Parametric and Non-parametric Approaches to Recover Regular Graph Partitions

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

ASMDA-2011, Rome
June, 2011

Outline

- Estimating the parameters and underlying partitions of a stochastic block model by means of the EM-algorithm. The sample is a weighted graph based on similarities between sites (e.g., social or metabolic networks).
- Non-parametric statistics: modularities. Minima or maxima over k-partitions of the vertices give modules/clusters with regular behavior of information flow within or between the clusters.
- Spectral characterization of a generalized random graph model: blown up structure + random noise.
- Deterministic case: volume-regularity and spectra.

Outline

- Estimating the parameters and underlying partitions of a stochastic block model by means of the EM-algorithm. The sample is a weighted graph based on similarities between sites (e.g., social or metabolic networks).
- Non-parametric statistics: modularities. Minima or maxima over k-partitions of the vertices give modules/clusters with regular behavior of information flow within or between the clusters.
- Spectral characterization of a generalized random graph model: blown up structure + random noise.
- Deterministic case: volume-regularity and spectra.

Outline

- Estimating the parameters and underlying partitions of a stochastic block model by means of the EM-algorithm. The sample is a weighted graph based on similarities between sites (e.g., social or metabolic networks).
- Non-parametric statistics: modularities. Minima or maxima over k-partitions of the vertices give modules/clusters with regular behavior of information flow within or between the clusters.
- Spectral characterization of a generalized random graph model: blown up structure + random noise.
- Deterministic case: volume-regularity and spectra.

Outline

- Estimating the parameters and underlying partitions of a stochastic block model by means of the EM-algorithm. The sample is a weighted graph based on similarities between sites (e.g., social or metabolic networks).
- Non-parametric statistics: modularities. Minima or maxima over k-partitions of the vertices give modules/clusters with regular behavior of information flow within or between the clusters.
- Spectral characterization of a generalized random graph model: blown up structure + random noise.
- Deterministic case: volume-regularity and spectra.

Stochastic block model

Bickel and Chen (PNAS 2009) introduced a random block model which is, in fact, a generalized random graph.

- For given k, vertices independently belong to cluster V_a with probability π_a, $a = 1, \dots, k$; $\sum_{a=1}^{k} \pi_a = 1$.
- Vertices of V_a and V_b are connected independently of each other with probabilities $\mathbb{P}(i \sim j | i \in V_a, j \in V_b) = p_{ab}$, $1 \le a, b \le k$.

The parameters are collected in the vector $\underline{\pi} = (\pi_1, \dots, \pi_k)$ and in the $k \times k$ symmetric matrix $\mathbf{P}$ of p_{ab}'s.
Our statistical sample is the $n \times n$ symmetric, 0-1 adjacency matrix $\mathbf{A} = (a_{ij})$ of a simple graph on n vertices. There are no loops, so the diagonal entries are zeroes. We want to estimate the parameters of the above block model.

By the theorem of mutually exclusive and exhaustive events, the likelihood function is

$$\frac{1}{2}\sum_{1\le a,b\le k}\pi_a\pi_b\prod_{i\in C_a,j\in C_b,i\neq j}p_{ab}^{a_{ij}}(1-p_{ab})^{(1-a_{ij})}$$
$$=\frac{1}{2}\sum_{1\le a,b\le k}\pi_a\pi_b\cdot p_{ab}^{e_{ab}}\cdot(1-p_{ab})^{(n_{ab}-e_{ab})}$$

reminiscent of the mixture of binomial distributions, where
e_{ab} is the number of edges connecting vertices of V_a and V_b $(a\neq b)$;
e_{aa} is twice the number of edges with endpoints in V_a;
$n_{ab}=|V_a|\cdot|V_b|$ if $a\neq b$, and $n_{aa}=|V_a|\cdot(|V_a|-1)$, $a=1,\dots,k$.

