No going back: eigenvalues and eigenvectors of the non-backtracking matrix and Laplacian; their usage for sparse spectral clustering

Marianna Bolla
BME Math. Inst. marib@math.bme.hu

Probability Seminar at the ELTE Prob. Stat. Dept.

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

October 17, 2025.

Outline

- Non-backtracking matrix of simple graphs.
- Sparse stochastic block model.
- Belief propagation.
- Inflation–deflation.
- k-means clustering with node representatives.
- Non-backtracking random walk, transition probability matrix and Laplacian.
- Bond percolation for simulated data.
- Application for real-world data.

Preliminaries

DENSE	SPARSE
Füredi-Komlós, Combinatorica (1981): $\mathbb{E}(a_{ij}) = \mu > 0 \quad (i \neq j)$	Percolated $G_n(p)$, $p = \frac{c}{n}$ $(n \to \infty)$ Erdős-Rényi
B, Wiley (2013): $\mathbb{E}(a_{ij}) = c_{ab}$ $(i \in V_a, j \in V_b)$, $1 \leq a, b \leq k$	Percolated SBM: $\mathbb{E}(a_{ij}) = \frac{c_{ab}}{n}$ $(i \in V_a, j \in V_b)$, $1 \leq a, b \leq k$
$\lambda_i(\mathbf{A})$ is aligned with $\lambda_i(\mathbb{E}\mathbf{A})$, $i = 1, \ldots, k \quad$ (F-K: $k = 1$)	$\lambda_i(\mathbf{B})$ is aligned with $\lambda_i(\mathbb{E}\mathbf{A})$, $i = 1, \ldots, k$ non-backtracking
Perturbation: Wigner-type matrix Subspace perturbation: Davis-Kahan	deformed Wigner matrix Bauer–Fike

Non-Backtracking (Hashimoto) matrix of simple graphs

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

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

where $e = \{i \to j\}$ and $f = \{s \to l\}$ are oriented edges, and $e \to f$ with $e = (e_1, e_2)$ and $f = (f_1, f_2)$ means that $e_2 = f_1$;
$e^{-1} = \{j \to i\}$.
Historically, it is $\mathbf{B}^T$ that is called non-backtracking matrix.

Relation to line-graphs

Proposition

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

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

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

In Lovász, Combinatorial Exercises: 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. But two graphs are isomorphic $\iff$ their non-backtracking matrices are the same (after relabeling).

Transpose (*), involution, and swapping

Though $\mathbf{B}$ 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: $b^*_{ef} = b_{e^{-1}\,f^{-1}}$.

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 $\mathbb{R}^{2m}$ ($\mathbf{V} = \mathbf{V}^* = \mathbf{V}^T$, $\mathbf{V}^2 = \mathbf{I}$): $\mathbf{V} = \begin{pmatrix} \mathbf{O} & \mathbf{I}_m \\ \mathbf{I}_m & \mathbf{O} \end{pmatrix}$. Then $\mathbf{V}\mathbf{x} = \check{\mathbf{x}}$ and $\mathbf{V}\check{\mathbf{x}} = \mathbf{x}$;
$\mathbf{B}^* = \mathbf{B}^T = \mathbf{V}\mathbf{B}\mathbf{V}$, $\mathbf{B}\mathbf{V}$ and $\mathbf{V}\mathbf{B}$ are symmetric (PT-invariance).

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

Eigenvalues of B (Ihara formula)

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

Ihara formula: **B** 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 $\lambda_{max}(\mathbf{B}) = \lambda_{max}(\mathbf{K}) \leq \lambda_{max}(\mathbf{A})$.

Real eigenvalues and eigenvectors of B (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.

Deflation

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

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

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

Calculation

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

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 $\mathbf{B}^*$ with (real) eigenvalue μ, i.e., $\mathbf{B}^*\mathbf{x} = \mu\mathbf{x}$, then $(\mathbf{B}^*\mathbf{x})^{out} = (\mu\mathbf{x})^{out} = \mu\mathbf{x}^{out}$ and $(\mathbf{B}^*\mathbf{x})^{in} = (\mu\mathbf{x})^{in} = \mu\mathbf{x}^{in}$. Therefore,

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

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 $\mathbf{B}^*$ 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,

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

According to the previous remarks, the vector $\mathbf{x}$ is a left eigenvector, and $\check{\mathbf{x}}$ is a right eigenvector of $\mathbf{B}$ with the same (real) eigenvalue. For 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.

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 $\mathbf{B}$.

