Joint Spectral Clustering in Multilayer Degree-Corrected Stochastic Blockmodels

Joshua Agterberg

Johns Hopkins Mathematical Institute for Data Science

February, 2023

Funding partially provided by Acheson J. Duncan Advancement of Research in Statistics Fund, MINDS Fellowship, and Counselman Fellowship

Joint Work With:

Zachary Lubberts (JHU)

Jesús Arroyo (Texas A&M)

Outline

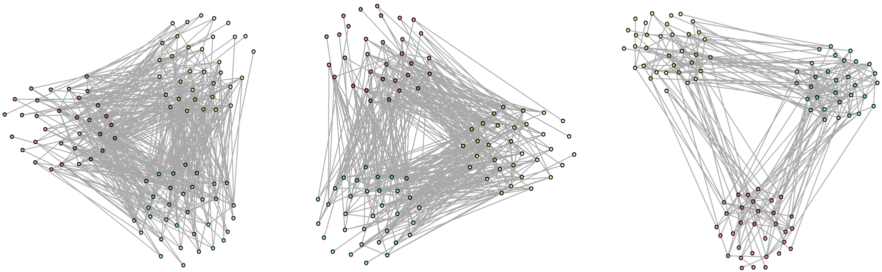
Outline

1. Motivation
2. Our Procedure
3. Theoretical Results
4. Numerical Results
5. Conclusion

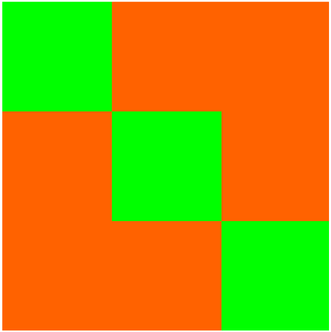
Multilayer Networks

Setting

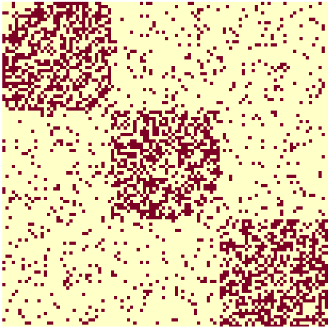
Observe L networks on same n vertices



Stochastic Blockmodels



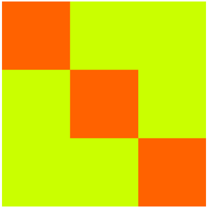
(a) Population Adjacency Matrix

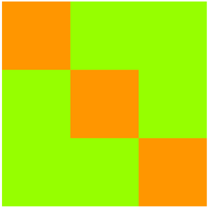


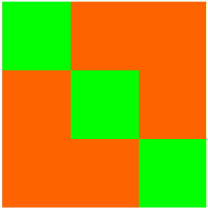
(b) Observed Adjacency Matrix

Multilayer Networks

Multilayer SBM (e.g. (Lei and Lin, 2022; Lei et al., 2020)) incorporates network-level idiosyncracies:

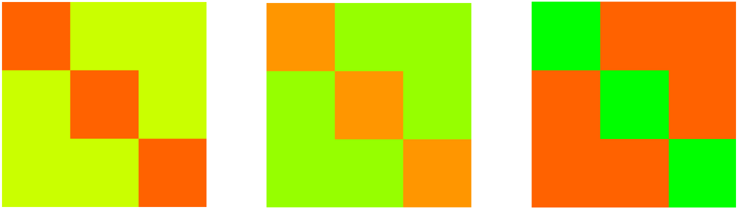






Multilayer Networks

Multilayer SBM (e.g. (Lei and Lin, 2022; Lei et al., 2020)) incorporates network-level idiosyncracies:

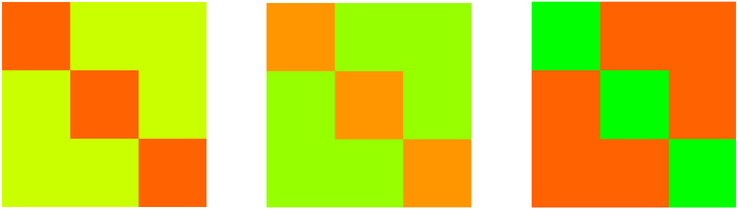


Other examples:

- Mixture of multilayer SBMs (Jing et al., 2021; Fan et al., 2021)
- DIMPLE model (Pensky and Wang, 2021; Noroozi and Pensky, 2022)
- COSIE model (Arroyo et al., 2021)
- More... (Young et al., 2022; Jones and Rubin-Delanchy, 2020)

