i.e., $\mathbf{N}^*\mathbf{x} = \mu\mathbf{x}$, then $(\mathbf{N}^*\mathbf{x})^{out} = (\mu\mathbf{x})^{out} = \mu\mathbf{x}^{out}$ and $(\mathbf{N}^*\mathbf{x})^{in} = (\mu\mathbf{x})^{in} = \mu\mathbf{x}^{in}$. Therefore,

$$\mu \begin{pmatrix} \mathbf{x}^{out} \\ \mathbf{x}^{in} \end{pmatrix} = \begin{pmatrix} (\mathbf{N}^*\mathbf{x})^{out} \\ (\mathbf{N}^*\mathbf{x})^{in} \end{pmatrix} = \begin{pmatrix} \mathbf{O} & \mathbf{D_A} - \mathbf{I}_n \\ -\mathbf{I}_n & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{x}^{out} \\ \mathbf{x}^{in} \end{pmatrix},$$

so

$$\mu \begin{pmatrix} \mathbf{x}^{out} \\ \mathbf{x}^{in} \end{pmatrix} = \begin{pmatrix} (\mathbf{D_A} - \mathbf{I}_n)\mathbf{x}^{in} \\ \mathbf{A}\mathbf{x}^{in} - \mathbf{x}^{out} \end{pmatrix} = \mathbf{K} \begin{pmatrix} \mathbf{x}^{out} \\ \mathbf{x}^{in} \end{pmatrix}.$$

In particular, if $\mathbf{x}$ is a right eigenvector of $\mathbf{N}^*$ with a real eigenvalue $\mu \neq 0$, then the $2n$-dimensional vector comprised of parts $\mathbf{x}^{out}$ and $\mathbf{x}^{in}$ is a right eigenvector of $\mathbf{K}$ with the same eigenvalue. Indeed,

$$\mu \begin{pmatrix} \mathbf{x}^{out} \\ \mathbf{x}^{in} \end{pmatrix} = \begin{pmatrix} (\mathbf{N}^*\mathbf{x})^{out} \\ (\mathbf{N}^*\mathbf{x})^{in} \end{pmatrix} = \mathbf{K} \begin{pmatrix} \mathbf{x}^{out} \\ \mathbf{x}^{in} \end{pmatrix}.$$

According to the previous remarks, the vector $\mathbf{x}$ is a left eigenvector, and $\check{\mathbf{x}}$ is a right eigenvector of $\mathbf{N}$ with the same eigenvalue. To both of them the two segments, $\mathbf{x}^{out}$ and $\mathbf{x}^{in}$ of the right eigenvector of $\mathbf{K}$ are responsible.
In view of the relation $\mathbf{x}^{out} = \frac{1}{\mu}(\mathbf{D_A} - \mathbf{I}_n)\mathbf{x}^{in}$, it suffices to consider only $\mathbf{x}^{in} \in \mathbb{R}^n$ for further clustering purposes.

Bond percolation

The bond percolation threshold for the giant component to appear in a sparse simple graph is $\beta > \frac{1}{\lambda_{max}(\mathbf{N})}$, where β is the edge retention probability, see Newman, M. E. J., Message passing methods on complex networks, Proc. R. Soc. London A (2023).
The proof uses the method of Belief Propagation (BP) (when the so-called message passing system of equations has a non-trivial solution).
In the dense case, it happens at $\frac{1}{\lambda_{max}(\mathbf{A})}$, see Bollobás, B., Borgs, C., Chayes, J., and Riordan, O., Percolation on dense graph sequences, Annals of Probability (2010).
More generally, we are looking for the number k, so that k strongly connected clusters (communities) can be detected (within the giant component) in a graph coming from the sparse stochastic block model. We are also looking for the clusters themselves.
The Erdős–Rényi graph $G_n(p)$ is a special case with $k = 1$, where the edges of the complete graph on n nodes are retained with probability $\beta = p$.

The sparse stochastic block model SBM_k

The $k \times k$ *probability matrix* $\mathbf{P}$ of the random graph $G_n \in SBM_k$ has entries $p_{ab} = \frac{c_{ab}}{n}$, where the $k \times k$ symmetric affinity matrix $\mathbf{C} = (c_{ab})$ stays constant as $n \to \infty$. An edge between $i < j$ comes into existence, independently of the others, with probability p_{ab} if $i \in V_a$ and $j \in V_b$, where $(V_1, \ldots, V_k)$ is a partition of the node-set V into k disjoint clusters; $a_{ji} := a_{ij}$. It can be extended to the $i = j$ case when self-loops are allowed, or else, the diagonal entries of the adjacency matrix are zeros.

$\bar{\mathbf{A}}$: the $n \times n$ inflated matrix of the $k \times k$ $\mathbf{P}$: $\bar{a}_{ij} = p_{ab}$ if $i \in V_a$ and $b \in V_b$. When loops are allowed, then $\mathbb{E}(a_{ij}) = \bar{a}_{ij}$ for all $1 \leq i, j \leq n$. In the loopless case, the expected adjacency matrix $\mathbb{E}\mathbf{A}$ differs from $\bar{\mathbf{A}}$ with respect to the the main diagonal, but the diagonal entries are negligeable.

Special cases

Sometimes $c_{ab} = c_{in}$ is the within-cluster ($a = b$) and $c_{ab} = c_{out}$ is the between-cluster ($a \neq b$) affinity. The network is called assortative if $c_{in} > c_{out}$, and disassortative if $c_{in} < c_{out}$. Of course, remarkable difference is needed between the two, to recognize the clusters.

The cluster sizes are $n_1, \ldots, n_k$ ($\sum_{i=1}^{k} n_i = n$), so the $k \times k$ diagonal matrix $\mathbf{R} := \mathrm{diag}(r_1, \ldots, r_k)$, where $r_a = \frac{n_a}{n}$ is the relative size of cluster a ($a = 1, \ldots, k$), is also a model parameter ($\sum_{a=1}^{k} r_a = 1$). It is nearly kept fixed as $n \to \infty$.