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

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

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

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

is unstable, if the spectral radius of the big block matrix $\mathbf{B} \otimes \mathbf{T}$ 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 $\varepsilon = (\mathbf{B} \otimes \mathbf{T})\varepsilon$, where ε is a $2mk$-dimensional vector and the $2mk \times 2mk$ matrix of the linear system is $\mathbf{B} \otimes \mathbf{T}$. Here $\mathbf{B}$ 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 $\mathbf{B} \otimes \mathbf{T}$ (products of eigenvalues of $\mathbf{B}$ and $\mathbf{T}$) that are greater than 1.

Sufficient condition for the percolation threshold

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

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

it suffices that $\lambda(\mathbf{B}) > \sqrt{c}$, as the eigenvalues of $\mathbf{B}$ 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 **B** 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 $\mathbf{B}$ 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 $\lambda_{max}(\bar{\mathbf{A}}) \sim \lambda_{max}(\mathbf{B}) > 1$, 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 $\mathbf{B}$ 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 $\mathbb{E}\mathbf{A}$, 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 $\frac{\mathbf{End}\,\mathbf{u}_i}{\|\mathbf{End}\,\mathbf{u}_i\|}$.
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 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.

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

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 $\mathbf{B}$ 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 transition probability matrix and Laplacian

If G is connected with $d_{min} \geq 2$, then for all eigenvalues of $\mathbf{B}$, $|\mu| \geq 1$ holds. In particular, if $d_{min} > 2$, then the eigenvalues of $\mathbf{B}$ with $|\mu| = 1$ are ± 1's.

If G is a connected graph that is not a cycle and $d_{\min} \geq 2$, then $\mathbf{B}$ is irreducible. Therefore, the Frobenius theorem is applicable to $\mathbf{B}$, and under the above conditions, it has a single positive real eigenvalue among its maximum absolute value ones with corresponding eigenvector of all positive real coordinates. It is also the spectral radius $\rho(\mathbf{B})$ of $\mathbf{B}$.

Notation

The non-backtracking graph $\mathcal{G}$ corresponding to the simple graph G is a special directed graph, with $2m \times 2m$ adjacency matrix $\mathbf{B}$, obeying the PT-invariance. The row-sums of $\mathbf{B}$ are put in the $2m \times 2m$ diagonal matrix $\mathcal{D}_{row}$.
The diagonal entries of the row-sums of $\mathbf{B}^T$, or equivalently, those of the column-sums of $\mathbf{B}$ are contained in the diagonal matrix $\mathcal{D}_{col}$; by the PT-invariance, $\mathcal{D}_{col} = \mathbf{V}\mathcal{D}_{row}\mathbf{V}$; so $\mathrm{diag}(\mathcal{D}_{row})$ and $\mathrm{diag}(\mathcal{D}_{col})$ are swappings of each other.
The diagonal entry of $\mathcal{D}_{row}$, corresponding to the oriented edge $[i,j]$ is $d_j - 1$. Since it has multiplicity d_j, the number of edges in the non-backtracking graph is $\sum_{j=1}^{n} d_j(d_j-1) = \sum_{j=1}^{n} d_j^2 - 2m$.

Non-backtracking random walk

The non-backtracking random walk on the original graph is not Markovian (has the memory that there is no way back in the next step), but it is Markovian on the graph of the directed edges with transition probability matrix $\mathcal{T} := \mathcal{D}_{row}^{-1}\mathbf{B}$. It means that the probability of going from the oriented edge e to the oriented edge f is $\mathrm{Prob}\,(e \to f) = \frac{1}{d_e} b_{ef}$.

It is 0 if $f = e^{-1}$ or if the end-node of e is not the start-node of f; otherwise, it is $\frac{1}{d_e}$, where d_e is the diagonal entry of $\mathcal{D}_{row}$, corresponding to the oriented edge $e = [i, j]$, i.e., it is $d_j - 1$.

This random walk on the oriented edges is already Markovian, as going back is prohibited on the non-backtracking graph (a forbidden transition corresponds to a 0 entry of $\mathbf{B}$, so it has 0 probability).

The non-backtracking Laplacian generates the random walk: $\mathcal{L} = \mathbf{I}_{2m} - \mathcal{T}$. Its eigenvalues are 1 minus the eigenvalues of $\mathcal{T}$ with the same eigenvectors.

Spectral properties of $\mathcal{L}$ and $\mathcal{T}$

It is known that 0 is an eigenvalue of $\mathbf{B}$ if and only if G contains nodes of degree one (for example, if it is a tree). Therefore, if $d_{min} \geq 2$, then 1 cannot be an eigenvalue of $\mathcal{L}$, and the spectral gap of the eigenvalues of $\mathcal{L}$ from 1 is investigated in Jost et al., Discrete Math. 2023, and it is bounded from below by $\frac{1}{d_{max}-1}$ and proved that this bound is sharp (attained at regular and certain circle graphs).

