Finding normalized and modularity cuts by spectral clustering

Marianna Bolla
Institute of Mathematics
Budapest University of Technology and Economics
marib@math.bme.hu

Ljubjana
2010, October

Outline

- Find community structure in networks. Antiexpanders, but the clusters are good expanders. How many clusters?
- Penalize clusters of extremely different sizes/volumes.
- Minimum multiway cut problems: ratio cut and normalized cut based on Laplacian and normalized Laplacian spectra. Communities with sparse between-cluster (and dense within-cluster) connections.
- Modularity cuts: spectral decomposition of the modularity and normalized modularity matrix. Communities with more within-cluster (and less between-cluster) connections than expected under independence.

Outline

- Find community structure in networks. Antiexpanders, but the clusters are good expanders. How many clusters?
- Penalize clusters of extremely different sizes/volumes.
- Minimum multiway cut problems: ratio cut and normalized cut based on Laplacian and normalized Laplacian spectra. Communities with sparse between-cluster (and dense within-cluster) connections.
- Modularity cuts: spectral decomposition of the modularity and normalized modularity matrix. Communities with more within-cluster (and less between-cluster) connections than expected under independence.

Outline

- Find community structure in networks. Antiexpanders, but the clusters are good expanders. How many clusters?
- Penalize clusters of extremely different sizes/volumes.
- Minimum multiway cut problems: ratio cut and normalized cut based on Laplacian and normalized Laplacian spectra. Communities with sparse between-cluster (and dense within-cluster) connections.
- Modularity cuts: spectral decomposition of the modularity and normalized modularity matrix. Communities with more within-cluster (and less between-cluster) connections than expected under independence.

Outline

- Find community structure in networks. Antiexpanders, but the clusters are good expanders. How many clusters?
- Penalize clusters of extremely different sizes/volumes.
- Minimum multiway cut problems: ratio cut and normalized cut based on Laplacian and normalized Laplacian spectra. Communities with sparse between-cluster (and dense within-cluster) connections.
- Modularity cuts: spectral decomposition of the modularity and normalized modularity matrix. Communities with more within-cluster (and less between-cluster) connections than expected under independence.

Notation

$G = (V, \mathbf{W})$: edge-weighted graph on n vertices,
$\mathbf{W}$: $n \times n$ symmetric matrix, $w_{ij} \geq 0$, $w_{ii} = 0$.
(w_{ij}: similarity between vertices i and j). Simple graph: 0/1 weights
W.l.o.g., $\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} = 1$, joint distribution with marginal entries:

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

(generalized vertex degrees) $\mathbf{D} = \mathtt{diag}\,(d_1, \ldots, d_n)$
Laplacian: $\mathbf{L} = \mathbf{D} - \mathbf{W}$
Normalized Laplacian $\mathbf{L}_D = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$

$1 \leq k \leq n$

$P_k = (V_1, \ldots, V_k)$: k-partition of the vertices

$V_1, \ldots, V_k$: disjoint, non-empty vertex subsets, clusters

$\mathcal{P}_k$: the set of all k-partitions

$e(V_a, V_b) = \sum_{i \in V_a} \sum_{j \in V_b} w_{ij}$: weighted cut between V_a and V_b

$\mathtt{Vol}\,(V_a) = \sum_{i \in V_a} d_i$: volume of V_a

Ratio cut of $P_k = (V_1, \ldots, V_k)$ given $\mathbf{W}$:

$$g(P_k, \mathbf{W}) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left(\frac{1}{|V_a|} + \frac{1}{|V_b|} \right) e(V_a, V_b) = \sum_{a=1}^{k} \frac{e(V_a, \bar{V}_a)}{|V_a|}$$

Normalized cut of $P_k = (V_1, \ldots, V_k)$ given $\mathbf{W}$:

$$f(P_k, \mathbf{W}) = \sum_{a=1}^{k-1} \sum_{b=a+1}^{k} \left(\frac{1}{\mathtt{Vol}(V_a)} + \frac{1}{\mathtt{Vol}(V_b)} \right) e(V_a, V_b)$$

$$= \sum_{a=1}^{k} \frac{e(V_a, \bar{V}_a)}{\mathtt{Vol}(V_a)} = k - \sum_{a=1}^{k} \frac{e(V_a, V_a)}{\mathtt{Vol}(V_a)}$$

Minimum k-way ratio cut and normalized cut of $G = (V, \mathbf{W})$:

$$g_k(G) = \min_{P_k \in \mathcal{P}_k} g(P_k, \mathbf{W}) \quad \text{and} \quad f_k(G) = \min_{P_k \in \mathcal{P}_k} f(P_k, \mathbf{W})$$

The k-means algorithm

The problem: given the points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ and an integer $1 \leq k \leq n$, find the k-partition of the index set $\{1, \ldots, n\}$ (or equivalently, the clustering of the points into k disjoint non-empty subsets) which minimizes the following k-variance:

$$\begin{aligned} S_k^2(\mathbf{x}_1, \ldots, \mathbf{x}_n) &= \min_{P_k \in \mathcal{P}_k} S_k^2(P_k, \mathbf{x}_1, \ldots, \mathbf{x}_n) \\ &= \min_{P_k = (V_1, \ldots, V_k)} \sum_{a=1}^{k} \sum_{j \in V_a} \|\mathbf{x}_j - \mathbf{c}_a\|^2, \end{aligned}$$

$$\mathbf{c}_a = \frac{1}{|V_a|} \sum_{j \in V_a} \mathbf{x}_j.$$

Usually, $d \leq k \ll n$.

To find the global minimum is NP-complete, but the iteration of the k-means algorithm, first described in MacQueen (1963) is capable to find a local minimum in polynomial time.
If there exists a well-separated k-clustering of the points (even the largest within-cluster distance is smaller than the smallest between-cluster one) the convergence of the algorithm to the global minimum is proved by Dunn (1973-74), with a convenient starting. Under relaxed conditions, the speed of the algorithm is increased by a filtration in Kanungo et al. (2002).
The algorithm runs faster if the separation between the clusters increases and an overall running time of $\mathcal{O}(kn)$ can be guaranteed.

