Fundamental Limits of Spectral Clustering in Stochastic Block Models

Anderson Ye Zhang

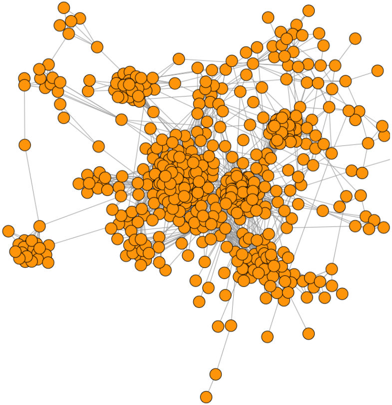
Department of Statistics and Data Science
University of Pennsylvania

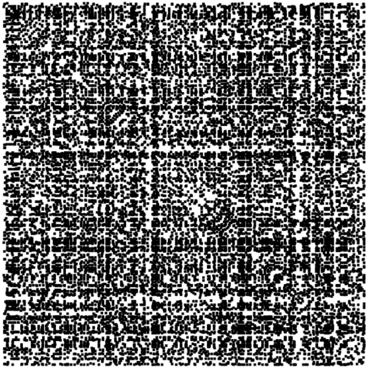
Outline

- Introduction to Community Detection and Spectral Clustering
- Sharp Statistical Analysis

Introduction to Community Detection and Spectral Clustering

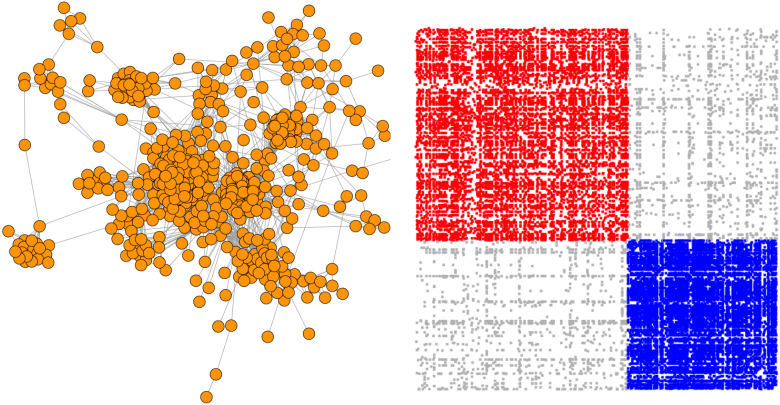
Networks and Community Detection





Human Gene-gene Co-association Network

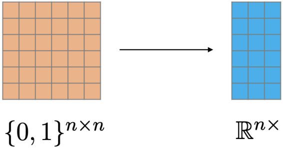
Networks and Community Detection



Human Gene-gene Co-association Network

Spectral Clustering

- Idea: dimension reduction and embedding.

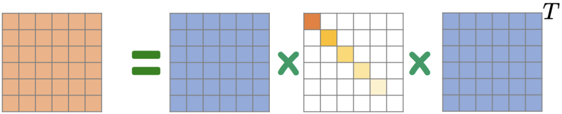


- Spectral Clustering: Spectral Decomposition + Clustering

Spectral Clustering

Input: Data matrix $A \in \{0,1\}^{n \times n}$, number of communities k

1. Perform eigendecomposition on A to have $A = \sum_i \lambda_i u_i u_i^T$.

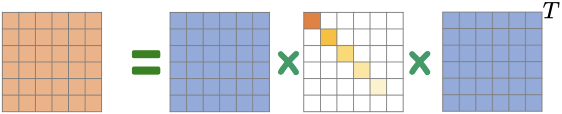


$A \in \{0,1\}^{n \times n}$

Spectral Clustering

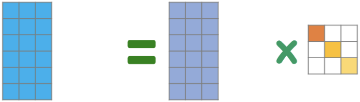
Input: Data matrix $A \in \{0,1\}^{n \times n}$, number of communities k

1. Perform eigendecomposition on A to have $A = \sum_i \lambda_i u_i u_i^T$.



$A \in \{0,1\}^{n \times n}$

2. Let $U = (u_1, \ldots, u_k) \in \mathbb{R}^{n \times k}$, $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_k\} \in \mathbb{R}^{k \times k}$.