The model SBM_k is called symmetric if $r_1 = \cdots = r_k = \frac{1}{k}$ and all diagonal entries of the affinity matrix are equal to c_{in}, whereas the off-diagonal ones to c_{out}.

Average degrees

The average degree of a real world graph on m edges and n nodes is $\frac{2m}{n}$. The expected average degree of the random graph $G_n \in SBM_k$ is

$$c = \frac{1}{n}\sum_{a=1}^{k}\sum_{b=1}^{k} n_a n_b p_{ab} = \frac{1}{n^2}\sum_{a=1}^{k}\sum_{b=1}^{k} n_a n_b c_{ab} = \sum_{a=1}^{k} r_a c_a,$$

where $c_a = \sum_{b=1}^{k} r_b c_{ab}$ is the average degree of cluster a. It is valid only if self-loops are allowed. Otherwise, c_a and c should be decreased with a term of order $\frac{1}{n}$, but it will not make too much difference in the subsequent calculations.

Kesten–Stigum threshold

In Bordenave, C., Lelarge, M., Massoulié, L., Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs, Ann. Probab. (2018), the case when $c_a = c$ for all a is considered. (This is the hardest case, as otherwise the clusters could be distinguished by sorting the node-degrees.) In this case $\frac{1}{c}\bar{\mathbf{A}}$ is a stochastic matrix, and so, the spectral radius of $\bar{\mathbf{A}}$ is c.

In the symmetric case, $c = \frac{c_{in}+(k-1)c_{out}}{k}$ and the separation of the clusters only depends on the c_{in}, c_{out} relation. If c_{in} is „close" to c_{out}, then the groups cannot be distinguished. The detectability Kesten–Stigum threshold in the symmetric case is

$$|c_{in} - c_{out}| > k\sqrt{c} \Longleftrightarrow \mu_2 = \cdots = \mu_k > \sqrt{c},$$

where $\mu_2 = \cdots = \mu_k$ is the second largest (real) eigenvalue of $\mathbf{N}$.

BP in the general sparse SBM_k model

Given the observed graph on n nodes,

$$\psi_i^a \propto \mathbb{P}(i \text{ is in the cluster } a), \quad a = 1, \dots, k$$

defines the marginal membership (state) distribution of node i. We assume that our neighbors are independent of each other, when conditioned on our own state. This can be modeled by having each node j send a message to i, which is an estimate of j's marginal if i were not there. Therefore, the conditional probability

$$\psi_{j\to i}^a := \mathbb{P}(j \text{ is in cluster } a \text{ when } i \text{ is not present})$$

can be computed through neighbors of j that are different from i:

$$\psi_{j\to i}^a = C_a^{ij} r_a \prod_{l\sim j,\, l\neq i} \sum_{b=1}^{k} \psi_{l\to j}^b p_{ab}, \quad a = 1, \dots, k,$$

where C_a^{ij} is a normalizing factor.

The above BP (message-passing) system of equations ($2mk$ non-linear equations with the same number of unknowns) can be solved by initializing messages randomly, then repeatedly updating them. This procedure usually converges quickly and the resulting fixed point gives a good estimate of the marginals:

$$\psi_i^a \propto r_a \prod_{j \sim i} \sum_{b=1}^{k} \psi_{j \to i}^b \, p_{ab},$$

where the constant of proportionality is chosen according to $\sum_{a=1}^{k} \psi_i^a = 1$. However, the system of equations contains the model parameters, so it can be solved only if the model parameters are known. For a given graph (n and k fixed), the parameters r_a's and c_{ab}'s can be estimated by the EM algorithm, see Bolla, M., Spectral clustering and biclustering, Wiley (2013).

Calculations

In Moore, C., The computer science and physics of community detection: Landscapes, phase transitions, and hardness, Bull. EATCS (2017), the symmetric case is treated, when BP has a trivial fixed point $\psi^a_{j\to i} = \frac{1}{k}$, for $a = 1, \ldots, k$. If it gets stuck there, then BP does no better than chance. It happens when this trivial fixed point of this discrete dynamical system is asymptotically stable.

In the generic case, we have an unstable fixed point via linearization:

$$\psi^a_{j\to i} := r_a + \varepsilon^a_{j\to i}.$$

We substitute it in the original BP system and expand it to first order in ε (vector of $2mk$ coordinates $\varepsilon^a_{j\to i}$'s):

$$\varepsilon^a_{j\to i} = \psi^a_{j\to i} - r_a = r_a \left\{ C_a^{ij} \prod_{l\sim j,\, l\neq i} \left[\sum_{b=1}^{k} \psi^b_{l\to j}\, p_{ab} \right] - 1 \right\}$$

$$= r_a \left\{ C_a^{ij} \prod_{l\sim j,\, l\neq i} \left[\sum_{b=1}^{k} (r_b + \varepsilon^b_{l\to j})\, p_{ab} \right] - 1 \right\}$$

$$= r_a \left\{ C_a^{ij} \prod_{l\sim j,\, l\neq i} \left[\sum_{b=1}^{k} r_b p_{ab} + \sum_{b=1}^{k} \varepsilon^b_{l\to j}\, p_{ab} \right] - 1 \right\}$$

$$= r_a \left\{ C_a^{ij} \prod_{l\sim j,\, l\neq i} \left[\frac{c_a}{n} + \sum_{b=1}^{k} \varepsilon^b_{l\to j}\, \frac{c_{ab}}{n} \right] - 1 \right\}$$

$$= r_a \left\{ C_a^{ij} (\frac{1}{n})^{s_j-1} \left[\sum_{b=1}^{k} \sum_{l\sim j,\, l\neq i} \varepsilon^b_{l\to j}\, c_{ab}\, c_a^{s_j-2} + c_a^{s_j-1} \right] - 1 \right\} + O(\varepsilon^2).$$