EM-algorithm for incomplete data

Here **A** is the incomplete data specification, as the cluster memberships are missing. Therefore we complete our data matrix **A** by latent membership vectors $\Delta_1, \dots, \Delta_n$ of the vertices that are k-dimensional i.i.d. $Poly(1, \underline{\pi})$ random vectors.
$\Delta_i = (\Delta_{1i}, \dots, \Delta_{ki})$, where $\Delta_{ai} = 1$ if $i \in V_a$ and zero otherwise.
Thus, the sum of the coordinates of any Δ_i is 1, and
$\mathbb{P}(\Delta_{ai} = 1) = \pi_a$.
The likelihood function above is

$$\frac{1}{2} \sum_{1 \le a,b \le k} \pi_a \pi_b \cdot p_{ab}^{\sum_{i,j:\, i \neq j} \Delta_{ai} \Delta_{bj} a_{ij}} \cdot (1 - p_{ab})^{\sum_{i,j:\, i \neq j} \Delta_{ai} \Delta_{bj} (1 - a_{ij})}$$

that is maximized in the alternating E and M steps of the EM-algorithm.

We remark that the complete likelihood would be the squareroot of

$$\prod_{1\le a,b\le k} p_{ab}^{e_{ab}} \cdot (1-p_{ab})^{(n_{ab}-e_{ab})}$$

$$= \prod_{a=1}^{k} \prod_{i=1}^{n} \prod_{b=1}^{k} [p_{ab}^{\sum_{j:\, j\ne i} \Delta_{bj} a_{ij}} \cdot (1-p_{ab})^{\sum_{j:\, j\ne i} \Delta_{bj}(1-a_{ij})}]^{\Delta_{ai}}$$

that is valid only in case of known cluster memberships.
Starting with initial parameter values $\underline{\pi}^{(0)}$, $\mathbf{P}^{(0)}$ and membership vectors $\Delta_1^{(0)}, \dots, \Delta_n^{(0)}$, the t-th step of the iteration is the following ($t = 1, 2, \dots$).

E-step

We calculate the conditional expectation of each Δ_i conditioned on the model parameters and on the other cluster assignments obtained in the $(t-1)$-th step (denoted by $M^{(t-1)}$).

By the Bayes theorem, the responsibility of vertex i for cluster a:

$$\pi_{ai}^{(t)} = \mathbb{E}(\Delta_{ai}|M^{(t-1)}) = \mathbb{P}(\Delta_{ai} = 1|M^{(t-1)})$$
$$= \frac{\mathbb{P}(M^{(t-1)}|\Delta_{ai} = 1) \cdot \pi_a^{(t-1)}}{\sum_{l=1}^{k} \mathbb{P}(M^{(t-1)}|\Delta_{li} = 1) \cdot \pi_l^{(t-1)}}$$

$(a = 1, \ldots, k;\ i = 1, \ldots, n)$. Thus, for each i, $\pi_{ai}^{(t)}$ is proportional to the numerator, where

$$\mathbb{P}(M^{(t-1)}|\Delta_{ai} = 1) = \prod_{b=1}^{k} (p_{ab}^{(t-1)})^{\sum_{j\neq i} \Delta_{bj}^{(t-1)} a_{ij}} \cdot (1 - p_{ab}^{(t-1)})^{\sum_{j\neq i} \Delta_{bj}^{(t-1)} (1-a_{ij})}$$

is the part of the likelihood affecting vertex i under the condition $\Delta_{ai} = 1$.

M-step

For all a, b pairs separately, we maximize the truncated binomial likelihood

$$p_{ab}^{\sum_{i,j:\, i\neq j} \pi_{ai}^{(t)} \pi_{bj}^{(t)} a_{ij}} \cdot (1 - p_{ab})^{\sum_{i,j:\, i\neq j} \pi_{ai}^{(t)} \pi_{bj}^{(t)} (1 - a_{ij})}$$

with respect to the parameter p_{ab}. Obviously, the maximum is attained by the following estimators of p_{ab}'s comprising the symmetric matrix $\mathbf{P}^{(t)}$:

$$p_{ab}^{(t)} = \frac{\sum_{i,j:\, i\neq j} \pi_{ai}^{(t)} \pi_{bj}^{(t)} a_{ij}}{\sum_{i,j:\, i\neq j} \pi_{ai}^{(t)} \pi_{bj}^{(t)}}, \quad 1 \leq a \leq b \leq k,$$

where edges connecting vertices of clusters a and b are counted fractionally, multiplied by the membership probabilities of their endpoints.