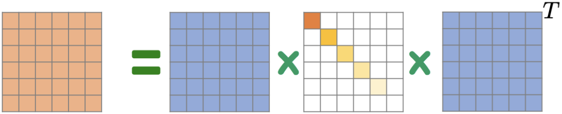


$U\Lambda \in \mathbb{R}^{n \times k}$ $\quad U \in \mathbb{R}^{n \times k}$ $\quad \Lambda \in \mathbb{R}^{k \times k}$

Spectral Clustering

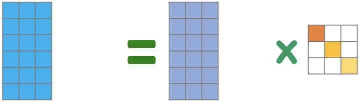
Input: Data matrix $A \in \{0,1\}^{n\times n}$, number of communities k

1. Perform eigendecomposition on A to have $A = \sum_i \lambda_i u_i u_i^T$.



$A \in \{0,1\}^{n\times n}$

2. Let $U = (u_1, \ldots, u_k) \in \mathbb{R}^{n\times k}$, $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_k\} \in \mathbb{R}^{k\times k}$.



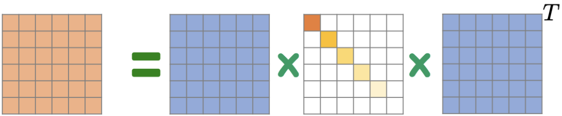
$U\Lambda \in \mathbb{R}^{n\times k}$ $\quad U \in \mathbb{R}^{n\times k}$ $\quad \Lambda \in \mathbb{R}^{k\times k}$

3. Apply k-means to rows of $U\Lambda \in \mathbb{R}^{n\times k}$.

Spectral Clustering

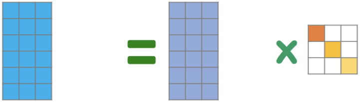
Input: Data matrix $A \in \{0,1\}^{n \times n}$, number of communities k

1. Perform eigendecomposition on A to have $A = \sum_i \lambda_i u_i u_i^T$.



$A \in \{0,1\}^{n \times n}$

2. Let $U = (u_1, \ldots, u_k) \in \mathbb{R}^{n \times k}$, $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_k\} \in \mathbb{R}^{k \times k}$.



$U\Lambda \in \mathbb{R}^{n \times k}$ $\quad U \in \mathbb{R}^{n \times k}$ $\quad \Lambda \in \mathbb{R}^{k \times k}$

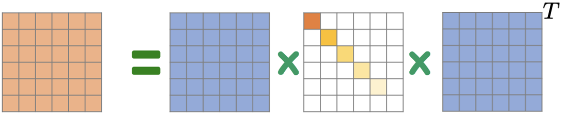
3. Apply k-means to rows of $U\Lambda \in \mathbb{R}^{n \times k}$.

   Remark: Eigenvectors are weighted as they are not equally important.

Spectral Clustering for Dense Networks

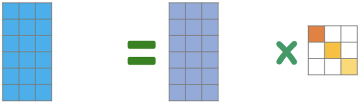
Input: Data matrix $A \in \{0,1\}^{n \times n}$, number of communities k

1. Perform eigendecomposition on A to have $A = \sum_i \lambda_i u_i u_i^T$.



$A \in \{0,1\}^{n \times n}$

2. Let $U = (u_1, \ldots, u_k) \in \mathbb{R}^{n \times k}$, $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_k\} \in \mathbb{R}^{k \times k}$.



$U\Lambda \in \mathbb{R}^{n \times k}$ $\quad U \in \mathbb{R}^{n \times k}$ $\quad \Lambda \in \mathbb{R}^{k \times k}$

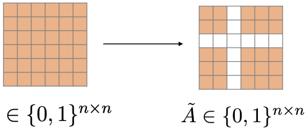
3. Apply k-means to rows of $U\Lambda \in \mathbb{R}^{n \times k}$.
   Remark: Needs an additional step for sparse networks.

Spectral Clustering for Sparse/Dense Networks

Input: Data matrix $A \in \{0,1\}^{n\times n}$, number of communities k

1. [Trim the network by removing high-degree nodes.] Let d_i be the degree of node i. Define $\tilde{A} \in \{0,1\}^{n\times n}$ such that

$$\tilde{A}_{i,j} = \begin{cases} A_{i,j}, & \text{if } d_i, d_j \le \tau, \\ 0, & \text{o.w.}. \end{cases}$$



2. Perform eigendecomposition on $\tilde{A}$ to have $\tilde{A} = \sum_i \lambda_i u_i u_i^T$.
3. Let $U = (u_1, \ldots, u_k) \in \mathbb{R}^{n\times k}$, $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_k\} \in \mathbb{R}^{k\times k}$.
4. Apply k-means to rows of $U\Lambda \in \mathbb{R}^{n\times k}$.