Here s_j denotes the number of neighbors of j and $s_j - 1$ is the number of its neighbors that are different from i (this number is frequently 0 or 1, as we have a sparse graph). If $s_j < 2$ happens, then the corresponding entry of the non-backtracking matrix is 0. To specify the normalizing factor C_a^{ij}, we substitute zeros for ε's that provide the trivial solution. This approximately yields

$$C_a^{ij} \left(\frac{1}{n}\right)^{s_j-1} c_a^{s_j-1} - 1 = 0,$$

so

$$C_a^{ij} = \left(\frac{n}{c_a}\right)^{s_j-1}.$$

Substituting this into the original equation, we get

$$\varepsilon^a_{j\to i} = r_a \left\{ \left(\frac{n}{c_a}\right)^{s_j-1} \left(\frac{1}{n}\right)^{s_j-1} c_a^{s_j-2} \left[\sum_{b=1}^{k} \sum_{l\sim j, l\neq i} \varepsilon^b_{l\to j}\, c_{ab} + c_a \right] - 1 \right\}$$

$$+ O(\varepsilon^2) = r_a \left\{ \frac{1}{c_a} \left[\sum_{b=1}^{k} \sum_{l\sim j, l\neq i} \varepsilon^b_{l\to j}\, c_{ab} + c_a \right] - 1 \right\} + O(\varepsilon^2).$$

The linear dynamical system approximating the above system of difference equations is

$$\varepsilon = (\mathbf{N} \otimes \mathbf{T})\varepsilon,$$

where $\mathbf{T} = \mathbf{GRC}$ is the transmission matrix with $\mathbf{G} = \mathrm{diag}(\frac{1}{c_1}, \ldots, \frac{1}{c_k})$.

The fixed point $\mathbf{0}$ of

$$\varepsilon^{(t+1)} = (\mathbf{N} \otimes \mathbf{T})\varepsilon^{(t)}$$

is unstable, if the spectral radius of the big block matrix $\mathbf{N} \otimes \mathbf{T}$ is grater than 1.

Note that $\mathbf{T}$ is a stochastic matrix, so its largest eigenvalue is 1, and the others are less than 1 and positive in the assortative case. In this way, we have proved the following.

Theorem

With arbitrary, but fixed positive integer k and $k \times k$ parameter matrices $\mathbf{R} = \mathrm{diag}(r_1, \ldots, r_k)$ (cluster proportions) and $\mathbf{C}$ (symmetric affinity matrix), the linear approximation of the BP system is $\varepsilon = (\mathbf{N} \otimes \mathbf{T})\varepsilon$, where ε is a $2mk$-dimensional vector and the $2mk \times 2mk$ matrix of the linear system is $\mathbf{N} \otimes \mathbf{T}$. Here $\mathbf{N}$ is the non-backtracking matrix of the graph and $\mathbf{T} = \mathbf{GRC}$ is the transmission matrix with $\mathbf{G} = \mathrm{diag}(\frac{1}{c_1}, \ldots, \frac{1}{c_k})$, where $c_a = \sum_{b=1}^{k} r_b c_{ab}$ is the average degree of cluster a, for $a = 1, \ldots, k$. The trivial $\mathbf{0}$ solution of the BP equation is unstable if there are eigenvalues of $\mathbf{N} \otimes \mathbf{T}$ (products of eigenvalues of $\mathbf{N}$ and $\mathbf{T}$) that are greater than 1.

Sufficient condition for the percolation threshold

If $c_1 = \cdots = c_k = c$, then for

$$\lambda(\mathbf{N} \otimes (c\mathbf{T})) = \lambda(\mathbf{N})\,\lambda(\mathbf{RC}) > c$$

it suffices that $\lambda(\mathbf{N}) > \sqrt{c}$, as the eigenvalues of $\mathbf{N}$ and $\mathbf{RC}$ are aligned, see
Bordenave, C., Lelarge, M., Massoulié, L., Non-backtracking spectrum of random graphs: Community detection and non-regular Ramanujan graphs, Ann. Prob. (2018).
They allow „small" fluctuations of the cluster membership proportions that causes the same order of fluctuations in the average degrees of the clusters. For the membership proportion of cluster a, denoted by $r_a^{(n)}$, the assumption

$$\max_{a \in \{1,\ldots,k\}} |r_a^{(n)} - r_a| = O(n^{-\gamma})$$

is made with some $\gamma \in (0, 1]$.

This assumption implies that in the $c_1 = \cdots = c_k = c$ case, $\max_{a\in\{1,\dots,k\}} |c_a^{(n)} - c| = O(n^{-\gamma})$.
They prove that if $\max_a c_a^{(n)} = c + O(n^{-\gamma})$ with some $\gamma \in (0, 1]$, and the relative proportions of the clusters converge, then w.h.p.

$$\mu_i = \nu_i + o(1) \quad (i = 1, \dots k_0) \quad \text{and} \quad \mu_i < \sqrt{c} + o(1) \quad (i > k_0),$$

where μ_i's and ν_i's $(i = 1, \dots, k_0)$ are the structural eigenvalues of **N** and **RC**, respectively, whereas $k_0 \le k$ is the positive integer for which $\nu_i^2 \ge \nu_1$ $(i = 1, \dots k_0)$ and $\nu_{k_0+1}^2 < \nu_1$ holds.
In particular, in the SBM_1 (Erdős–Rényi) model, $\mu_1 = c + o(1)$ and $\mu_2 \le \sqrt{c} + o(1)$.

Even if the average degrees of the clusters are not the same, in the next (Inflation–Deflation) slide we will show that the non-zero eigenvalues of $\bar{\mathbf{A}}$ are the same as those of $\mathbf{RC}$, so they are in the neighborhood of the leading eigenvalues of $\mathbf{N}$ within a factor between u and v, where

$$u = \min_a \frac{c}{c_a^{(n)}} \quad \text{and} \quad v = \max_a \frac{c}{c_a^{(n)}}.$$

However, the leading eigenvalues of $\bar{\mathbf{A}}$ and $\mathbf{A}$ are farther apart, seemingly contradicting to the laws of large numbers.
Also see the theory of deformed Wigner matrices: Capitaine, M., Donati-Martin, C., Féral, D., The largest eigenvalues of finite rank deformation of large Wigner matrices,. . . , Ann. Prob. (2009).