The ML-estimator of $\underline{\pi}$ in the t-th step is $\underline{\pi}^{(t)}$ of coordinates $\pi_a^{(t)} = \frac{1}{n}\sum_{i=1}^{n} \pi_{ai}^{(t)}$ $(a = 1, \dots, k)$, while that of the membership vector Δ_i is obtained by discrete maximization: $\Delta_{ai}^{(t)} = 1$, if $\pi_{ai}^{(t)} = \max_{b\in\{1,\dots,k\}} \pi_{bi}^{(t)}$ and 0, otherwise. (In case of ambiguity, the cluster with the smallest index is selected.) This choice of $\underline{\pi}$ will increase the likelihood.

The above algorithm is a special case of so-called Collaborative Filtering, see Hoffman, T., Puzicha, J., Ungar, L., Foster, D.

According to the general theory of EM-algorithm (Dempster, Laird, Rubin, J. R. Statist. Soc B 39, 1977), in exponential families (as in the present case), convergence to a local maximum can be guaranteed (depending on the starting values), but it runs in polynomial time in n.

Multiway cuts, modularities

Find community structure in large networks.
Communities/clusters/modules: inter- and intra-cluster connections mainly depend on the cluster memberships. They are are strongly or loosely connected subsets of vertices that can be identified with social groups or interacting enzymes in social or metabolic networks.
Modularities are non-parametric statistics calculated on the graph based on its adjacency matrix and maximized/minimized over k-partitions of the vertices.

Minimum multiway cut problems; ratio cut and normalized cut: communities with sparse between-cluster (and dense within-cluster) connections.

Modularity cuts: communities with more within-cluster (and less between-cluster) connections than expected under independence.

Spectral methods: looking for spectral gap in the Laplacian or modularity spectrum, then find the clusters by means of the eigenvectors, corresponding to the structural eigenvalues.

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:

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

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

$1 \le k \le n$
$P_k = (V_1, \dots, V_k)$: k-partition of the vertices
$V_1, \dots, V_k$: 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 $P_k = (V_1, \dots, V_k)$ 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 $P_k = (V_1, \dots, V_k)$ 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 $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$ and an integer $1 \le k \le n$, find the k-partition of the index set $\{1, \dots, n\}$ (or equivalently, the clustering of the points into k disjoint non-empty subsets) which minimizes the following k-variance:

$$S_k^2(\mathbf{x}_1, \dots, \mathbf{x}_n) = \min_{P_k \in \mathcal{P}_k} S_k^2(P_k, \mathbf{x}_1, \dots, \mathbf{x}_n)$$

$$= \min_{P_k = (V_1, \dots, V_k)} \sum_{a=1}^{k} \sum_{j \in V_a} \|\mathbf{x}_j - \mathbf{c}_a\|^2,$$

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

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

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

$$\tilde{S}_k^2(\mathbf{x}_1, \dots, \mathbf{x}_n) = \min_{P_k \in \mathcal{P}_k} \tilde{S}_k^2(P_k, \mathbf{x}_1, \dots, \mathbf{x}_n)$$

$$= \min_{P_k = (V_1, \dots, V_k)} \sum_{a=1}^k \sum_{j \in V_a} d_j \|\mathbf{x}_j - \mathbf{c}_a\|^2,$$

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

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

Ratio cut, partition matrices

P_k: $n \times k$ balanced partition matrix $\mathbf{Z}_k = (\mathbf{z}_1, \dots, \mathbf{z}_k)$
k-partition vector: $\mathbf{z}_a = (z_{1a}, \dots, z_{na})^T$, where
$z_{ia} = \frac{1}{\sqrt{|V_a|}}$, 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 $\mathbf{W}$:

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

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 $\mathbf{L}$: $0 = \lambda_1 < \lambda_2 \le \cdots \le \lambda_n$
unit-norm, pairwise orthogonal eigenvectors: $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$;
$\mathbf{u}_1 = \mathbf{1}/\sqrt{n}$

The discrete problem is relaxed to a continuous one:

$\mathbf{r}_1, \dots, \mathbf{r}_n \in \mathbb{R}^k$: representatives of the vertices
$\mathbf{X} = (\mathbf{r}_1, \dots, \mathbf{r}_n)^T = (\mathbf{x}_1, \dots, \mathbf{x}_k)$

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

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

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

and equality can be attained only in the $k = 1$ trivial case, otherwise the eigenvectors $\mathbf{u}_i$ $(i = 2, \dots, k)$ 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?

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

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 $\mathcal{F}_k = \operatorname{Span}\{\mathbf{z}_1, \dots, \mathbf{z}_k\}$.

Minimizing the normalized cut

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

$$f(P_k, \mathbf{W}) = \mathrm{tr}\, \mathbf{Z}_k^T \mathbf{L} \mathbf{Z}_k = \mathrm{tr}\, (\mathbf{D}^{1/2}\mathbf{Z}_k)^T \mathbf{L}_D (\mathbf{D}^{1/2}\mathbf{Z}_k) \qquad (4)$$

Normalized Laplacian eigenvalues (G is connected):
$0 = \lambda'_1 < \lambda'_2 \le \dots \le \lambda'_n \le 2$
eigenvalues of $\mathbf{L}_D$ with corresponding unit-norm, pairwise orthogonal eigenvectors $\mathbf{u}_1, \dots, \mathbf{u}_n$,
$\mathbf{u}_1 = (\sqrt{d_1}, \dots, \sqrt{d_n})^T$.
Continuous relaxation: $\mathbf{X} = (\mathbf{r}_1, \dots, \mathbf{r}_n)^T = (\mathbf{x}_1, \dots, \mathbf{x}_k)$

$$\min_{\mathbf{X}^T\mathbf{D}\mathbf{X}=\mathbf{I}_k} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_{ij} \|\mathbf{r}_i - \mathbf{r}_j\|^2 = \min_{\mathbf{X}^T\mathbf{D}\mathbf{X}=\mathbf{I}_k} \mathrm{tr}\, \mathbf{X}^T \mathbf{L} \mathbf{X} = \sum_{i=1}^{k} \lambda'_i$$

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

$$f_k(G) = \min_{\mathbf{Z}_k \in \mathcal{Z}_k^N} \operatorname{tr} \mathbf{Z}_k^T \mathbf{L} \mathbf{Z}_k \geq \sum_{i=1}^{k} \lambda_i'$$

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.

Spectral gap and variance

In B, Tusnády, Discrete Math., 1994

Theorem

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

$$\tilde{S}_2^2(X_2) \le \frac{\lambda_2'}{\lambda_3'}$$

Can it be generalized for $k > 2$?

Newman–Girvan modularity for edge-weighted graphs

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

Definition

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

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

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 given k we maximize $Q(P_k, \mathbf{W})$ over $\mathcal{P}_k$.
We want to penalize partitions with clusters of extremely different sizes or volumes

Definition

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

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

Definition

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

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

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

Maximizing the balanced Newman–Girvan modularity

$\mathbf{B} = \mathbf{W} - \mathbf{d}\mathbf{d}^T$: modularity matrix
$\mathbf{d} = (d_1, \dots, d_n)^T$
Spectrum: $\beta_1 \geq \cdots \geq \beta_p > 0 = \beta_{p+1} \geq \cdots \geq \beta_n$

$$\max_{P_k \in \mathcal{P}_k} Q_B(P_k, \mathbf{W}) \leq \sum_{a=1}^{k} \beta_a \leq \sum_{a=1}^{p+1} \beta_a .$$