Spectral Clustering

- is computationally appealing
- often has remarkably good performance
- has been widely used in various problems

Q: Why does spectral clustering work? How well does it perform?

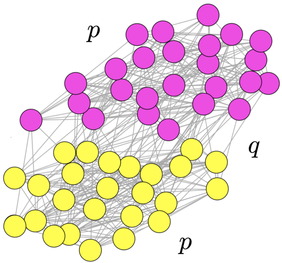
Sharp Statistical Analysis

Stochastic Block Model

- $A \in \{0,1\}^{n \times n}$
- k communities
- $z^* \in [k]^n$: underlying true community assignment vector
- Each edge is generated independently as follows:

$$\mathbb{E}A_{i,j} \sim \begin{cases} p, & \text{if } z_i^* = z_j^*, \\ q, & \text{o.w.} \end{cases}$$

  where $p > q$.
- Goal: Recover the community assignment z^*

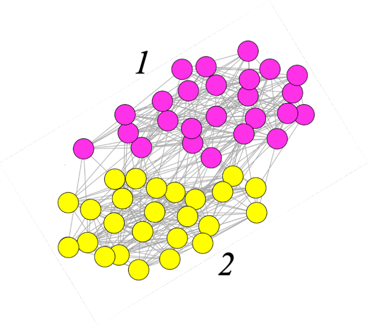


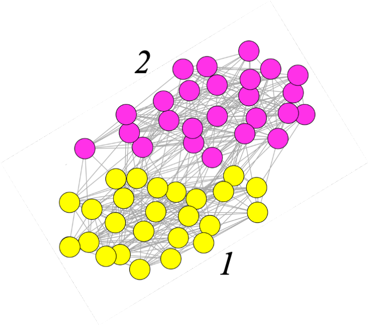
Stochastic Block Model

- Loss $\ell(\hat{z}, z^*)$: the proportion of nodes misclustered, considering all label permutations:

$$\ell(\hat{z}, z^*) = \frac{1}{n} \min_{\phi \in \Phi} \sum_{i \in [n]} \mathbb{I}\left\{\phi\left(\hat{z}_i\right) \neq z_i^*\right\},$$

where $\Phi = \{\phi :$ bijection from $[k]$ to $[k]\}$.





Assumptions

For simplicity, in this talk we assume

- The number of communities k is finite
- The communities sizes are all in the same order
- The probabilities p, q are in the same order

Polynomial Error Rate

Proposition

Assume $\frac{n(p-q)^2}{p} \to \infty$. We have w.h.p.

$$\ell(\hat{z}, z^*) \leq C\frac{p}{n(p-q)^2},$$

for some constant $C > 0$.

Remarks:

- $\frac{n(p-q)^2}{p}$ can be understood as the signal-to-noise ratio (SNR).
- $\ell(\hat{z}, z^*) \lesssim 1/\text{SNR}$.

Polynomial Error Rate

$$\ell(\hat{z}, z^*) \lesssim 1/\text{SNR}$$

Minimax Rate for Community Detection: If we consider all possible methods, how small the community detection error can be?

$$\exp\left(-c\text{SNR}\right)$$

In literature, spectral clustering is often used as an initialization for sophisticated algorithms to achieve the minimax rate.

Puzzling: But numerically such improvement is often marginal.

Q: Can we obtain a sharp upper bound for spectral clustering?

Exponential Error Rate

Theorem (Abbe, Fan, Wang, Zhong, 2020)

Assume $p = a\frac{\log n}{n}$ and $q = b\frac{\log n}{n}$ where a, b are constants. Assume the SBM has two equal-sized communities. Then

$$\mathbb{E}\ell(\hat{z}, z^*) \leq \exp\left(-(1+o(1))(\sqrt{a}-\sqrt{b})^2(\log n)/2\right)$$

Q: Can we study sparse SBMs?

Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong. Entrywise eigenvector analysis of random matrices with low expected rank. Annals of statistics, 48(3):1452, 2020

Main Result

Theorem (Z. 2023)

Assume $\frac{n(p-q)^2}{p} \to \infty$. We have

$$\mathbb{E}\ell(\hat{z}, z^*) \leq \exp\left(-(1-o(1))J_{\mathsf{min}}\right) + 2n^{-3},$$
$$\mathbb{E}\ell(\hat{z}, z^*) \geq \exp\left(-(1+o(1))J_{\mathsf{min}}\right) - 2n^{-3},$$

where J_{min} is a function of p, q, and the community sizes $n_1, n_2, \ldots, n_k$.