The parity time symmetry for $\mathcal{T}$ is also true: $\mathcal{T}\mathbf{V}$ and $\mathbf{V}\mathcal{T}$ are symmetric; further, $\mathcal{T}^* = \mathcal{T}^T = \mathbf{V}\mathcal{T}\mathbf{V}$. Because the diagonal entries of $\mathcal{D}_{row}$ are the numbers $d_i - 1$, to have the inverse we assume that $d_i \geq 2$ in the original graph, and to preserve the number of connected components in G and $\mathcal{G}$ we assume that G is not the cycle graph.

In Bauer it is proved that the eigenvalues of $\mathcal{L}$ are contained in the complex disc of center 1 and radius 1. In particular, its real eigenvalues are in the [0,2] interval. Analogously, the eigenvalues of $\mathcal{T}$ are in the complex disc of center 0 and radius 1; the real ones are in the $[-1, 1]$ interval. The number 2 is an eigenvalue of $\mathcal{L}$, or equivalently -1 is an eigenvalue of $\mathcal{T}$ if and only if the underlying simple graph is bipartite. Furthermore, 0 is always an eigenvalue of $\mathcal{L}$, and its multiplicity is equal to the number of the connected components of $\mathcal{G}$, which is the same as the number of the connected components of G, whenever none of them is the cycle graph.

Observe that $\mathcal{L}$ rather resembles the normalized Laplacian that is in the simple graph case $\mathbf{L_D} = \mathbf{I}_n - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$, where $\mathbf{A}$ is the adjacency and $\mathbf{D}$ is the diagonal degree matrix, and it has the same eigenvalues as $\mathbf{I}_n - \mathbf{D}^{-1}\mathbf{A}$. In the present, unsymmetric situation there are complex eigenvalues too and we distinguish between right and left eigenvectors as follows.

Right and left eigenvectors in general

Let $\mathbf{A}$ be $n \times n$ with possible complex entries, but diagonalizable and irreducible. $\mathbf{u}$ is right eigenvector of $\mathbf{A}$ with eig.val. λ if $\mathbf{Au} = \lambda\mathbf{u}$. $\mathbf{v}$ is left eigenvector of $\mathbf{A}$ with the same eig.val. λ if $\mathbf{v}^*\mathbf{A} = \lambda\mathbf{v}^*$ (equivalently, $\mathbf{A}^*\mathbf{v} = \bar{\lambda}\mathbf{v}$). If $\mathbf{A}$ is diagonalizable then

$$\mathbf{A} = \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{v}_i^* = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{-1},$$

where $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ columnwise and $\mathbf{U}^{-1}$ contains the vectors $\mathbf{v}_i^*$ rowvise.
Note that the matrices $\mathbf{u}_i\mathbf{v}_i^*$ in the above dyadic decomposition are (skew) projections (idempotent) as the right- and left eigenvectors form a biorthonormal system: $\mathbf{v}_i^*\mathbf{u}_j = \mathbf{u}_j^*\mathbf{v}_i = \delta_{ij}$ and $\mathbf{v}_i^*\mathbf{A}\mathbf{u}_j = \lambda_i\delta_{ij}$.

Eigenvalues, left and right eigenvectors of $\mathcal{T}$

$\mathcal{T}$: eigenval.	$\lambda \in \mathbb{R}$	$\lambda \in \mathbb{C}$,	$\bar{\lambda} \neq \lambda$
right eigenvect.	$\mathbf{z} \in \mathbb{R}^{2m}$	$\mathbf{z} \in \mathbb{C}^{2m}$,	$\bar{\mathbf{z}} \in \mathbb{C}^{2m}$
left eigenvect.	$-\frac{1}{\lambda}\check{\mathbf{z}} \in \mathbb{R}^{2m}$	$c\check{\bar{\mathbf{z}}} = c\bar{\check{\mathbf{z}}}$,	$c\check{\mathbf{z}} \in \mathbb{C}^{2m}$

$\mathcal{T}^T$: eigenval.	$\lambda \in \mathbb{R}$	$\bar{\lambda} \in \mathbb{C}$,	$\lambda \neq \bar{\lambda}$
right eigenvect.	$-\frac{1}{\lambda}\check{\mathbf{z}} \in \mathbb{R}^{2m}$	$c\check{\bar{\mathbf{z}}} = c\bar{\check{\mathbf{z}}}$,	$c\check{\mathbf{z}} \in \mathbb{C}^{2m}$
left eigenvect.	$\mathbf{z} \in \mathbb{R}^{2m}$	$\mathbf{z}$,	$\bar{\mathbf{z}} \in \mathbb{C}^{2m}$

Important: If $\mathbf{z}$ is an eigenvector of $\mathcal{T}$ corresponding to an eigenvalue not 1, then $\mathbf{z} \perp \mathbf{1}$, where $\mathbf{1}$ is the eigenvector corresponding to the eigenvalue 1.
Note that usually there is no orthogonality of the eigenvectors ($\mathcal{T}$ is not a normal matrix in general).