Sometimes the points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are endowed with the positive weights $d_1, \ldots, d_n$, w.l.o.g., $\sum_{i=1}^n d_i = 1$.
Weighted k-variance of the points:

$$\begin{aligned}\tilde{S}_k^2(\mathbf{x}_1, \ldots, \mathbf{x}_n) &= \min_{P_k \in \mathcal{P}_k} \tilde{S}_k^2(P_k, \mathbf{x}_1, \ldots, \mathbf{x}_n) \\ &= \min_{P_k=(V_1,\ldots,V_k)} \sum_{a=1}^{k} \sum_{j \in V_a} d_j \|\mathbf{x}_j - \mathbf{c}_a\|^2, \\ \mathbf{c}_a &= \frac{1}{\sum_{j \in V_a} d_j} \sum_{j \in V_a} d_j \mathbf{x}_j.\end{aligned}$$

E.g., $d_1, \ldots, d_n$ is a discrete probability distribution and a random vector takes on values $\mathbf{x}_1, \ldots, \mathbf{x}_n$ with these probabilities; e.g., in a MANOVA (Multivariate Analysis of Variance) setup. The above algorithms can be easily adapted to this situation.

Partition matrices

P_k: $n \times k$ balanced partition matrix $\mathbf{Z}_k = (\mathbf{z}_1, \ldots, \mathbf{z}_k)$
k-partition vector: $\mathbf{z}_a = (z_{1a}, \ldots, z_{na})^T$, where
$z_{ia} = \frac{1}{|V_a|}$, if $i \in V_a$ and 0, otherwise.
$\mathbf{Z}_k$ is suborthogonal: $\mathbf{Z}_k^T \mathbf{Z}_k = \mathbf{I}_k$
The ratio cut of the k-partition P_k given $\mathbf{W}$:

$$g(P_k, \mathbf{W}) = \mathtt{tr}\, \mathbf{Z}_k^T \mathbf{L} \mathbf{Z}_k = \sum_{a=1}^{k} \mathbf{z}_a^T \mathbf{L} \mathbf{z}_a. \tag{1}$$

We want to minimize it over balanced k-partition matrices $\mathbf{Z}_k \in \mathcal{Z}_k^B$.

Estimation by Laplacian eigenvalues

G is connected, the spectrum of $\mathbf{L}$: $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_n$
unit-norm, pairwise orthogonal eigenvectors: $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$,
$\mathbf{u}_1 = \mathbf{1}$
The discrete problem is relaxed to a continuous one:
$\mathbf{r}_1, \ldots, \mathbf{r}_n \in \mathbb{R}^k$: representatives of the vertices
$\mathbf{X} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)^T = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$

$$\min_{\mathbf{X}^T\mathbf{X}=\mathbf{I}_k} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \|\mathbf{r}_i - \mathbf{r}_j\|^2 = \min_{\mathbf{X}^T\mathbf{X}=\mathbf{I}_k} \operatorname{tr} \mathbf{X}^T\mathbf{L}\mathbf{X} = \sum_{i=1}^{k} \lambda_i$$

and equality is attained with $\mathbf{X} = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$.

$$g_k(G) = \min_{\mathbf{Z}_k \in \mathcal{Z}_k^B} \operatorname{tr} \mathbf{Z}_k^T \mathbf{L} \mathbf{Z}_k \geq \sum_{i=1}^{k} \lambda_i \tag{2}$$

and equality can be attained only in the $k = 1$ trivial case, otherwise the eigenvectors $\mathbf{u}_i$ $(i = 2, \ldots, k)$ cannot be partition vectors, since their coordinates sum to 0 because of the orthogonality to the $\mathbf{u}_1 = \mathbf{1}$ vector.

Optimum choice of k?

$$\operatorname{tr} \mathbf{Z}_k^T \mathbf{L} \mathbf{Z}_k = \sum_{i=1}^{n} \lambda_i \sum_{a=1}^{k} (\mathbf{u}_i^T \mathbf{z}_a)^2. \tag{3}$$

This sum is the smallest possible if the largest $(\mathbf{u}_i^T \mathbf{z}_a)^2$ terms correspond to eigenvectors belonging to the smallest eigenvalues. Thus, the above sum is the most decreased by keeping only the k smallest eigenvalues in the inner summation and the corresponding eigenvectors are close to the subspace $\mathcal{F}_k = \operatorname{Span}\{\mathbf{z}_1, \ldots, \mathbf{z}_k\}$.

$$g_{k,k}(\mathbf{Z}_k, \mathbf{L}) := \sum_{i=1}^{k} \sum_{a=1}^{k} \lambda_i (\mathbf{u}_i^T \mathbf{z}_a)^2,$$

we maximize $g_{k,k}(\mathbf{Z}_k, \mathbf{L})$ over $\mathcal{Z}_k^B$ for given $\mathbf{L}$.
The vectors $\sqrt{\lambda_i}\mathbf{u}_i$ are projected onto the subspace $\mathcal{F}_k$:

$$\sqrt{\lambda_i}\mathbf{u}_i = \sum_{a=1}^{k} \sqrt{\lambda_i}(\mathbf{u}_i^T \mathbf{z}_a)\mathbf{z}_a + \text{ort}_{\mathcal{F}_k}(\sqrt{\lambda_i}\mathbf{u}_i), \quad i = 1, \ldots, k.$$

As $\sqrt{\lambda_1}\mathbf{u}_1 = \mathbf{0}$, there is no use of projecting it.
By the Pythagorean equality:

$$\lambda_i = \|\sqrt{\lambda_i}\mathbf{u}_i\|^2 = \sum_{a=1}^{k} \lambda_i (\mathbf{u}_i^T \mathbf{z}_a)^2 + \texttt{dist}^2(\sqrt{\lambda_i}\mathbf{u}_i, \mathcal{F}_k), \quad i = 1, \ldots, k.$$

$$\mathbf{X}'_k = (\sqrt{\lambda_1}\mathbf{u}_1, \ldots, \sqrt{\lambda_k}\mathbf{u}_k) = (\mathbf{r}'_1, \ldots, \mathbf{r}'_n)^T.$$

$$S_k^2(\mathbf{X}'_k) = \sum_{i=1}^{k} \mathtt{dist}^2(\sqrt{\lambda_i}\mathbf{u}_i, \mathcal{F}_k).$$

$$\sum_{i=1}^{k} \lambda_i = g_{k,k}(\mathbf{Z}_k, \mathbf{L}) + S_k^2(P_k, \mathbf{X}'_k), \tag{4}$$

where the partition matrix $\mathbf{Z}_k$ corresponds to the k-partition P_k.