Proposition

The matrix $\bar{\mathbf{A}}$ *has rank* k *and its non-zero eigenvalues (*ν*'s) with unit norm eigenvectors (*$\mathbf{u}$*'s) satisfy* $\bar{\mathbf{A}}\mathbf{u} = \nu\mathbf{u}$*, where* $\mathbf{u}$ *is the inflated vector of* $\tilde{\mathbf{u}} = (u(1), \ldots, u(k))^*$ *with block-sizes* $n_1, \ldots, n_k$*. With the notation* $\mathbf{R} = \frac{1}{n}\mathrm{diag}(n_1, \ldots, n_k) = \mathrm{diag}(r_1, \ldots, r_k)$*, the deflated equation is equivalent to*

$$\mathbf{R}^{\frac{1}{2}}\mathbf{C}\mathbf{R}^{\frac{1}{2}}\mathbf{v} = \nu\mathbf{v},$$

where $\mathbf{v} = \sqrt{n}\mathbf{R}^{\frac{1}{2}}\tilde{\mathbf{u}}$*. Further, if* $\mathbf{u}_1, \ldots \mathbf{u}_k$ *is the set of orthonormal eigenvectors of* $\bar{\mathbf{A}}$*, then* $\mathbf{v}_i = \sqrt{n}\mathbf{R}^{\frac{1}{2}}\tilde{\mathbf{u}}_i$ $(i = 1, \ldots, k)$ *is the set of orthonormal eigenvectors of* $\mathbf{R}^{\frac{1}{2}}\mathbf{C}\mathbf{R}^{\frac{1}{2}}$*. Also,* $\mathbf{R}^{\frac{1}{2}}\mathbf{v}_i = \sqrt{n}\mathbf{R}\tilde{\mathbf{u}}_i$ *are right eigenvectors of* $\mathbf{RC}$ *and* $\mathbf{R}^{-\frac{1}{2}}\mathbf{v}_i = \sqrt{n}\tilde{\mathbf{u}}_i$ *are left eigenvectors of* $\mathbf{RC}$ *with the same eigenvalues* ν_i*, for* $i = 1, \ldots, k$*.*

Deformed Wigner matrices

The (random) adjacency matrix $\mathbf{A}$ of (the random graph) G_n coming from the SBM_k model is $\mathbf{A} = \bar{\mathbf{A}} + \mathbf{E}$, where $\mathbf{E}$ is an appropriate (random) error matrix and all the matrices are $n \times n$ symmetric. We can achieve that the matrix $\mathbf{A}$ contains 1's in the (a, b)-th block with probability p_{ab}, and 0's otherwise. Indeed, for indices $1 \leq a \leq b \leq k$ and $i \in V_a$, $j \in V_b$ let

$$e_{ji} = e_{ij} := \begin{cases} 1 - p_{ab} & \text{with probability} \quad p_{ab} \\ -p_{ab} & \text{with probability} \quad 1 - p_{ab} \end{cases}$$

where e_{ji} (entries of $\mathbf{E}$) be independent random variables. This $\mathbf{E}$ is not a Wigner noise as it does not have a nested structure. However, it is approximately $\frac{1}{\sqrt{n}} \times$Wigner noise, and a „semicircle law" is also valid with radious of constant order:
$2\sigma = 2 \max_{a,b} \sqrt{p_{ab}(1 - p_{ab})} \leq 1$.
Now $\bar{\mathbf{A}}$ is the finite rank (k) perturbation, and if $\lambda_{max}(\bar{\mathbf{A}}) \sim \lambda_{max}(\mathbf{N}) > 1$, then the spectrum of $\mathbf{A}$ is out of the semicircle.

Finding the clusters

Proposition(Based on Theorem 1 of Stephan, L., Massoulié, Non-backtracking spectra of inhomogeneous random graphs, Mathematical Statistics and Learning (2022).)
Let $\mathbb{E}\mathbf{A}$ be the expected adjacency matrix of a random simple graph. Assume that $k = \operatorname{rank}(\mathbb{E}\mathbf{A}) = n^{o(1)}$, the graph is sparse enough, and the eigenvectors corresponding to the non-zero eigenvalues of the matrix $\mathbb{E}\mathbf{A}$ are sufficiently delocalized. Let k_0 denote the number of eigenvalues of $\mathbb{E}\mathbf{A}$ whose absolute value is larger than $\sqrt{\rho}$, where ρ is the spectral radius of $\mathbb{E}\mathbf{A}$: these are $\nu_1 \geq \cdots \geq \nu_{k_0}$ with corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k_0}$ (they form an orthonormal system as $\mathbb{E}\mathbf{A}$ is a real symmetric matrix). Then for $i \leq k_0 \leq k$, the ith largest eigenvalue μ_i of $\mathbf{N}$ is asymptotically (as $n \to \infty$) equals to ν_i and all the other eigenvalues of $\mathbf{N}$ are constrained to the circle (in the complex plane) of center 0 and radius $\sqrt{\rho}$.

Proposition continued (eigenvectors of N)

Further, if $i \leq k_0$ is such that ν_i is a sufficiently isolated eigenvalue of $\mathbb{E}\mathbf{A}$, then the standardized eigenvector of $\mathbf{N}$ corresponding to μ_i has inner product close to 1 with the standardized inflated version of $\mathbf{u}_i$, namely, with $\frac{\mathbf{End}\,\mathbf{u}_i}{\|\mathbf{End}\,\mathbf{u}_i\|}$.
Let $\mathbf{x}$ be a unit-norm eigenvector of $\mathbf{N}$, corresponding to the eigenvalue μ that is close to the eigenvalue ν of the expected adjacency matrix, with corresponding eigenvector $\mathbf{u} \in \mathbb{R}^n$. If our graph is from the SBM_k model, then (without knowing its parameters) we know that $\mathbf{u}$ is a step-vector with at most k different coordinates. Then by the above Proposition,

$$\left\langle \mathbf{x}, \frac{\mathbf{End}\,\mathbf{u}}{\|\mathbf{End}\,\mathbf{u}\|} \right\rangle \geq \sqrt{1-\varepsilon} \geq 1 - \frac{1}{2}\varepsilon,$$

where ε can be arbitrarily „small" with increasing n.