More about these eigenvectors

Theorem

Assume that $\mathbf{B}$ *is irreducible and diagonalizable. Then the eigenvalues of* $\mathcal{T} = \mathcal{D}_{row}^{-1}\mathbf{B}$ *are allocated within the closed circle of center* $\mathbf{0}$ *and radius 1 of the complex plane* $\mathbf{C}$*, and 1 is a single real eigenvalue. Furthermore, the right eigenvectors* $\mathbf{z}_i$*'s corresponding to the positive real eigenvalues* λ_i*'s of* $\mathcal{T}$ *can be normalized so that they form a* $\mathcal{D}_{row}$*-orthonormal system. The left eigenvectors* $\check{\mathbf{z}}_i$*s (with the same positive real eigenvalues) form a* $\mathcal{D}_{col}$*-orthonormal system at the same time; i.e.,* $\mathbf{z}_i^T \mathcal{D}_{row} \mathbf{z}_j = \check{\mathbf{z}}_i^T \mathcal{D}_{col} \check{\mathbf{z}}_j = \delta_{ij}$ *for* i, j *such that* $\lambda_i > 0$ *and* $\lambda_j > 0$ *are positive reals. (Note that we use the notation* $\mathbf{z}_i^T$ *as the eigenvectors, corresponding to real eigenvalues of a matrix of real entries, can also have real coordinates.) Further,* $\mathbf{w}_j = -\frac{1}{\lambda_j}\check{\mathbf{z}}_j$ *is the left eigenvector of* $\mathcal{T}$ *such that* $\mathbf{z}_i^T \mathbf{w}_j = \delta_{ij}$ *for* $i, j = 1, \ldots, k$*, where* k *is the number of real eigenvalues of* $\mathcal{T}$*.*

Proof (sketch)

The transition probability matrix $\mathcal{T}$ is a doubly stochastic matrix as its row-sums are 1s; further, its transpose

$$(\mathcal{D}_{row}^{-1}\mathbf{B})^T = \mathbf{B}^T\mathcal{D}_{row}^{-1} = \mathcal{D}_{col}^{-1}\mathbf{B}^T$$

is also a stochastic matrix (it is the transition probability matrix of the reversed random walk along to the inverses of the oriented edges). So the largest modulus real eigenvalue of both $\mathcal{T}$ and $\mathcal{T}^T$ is 1 with eigenvector $\mathbf{1}$. This also means that the stationary distribution of the corresponding ergodic Markov chain is uniform. The left and right eigenvectors, corresponding to the (same) real eigenvalues of $\mathcal{T}$, form a bi-orthogonal system as the matrix $\mathcal{T}$ is diagonalizable.
Therefore, 1 is a single real eigenvalue of the irreducible matrix $\mathcal{D}_{row}^{-1}\mathbf{B}$ of nonnegative entries, by the Frobenius theorem; also, the moduli of the other (possibly complex) eigenvalues are at most 1.

A right eigenvector $\mathbf{z}$ with eigenvalue λ of $\mathcal{T}$ satisfies the equation

$$\mathcal{D}_{row}^{-1}\mathbf{B}\mathbf{z} = \lambda\mathbf{z}.$$

We will use the special structure of $\mathbf{B}$ when we consider a positive real eigenvalue λ and corresponding right eigenvector $\mathbf{z}$ of $\mathcal{T}$. By $\mathbf{V}^2 = \mathbf{I}_{2m}$, $\mathbf{V}^T = \mathbf{V}$, this is equivalent to

$$(\mathbf{V}\mathcal{D}_{row}^{-1}\mathbf{V})(\mathbf{V}\mathbf{B}\mathbf{V})(\mathbf{V}\mathbf{z}) = \lambda(\mathbf{V}\mathbf{z}),$$

so

$$\mathcal{D}_{col}^{-1}\mathbf{B}^T\check{\mathbf{z}} = \lambda\check{\mathbf{z}}.$$