Multilayer Networks

Multilayer SBM (e.g. (Lei and Lin, 2022; Lei et al., 2020)) incorporates network-level idiosyncracies:



Other examples:

- Mixture of multilayer SBMs (Jing et al., 2021; Fan et al., 2021)
- DIMPLE model (Pensky and Wang, 2021; Noroozi and Pensky, 2022)
- COSIE model (Arroyo et al., 2021)
- More... (Young et al., 2022; Jones and Rubin-Delanchy, 2020)

Problem

Vertices act similarly between networks!

Multilayer DCSBM

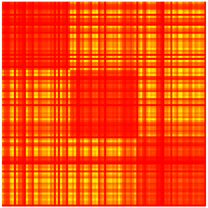
Solution

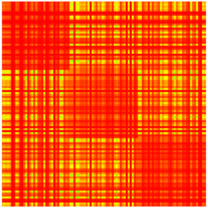
Introduce network-specific degree corrections

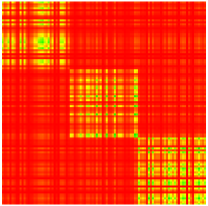
Multilayer DCSBM

Solution

Introduce network-specific degree corrections



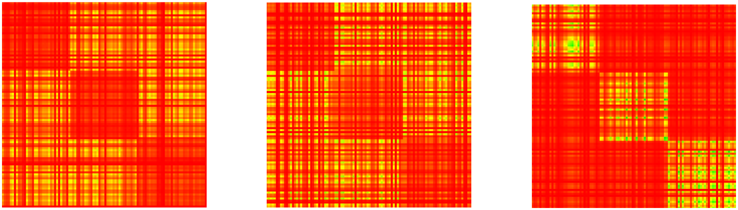




Multilayer DCSBM

Solution

Introduce network-specific degree corrections



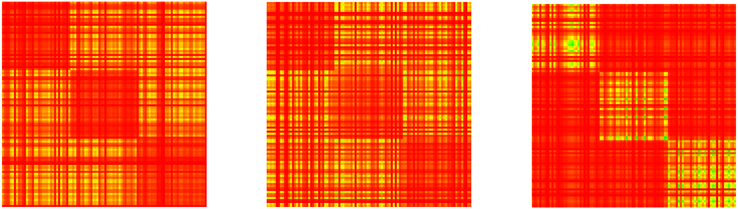
Maintains:

1. Global community structure
2. Network-level idiosyncrasies
3. Vertex-level idiosyncrasies

Multilayer DCSBM

Solution

Introduce network-specific degree corrections



Maintains:

1. Global community structure
2. Network-level idiosyncrasies
3. Vertex-level idiosyncrasies

Inference Task

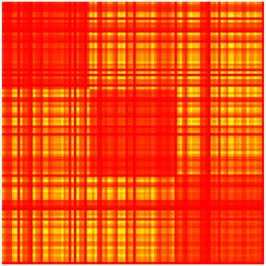
Recover communities!

Outline

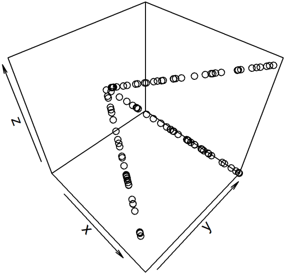
Spectral Geometry

Observation 1

Each population network is rank K, with rows of scaled eigenvectors supported on one of K rays, where K is the number of communities.



Population adjacency matrix

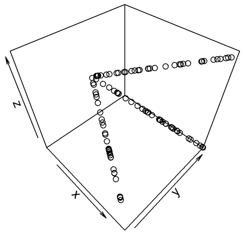


Rows of scaled eigenvectors of population adjacency matrix, viewed as points in dimension $K = 3$

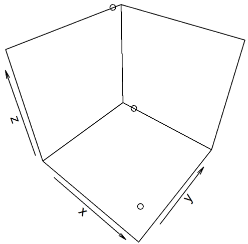
Spectral Geometry

Observation 2

Projecting each ray to the sphere results in community memberships for a single network.