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

Normalized modularity matrix

$\mathbf{B}_D = \mathbf{D}^{-1/2}\mathbf{B}\mathbf{D}^{-1/2} = \mathbf{I} - \mathbf{L}_D - \sqrt{\mathbf{d}}\sqrt{\mathbf{d}}^T$

$1 \geq \beta'_1 \geq \cdots \geq \beta'_n \geq -1$: spectrum of $\mathbf{B}_D$ (1 is not an eigenvalue if G is connected)

$\mathbf{u}'_1, \dots, \mathbf{u}'_n$: unit-norm, pairwise orthogonal eigenvectors

$\mathbf{u}'_1 = (\sqrt{d_1}, \dots, \sqrt{d_n})^T =: \sqrt{\mathbf{d}}$

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

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 $n_1, \dots, n_k$ 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} $(1 \leq i, j \leq n)$.

Wigner-noise

Definition

The $n \times n$ symmetric real matrix $\mathbf{W}$ is a Wigner-noise if its entries w_{ij}, $1 \le i \le j \le n$, are independent random variables, $\mathbf{E} w_{ij} = 0$, $\mathrm{Var}\, w_{ij} \le \sigma^2$ with some $0 < \sigma < \infty$ and the w_{ij}'s are uniformly bounded (there is a constant $K > 0$ such that $|w_{ij}| \le K$).

Füredi, Komlós (Combinatorica, 1981):

$$\max_{1 \le i \le n} |\lambda_i(\mathbf{W})| \le 2\sigma\sqrt{n} + O(n^{1/3} \log n)$$

with probability tending to 1 as $n \to \infty$.

Perturbation results for weighted graphs

$\mathbf{A} = \mathbf{B} + \mathbf{W}$, where
$\mathbf{W}$: $n \times n$ Wigner-noise
$\mathbf{B}$: $n \times n$ blown-up matrix of $\mathbf{P}$ with blow-up sizes $n_1, \dots, n_k$, $\sum_{i=1}^{k} n_i = n$.
$\mathbf{P}$: $k \times k$ pattern matrix
k is kept fixed as $n_1, \dots, n_k \to \infty$ "at the same rate": there is a constant c such that
$\frac{n_i}{n} \geq c, \quad i = 1, \dots k.$
growth rate condition: g.r.c.

Spectrum of a noisy graph

$G_n = (V, \mathbf{A})$, $\mathbf{A} = \mathbf{B} + \mathbf{W}$ is $n \times n$, $n \to \infty$
$\mathbf{B}$ induces a planted partition $P_k = (V_1, \dots, V_k)$ of V.
Weyl's perturbation theorem $\Longrightarrow$
Adjacency spectrum of G_n: under g.r.c. there are k structural eigenvalues of order n (in absolute value) and the others are $\mathcal{O}(\sqrt{n})$, almost surely.
The eigenvectors $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_k)$ corresponding to the structural eigenvalues are "not far" from the subspace of stepwise constant vectors on $P_k \Longrightarrow$

$$S_k^2(\mathbf{X}) \le S_k^2(P_k, \mathbf{X}) = \mathcal{O}(\frac{1}{n}), \quad \text{almost surely.}$$

Spectrum of the normalized Laplacian

$G_n = (V, \mathbf{A})$, $\mathbf{A} = \mathbf{B} + \mathbf{W}$ is $n \times n$, $n \to \infty$
$\mathbf{L}_D = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$

Theorem

There exists a positive number $\delta \in (0,1)$, independent of n, such that for every $0 < \tau < 1/2$ the following statement holds with probability tending to 1 as $n \to \infty$, under the g.r.c.:
there are exactly k eigenvalues of $\mathbf{L}_D$ that are located in the union of intervals $[-n^{-\tau}, 1-\delta+n^{-\tau}]$ and $[1+\delta-n^{-\tau}, 2+n^{-\tau}]$, while all the others are in the interval $(1-n^{-\tau}, 1+n^{-\tau})$.