Consequently, if $\mathbf{z}$ is a right eigenvector of $\mathcal{D}_{row}^{-1}\mathbf{B}$ with the real eigenvalue λ, then $\check{\mathbf{z}}$ is a right eigenvector of $\mathcal{D}_{col}^{-1}\mathbf{B}^T$, with the same real eigenvalue λ; and it is also a left eigenvector of $\mathcal{D}_{row}^{-1}\mathbf{B}$ with the real eigenvalue λ.

continued with complex λ's that are not real

Now $\bar{\mathbf{z}}$ is a left eigenvector of $\mathcal{T}$ with the same eigenvalue. Indeed,

$$\mathcal{T}^*\bar{\mathbf{z}} = \mathcal{T}^T\bar{\mathbf{z}} = \mathcal{D}_{col}^{-1}\mathbf{B}^T\bar{\mathbf{z}} = (\mathbf{V}\mathcal{D}_{row}^{-1}\mathbf{V})(\mathbf{V}\mathbf{B}\mathbf{V})\bar{\mathbf{z}} = \mathbf{V}(\mathcal{D}_{row}^{-1}\mathbf{B}\bar{\mathbf{z}}) =$$

$$= \mathbf{V}\overline{\mathcal{D}_{row}^{-1}\mathbf{B}\mathbf{z}} = \mathbf{V}\overline{\lambda\mathbf{z}} = \mathbf{V}\bar{\lambda}\bar{\mathbf{z}} = \bar{\lambda}(\mathbf{V}\bar{\mathbf{z}}) = \bar{\lambda}\bar{\bar{\mathbf{z}}} = \bar{\lambda}\bar{\mathbf{z}},$$

so $\bar{\mathbf{z}}$ is a right eigenvector of $\mathcal{T}^*$ with eigenvalue $\bar{\lambda}$ which means that $\bar{\mathbf{z}}$ is a left eigenvector of $\mathcal{T}$ with eigenvalue λ.

Generalized eigenvalue problem

$$\mathbf{B}\mathbf{z}_i = \lambda_i \mathcal{D}_{row} \mathbf{z}_i$$

By the theory of the generalized eigenvalue problem, since $\mathcal{D}_{row}$ is positive definite and $\mathbf{B}$ is diagonalizable, there exists a basis (usually not ortogonormal) in $\mathbf{C}^{2m}$ which simultaneously diagonalizes $\mathbf{B}$ and $\mathcal{D}_{row}$. As $\mathcal{D}_{row}$ is also diagonal with positive diagonal entries, we can choose a $\mathcal{D}_{row}$-orthonormal basis, i.e.,

$$\mathbf{z}_i^* \mathcal{D}_{row} \mathbf{z}_j = \delta_{ij}, \quad i, j = 1, \ldots, k.$$

With the notation $\mathbf{Z}_k := (\mathbf{z}_1, \ldots, \mathbf{z}_k)$ and $\mathbf{\Lambda}_k := \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$, we have that

$$\mathbf{Z}_k^T \mathbf{B} \mathbf{Z}_k = \mathbf{Z}_k^* \mathbf{B} \mathbf{Z}_k = \mathbf{\Lambda}_k.$$

Hence, the first k (positive real) generalized eigenvalues of $\mathbf{B}$ and $\mathcal{D}_{row}$ are $\lambda_1, \ldots, \lambda_k$, that are the first k eigenvalues of $\mathcal{T}$, and these are real numbers. The other λ's can be complex.

Consider the left eigenvectors $\mathbf{w}_i$'s of $\mathcal{T}$ that constitute a biorthonormal system with $\mathbf{z}_i$'s: $\mathbf{z}_i^* \mathbf{w}_j = \delta_{ij}$. (In particular, $\mathbf{w}_1 \parallel \check{\mathbf{z}}_1 \parallel \mathbf{1}$ will do, which is the scalar multiple of the vector $\mathbf{1}$).
With the notation $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_{2m})$ and by the biorthogonality, this means that $\mathbf{W}^* = \mathbf{Z}^{-1}$.
In this way, we have the system of equations

$$\mathcal{D}_{col}^{-1} \mathbf{B}^T \mathbf{w}_i = \bar{\lambda}_i \mathbf{w}_i,$$

because a left eigenvector of $\mathcal{D}_{row}^{-1}\mathbf{B}$ is a right eigenvector of its adjoint (transpose as real) $\mathcal{D}_{col}^{-1}\mathbf{B}^T$ with eigenvalue $\bar{\lambda}_i$.
This is exactly the problem of finding the generalized eigenvalues of the matrices $\mathbf{B}^T$ and $\mathcal{D}_{col}$. Indeed,

$$\mathbf{B}^T \mathbf{w}_i = \bar{\lambda}_i \mathcal{D}_{col} \mathbf{w}_i \tag{1}$$

that is used for $i = 1, \ldots, k$.