We are looking for the k-partition maximizing the first term. In view of (4), increasing $g_{k,k}(\mathbf{Z}_k, \mathbf{L})$ can be achieved by decreasing $S_k^2(\mathbf{X}'_k)$; latter one is obtained by appying the k-means algorithm with k clusters for the k-dimensional representatives $\mathbf{r}'_1, \ldots, \mathbf{r}'_n$.
As the first column of $\mathbf{X}'_k$ is $\mathbf{0}$, it is equivalent to apply the k-means algorithm with k clusters for the $(k-1)$-dimensional representatives that are the row vectors of the $n \times (k-1)$ matrix obtained from $\mathbf{X}_k$ by deleting its first column.
$\lambda_k \ll \lambda_{k+1}$: we gain the most by omitting the $n-k$ largest eigenvalues.

Perturbation results

$\mathbf{W} = \mathbf{W}_w + \mathbf{W}_b$: within- and between-cluster edges w.r.t. P_k

$\mathbf{D} = \mathbf{D}_w + \mathbf{D}_b$ and $\mathbf{L} = \mathbf{L}_w + \mathbf{L}_b$

ρ: the smallest positive eigenvalue of $\mathbf{L}_w$

ε: the largest eigenvalue of $\mathbf{L}_b$

Suppose $\varepsilon(P_k) < \rho(P_k)$

$\rho = \min_{\rho_i}$ is "large" if the clusters are good expanders

$\varepsilon \leq 2 \max_{i \in \{1,\ldots,n\}} \sum_{j:\, c(j) \neq c(i)} w_{ij}$,

where $c(i)$ is the cluster membership of vertex i.

It is small, if from each vertex there are few, small-weight edges emaneting to clusters different of the vertex's own cluster

We proved (B, Tusnády, Discrete Math. 1994) that

$$S_k^2(\mathbf{X}_k) = \min_{P_k \in \mathcal{P}_k} S_k^2(P_k, \mathbf{X}_k) \leq k \min_{P_k \in \mathcal{P}_k} \frac{\varepsilon(P_k)}{\rho(P_k)}$$

By the Weyl's perturbation theorem:

$$0 < \lambda_k \leq \varepsilon < \rho \leq \lambda_{k+1} \leq \rho + \varepsilon \tag{5}$$

for the ε and ρ of any k-partition with $\varepsilon < \rho$.
Consequently,

$$\frac{\lambda_k}{\lambda_{k+1}} \leq \min_{P_k \in \mathcal{P}_k} \frac{\varepsilon(P_k)}{\rho(P_k)}, \tag{6}$$

Theorem

If for a given k-partition $\frac{\varepsilon(P_k)}{\rho(P_k)} < 1$, *then*

$$S_k^2(P_k, \mathbf{X}_k) \leq k \frac{[\varepsilon(P_k)]^2}{[\rho(P_k) - \varepsilon(P_k)]^2},$$

where $\mathbf{X}_k = (\mathbf{u}_1, \ldots, \mathbf{u}_k)$.

Theorem

$$S_k^2(P_k, \mathbf{X}'_k) \leq \frac{[\varepsilon(P_k)]^2}{[\rho(P_k) - \varepsilon(P_k)]^2} \sum_{i=1}^{k} \lambda_i,$$

where $\mathbf{X}'_k = (\sqrt{\lambda_1}\mathbf{u}_1, \ldots, \sqrt{\lambda_k}\mathbf{u}_k)$. *Consequently,*

$$S_k^2(\mathbf{X}'_k) \leq \sum_{i=1}^{k} \lambda_i \cdot \min_{P_k \in \mathcal{P}_k} \frac{\varepsilon(P_k)^2}{(\rho(P_k) - \varepsilon(P_k))^2}$$

Minimizing the normalized cut

$n \times k$ normalized partition matrix: $\mathbf{Z}_k = (\mathbf{z}_1, \ldots, \mathbf{z}_k)$
$\mathbf{z}_a = (z_{1a}, \ldots, z_{na})^T$, where $z_{ia} = \frac{1}{\sqrt{\mathtt{Vol}\,(V_a)}}$, if $i \in V_a$ and 0, otherwise.
The normalized cut of the k-partition P_k given $\mathbf{W}$:

$$f(P_k, \mathbf{W}) = \mathtt{tr}\,\mathbf{Z}_k^T \mathbf{L} \mathbf{Z}_k = \mathtt{tr}\,(\mathbf{D}^{1/2}\mathbf{Z}_k)^T \mathbf{L}_D (\mathbf{D}^{1/2}\mathbf{Z}_k) \qquad (7)$$

or equivalently,

$$f(P_k, \mathbf{W}) = k - \mathtt{tr}\,(\mathbf{D}^{1/2}\mathbf{Z}_k)^T (\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})(\mathbf{D}^{1/2}\mathbf{Z}_k).$$

We want to minimize $f(P_k, \mathbf{W})$ over the k-partitions. It is equivalent to maximizing $\mathtt{tr}\,\mathbf{Z}_k^T \mathbf{W} \mathbf{Z}_k$ over normalized k-partition matrices $\mathbf{Z}_k \in \mathcal{Z}_k^N$.

Normalized Laplacian eigenvalues

G is connected,
$0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_p < 1 \leq \lambda_{p+1} \leq \cdots \leq \lambda_n \leq 2$
eigenvalues of $\mathbf{L}_D$ with corresponding unit-norm, pairwise orthogonal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$,
$\mathbf{u}_1 = (\sqrt{d_1}, \ldots, \sqrt{d_n})^T$
Continuous relaxation: $\mathbf{X} = (\mathbf{r}_1, \ldots, \mathbf{r}_n)^T = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$

$$\min_{\mathbf{X}^T\mathbf{D}\mathbf{X}=\mathbf{I}_k} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \|\mathbf{r}_i - \mathbf{r}_j\|^2 = \min_{\mathbf{X}^T\mathbf{D}\mathbf{X}=\mathbf{I}_k} \operatorname{tr} \mathbf{X}^T\mathbf{L}\mathbf{X} = \sum_{i=1}^{k} \lambda_i$$

and the minimum is attained with $\mathbf{x}_i = \mathbf{D}^{-1/2}\mathbf{u}_i$ $(i = 1, \ldots, k)$.
Especially, $\mathbf{x}_1 = \mathbf{1}$.