Rows of scaled eigenvectors of population adjacency matrix, viewed as points in dimension $K = 3$



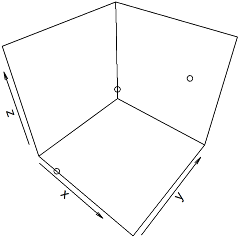
Row-normalized scaled eigenvectors of population adjacency matrix, viewed as points in dimension $K = 3$

Spectral Geometry

Observation 3

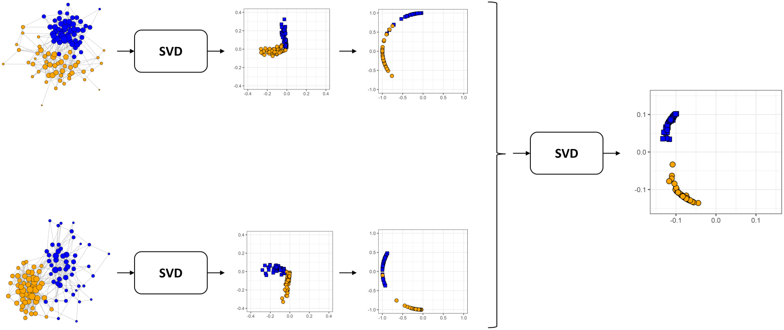
$n \times LK$ matrix of concatenated row-normalized embedding has left singular subspace that reveals community memberships for all networks.

$n \times LK$ matrix of concatenated row-normalized embedding.

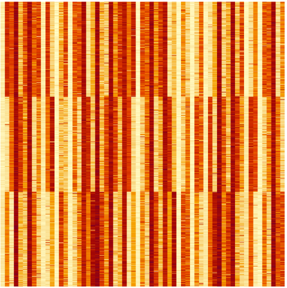


Rows of left singular vectors viewed as points in dimension $K = 3$

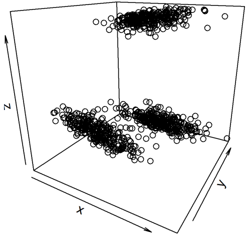
DC-MASE: Degree-Corrected Multiple Adjacency Spectral Embedding



DC-MASE: Degree-Corrected Multiple Adjacency Spectral Embedding



$n \times LK$ matrix of concatenated row-normalized embedding of each observed adjacency matrix.



Rows of left singular vectors of the matrix on the left, viewed as points in dimension $K = 3$

Outline

Theoretical Guarantees

Theorem (Informal Restatement of Theorem 1 of Agterberg et al. (2022))

*Consider a multilayer DCSBM with each network having **the same signal strength**, and suppose each edge probability matrix has rank K and that each node has degree correction parameter θ_i. Let $\widehat{z}$ denote the output of clustering with K-means on the rows of the output of DC-MASE, and define*

$$\ell(\widehat{z}, z) := \frac{\#\text{misclustered nodes}}{n}.$$

Then

$$\mathbb{E}\ell(\widehat{z}, z) \leq \frac{2K}{n} \sum_{i=1}^{n} \exp\left(-cL\theta_i \times (\mathrm{SNR}\text{-like term}) \right).$$

Note: the same signal strength condition is not required in the main result.

Improving Estimation?

- Has been demonstrated that vanilla spectral clustering (using a slightly different procedure) achieves the error rate:

$$\mathbb{E}\ell(\widehat{z}, z) \leq \frac{2K}{n} \sum_{i=1}^{n} \exp\Big(- c\theta_i \times (\mathrm{SNR}\text{-like term})\Big).$$

- In contrast, our results show that

$$\mathbb{E}\ell(\widehat{z}, z) \leq \frac{2K}{n} \sum_{i=1}^{n} \exp\Big(- cL\theta_i \times (\mathrm{SNR}\text{-like term})\Big).$$

Improving Estimation?

- Has been demonstrated that vanilla spectral clustering (using a slightly different procedure) achieves the error rate:

$$\mathbb{E}\ell(\widehat{z}, z) \leq \frac{2K}{n} \sum_{i=1}^{n} \exp\Big(- c\theta_i \times (\text{SNR-like term})\Big).$$

- In contrast, our results show that

$$\mathbb{E}\ell(\widehat{z}, z) \leq \frac{2K}{n} \sum_{i=1}^{n} \exp\Big(- cL\theta_i \times (\text{SNR-like term})\Big).$$

Key Takeaway

Multiple networks *improve* estimation guarantees (relative to the single network setting).