Continued

$$\left\| \mathbf{x} - \frac{\mathbf{End}\,\mathbf{u}}{\|\mathbf{End}\,\mathbf{u}\|} \right\|^2 \leq 2 - 2(1 - \frac{1}{2}\varepsilon) = \varepsilon$$

and by $\mathbf{x}^{in} = \mathbf{End}^*\mathbf{x}$ and $\mathbf{End}^*\mathbf{End} = \mathbf{D_A}$,

$$\left\| \mathbf{End}^*\mathbf{x} - \mathbf{End}^* \frac{\mathbf{End}\,\mathbf{u}}{\|\mathbf{End}\,\mathbf{u}\|} \right\|^2 = \left\| \mathbf{x}^{in} - \mathbf{D_A} \frac{\mathbf{u}}{\|\mathbf{End}\,\mathbf{u}\|} \right\|^2.$$

Consequently,

$$\left\| \mathbf{D_A}^{-1}\mathbf{x}^{in} - \frac{\mathbf{u}}{\|\mathbf{End}\,\mathbf{u}\|} \right\|^2 \leq \|\mathbf{D_A}^{-1}\mathbf{End}^*\|^2\varepsilon \leq \varepsilon$$

as $\|\mathbf{D_A}^{-1}\mathbf{End}^*\|^2 \leq \max_i \frac{1}{d_i} = \frac{1}{\min_i d_i} \leq 1$.

Summarizing

Theorem

Assume that the expected adjacency matrix of the underlying random graph on n nodes and m edges has rank k with k single eigenvalues and corresponding unit-norm eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_k \in \mathbb{R}^n$. Assume that the non-backtracking matrix $\mathbf{N}$ of the random graph has k structural eigenvalues (aligned with those of the expected adjacency matrix) with eigenvectors $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathbb{R}^{2m}$ such that

$$\left\langle \mathbf{x}_j, \frac{\mathbf{End}\,\mathbf{u}_j}{\|\mathbf{End}\,\mathbf{u}_j\|} \right\rangle \geq \sqrt{1-\varepsilon}, \quad j = 1, \dots, k.$$

Then for the transformed vectors $\mathbf{D}_{\mathbf{A}}^{-1}\mathbf{x}_j^{in} \in \mathbb{R}^n$, the relation $\sum_{j=1}^{k} \left\| \mathbf{D}_{\mathbf{A}}^{-1}\mathbf{x}_j^{in} - \frac{\mathbf{u}_j}{\|\mathbf{End}\,\mathbf{u}_j\|} \right\|^2 \leq k\varepsilon$ holds.

Consequences

Corollary: If $\mathbf{u}_j$'s are step-vectors on k steps (e.g., if our graph comes from the SBM_k model), then the k-variance of the node representatives (objective function of the k-means algorithm)

$$(\frac{1}{d_i}x_{1i}^{in},\ldots,\frac{1}{d_i}x_{ki}^{in}),\quad i=1,\ldots,n$$

is estimated from above with $k\varepsilon$ too.

Remark: In case of a simple graph, the n-dimensional vectors $\mathbf{x}_j^{in}$ $(j=1,\ldots,k)$ are the first segments of the right eigenvectors of the matrix $\mathbf{K}$. So, we have to perform the spectral decomposition of a $2n\times 2n$ matrix only instead of a $2m\times 2m$ one, which fact has further computational benefit (except for trees, $n\leq m$, but usually n is much smaller than m).

Non-Backtracking matrix of edge-weighted graphs

Let $G = (V, E)$ be the skeleton of an edge-weighted graph, $|V| = n$, $|E| = m$; the weight of edge $e = \{i, j\}$ is $W_e = w_{ij} = w_{ji} > 0$, where the remaining entries of the the $n \times n$ symmetric edge weight matrix **W** are zeros (including the diagonal).
Let the $2m \times 2m$ diagonal matrix **D** contain the positive edge-weights in its main diagonal (the first m diagonal entries are the same as the second m ones as $W_e = W_{e^{-1}}$). With them,

$$\mathbf{B} = \mathbf{ND} \quad \text{and} \quad \mathbf{B}^* = \mathbf{DN}^*.$$

The general entry of the $2m \times 2m$ non-backtracking matrix **B** is

$$b_{ef} = W_f \delta_{e \to f} \delta_{f \neq e^{-1}}.$$

Notation

We assume that there are constants C_1 and C_2 (independent of n):

$$C_1 \leq w_{ij} \leq C_2, \quad \text{for} \quad w_{ij} \neq 0.$$

Further, we assume that the skeleton's node degrees

$$d_i = |\{i : w_{ij} > 0, \quad j = 1, \ldots, n\}|, \quad i = 1, \ldots, n$$

are of constant order (it is the case in the k-cluster stochastic block model (SBM_k)).
Let $\mathbf{D}^{\mathbf{W}}$ denote the $n \times n$ diagonal matrix of diagonal entries

$$d_i^{\mathbf{W}} = \sum_{j=1}^{n} w_{ij}, \quad i = 1, \ldots, n,$$

that are the so-called generalized degrees. In the unweighted case (0-1 weights), $d_i^{\mathbf{W}} = d_i$ and $C_1 = C_2 = 1$; in general,

$$C_1 d_i \leq d_i^{\mathbf{W}} \leq C_2 d_i, \quad i = 1, \ldots, n.$$

Start- and End-matrices, in- and out-vectors

The **End** and **Start** matrices are defined as in the unweighted case:

$$\mathbf{End}^* \, \mathbf{D} \, \mathbf{End} = \mathbf{Start}^* \, \mathbf{D} \, \mathbf{Start} = \mathbf{D}^{\mathbf{W}} \quad \text{and} \quad \mathbf{Start}^* \, \mathbf{D} \, \mathbf{End} = \mathbf{W}.$$