$$f_k(G) = \min_{\mathbf{Z}_k \in \mathcal{Z}_k^N} \operatorname{tr} \mathbf{Z}_k^T \mathbf{L} \mathbf{Z}_k \geq \sum_{i=1}^{k} \lambda_i$$

and equality can be attained only in the $k = 1$ trivial case, otherwise the transformed eigenvectors $\mathbf{D}^{-1/2}\mathbf{u}_i$ $(i = 2, \ldots, k)$ cannot be partition vectors, since their coordinates sum to 0 due to the orthogonality of the $\mathbf{1}$ vector. Equivalently,

$$\max_{\mathbf{Z}_k \in \mathcal{Z}_k^N} \operatorname{tr} \mathbf{Z}_k^T \mathbf{W} \mathbf{Z}_k \leq k - \sum_{i=1}^{k} \lambda_i = \sum_{i=1}^{k} (1 - \lambda_i).$$

Optimum choice of k

$$\mathrm{tr}\,\mathbf{Z}_k^T\mathbf{W}\mathbf{Z}_k) = \sum_{i=1}^{n}(1-\lambda_i)\sum_{a=1}^{k}[(\mathbf{u}_i)^T(\mathbf{D}^{1/2}\mathbf{z}_a)]^2$$

increased if we neglect the terms belonging to the eigenvalues at least 1, hence, the outer summation stops at p. The inner sum is the largest in the $k = p$ case, when the unit-norm, pairwise orthogonal vectors $\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_p$ are close to the orthonormal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_p$, respectively.
$\mathcal{F}_p := \mathrm{Span}\,\{\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_p\}$

$$f_{p,p}(\mathbf{Z}_p, \mathbf{W}) := \sum_{i=1}^{p}(1-\lambda_i)\sum_{a=1}^{p}[(\mathbf{u}_i')^T(\mathbf{D}^{1/2}\mathbf{z}_a)]^2$$

$$\mathrm{tr}\,(\mathbf{D}^{1/2}\mathbf{Z}_p)^T(\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2})(\mathbf{D}^{1/2}\mathbf{Z}_p) \le f_{p,p}(\mathbf{Z}_p, \mathbf{W})$$

For given $\mathbf{W}$, we maximize $f_{p,p}(\mathbf{Z}_p, \mathbf{W})$ over $\mathcal{Z}_p^N$.
The vectors $\sqrt{1-\lambda_i}\,\mathbf{u}_i$ are projected onto the subspace $\mathcal{F}_p$:

$$\sqrt{1-\lambda_i}\,\mathbf{u}_i = \sum_{a=1}^{p}[(\sqrt{1-\lambda_i}\,\mathbf{u}_i)^T\mathbf{D}^{1/2}\mathbf{z}_a]\,\mathbf{D}^{1/2}\mathbf{z}_a + \mathrm{ort}_{\mathcal{F}_p}(\sqrt{1-\lambda_i}\,\mathbf{u}_i)$$

$(i = 1, \ldots, p)$
As $\sqrt{1-\lambda_1}\,\mathbf{u}_1 = \mathbf{u}_1$ is in $\mathcal{F}_p$, its orthogonal component is $\mathbf{0}$.
By the Pythagorean equality:

$$1-\lambda_i = \sum_{a=1}^{p}[(\sqrt{1-\lambda_i}\,\mathbf{u}_i)^T\mathbf{D}^{1/2}\mathbf{z}_a]^2 + \mathtt{dist}^2(\sqrt{1-\lambda_i}\,\mathbf{u}_i, \mathcal{F}_p)$$

$(i = 1, \ldots, p)$

Representation

$$\mathbf{X}'_p = (\mathbf{r}_1, \ldots, \mathbf{r}_n)^T = (\sqrt{1-\lambda_1}\mathbf{D}^{-1/2}\mathbf{u}_1, \ldots, \sqrt{1-\lambda_p}\mathbf{D}^{-1/2}\mathbf{u}_p)$$

$$\mathtt{dist}^2(\sqrt{1-\lambda_i}\,\mathbf{u}_i, \mathcal{F}_p) = \sum_{j=1}^{n} d_j (r_{ji} - c_{ji})^2, \quad i = 1, \ldots, p$$

where r_{ji} is the ith coordinate of the vector $\mathbf{r}_j$ and c_{ji} is the same for vector $\mathbf{c}_j \in \mathbb{R}^p$; further, there are at most p different ones among the centers $\mathbf{c}_1, \ldots, \mathbf{c}_n$ assigned to the vertex representatives. Namely,

$$c_{ji} = \frac{1}{\mathtt{Vol}\,(V_a)} \sum_{\ell \in V_a} d_\ell r_{\ell i}, \quad j \in V_a, \quad i = 1, \ldots, p.$$

$$\tilde{S}_p^2(P_p, \mathbf{X}_p) = \sum_{i=1}^{p} \mathtt{dist}^2(\sqrt{1-\lambda_i}\,\mathbf{u}_i, \mathcal{F}_p) = \sum_{j=1}^{n} d_j \|\mathbf{r}_j - \mathbf{c}_j\|^2.$$

$$\sum_{i=1}^{p}(1-\lambda_i) = f_{p,p}(\mathbf{Z}_p, \mathbf{W}) + \tilde{S}_p^2(P_p, \mathbf{X}'_p).$$

We are looking for the p-partition maximizing the first term. In view of the above formula, increasing $f_{p,p}(\mathbf{Z}_p, \mathbf{W})$ can be achieved by decreasing $\tilde{S}_p^2(\mathbf{X}_p)$; latter one is obtained by appying the k-means algorithm with p clusters for the p-dimensional representatives $\mathbf{r}_1, \ldots, \mathbf{r}_n$ with respective weights $d_1, \ldots, d_n$. As the first column of $\mathbf{X}_p$ is $\mathbf{1}$, it is equivalent to apply the k-means algorithm with p clusters for the $(p-1)$-dimensional representatives that are the row vectors of the $n \times (p-1)$ matrix obtained from $\mathbf{X}_p$ by deleting its first column.