Outline

Simulated Data

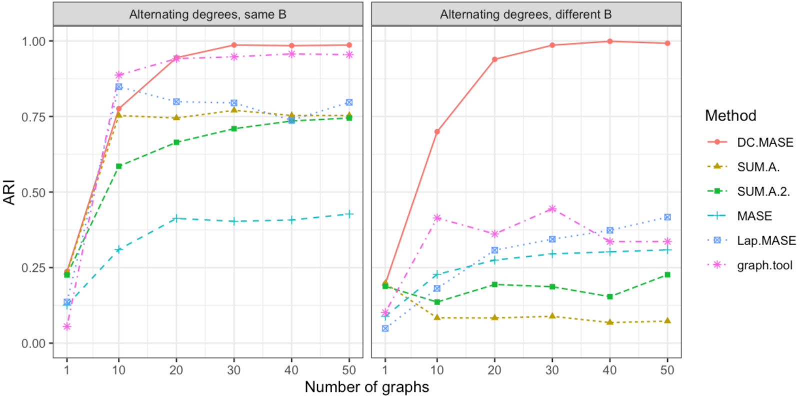


Figure: Performance of DC-MASE on networks with severe degree heterogeneity between networks. In each community, half of the degree corrections are one value, and half of the degrees are another value, and these switch between networks.

Flight Network Data

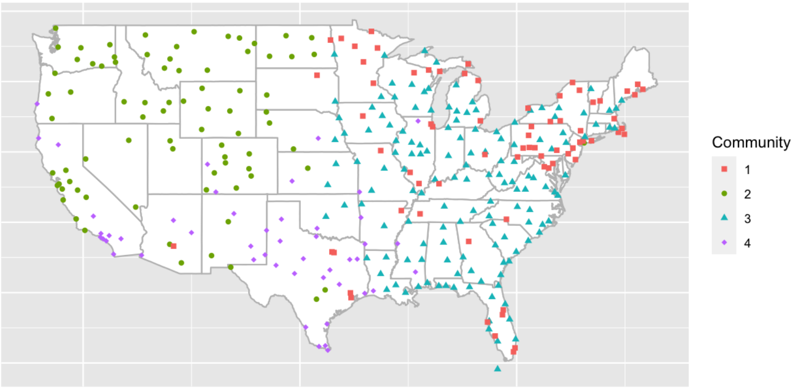


Figure: Network of flights– each network is a month of flights between airports

Tracking Flights

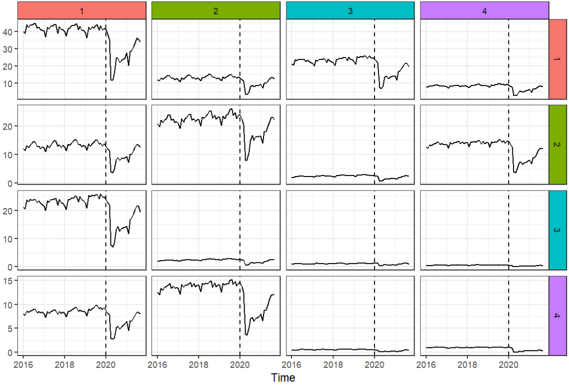


Figure: Average flights in each community over time

Tracking Degree-Corrections

Figure: Estimated degree-corrections within each community over time

Tracking Degree-Corrections

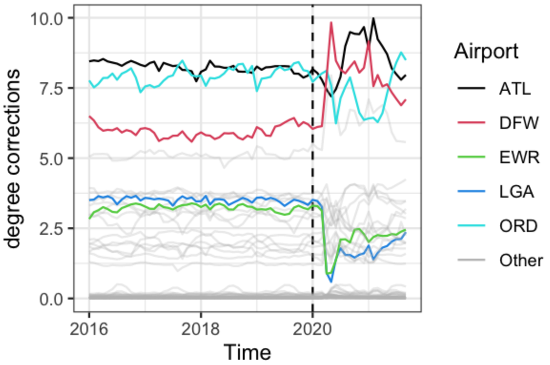


Figure: Estimated Degree-Corrections for some airports in the first community over time

Outline

Conclusion

- The Multilayer DCSBM model allows for:
  - Global community structure
  - Layer heterogeneity in the connection probabilities between clusters
  - Vertex heterogeneity within and between networks

Conclusion

- The Multilayer DCSBM model allows for:
    - Global community structure
    - Layer heterogeneity in the connection probabilities between clusters
    - Vertex heterogeneity within and between networks
- DC-MASE accounts for vertex and layer heterogeneity but still benefits from multiple observations and shared community structure

Conclusion

- The Multilayer DCSBM model allows for:
  - Global community structure
  - Layer heterogeneity in the connection probabilities between clusters
  - Vertex heterogeneity within and between networks
- DC-MASE accounts for vertex and layer heterogeneity but still benefits from multiple observations and shared community structure

References I

Joshua Agterberg, Zachary Lubberts, and Jesús Arroyo. Joint Spectral Clustering in Multilayer Degree-Corrected Stochastic Blockmodels, December 2022. arXiv:2212.05053 [math, stat].