For any vector $\mathbf{x} \in \mathbb{R}^{2m}$, the following n-dimensional vectors are introduced:

$$\mathbf{x}^{out} := \mathbf{Start}^* \mathbf{D} \mathbf{x} \quad \text{and} \quad \mathbf{x}^{in} := \mathbf{End}^* \mathbf{D} \mathbf{x}.$$

Coordinatewise, for $i = 1, \ldots, n$,

$$x_i^{out} = \sum_{j \colon j \sim i} w_{ij} x_{i \to j} = \sum_{e \colon e_1 = i} W_e x_e, \quad x_i^{in} = \sum_{j \colon j \sim i} w_{ij} x_{j \to i} = \sum_{e \colon e_2 = i} W_e x_e.$$

Tracing back the problem to lower order matrices

No counterpart of matrix $\mathbf{K}$ works here, but if we know a real eigenvalue μ of $\mathbf{B}$, we are able to find a linear system of equations for the *out*-transform of the corresponding eigenvector that is necessary for spectral clustering. With a Lapacian type equation, μ can also be concluded.

Proposition Let $\mathbf{x}$ be a (right) eigenvector of $\mathbf{B}$ corresponding to a single positive real eigenvalue μ such that $\mu \neq w_{ij}$, $\forall i, j \in \{1, \ldots, n\}$. Then $\mathbf{y} = \mathbf{x}^{out}$ satisfies the homogeneous system of linear equations

$$[\mathbf{I}_n - \tilde{\mathbf{A}}(\mu) + \tilde{\mathbf{D}}(\mu)]\mathbf{y} = \mathbf{0}$$

with a Laplacian type coefficient matrix, where

$$\tilde{\mathbf{A}}(\mu)_{ij} = \frac{\mu w_{ij}}{\mu^2 - w_{ij}^2} \quad \text{and} \quad \tilde{\mathbf{D}}(\mu)_{ii} = \sum_{j=1}^{n} \frac{w_{ij}^2}{\mu^2 - w_{ij}^2},$$

with the understanding that $w_{ij} = 0$ whenever $i \not\sim j$.

Proof of the Proposition

If $\mathbf{x}$ is a (right) eigenvector of $\mathbf{B}$ with corresponding (real) eigenvalue μ, then

$$\mu x_e = \sum_{e \to f,\, f \neq e^{-1}} W_f x_f = \sum_{f : f_1 = e_2} W_f x_f - W_{e^{-1}} x_{e^{-1}} = y_{e_2} - W_e x_{e^{-1}}.$$

Likewise,

$$\mu x_{e^{-1}} = \sum_{e^{-1} \to f,\, f \neq e} W_f x_f = \sum_{f : f_1 = e_1} W_f x_f - W_e x_e = y_{e_1} - W_e x_e.$$

From here,

$$\mu^2 x_e = \mu y_{e_2} - \mu W_e x_{e^{-1}} = \mu y_{e_2} - W_e y_{e_1} + W_e^2 x_e,$$

and so,

$$x_e = \frac{\mu y_{e_2} - W_e y_{e_1}}{\mu^2 - W_e^2}$$

which shows that $\mathbf{y} \neq \mathbf{0}$ as $\mathbf{x} \neq 0$.

Proof continued

Substituting this formula for x_e in the original equation, we get that for any edge $e = \{j \to i\}$,

$$\frac{\mu^2 y_i - \mu w_{ij} y_j}{\mu^2 - w_{ij}^2} = \sum_{l : l \sim i,\, l \neq j} w_{li} \frac{\mu y_l - w_{li} y_i}{\mu^2 - w_{li}^2}.$$

Further developing,

$$\frac{\mu^2 y_i}{\mu^2 - w_{ij}^2} - \frac{\mu w_{ij} y_j}{\mu^2 - w_{ij}^2} = \sum_{l : l \sim i} \frac{\mu w_{li}}{\mu^2 - w_{li}^2} y_l - \sum_{l : l \sim i} \frac{w_{li}^2}{\mu^2 - w_{li}^2} y_i - w_{ji} \frac{\mu y_j - w_{ji} y_i}{\mu^2 - w_{ji}^2},$$

which provides

$$y_i = \frac{\mu^2 y_i}{\mu^2 - w_{ij}^2} - \frac{w_{ij}^2 y_i}{\mu^2 - w_{ij}^2} = \sum_{l : l \sim i} \frac{\mu w_{li}}{\mu^2 - w_{li}^2} y_l - \sum_{l : l \sim i} \frac{w_{li}^2}{\mu^2 - w_{li}^2} y_i.$$

Consequences

The above homogeneous system of linear equations for the coordinates of $\mathbf{y}$ must have a non-trivial solution, so

$$|\mathbf{I}_n - \tilde{\mathbf{A}}(\mu) + \tilde{\mathbf{D}}(\mu)| = 0.$$

This is not a polynomial (characteristic) equation, but it is a rational function of μ. By the assumptions of the Proposition, the denominators are not zeros, so we can multiply the determinant equations with them, and we obtain an high-degree (higher than n) polynomial of μ.
The leading positive real solutions $\mu_1 \geq \cdots \geq \mu_k$ are the same as the structural eigenvalues of $\mathbf{B}$. Their number will be denoted by k. The corresponding $\mathbf{y}_1, \ldots, \mathbf{y}_k$ can be obtained by solving the system of the above homogeneous linear equations (with only an $n \times n$ coefficient matrix).