$2n^{-3}$ can be replaced by n^{-C} for an arbitrarily large constant $C > 0$, and in general is negligible.

Main Result

Theorem (Z. 2023)

Assume $\frac{n(p-q)^2}{p} \to \infty$. We have

$$\mathbb{E}\ell(\hat{z}, z^*) \leq \exp\left(-(1-o(1))J_{\text{min}}\right) + 2n^{-3},$$
$$\mathbb{E}\ell(\hat{z}, z^*) \geq \exp\left(-(1+o(1))J_{\text{min}}\right) - 2n^{-3},$$

where J_{min} is a function of p, q, and the community sizes $n_1, n_2, \ldots, n_k$.

Remarks:

- Holds for both sparse ($np \ll \log n$) and dense networks.
- Holds for multi-community and imbalanced SBMs.
- Matching lower and upper bounds.
- Case-wise analysis.

Exponents

J_{min} is a function of p, q, and community sizes $\{n_i\}_{i\in[k]}$.

- Definition:

$$\min_{1\le a\ne b\le k} \max_t \left((n_a - n_b)t\frac{p+q}{2} - n_a \log\left(qe^t + 1 - q\right) - n_b \log\left(pe^{-t} + 1 - p\right)\right)$$

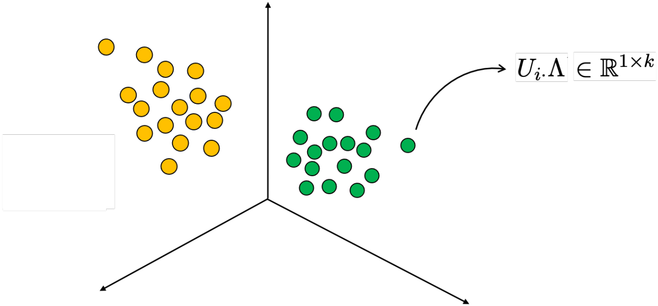
- Interpretation: tail probability of Bernoulli random variables

$$\min_{1\le a\ne b\le k} -\log \mathbb{P}\left(\sum_{i\in[n_a]} X_i - \sum_{j\in[n_b]} Y_j \ge (n_a - n_b)\frac{p+q}{2}\right) = (1+o(1))J_{\text{min}}$$

where $\{X_i\} \overset{iid}{\sim} \text{Ber}(q)$ and $\{Y_j\} \overset{iid}{\sim} \text{Ber}(p)$.

Intuition

Recall: Apply k-means to rows of $U\Lambda \in \mathbb{R}^{n\times k}$, ie., $\{U_{i\cdot}\Lambda\}_{i\in[n]} \in \mathbb{R}^{1\times k}$.



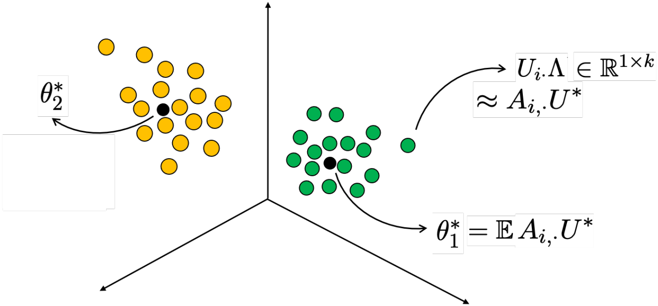
Intuition

Recall: Apply k-means to rows of $U\Lambda \in \mathbb{R}^{n\times k}$, ie., $\{U_{i\cdot}\Lambda\}_{i\in[n]} \in \mathbb{R}^{1\times k}$.

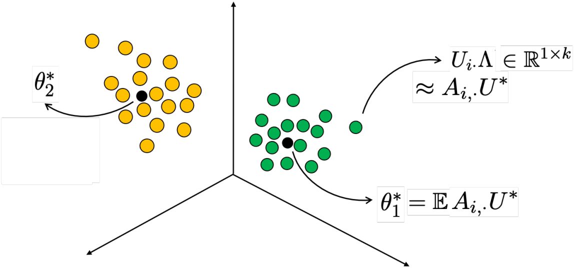
Since $U\Lambda = \tilde{A}U$, then

$$U_{i\cdot}\Lambda = \tilde{A}_{i,\cdot}U = A_{i,\cdot}U \approx A_{i,\cdot}U^*,$$

where U^* the leading eigenspace of of $\mathbb{E}A$.