Jesús Arroyo, Avanti Athreya, Joshua Cape, Guodong Chen, Carey E. Priebe, and Joshua T. Vogelstein. Inference for Multiple Heterogeneous Networks with a Common Invariant Subspace. *Journal of Machine Learning Research*, 22(142): 1–49, 2021. ISSN 1533-7928.

Xing Fan, Marianna Pensky, Feng Yu, and Teng Zhang. ALMA: Alternating Minimization Algorithm for Clustering Mixture Multilayer Network, October 2021. arXiv:2102.10226 [cs, math, stat].

References II

Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Improvements on SCORE, Especially for Weak Signals. *Sankhya A*, March 2021. ISSN 0976-836X, 0976-8378. doi: 10.1007/s13171-020-00240-1.

Bing-Yi Jing, Ting Li, Zhongyuan Lyu, and Dong Xia. Community detection on mixture multilayer networks via regularized tensor decomposition. *The Annals of Statistics*, 49(6):3181–3205, December 2021. ISSN 0090-5364, 2168-8966. doi: 10.1214/21-AOS2079.

Andrew Jones and Patrick Rubin-Delanchy. The multilayer random dot product graph. *arXiv:2007.10455 [cs, stat]*, July 2020.

References III

Jing Lei and Kevin Z. Lin. Bias-Adjusted Spectral Clustering in Multi-Layer Stochastic Block Models. *Journal of the American Statistical Association*, 0(0):1–13, March 2022. ISSN 0162-1459. doi: 10.1080/01621459.2022.2054817.

Jing Lei, Kehui Chen, and Brian Lynch. Consistent community detection in multi-layer network data. *Biometrika*, 107(1): 61–73, March 2020. ISSN 0006-3444. doi: 10.1093/biomet/asz068.

Majid Noroozi and Marianna Pensky. Sparse Subspace Clustering in Diverse Multiplex Network Model, June 2022. arXiv:2206.07602 [cs, stat].

Marianna Pensky and Yaxuan Wang. Clustering of Diverse Multiplex Networks. *arXiv:2110.05308 [stat]*, October 2021.

References IV

Jean-Gabriel Young, Alec Kirkley, and M. E. J. Newman. Clustering of heterogeneous populations of networks. *Physical Review E*, 105(1):014312, January 2022. doi: 10.1103/PhysRevE.105.014312.

Thank you!

: @JAgterberger

Multilayer DCSBM

Formally:

- Observe L (symmetric) adjacency matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)} \in \{0,1\}^{n \times n}$ on same n vertices

Multilayer DCSBM

Formally:

- Observe L (symmetric) adjacency matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)} \in \{0,1\}^{n \times n}$ on same n vertices
- Each vertex belongs to one of K communities; set $z : [n] \to [K]$ as the assignment vector

Multilayer DCSBM

Formally:

- Observe L (symmetric) adjacency matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)} \in \{0,1\}^{n \times n}$ on same n vertices
- Each vertex belongs to one of K communities; set $z : [n] \to [K]$ as the assignment vector
- Upper triangles of each adjacency matrix satisfy

$$\mathbf{A}_{ij}^{(l)} \sim \mathrm{Bernoulli}(\theta_i^{(l)} \theta_j^{(l)} \mathbf{B}_{z(i),z(j)}^{(l)})$$

  independently

Multilayer DCSBM

Formally:

- Observe L (symmetric) adjacency matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)} \in \{0,1\}^{n \times n}$ on same n vertices
- Each vertex belongs to one of K communities; set $z : [n] \to [K]$ as the assignment vector
- Upper triangles of each adjacency matrix satisfy