We know that $\check{\mathbf{Z}}_k^* \mathcal{D}_{col} \check{\mathbf{Z}}_k = \mathbf{I}_k$ and $\mathbf{W}_k = \check{\mathbf{Z}}_k \mathbf{C}_k$. We want to prove that for the $k \times k$ diagonal real matrix $\mathbf{C}_k$: $\mathbf{C}_k = -\mathbf{\Lambda}_k^{-1}$.

$$\mathbf{W}_k^* \mathbf{B}^T \mathbf{W}_k = \mathbf{C}_k^* \check{\mathbf{Z}}_k^* \mathcal{D}_{col} \check{\mathbf{Z}}_k \mathbf{C}_k \mathbf{\Lambda}_k = \mathbf{C}_k^2 \mathbf{\Lambda}_k.$$

As $\mathbf{V}\mathbf{W}_k = \check{\mathbf{Z}}_k$:

$$\check{\mathbf{W}}_k^T \mathbf{B}^{-1} \check{\mathbf{W}}_k = \mathbf{\Lambda}_k^{-1}$$

so by the equality of a part of SVD in $\mathbf{B}^T$ and $\mathbf{B}^{-1}$, $\mathbf{\Lambda}_k^{-1} = \mathbf{C}_k^2 \mathbf{\Lambda}_k$, so $\mathbf{C}_k^2 = {\mathbf{\Lambda}_k}^{-2}$ and $c_i = -\frac{1}{\lambda_i}$, $i = 1, \ldots, k$.

Relation between the eigenvalues of B and $\mathcal{L}$ or $\mathcal{T}$

Our ultimate goal is to find clusters of the nodes by means of the structural non-backtracking eigenvalues that are real ones, separated from the bulk of the spectrum (they are positive in assortative networks). To conclude for them it is more convenient and customary to consider the non-backtracking Laplacian ($\mathcal{L}$) eigenvalues separated from 1, or equivalently, the eigenvalues of the transition probability matrix ($\mathcal{T}$), separated from 0. All these eigenvalues are confined to a circle of radius 1 in the complex plane, but we are interested only in the „structural" (outstanding) real ones. For this purpose, we consider the following equivalent version:

$$(\mathcal{D}_{row}^{-1/2}\mathbf{B}\mathcal{D}_{row}^{-1/2})(\mathcal{D}_{row}^{1/2}\mathbf{z}) = \lambda(\mathcal{D}_{row}^{1/2}\mathbf{z}).$$

The structural positive real eigenvalues

$$\mu_1 \geq \cdots \geq \mu_k > 0$$

of $\mathbf{B}$ and the largest modulus eigenvalues of $\mathcal{T}$ are

$$1 = \lambda_1, \lambda_2, \ldots, \lambda_k.$$

To a real eigenvalue λ_i an eigenvector $\mathbf{z}_i$ of $\mathcal{T}$ corresponds ($\mathbf{z}_i$ also has real coordinates). By the above equivalent version, the matrix $\mathcal{D}_{row}^{-1/2}\mathbf{B}\mathcal{D}_{row}^{-1/2}$ has the same eigenvalues with eigenvectors $\mathbf{x} = \mathcal{D}_{row}^{1/2}\mathbf{z}$. We know that $\mathbf{z}_1 = \mathbf{1}$ and $\mathbf{x}_1 = \mathcal{D}_{row}^{1/2}\mathbf{1}$.
These eigenvectors form a $\mathcal{D}_{row}$-orthonormal system, i.e. $\mathbf{z}_i^T \mathcal{D}_{row} \mathbf{z}_j = \delta_{ij}$ for $i, j = 1, \ldots, k$. Consequently, the vectors $\mathbf{x}_i = \mathcal{D}_{row}^{1/2}\mathbf{z}_i$ are orthonormal: $\mathbf{x}_i^T \mathbf{x}_j = \delta_{ij}$ for $i, j = 1, \ldots, k$. They are the **right eigenvectors** of $\mathcal{D}_{row}^{-1/2}\mathbf{B}\mathcal{D}_{row}^{-1/2}$ with eigenvalues λ_i.