Similarly, for $d < k \leq p$:

$$\sum_{i=1}^{d}(1-\lambda_i) = \sum_{i=1}^{d}\sum_{a=1}^{k}[(\sqrt{1-\lambda_i}\,\mathbf{u}_i)^T\mathbf{D}^{1/2}\mathbf{z}_a]^2 + \sum_{i=1}^{d}\mathtt{dist}^2(\sqrt{1-\lambda_i}\,\mathbf{u}_i, \mathcal{F}_k)$$

$$:= f_{k,d}(\mathbf{Z}_k, \mathbf{W}) + \tilde{S}_k^2(P_k, \mathbf{X}'_d),$$

$\mathbf{X}'_d = (\sqrt{1-\lambda_1}\,\mathbf{D}^{-1/2}\mathbf{u}_1, \ldots, \sqrt{1-\lambda_d}\,\mathbf{D}^{-1/2}\mathbf{u}_d)$.
In the presence of a spectral gap between λ_d and $\lambda_{d+1} < 1$ neither $\sum_{i=1}^{d}(1-\lambda_i)$ nor $\tilde{S}_k^2(\mathbf{X}'_d)$ is increased significantly by introducing one more eigenvalue-eigenvector pair (by using $(d+1)$-dimensional representatives instead of d-dimensional ones). Consequently, $f_{k,d}(\mathbf{Z}_k, \mathbf{W})$ would not change much, and $k = d$ clusters based on d-dimensional representatives will suffice. Increasing the number of clusters $(k+1, \ldots, p)$ will decrease the sum of the inner variances $\tilde{S}_{k+1}^2(\mathbf{X}'_d) \leq \tilde{S}_k^2(\mathbf{X}'_d)$, but not significantly.

In the $k = p$ special case

$$\mathtt{tr}\, \mathbf{Z}_p^T \mathbf{L} \mathbf{Z}_p \geq p - f_{p,p}(\mathbf{Z}_p, \mathbf{W}) = \sum_{i=1}^{p} \lambda_i + \tilde{S}_p^2(P_p, \mathbf{X}'_p)$$

that is true for the minima too:

$$f_p(G) \geq \sum_{i=1}^{p} \lambda_i + \tilde{S}_p^2(\mathbf{X}'_p),$$

a sharper lower estimator for the minimum normalized p-way cut than $\sum_{i=1}^{p} \lambda_i$.

Spectral gap and variance

In B, Tusnády, Discrete Math., 1994

Theorem

In the representation $\mathbf{X}_2 = (\mathbf{D}^{-1/2}\mathbf{u}_1, \mathbf{D}^{-1/2}\mathbf{u}_2) = (\mathbf{1}, \mathbf{D}^{-1/2}\mathbf{u}_2)$*:*

$$S_k^2(X_2) \leq \frac{\lambda_2}{\lambda_3}$$

Similarly,

Theorem

In the representation $\mathbf{X}_2' = (\sqrt{1-\lambda_1}\mathbf{D}^{-1/2}\mathbf{u}_1, \sqrt{1-\lambda_2}\mathbf{D}^{-1/2}\mathbf{u}_2) = (\mathbf{1}, \sqrt{1-\lambda_2}\mathbf{D}^{-1/2}\mathbf{u}_2)$*:*

$$\tilde{S}_k^2(X_2') \leq \frac{\lambda_2(1-\lambda_2)}{\lambda_3}$$

Can it be generalized for $k > 2$?

Isoperimetric number

Definition

The Cheeger constant of the weighted graph $G = (V, W)$ is

$$h(G) = \min_{\substack{U \subset V \\ \mathrm{Vol}\,(U) \leq 1/2}} \frac{e(U, \bar{U})}{\mathrm{Vol}\,(U)}$$

Theorem

(B, M-Sáska, Discrete Math. 2004). Let λ_2 be the smallest positive eigenvalue of $\mathbf{L}_D$. Then

$$\frac{\lambda_2}{2} \leq h(G) \leq \min\{1, \sqrt{2\lambda_2}\}$$

. If $\lambda_2 \leq 1$ then the upper estimate can be improved to

$$h(G) \leq \sqrt{\lambda_2(2 - \lambda_2)}.$$

$f_2(G)$ is the symmetric version of $h(G)$: $f_2(G) \leq 2h(G) \Longrightarrow$

$$f_2(G) \leq \sqrt{\lambda_2(2-\lambda_2)}, \quad \lambda_2 \leq 1.$$

$h(G)$ and $f_2(G)$ were estimated by Mohar,J. Comb. Theory B. 1989 for simple graphs (similar formula by max-degree)

Theorem

Suppose that $G = (V, \mathbf{W})$ is connected, and λ_i's are the eigenvalues of $\mathbf{L}_D$. Then $\sum_{i=1}^{k} \lambda_i \leq f_k(G)$ and in the case when the optimal k-dimensional representatives can be classified into k well-separated clusters in such a way that the maximum cluster diameter ε satisfies the relation $\varepsilon \leq \min\{1/\sqrt{2k}, \sqrt{2}\min_i \sqrt{\mathtt{Vol}(V_i)}\}$ with k-partiton $(V_1, \ldots, V_k)$ induced by the clusters above, then

$$f_k(G) \leq c^2 \sum_{i=1}^{k} \lambda_i,$$

where $c = 1 + \varepsilon c'/(\sqrt{2} - \varepsilon c')$ and $c' = 1/\min_i \sqrt{\mathtt{Vol}(V_i)}$.

The Newman–Girvan modularity for simple graphs

$G = (V, \mathbf{A})$: simple graph, $|V| = n$, $\mathbf{A}$ is 0/1 adjacency matrix
number of edges: $e = \frac{1}{2} \sum i = 1^n \sum i = 1^n a_{ij}$
$d_i = \sum_{j=1}^{n} a_{ij}$: the degree of vertex i

$h_{ij} := \frac{d_i d_j}{2e}$: expected number of $i \sim j$ edges by random attachment

$P_k = (V_1, \ldots, V_k)$: k-partition of the vertices (modules)

Newman and Girvan, Phys. Rev. E, 2004 introduced a modularity with large values for stronger (than expected) intra-module connections.

Definition

The Newman–Girvan modularity of the k-partition P_k given $\mathbf{A}$:

$$Q_{NG}(P_k, \mathbf{A}) = \frac{1}{2e} \sum_{a=1}^{k} \sum_{i,j \in V_a} (a_{ij} - h_{ij})$$

Obviously, Q_{NG} takes on values less than 1, and due to $\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} = \sum_{i=1}^{n} \sum_{j=1}^{n} h_{ij}$ it is 0 for $k = 1$ (no community structure at all); therefore, only integers $k \in [2, N]$ are of interest. For given $\mathbf{A}$ and k, we are looking for $\max_{P_k \in \mathcal{P}_k} Q_{NG}(P_k, \mathbf{A})$, more precisely, for the optimum k-partition giving the arg max of it; and eventually, for the optimum k too (there may be several local maxima).

$\mathbf{H} = (h_{ij})$: matrix of rank 1

$\mathbf{B} = \mathbf{A} - \mathbf{H}$: modularity matrix

The row sums of $\mathbf{B}$ are zeros, therefore it always has a 0 eigenvalue with the trivial eigenvector $\mathbf{1} = (1, 1, \ldots, 1)^T$.
If G is the complete graph, then $\mathbf{B}$ is negative semidefinite; but typically, it is indefinite.