Intuition



$$\mathbb{P}\left(i\text{th node wrongly clustered}\right) \approx \mathbb{P}\left(A_{i,\cdot}U^* \text{ is closer to } \theta_2^* \text{ than to } \theta_1^*\right)$$

$$=\mathbb{P}\left(\sum_{j:z_j^*=1} A_{ij} - \sum_{j\neq i:z_j^*=2} A_{ij} \geq (n_1-n_2)\frac{p+q}{2}\right)$$

$$=\mathbb{P}\left(\sum_{l\in[n_1]} X_l - \sum_{j\in[n_2]} Y_j \geq (n_1-n_2)\,\frac{p+q}{2}\right),$$

where $\{X_i\} \overset{iid}{\sim} \text{Ber}(q), \{Y_j\} \overset{iid}{\sim} \text{Ber}(p)$.

Technical Tool

Key technical tool: entrywise perturbation analysis for eigenvector/eigenspaces.

From previous slide:

$$U_{i\cdot}\Lambda = \tilde{A}_{i,\cdot}U = A_{i,\cdot}U \approx A_{i,\cdot}U^*,$$

Q: How to make $\approx$ rigorous?
For simplicity, consider the vector case

$$A_{i,\cdot}u \approx A_{i,\cdot}u^*$$

where u, u^* are the leading sample and population eigenvector.

Challenge: $A_{i,\cdot}$ and u are not independent.
Remedy: Use the leave-one-out technique

Technical Tool

For simplicity, consider the vector case

$$A_{i,\cdot}u \approx A_{i,\cdot}u^*$$

where u, u^* are the leading sample and population eigenvector.

If $A_{i,\cdot}$ and u were independent, then

$$A_{i,\cdot}u = \sum_j A_{i,j}u_j$$

would be a weighted average of Bernoulli random variables.
Challenge: Sharp tail probability for $A_{i,\cdot}u$.
The use of Bernstein inequality / Chernoff bound involves $\|u\|_\infty$, which comes with an $\log n$ factor, resulting in the assumption $np \geq \log n$ as in Abbe, Fan, Wang, Zhong, 2020.

Technical Tool

$$A_{i,\cdot}u = \sum_j A_{i,j} u_j.$$

To avoid the appearance of $\|u\|_\infty$, we truncate the eigenvectors:

$$u_j = u_j \mathbb{I}\left\{|u_j| \le t\right\} + \left(u_j - u_j \mathbb{I}\left\{|u_j| \le t\right\}\right).$$

Then

$$\begin{aligned} A_{i,\cdot}u = & \sum_{j \in [n]} A_{i,j} u_j \mathbb{I}\left\{|u_j| \le t\right\} \\ & + \sum_{j \in [n]} A_{i,j} \left(u_j - u_j \mathbb{I}\left\{|u_j| \le t\right\}\right) \end{aligned}$$

Technical Tool

$$A_{i,\cdot}u = \sum_{j\in[n]} A_{i,j}u_j\mathbb{I}\left\{|u_j| \le t\right\}$$

Chernoff bound can now be applied.

The ℓ_∞ norm of the truncated eigenvector is t.

$$+\sum_{j\in[n]} A_{i,j}\left(u_j - u_j\mathbb{I}\left\{|u_j| \le t\right\}\right)$$

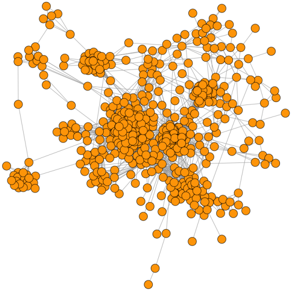
Related to $\sum_{j\in[n]} u_j^2\mathbb{I}\left\{|u_j| > t\right\}$, a truncated

ℓ_2 norm of u. Can be shown to be negligible.

$\approx A_{i,\cdot}u^*$

Novelty: an "eigenvector truncation" idea and a truncated ℓ_2 perturbation analysis.

Summary



- Sharp analysis for the performance of spectral clustering under SBMs
- Works for sparse networks
- Exponential error rates

Anderson Ye Zhang. Fundamental limits of spectral clustering in stochastic block models.
arXiv preprint arXiv:2301.09289, 2023

Thank You