Representation: $\mathbf{x}_i = \mathbf{D}^{-1/2}\mathbf{u}_i, \quad (i = 1, \dots, k)$

$$\tilde{S}_k^2(P_k, \mathbf{X}) \le \frac{k}{(\frac{\delta}{n^{-\tau}} - 1)^2} \quad \text{w. p. to 1 as } n \to \infty, \quad \text{under g.r.c.}$$

Noisy graph is a random simple graph with appropriate noise

The uniform bound K on the entries of $\mathbf{W}$ is such that $\mathbf{A} = \mathbf{B} + \mathbf{W}$ has entries in [0,1].
With an appropriate Wigner-noise the noisy matrix $\mathbf{A}$ is a generalized random graph: edges between V_i and V_j exist with probability $0 < p_{ij} < 1$.
For $1 \le i \le j \le k$ and $l \in V_i$, $m \in V_j$:

$$w_{lm} := \begin{cases} 1 - p_{ij}, & \text{with probability} \quad p_{ij} \\ -p_{ij} & \text{with probability} \quad 1 - p_{ij} \end{cases}$$

be independent random variables, otherwise $\mathbf{W}$ is symmetric. The entries have zero expectation and bounded variance:

$$\sigma^2 = \max_{1 \le i \le j \le k} p_{ij}(1 - p_{ij}) \le \frac{1}{4}.$$

Szemerédi's Regularity Lemma

For any graph on n vertices there exist a partition $(V_0, V_1, \ldots, V_k)$ of the vertices (here V_0 is a "small" exceptional set) such that "most" of the V_i, V_j pairs $(1 \le i < j \le k)$ are ε-regular with $\varepsilon > 0$ fixed in advance.

The pair V_i, V_j $(i \neq j)$ is ε-regular, if for any $A \subset V_i$, $B \subset V_j$ with $|A| > \varepsilon |V_i|$, $|B| > \varepsilon |V_j|$:

$$|\mathrm{dens}\,(A, B) - \mathrm{dens}\,(V_i, V_j)| < \varepsilon,$$

where

$$\mathrm{dens}\,(A, B) = \frac{e(A, B)}{|A| \cdot |B|}$$

is the edge-density between the disjoint vertex-sets A and B.

Informally, ε-regularity means that the edge-densities between the V_i, V_j pairs are homogeneous.

If the graph is sparse, then $k = 1$, otherwise k can be arbitrarily large (but it depends only on ε).

The planted partition is ε-regular almost surely

With the above Wigner-noise, $e(V_i, V_j)$ is the sum of $|V_i| \cdot |V_j|$ independent, identically distributed Bernoulli variables with parameter p_{ij} $(1 \le i, j \le k)$. Hence, $e(A, B)$ is binomially distributed with expectation $|A| \cdot |B| \cdot p_{ij}$ and variance $|A| \cdot |B| \cdot p_{ij}(1 - p_{ij})$.
By Chernoff's inequality for large deviations:

$$\begin{aligned}\mathbb{P}\left(|\mathrm{dens}\,(A, B) - p_{ij}| > \varepsilon\right) &\le e^{-\frac{\varepsilon^2 |A|^2 |B|^2}{2[|A||B| p_{ij}(1-p_{ij}) + \varepsilon |A||B|/3]}} \\ &= e^{-\frac{\varepsilon^2 |A||B|}{2[p_{ij}(1-p_{ij}) + \varepsilon/3]}} \\ &\le e^{-\frac{\varepsilon^4 |V_i||V_j|}{2[p_{ij}(1-p_{ij}) + \varepsilon/3}}\end{aligned}$$

that tends to 0, as $|V_i| = n_i \to \infty$ and $|V_j| = n_j \to \infty$. Hence, any pair V_i, V_j is ε-regular with probability tending to 1 if $n_1, \dots, n_k \to \infty$ under the g.r.c. (weaker than the structure guaranteed by Szemerédi's Lemma)

Recognizing the structure

Theorem

Let $\mathbf{A}_n$ be a sequence of $n \times n$ matrices, where $n \to \infty$. Assume that $\mathbf{A}_n$ has exactly k eigenvalues of order greater than $\sqrt{n}$, and there is a k-partition of the vertices such that the k-variance of the representatives is $\mathcal{O}(\frac{1}{n})$, in the representation with the corresponding eigenvectors. Then there is a blown-up matrix $\mathbf{B}_n$ such that $\mathbf{A}_n = \mathbf{B}_n + \mathbf{E}_n$ with $\|\mathbf{E}_n\| = \mathcal{O}(\sqrt{n})$.