$\beta_1 \geq \beta_2 \geq \cdots \geq \beta_n$ eigenvalues of $\mathbf{B}$ with unit norm, pairwise orthogonal eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$.

Newman, Phys. Rev. E, 2006 uses the eigenvectors belonging to the positive eigenvalues of $\mathbf{B}$.

for edge-weighted graphs

$G = (V, \mathbf{W})$, w.l.o.g. $\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} = 1$ supposed

$$h_{ij} = d_i d_j, \quad i, j = 1, \ldots, n$$

Definition

the Newman-Girvan modularity of P_k given $\mathbf{W}$:

$$\begin{aligned} Q(P_k, \mathbf{W}) &= \sum_{a=1}^{k} \sum_{i,j \in V_a} (w_{ij} - h_{ij}) \\ &= \sum_{a=1}^{k} [e(V_a, V_a) - \mathtt{Vol}^2(V_a)], \end{aligned}$$

Under the null-hypothesis, vertices i and j are connected to each other independently, with probabilities proportional (actually, because of the normalizing condition, equal) to their generalized degrees.

For or given k we maximize $Q(P_k, \mathbf{W})$ over $\mathcal{P}_k$.
This task is equivalent to minimizing

$$\sum_{a \neq b} \sum_{i \in V_a, j \in V_b} (w_{ij} - h_{ij}).$$

We want to penalize partitions with clusters of extremely different sizes or volumes

Definition

Balanced Newman–Girvan modularity of P_k given $\mathbf{W}$:

$$\begin{aligned} Q_B(P_k, \mathbf{W}) &= \sum_{a=1}^{k} \frac{1}{|V_a|} \sum_{i,j \in V_a} (w_{ij} - h_{ij}) \\ &= \sum_{a=1}^{k} \left[\frac{e(V_a, V_a)}{|V_a|} - \frac{\mathtt{Vol}^2(V_a)}{|V_a|} \right], \end{aligned}$$

Definition

Normalized Newman–Girvan modularity of P_k given $\mathbf{W}$:

$$Q_N(P_k, \mathbf{W}) = \sum_{a=1}^{k} \frac{1}{\mathtt{Vol}\,(V_a)} \sum_{i,j \in V_a} (w_{ij} - h_{ij})$$

$$= \sum_{a=1}^{k} \frac{e(V_a, V_a)}{\mathtt{Vol}\,(V_a)} - 1,$$

Maximizing the normalized Newman–Girvan modularity over $\mathcal{P}_k$ is equivalent to minimizing the normalized cut.

Maximizing the balanced Newman–Girvan modularity

$\mathbf{B} = \mathbf{W} - \mathbf{H}$: modularity matrix

Spectrum: $\beta_1 \geq \cdots \geq \beta_p > 0 = \beta_{p+1} \geq \cdots \geq \beta_n$

Unit-norm, pairwise orthogonal eigenvectors: $\mathbf{u}_1, \ldots, \mathbf{u}_n$, $\mathbf{u}_{p+1} = \mathbf{1}/\sqrt{n}$.

$$Q_B(P_k, \mathbf{W}) = Q_B(\mathbf{Z}_k, \mathbf{B}) = \sum_{a=1}^{k} \mathbf{z}_a^T \mathbf{B} \mathbf{z}_a = \mathtt{tr}\, \mathbf{Z}_k^T \mathbf{B} \mathbf{Z}_k.$$

We want to maximize $\mathtt{tr}\, \mathbf{Z}_k^T \mathbf{B} \mathbf{Z}_k$ over balanced k-partition matrices $\mathbf{Z}_k \in \mathcal{Z}_k^B$.

Continuous relaxation:

$$\max_{\mathbf{Y}^T\mathbf{Y}=\mathbf{I}_k} \mathtt{tr}\left(\mathbf{Y}^T\mathbf{B}\mathbf{Y}\right) = \max_{\mathbf{y}_a^T\mathbf{y}_b=\delta_{ab}} \sum_{a=1}^{k} \mathbf{y}_a^T\mathbf{B}\mathbf{y}_a = \sum_{a=1}^{k} \beta_a$$

and equality is attained when $\mathbf{y}_1, \ldots, \mathbf{y}_k$ are eigenvectors of $\mathbf{B}$ corresponding to $\beta_1, \ldots, \beta_k$. Though the vectors themselves are not necessarily unique (e.g., in case of multiple eigenvalues), the subspace $\mathtt{Span}\left\{\mathbf{y}_1, \ldots, \mathbf{y}_k\right\}$ is unique if $\beta_k > \beta_{k+1}$.

$$\max_{\mathbf{Z}_k \in \mathcal{Z}_k^B} Q_B(\mathbf{Z}_k, \mathbf{B}) \leq \sum_{a=1}^{k} \beta_a \leq \sum_{a=1}^{p+1} \beta_a. \tag{8}$$

Both inequalities can be attained by equality only in the $k = 1$, $p = 0$ case, when our underlying graph is the complete graph: $\mathbf{A} = \mathbf{1}\mathbf{1}^T - \mathbf{I}$, $\mathbf{W} = \frac{1}{n(n-1)}\mathbf{A}$. In this case there is only one cluster with partition vector of equal coordinates (balanced eigenvector belonging to the single 0 eigenvalue).

The maximum with respect to k of the maximum in (8) is attained with the choice of $k = p + 1$.

$$Q_B(\mathbf{Z}_k, \mathbf{B}) = \mathtt{tr}\, \mathbf{Z}_k^T \mathbf{B} \mathbf{Z}_k = \sum_{a=1}^{k} \mathbf{z}_a^T \Big(\sum_{i=1}^{n} \beta_i \mathbf{u}_i \mathbf{u}_i^T\Big) \mathbf{z}_a$$

$$= \sum_{i=1}^{n} \beta_i \sum_{a=1}^{k} (\mathbf{u}_i^T \mathbf{z}_a)^2.$$

We can increase the last sum if we neglect the terms belonging to the negative eigenvalues, hence, the outer summation stops at p, or equivalently, at $p + 1$. In this case the inner sum is the largerst in the $k = p + 1$ case, when the partition vectors $\mathbf{z}_1, \ldots, \mathbf{z}_{p+1}$ are "close" to the eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_{p+1}$, respectively. As both systems consist of orthonormal sets of vectors, the two subspaces spanned by them should be close to each other. The subspace $\mathcal{F}_{p+1} = \mathtt{Span}\,\{\mathbf{z}_1, \ldots, \mathbf{z}_{p+1}\}$ consists of stepwise constant vectors on $p + 1$ steps, therefore $\mathbf{u}_{p+1} \in \mathcal{F}_{p+1}$, and it suffices to process only the first p eigenvectors.