More general setup of Stephan and Massoulié

The weighted adjacency matrix is biult up as follows: $a_{ji} = a_{ij} \sim Bernoulli(p_{ij})$ for $1 \leq i \leq j$ independently of each other, where the $n \times n$ symmetric probability matrix $\mathbf{P}$ is a parameter (this forms the skeleton of the graph). Independently of a_{ij}s, $\mathbf{W}$ is an $n \times n$ symmetric weight matrix of independent random entries. Then an inhomogeneous undirected random graph $G = (V, E)$ is associated with the couple $(\mathbf{P}, \mathbf{W})$ such that each edge ij in the skeleton (randomized according to $\mathbf{P}$) holds weight W_{ij}.

In the SBM_k model the weights were constantly 1, and in the edge-weighted case we used a deterministic $\mathbf{W}$, where the ij entry of the weighted adjacency matrix was $w_{ij} \times Bernoulli(p_{ij})$ for $1 \leq i \leq j \leq n$; otherwise it was symmetric.

General setup with the couple $(\mathbf{P}, \mathbf{W})$

Define

$$\mathbf{Q} := \mathbf{P} \circ \mathbb{E}\mathbf{W} \quad \text{and} \quad \mathbf{K} := \mathbf{P} \circ \mathbb{E}(\mathbf{W} \circ \mathbf{W}).$$

Then $\mathbf{Q} \approx \mathbb{E}\mathbf{A}$ (diagonal negligible) and $\mathbf{K} \approx \mathrm{Var}\mathbf{A}$.

Proposition(Based on Theorem 1 of Stephan, L., Massoulié, Non-backtracking spectra of inhomogeneous random graphs, Mathematical Statistics and Learning (2022).)

Assume that $\max W_{ij} \leq 1$ and $k = \mathrm{rank}\,(\mathbf{Q}) = n^{o(1)}$, the graph is sparse enough, and the eigenvectors corresponding to the non-zero eigenvalues of the matrix $\mathbf{Q}$ are sufficiently delocalized. Let k_0 denote the number of eigenvalues of $\mathbf{Q}$ whose absolute value is larger than $\sqrt{\rho}$, where $\rho \geq 1$ is the spectral radius of $\mathbf{K}$: these are $\nu_1 \geq \cdots \geq \nu_{k_0}$ with corresponding eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{k_0}$ (they form an orthonormal system as $\mathbf{Q}$ is a real symmetric matrix). Then for $i \leq k_0 \leq k$, the ith largest eigenvalue μ_i of $\mathbf{B}$ is asymptotically (as $n \to \infty$) equals to ν_i and all the other eigenvalues of $\mathbf{B}$ are constrained to the circle (in the complex plane) of center 0 and radius $\sqrt{\rho}$.

Proposition continued (eigenvectors of B)

Further, if $i \leq k_0$ is such that ν_i is a sufficiently isolated eigenvalue of $\mathbf{Q}$, then the standardized eigenvector of $\mathbf{B}$ corresponding to μ_i has inner product close to 1 with the standardized inflated version of $\mathbf{u}_i$, namely, with $\frac{\mathbf{End}\,\mathbf{u}_i}{\|\mathbf{End}\,\mathbf{u}_i\|}$.

Application: Let $\mathbf{x}$ be a unit-norm eigenvector of $\mathbf{B}$, corresponding to the eigenvalue μ that is close to the eigenvalue ν of the matrix $\mathbf{Q}$, with corresponding eigenvector $\mathbf{u} \in \mathbb{R}^n$. If our graph is from the SBM_k model, then (without knowing its parameters) we know that $\mathbf{u}$ is a step-vector with at most k different coordinates. Then by the above Proposition,

$$\left\langle \mathbf{x}, \frac{\mathbf{End}\,\mathbf{u}}{\|\mathbf{End}\,\mathbf{u}\|} \right\rangle \geq \sqrt{1-\varepsilon} \geq 1 - \frac{1}{2}\varepsilon,$$

where ε can be arbitrarily „small" with increasing n.

Continued

$$\left\| \mathbf{x} - \frac{\mathbf{End}\,\mathbf{u}}{\|\mathbf{End}\,\mathbf{u}\|} \right\|^2 \leq 2 - 2(1 - \frac{1}{2}\varepsilon) = \varepsilon$$

and

$$\left\| \mathbf{x}^{out} - \mathbf{W}\frac{\mathbf{u}}{\mathbf{Endu}} \right\|^2 = \left\| \mathbf{Start}^* \mathbf{D} \left(\mathbf{x} - \frac{\mathbf{Endu}}{\|\mathbf{Endu}\|} \right) \right\|^2 \leq \|\mathbf{Start}^* \mathbf{D}\|^2 \varepsilon.$$

Consequently,

$$\left\| \mathbf{W}^{-1}\mathbf{x}^{out} - \frac{\mathbf{u}}{\|\mathbf{End}\,\mathbf{u}\|} \right\|^2 \leq \|\mathbf{W}^{-1}\mathbf{Start}^* \mathbf{D}\|^2 \varepsilon \leq \left(\frac{C_2}{C_1} \right)^2 \varepsilon.$$

Consequences

Assume that the non-backtracking matrix $\mathbf{B}$ has k structural (real) eigenvalues with eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_k \in \mathbb{R}^{2m}$. Then for the transformed vectors $\mathbf{W}^{-1}\mathbf{x}_j^{out} \in \mathbb{R}^n$, the relation

$$\sum_{j=1}^{k} \left\| \mathbf{W}^{-1}\mathbf{x}_j^{out} - \frac{\mathbf{u}_j}{\|\mathbf{End}\,\mathbf{u}_j\|} \right\|^2 \leq k\varepsilon \frac{C_2^2}{C_1^2}$$

holds.
If $\mathbf{u}_j$'s are step-vectors on k steps (e.g., if our graph comes from the SBM_k model), then the left-hand side estimates from above the k-variance of the node representatives that are row vectors of

$$(\mathbf{W}^{-1}\mathbf{x}_1^{out}, \ldots, \mathbf{W}^{-1}\mathbf{x}_k^{out}).$$

To get the $\mathbf{x}_j^{out} \in \mathbb{R}^n$ vectors we do not need the $2m$-dimensional eigenvectors $\mathbf{x}_j$'s of $\mathbf{N}$, but the previous calculations can be used.