Bauer–Fike perturbations

Proposition

$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^{-1}$ is diagonalizable with eigenvalues α's, $\mathbf{B}$ is arbitrary with eigenvalues β's (both are $n \times n$ with possibly complex entries). Then for any β there is an $i \in \{1, \dots, n\}$ such that

$$|\beta - \alpha_i| \leq \kappa(\mathbf{U})\|\mathbf{B} - \mathbf{A}\| =: R.$$

There can be more than one such i, but we can tell the following. Let C_i be the circle centered at α_i with radius R (in $\mathbf{C}$). For any union of some C_i's, which is disjoint of the union of the remaining C_i's, the number of β's within this union is equal to the number of C_i's in the union (or to the number of α's in the union). In particular, if a C_i is disjoint of the others, then there is exactly one β in it (so, β's and α's cannot be too far apart).

Here $\kappa(\mathbf{U}) = \|\mathbf{U}\| \cdot \|\mathbf{U}^{-1}\| = \frac{s_{max}(\mathbf{U})}{s_{min}(\mathbf{U})}$.

Application of te Bauer–Fike theorem

$$\|\mathcal{T} - \frac{1}{\mu_1}\mathbf{B}\| = \|(\mathcal{D}_{row}^{-1} - \frac{1}{\mu_1}\mathbf{I})\mathbf{B}\| \leq \|\mathcal{D}_{row}^{-1} - \frac{1}{\mu_1}\mathbf{I}\| \cdot \|\mathbf{B}\|,$$

where $\|\mathbf{B}\| = \max d_i - 1$ (the maximal singular value of $\mathbf{B}$) and

$$\|\mathcal{D}_{row}^{-1} - \frac{1}{\mu_1}\mathbf{I}\| = \max_i |\frac{1}{d_i - 1} - \frac{1}{\mu_1}| = \frac{1}{\min d_i - 1} - \frac{1}{\mu_1}.$$

Here d_i is around c, but $\min d_i < c$. So

$$\frac{R}{\kappa(\mathbf{Z})} \leq \frac{\max d_i - 1}{\min d_i - 1} - \frac{\max d_i - 1}{\mu_1}.$$

For diagonalizable matrices of nonnegative entries we know that $\min d_i - 1 \leq \mu_1 \leq \max d_i - 1$. Applying this to μ_1 of $\mathbf{B}$:

$$0 \leq \frac{R}{\kappa(\mathbf{Z})} \leq \frac{\max d_i - 1}{\min d_i - 1} - 1.$$

Jost et al., Discrete Math. 2023: the spectral gap between the transition probability spectrum and zero is at least $\frac{1}{\max_i d_i - 1}$. Also, taking into consideration the relation between the right–left eigenvectors, $\kappa(\mathbf{Z}) \leq \frac{\max d_i - 1}{\min d_i - 1}$ and

$$R \leq \frac{\max d_i - 1}{\min d_i - 1}\left(\frac{\max d_i - 1}{\min d_i - 1} - 1\right)$$

that is the closer to zero as $\max d_i$ is closer to $\min d_i$ which is supported by the SBM_k model.

Possibly, it is better to consider only $\mathbf{Z}_k$ and work with the k-rank approximation of $\mathcal{T}$. Jost et al., Discrete Math. 2023 (for $k = 1$): In the assortative sparse SBM_k model there are k positive real eigenvalues $\geq \frac{1}{\sqrt{c-1}}$, the others (partly complex conjugate pairs) are less in absolute value.
Good news, as for the clustering we use only (real) eigenvectors, corresponding to real eigenvalues of a real matrix.

Clustering

Spectral clustering algorithms use the heuristic that eigenvectors corresponding to the structural eigenvalues of a suitable matrix are applicable to the k-means clustering. In case of the dense stochastic block model, this is supported by subspace perturbation theorems. In the sparse case, similar arguments can be used for some deflated (n-dimensional) versions of the $2m$-dimensional $\mathbf{B}$- or $\mathcal{T}$-eigenvectors.

If G, and hence, $\mathcal{G}$, come from the sparse stochastic block model, we can cluster the vertices with the deflated $\mathbf{z}$ or $\mathbf{x}$ vectors.

As the eigenvectors, corresponding to the k leading eigenvalues of $\mathbf{B}$ are close to the leading inflated eigenvectors of $\bar{\mathbf{A}}$ (with an inner product approaching 1 as $n \to \infty$), and latter ones are step-vectors.