$$Q_B(\mathbf{Z}_{p+1}, \mathbf{B}) \leq Q'_{p+1,p}(\mathbf{Z}_{p+1}, \mathbf{B}) := \sum_{i=1}^{p} \beta_i \sum_{a=1}^{p+1} (\mathbf{u}_i^T \mathbf{z}_a)^2,$$

and in the sequel, for given $\mathbf{B}$, we want to maximize $Q'_{p+1,p}(\mathbf{Z}_{p+1}, \mathbf{B})$ over $\mathcal{Z}_{p+1}^{B}$.
Project the vectors $\sqrt{\beta_i}\mathbf{u}_i$ onto the subspace $\mathcal{F}_{p+1}$:

$$\sqrt{\beta_i}\mathbf{u}_i = \sum_{a=1}^{p+1} [(\sqrt{\beta_i}\mathbf{u}_i)^T \mathbf{z}_a]\mathbf{z}_a + \text{ort}_{\mathcal{F}_{p+1}}(\sqrt{\beta_i}\mathbf{u}_i), \qquad (9)$$

$$i = 1, \ldots, p.$$

$$\beta_i = \|\sqrt{\beta_i}\mathbf{u}_i\|^2 = \sum_{a=1}^{p+1} [(\sqrt{\beta_i}\mathbf{u}_i)^T \mathbf{z}_a]^2$$

$$+ \texttt{dist}^2(\sqrt{\beta_i}\mathbf{u}_i, \mathcal{F}_{p+1}), \quad i = 1, \ldots, p.$$

$$\sum_{i=1}^{p} \beta_i = \sum_{i=1}^{p} \sum_{a=1}^{p+1} [(\sqrt{\beta_i}\mathbf{u}_i)^T \mathbf{z}_a]^2 + \sum_{i=1}^{p} \mathtt{dist}^2(\sqrt{\beta_i}\mathbf{u}_i, \mathcal{F}_{p+1})$$
$$= Q'_{p+1,p}(\mathbf{Z}_{p+1}, \mathbf{B}) + S^2_{p+1}(\mathbf{X}_p),$$

where the rows of $\mathbf{X}_p = (\sqrt{\beta_1}\mathbf{u}_1, \ldots, \sqrt{\beta_p}\mathbf{u}_p)$ are regarded as p-dimensional representatives of the vertices.

We could as well take $(p+1)$-dimensional representatives as the last coordinates are zeros, and hence, $S^2_{p+1}(\mathbf{X}_p) = S^2_{p+1}(\mathbf{X}_{p+1})$.

Maximizing $Q'_{p+1,p}$ is equivalent to minimizing $S^2_{p+1}(\mathbf{X}_p)$ that can be obtained by applying the k-means algorithm for the p-dimensional representatives with $p+1$ clusters.

More generally, if there is a gap between $\beta_k \gg \beta_{k+1} > 0$, then we look for $k+1$ clusters based on k-dimensional representatives of the vertices.
k leading eigenvectors are projected onto the k-dimensional subspace of $\mathcal{F}_{k+1}$ orthogonal to $\mathbf{1}$.
Calculating eigenvectors is costy; the Lánczos method performs well if we calculate only eigenvectors belonging to some leading eigenvalues followed by a spectral gap. Some authors suggest to use as many eigenvectors as possible. In fact, using more eigenvectors (up to p) is better from the point of view of accuracy, but using less eigenvectors (up to a gap in the positive part of the spectrum) is better from the computational point of view. We have to compromise.

Normalized modularity matrix

$$\mathbf{B}_D = \mathbf{D}^{-1/2}\mathbf{B}\mathbf{D}^{-1/2}$$

$$\begin{aligned} Q_N(P_k, \mathbf{W}) = Q_N(\mathbf{Z}_k, \mathbf{B}) &= \sum_{a=1}^{k} \mathbf{z}_a^T \mathbf{B} \mathbf{z}_a \\ &= \mathtt{tr}\,(\mathbf{D}^{1/2}\mathbf{Z}_k)^T \mathbf{B}_D (\mathbf{D}^{1/2}\mathbf{Z}_k) \end{aligned}$$

$1 = \beta'_1 \geq \cdots \geq \beta'_n \geq -1$: spectrum of $\mathbf{B}_D$
$\mathbf{u}'_1, \ldots, \mathbf{u}'_n$: unit-norm, pairwise orthogonal eigenvectors
$\mathbf{u}'_1 = (\sqrt{d_1}, \ldots, \sqrt{d_n})^T$
p: number of positive eigenvalues of $\mathbf{B}_D$ (this p not necessarily coincides with that of $\mathbf{B}$)

$$\max_{\mathbf{Z}_k \in \mathcal{Z}_k^N} Q_N(\mathbf{Z}_k, \mathbf{B}) \leq \sum_{a=1}^{k} \beta'_a \leq \sum_{a=1}^{p+1} \beta'_a.$$

$$Q_N(\mathbf{Z}_k, \mathbf{B}) = \sum_{i=1}^{n} \beta'_i \sum_{a=1}^{k} [(\mathbf{u}'_i)^T (\mathbf{D}^{1/2} \mathbf{z}_a)]^2.$$

We can increase this sum if we neglect the terms belonging to the negative eigenvalues, hence, the outer summation stops at p, or equivalently, at $p+1$. The inner sum is the largest in the $k = p+1$ case, when the unit-norm, pairwise orthogonal vectors $\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_{p+1}$ are close to the eigenvectors $\mathbf{u}'_1, \ldots, \mathbf{u}'_{p+1}$, respectively. In fact, the two subspaces spanned by them should be close to each other.

$\mathcal{F}_{p+1} = \mathtt{Span}\left\{\mathbf{D}^{1/2}\mathbf{z}_1, \ldots, \mathbf{D}^{1/2}\mathbf{z}_{p+1}\right\}$ not stepwise constant vectors!