Special unweighted cases and β-percolation

In the most special „symmetric case", the transition matrix is $\mathbf{T} = \mathbf{GRC} = \frac{1}{ck}\mathbf{C}$, where $\mathbf{C}$ contains c_{in} in its main diagonal and c_{out}, otherwise; see the BP method. $\mathbf{T}$ is a stochastic matrix, so its largest eigenvalue is 1 with corresponding eigenvector $\mathbf{1}$ (the all 1's vector). The other eigenvalue is

$$\lambda = \frac{c_{in} - c_{out}}{kc}$$

with multiplicity $k - 1$. In the assortative case, $\lambda > 0$; further, $\lambda < 1$, as in the symmetric case

$$c = \frac{c_{in} + (k-1)c_{out}}{k}$$

holds. Consequently, the eigenvalues of $\mathbf{RC} = c\mathbf{T}$ are c and $c\lambda$, latter one has multiplicity $k - 1$.

With the BP method for this special case, the eigenvalues of $\mathbf{N} \otimes \mathbf{T}$ greater than 1 are considered (giving a non-trivial solution) that boils down to the condition $c\lambda^2 > 1$, which gives the Kesten–Stigum threshold $|c_{in} - c_{out}| > k\sqrt{c}$.

The SBM_k^β model: edges are retained with prob. β

The $k \times k$ probability matrix is $\frac{\beta \mathbf{C}}{n}$: $\mathbf{C}$ and c is multiplied by β, but $\mathbf{T} = \mathbf{GRC}$ remains unchanged.
So we consider $\beta c\mathbf{T}$ as for the model side, but the underlying graph and its $\mathbf{N}$ is the same as before. Therefore, the eigenvalues of $\mathbf{N} \otimes \beta c\mathbf{T} = \beta c(\mathbf{N} \otimes \mathbf{T})$ should be greater than c if a non-stable solution is required:

$$\beta\, \lambda(\mathbf{N} \otimes c\mathbf{T}) > c.$$

If the eigenvalues of $\mathbf{N}$ and $c\mathbf{T}$ are aligned, then this gives that $\lambda(\mathbf{N}) > \frac{\sqrt{c}}{\sqrt{\beta}}$ is needed for detectability; equivalently,

$$\beta > \frac{c}{\lambda^2(\mathbf{N})} = \left(\frac{\sqrt{c}}{\lambda(\mathbf{N})}\right)^2.$$

Remarks

This is in accord with the fact, that in the $k = 1$ case, in the Erdős–Rényi model, when the largest eigenvalue of $\mathbf{N}$ is around c, then $\beta = \frac{c}{\mu_1^2} = \frac{1}{\mu_1}$ is the percolation threshold, see Newman, M. E. J., Message passing methods on complex networks, Proc. R. Soc. London A (2023).

In the multiclass scenario, $\frac{c}{\mu_i^2}$ are further phase transitions, leading to i clusters, for $i = 1, \ldots, k_0$ until $\mu_{k_0} \geq \sqrt{c}$, but $\mu_{k_0+1} < \sqrt{c}$.

Also note that this has relevance only if $\lambda_{max}(\mathbf{N}) > \sqrt{c}$, so eigenvalues of $\mathbf{N}$ greater than $\sqrt{c}$ give the phase transitions.

Since $\mu_1 \geq \mu_2 \geq \ldots$, with larger β, larger number of clusters can be detected.

The SBM_2 symmetric model

$r_1 = r_2 = \frac{1}{2}$, $c_1 = c_2 = \frac{c_{in}+c_{out}}{2} = c$. In the assortative case,

$$c_{in} - c_{out} = |c_{in} - c_{out}| > 2\sqrt{c} = \sqrt{2}\sqrt{c_{in} + c_{out}}.$$

If $\mathbf{C}$ and c are multiplied with β, we get $\beta(c_{in} - c_{out}) > \sqrt{2}\sqrt{\beta(c_{in} + c_{out})}$. This means that

$$c_{in} - c_{out} > \sqrt{2}\frac{\sqrt{c_{in} + c_{out}}}{\sqrt{\beta}}. \tag{1}$$

Since $\beta < 1$, the right hand side gives a higher lower threshold than $\beta = 1$. This is also equivalent to

$$\beta > \frac{2(c_{in} + c_{out})}{(c_{in} - c_{out})^2},$$

which makes sense if $\frac{2(c_{in}+c_{out})}{(c_{in}-c_{out})^2} < 1$, i.e., if $c_{in} - c_{out} > 2\sqrt{c} = \sqrt{2}\sqrt{c_{in} + c_{out}}$. So, until equality is attained in (1), additional β-percolation can give detectable clusters too.

When $c_1 = \cdots = c_k = c$

Then $\mu_1 = c$, $c\mathbf{T} = \mathbf{RC}$, and its eigenvalues are closely aligned with the eigenvalues $\mu_2, \ldots, \mu_k$ of $\mathbf{N}$.
But $c_1 = \cdots = c_k = c$ is only approximately holds when $n \to \infty$.
Even then, the approximate bound $\beta > \frac{c}{\mu_i^2}$ has a leverage as k is increased.
In the c_{in} versus c_{out} scenario, $c_1 = \cdots = c_k = c$ is equivalent to $r_1 = \cdots = r_k$, and so,

$$\beta|c_{in} - c_{out}| > k\sqrt{\beta c}, \quad |c_{in} - c_{out}| > k\frac{\sqrt{c}}{\sqrt{\beta}}$$

which allows more detectable clusters if β is increased, see Figures.
$SBM_k^{\mathbf{W}}$ is a generalization of the SBM_k^{β} model, where the edges may have different edge-retention probability (w_{ij} for the connected node-pair i, j). Then the eigenvalues of $\mathbf{B} = \mathbf{ND}$ are used.