Proof: construction by the cluster centers.

Results with planted partitions and cut-matrices or low-rank approximation of the column space of $\mathbf{A}$:

- Frieze, A., Kannan, R.
- McSherry, F.
- Amin Coja-Oghlan

Volume regularity

Definition

Let $G = (V, \mathbf{W})$ be weighted graph with $\mathtt{Vol}\,(V) = 1$. The disjoint pair (A, B) is α-*volume regular* if for all $X \subset A$, $Y \subset B$ we have

$$|e(X, Y) - \rho(A, B)\mathtt{Vol}\,(X)\mathtt{Vol}\,(Y)| \leq \alpha\sqrt{\mathtt{Vol}\,(A)\mathtt{Vol}\,(B)},$$

where $\rho(A, B) = \frac{e(A,B)}{\mathtt{Vol}\,(A)\mathtt{Vol}\,(B)}$ is the relative inter-cluster density of (A, B).

Result

Let $G = (V, \mathbf{W})$ be an edge-weighted graph on n vertices, with generalized degrees $d_1, \dots, d_n$ and degree matrix $\mathbf{D}$. Suppose that $\texttt{Vol}\,(V) = 1$ and there are no dominant vertices: $d_i = \Theta(1/n)$, $i = 1, \dots, n$ as $n \to \infty$. Let the eigenvalues of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, enumerated in decreasing absolute values, be

$$1 = \rho_1 > |\rho_2| \geq \cdots \geq |\rho_k| > \varepsilon \geq |\rho_i|, \quad i \geq k + 1.$$

The partition $(V_1, \dots, V_k)$ of V is defined so that it minimizes the weighted k-variance $s^2 = \tilde{S}_k^2(\mathbf{X})$ of the vertex representatives obtained as row vectors of the $n \times (k-1)$ matrix $\mathbf{X}$ of column vectors $\mathbf{D}^{-1/2}\mathbf{u}_i$, where $\mathbf{u}_i$ is the unit-norm eigenvector belonging to ρ_i $(i = 2, \dots, k)$. Then the (V_i, V_j) pairs are $\mathcal{O}(\sqrt{k}s + \varepsilon)$-volume regular $(i \neq j)$.

Further, for the clusters V_i $(i = 1, \dots, k)$ the following holds. For all $X, Y \subset V_i$:

$$|e(X,Y) - \rho(V_i)\texttt{Vol}\,(X)\texttt{Vol}\,(Y)| = \mathcal{O}(\sqrt{k}s + \varepsilon)\texttt{Vol}\,(V_i),$$

where $\rho(V_i) = \frac{e(V_i,V_i)}{\texttt{Vol}^{\,2}(V_i)}$ is the relative intra-cluster density of V_i.

In the $k = 2$ case, due to the relation between the 2-variance and the spectral gap, we are able to prove the following.
Let the eigenvalues of $\mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$, enumerated in decreasing absolute values, be

$$1 = \rho_1 > |\rho_2| = \delta > \varepsilon = |\rho_3| \geq |\rho_i|, \quad i \geq 4.$$

The partition (A, B) of V is defined in such a way that it minimizes the weighted 2-variance of the coordinates of $\mathbf{D}^{-1/2}\mathbf{u}_2$, where $\mathbf{u}_2$ is the unit-norm eigenvector belonging to ρ_2. Then the (A, B) pair is $\mathcal{O}(\sqrt{\frac{1-\delta}{1-\varepsilon}})$-volume regular.