$\mathbf{X}'_p = (\sqrt{\beta'_1}\,\mathbf{D}^{-1/2}\mathbf{u}'_1, \ldots, \sqrt{\beta'_p}\,\mathbf{D}^{-1/2}\mathbf{u}'_p)$

$$\tilde{S}^2_{p+1}(P_k, \mathbf{X}'_p) = \sum_{i=1}^{p} \mathtt{dist}^2(\sqrt{\beta'_i}\,\mathbf{u}'_i, \mathcal{F}_{p+1})$$

$$= \sum_{j=1}^{n} d_j \|\mathbf{x}'_j - \mathbf{c}_j\|^2 \to min.$$

weighted k-means algorithm

gap history

Pure community structure

G: disjoint union of k complete graphs on $n_1, \ldots, n_k$ vertices
Eigenvalues of the modularity matrix:

- $k-1$ positive β_i's (>1)
  In the $n_1 = \cdots = n_k$ special case it is a multiple eigenvalue.
- 0: single eigenvalue
- -1: eigenvalue with multiplicity $n-k$

Eigenvalues of the normalized modularity matrix:

- 1: with multiplicity $k-1$
- 0: single eigenvalue
- $n-k$ negative eigenvalues in (-1,0)
  In the $n_1 = \cdots = n_k$ special case there is one negative eigenvalue with multiplicity $n-k$

Piecewise constant eigenvectors belonging to the positive eigenvalues $\Longrightarrow$ k clusters

Pure anticommunity structure

G: complete k-partite graph on $n_1, \ldots, n_k$ vertices

Eigenvalues of the modularity matrix:

- $k-1$ negative eigenvalues
- 0: all the other eigenvalues

Eigenvalues of the normalized modularity matrix:

- $k-1$ negative eigenvalues in (-1,0)
- 0: all the other eigenvalues

MINIMIZE THE MODULARITY!

Piecewise constant eigenvectors belonging to the negative eigenvalues $\Longrightarrow$ k clusters

Block matrices

Definition

The $n \times n$ symmetric real matrix $\mathbf{B}$ is a blown-up matrix, if there is a $k \times k$ symmetric so-called pattern matrix $\mathbf{P}$ with entries $0 \leq p_{ij} \leq 1$, and there are positive integers $n_1, \ldots, n_k$ with $\sum_{i=1}^{k} n_i = n$, such that – after rearranging its rows and columns – the matrix $\mathbf{B}$ can be divided into $k \times k$ blocks, where block (i, j) is an $n_i \times n_j$ matrix with entries all equal to p_{ij} $(1 \leq i, j \leq n)$.

$G = (V, \mathbf{B})$ is a weighted graph with possible loops

Generalized random graph: G is a random graph whose edges exist independently, with probabilities p_{ij}.

Modularity matrix: $k - 1$ structural (large absolute value) eigenvalues $\Longrightarrow$ k clusters by the corresponding eigenvectors

Pure community structure: $p_{ii} = 1$, $p_{ij} = 0$ $(i \neq j)$

Pure anticommunity structure: $p_{ii} = 0$, $p_{ij} = 1$ $(i \neq j)$

Eigenvalues of blown-up matrices and their Laplacians

GENERAL CASE p_{ij}'s are arbitrary, $\mathbf{P} \Longrightarrow \mathbf{B}$
$\Delta := \mathtt{diag}\,(n_1, \ldots, n_k)$ with $n_i \geq cn$ $(i = 1, \ldots, k)$
$\mathtt{rank}\,(\mathbf{B}) = k$, non-zero eigenvalues: $n \cdot \lambda_i = \Theta(n)$
with stepwise constant eigenvectors
λ_i's are eigenvalues of $\Delta^{1/2}\mathbf{P}\Delta^{1/2}$
Laplacian eigenvalues:

- 0: single eigenvalue
- $\lambda_1, \ldots, \lambda_{k-1} = \Theta(n)$ with stepwise constant eigenvectors
- $\gamma_i = \Theta(n)$ with multiplicity $n_i - 1$, eigenspace: vectors with coordinates 0 except block i, where the sum of the coordinates is 0 $(i = 1, \ldots, k)$

Normalized Laplacian eigenvalues:
There exists $\delta > 0$ (independent of n) such that there are
k eigenvalues $\in [0, 1-\delta] \cup [1+\delta, 2]$
with stepwise constant eigenvectors.
All the other eigenvalues are 1.

COMPLETE k-PARTITE GRAPH $p_{ii} = 0$, $p_{ij} = 1$ $(i \neq j)$

non-zero eigenvalues of $\mathbf{B}$: $|\lambda_1|, \ldots, |\lambda_k| = \Theta(n)$

with stepwise constant eigenvectors

$\lambda_1, \ldots, \lambda_k \in [-\max_i n_i, -\min_i n_i] \cup [n - \max_i n_i, n - \min_i n_i]$

Laplacian eigenvalues:

- 0: single eigenvalue
- $\lambda_1 = \cdots = \lambda_{k-1} = n$ with stepwise constant eigenvectors
- $\gamma_i = n - n_i$ with multiplicity $n_i - 1$, eigenspace: vectors with coordinates 0 except block i, where the sum of the coordinates is 0 $(i = 1, \ldots, k)$

Normalized Laplacian eigenvalues:

There exists $\delta > 0$ (independent of n) such that there are $k - 1$ eigenvalues $\in [1 + \delta, 2]$

with stepwise constant eigenvectors.

0 is a single eigenvalue,

all the other eigenvalues are 1.

k DISJOINT CLUSTERS

$\mathbf{A} = \oplus_{i=1}^{k} A_i$ block-diagonal

Diagonal/non-diagonal entries of the $n_i \times n_i$ matrix $\mathbf{A}_i$: ν_i/μ_i

Large eigenvalues of $\mathbf{A}$: $\lambda_i = (n_i - 1)\mu_i + \nu_i = \Theta(n)$ with stepwise constant eigenvectors.

Small eigenvalues of $\mathbf{A}$: $\nu_i - \mu_i$ with multiplicity $n_i - 1$ $(i = 1, \ldots, k)$.

Laplacian eigenvalues:

- 0: with multiplicity k, eigenspace: stepwise constant vectors.
- $n_i\mu_i$ with with multiplicity $n_i - 1$ $(i = 1, \ldots, k)$

Normalized Laplacian eigenvalues:

- 0: with multiplicity k.
- $\frac{n_i\mu_i}{\nu_i - (n_i - 1)\mu_i} \sim 1$ with with multiplicity $n_i - 1$ $(i = 1, \ldots, k)$.

Union of k complete graphs ($\nu_i = 0$). Non-0 e.v.'s: $\frac{n_i}{n_i - 1} > 1$.