So the leading λs are within a constant factor of μs. The corresponding eigenvectors (if the structural eigenvalues are single) are continuous functions of the matrices. As the eigenvectors $\mathbf{x}_i$s of $\mathbf{B}$, corresponding to its structural eigenvalues $\mu_1, \ldots, \mu_k$ are close to the inflated versions of the eigenvalues of $\bar{\mathbf{A}}$, they are close to step-vectors if our graph comes from the sparse SBM_k model. Therefore, between the $\mathbf{z}_i$s, as eigenvectors of the transition probability matrix $\mathcal{T} = \mathcal{D}_{row}^{-1}\mathbf{B}$ and the inflated eigenvectors $\mathbf{v}_i$s of the matrix $\mathbf{D}_{\bar{\mathbf{A}}}^{-1}\bar{\mathbf{A}}$ (which are step-vectors), the following relation holds true, as the norms of the matrices $\mathcal{D}_{row}^{-1}$ and $\mathbf{D}_{\bar{\mathbf{A}}}^{-1}$ do not depend on n. Here the diagonal matrix $\mathbf{D}_{\bar{\mathbf{A}}}$ contains entries c_i in the ith block for $i = 1, \ldots, k$.

The structural eigenvalues of $\mathcal{D}_{row}^{-1/2}\mathbf{B}\mathcal{D}_{row}^{-1/2}$ are also λ_is with orthonormal eigenvectors $\mathcal{D}_{row}^{1/2}\mathbf{z}_i$ and those are aligned with the structural eigenvalues of $\mathbf{D}_{\bar{\mathbf{A}}}^{-1/2}\bar{\mathbf{A}}\mathbf{D}_{\bar{\mathbf{A}}}^{-1/2}$. Also, the unit-norm eigenvectors $\mathcal{D}_{row}^{1/2}\mathbf{z}_i$ are close to the inflated versions of the unit-norm eigenvectors of this matrix, which are step-vectors, say $\mathbf{v}_1, \ldots, \mathbf{v}_k$ (form an orthonormal system as the matrix is symmetric). So

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

where ε can be any small with increasing n. Therefore,

$$\left\| \mathcal{D}_{row}^{1/2}\mathbf{z} - \frac{\mathbf{End}\,\mathbf{v}}{\|\mathbf{End}\,\mathbf{v}\|} \right\|^2 \leq \varepsilon',$$

where the relation between ε' and ε does not depend on n, but both tend to 0 as $n \to \infty$.

Since $\mathcal{D}_{row}^{1/2}\mathbf{z}^{in} = \mathbf{End}^*\mathcal{D}_{row}^{1/2}\mathbf{z}$ and $\mathbf{End}^*\mathbf{End} = \mathbf{D}$,

$$\left\| \mathbf{End}^*\mathcal{D}_{row}^{1/2}\mathbf{z} - \mathbf{End}^* \frac{\mathbf{End}\,\mathbf{v}}{\|\mathbf{End}\,\mathbf{v}\|} \right\|^2 = \left\| (\mathcal{D}_{row}^{1/2}\mathbf{z})^{in} - \mathbf{D}\frac{\mathbf{v}}{\|\mathbf{End}\,\mathbf{v}\|} \right\|^2$$

also holds. Consequently,

$$\left\| \mathbf{D}^{-1}(\mathcal{D}_{row}^{1/2}\mathbf{z})^{in} - \frac{\mathbf{v}}{\|\mathbf{End}\,\mathbf{v}\|} \right\|^2 \leq \|\mathbf{D}^{-1}\mathbf{End}^*\|^2\varepsilon' \leq \varepsilon'.$$

Indeed, the largest eigenvalue of $(\mathbf{D}^{-1}\mathbf{End}^*)(\mathbf{End}\,\mathbf{D}^{-1}) = \mathbf{D}^{-1}\mathbf{D}\mathbf{D}^{-1} = \mathbf{D}^{-1}$ is $\max_i \frac{1}{d_i}$, so the largest singular value (spectral norm) of $\mathbf{D}^{-1}\mathbf{End}^*$ is estimated from above with $\left(\max_i \frac{1}{d_i}\right)^{\frac{1}{2}}$. Therefore,

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

Now we apply this to the k leading normalized eigenvectors $\mathbf{z}_1, \ldots, \mathbf{z}_k$ of $\mathcal{T}$, for which

$$\sum_{j=1}^{k} \|\mathbf{D}^{-1}(\mathcal{D}_{row}^{1/2}\mathbf{z}_j)^{in} - \frac{\mathbf{v}_j}{\|\mathbf{E}nd\,\mathbf{v}_j\|}\|^2 \leq k\varepsilon'.$$

As $\mathbf{v}_j$s are step-vectors with k different coordinates on the same k steps, the above sum of the squares estimates from above the objective function of the k-means algorithm. Without knowing the $\mathbf{v}_j$s, we minimize it with the k-dimensional node representatives that are row vectors of the $n \times k$ matrix, whose column vectors are

$$\mathbf{D}^{-1}(\mathcal{D}_{row}^{1/2}\mathbf{z}_j)^{in} \quad (j = 1, \ldots, k).$$