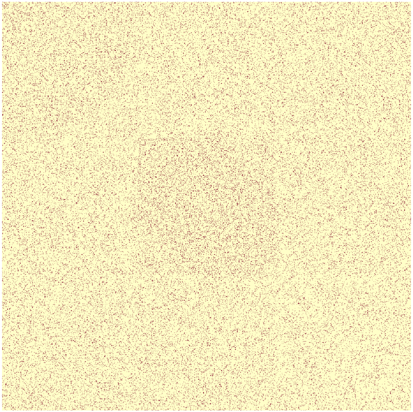
$$\mathbf{A}_{ij}^{(l)} \sim \text{Bernoulli}(\theta_i^{(l)} \theta_j^{(l)} \mathbf{B}_{z(i),z(j)}^{(l)})$$

  independently
- Each matrix $\mathbf{A}^{(l)}$ satisfies

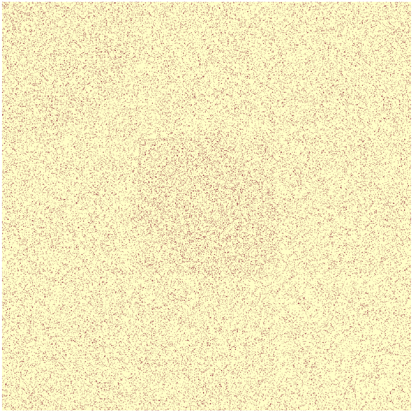
$$\mathbb{E}\mathbf{A}^{(l)} = \Theta^{(l)} \mathbf{Z} \mathbf{B}^{(l)} \mathbf{Z}^\top \Theta^{(l)},$$

  where $\mathbf{Z}_{ik} = 1$ if $z(i) = k$

Community Separation



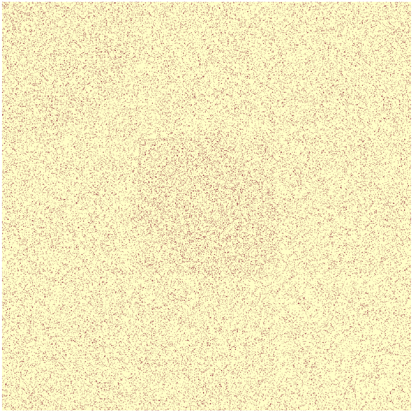
Community Separation



Two important quantities:

- $\lambda_{\min}^{(l)}$ = absolute value of smallest nonzero eigenvalue of $\mathbf{B}^{(l)}$.

Community Separation

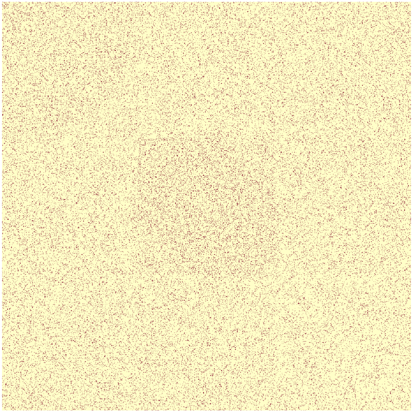


Two important quantities:

- $\lambda_{\min}^{(l)}$ = absolute value of smallest nonzero eigenvalue of $\mathbf{B}^{(l)}$. Example:

$$\mathbf{B}^{(l)} = \begin{pmatrix} 1 & 1-\lambda \\ 1-\lambda & 1 \end{pmatrix}$$

Community Separation



Two important quantities:

- $\lambda_{\min}^{(l)}$ = absolute value of smallest nonzero eigenvalue of $\mathbf{B}^{(l)}$. Example:

$$\mathbf{B}^{(l)} = \begin{pmatrix} 1 & 1-\lambda \\ 1-\lambda & 1 \end{pmatrix}$$

- Average community separation

$$\bar{\lambda} = \frac{1}{L} \sum_{l=1}^{L} \lambda_{\min}.$$

Degree Heterogeneity

- $\frac{\|\theta^{(l)}\|_3^3}{\|\theta^{(l)}\|^4}$ and $\frac{\theta^{(l)}_{\max}}{\|\theta^{(l)}\|^2}$ measure *degree heterogeneity relative to sparsity*

Degree Heterogeneity

- $\frac{\|\theta^{(l)}\|_3^3}{\|\theta^{(l)}\|^4}$ and $\frac{\theta^{(l)}_{\max}}{\|\theta^{(l)}\|^2}$ measure *degree heterogeneity relative to sparsity* Why?
    - First term satisfies

$$\frac{\theta^{(l)}_{\min}}{\|\theta^{(l)}\|^2} \leq \frac{\|\theta^{(l)}\|_3^3}{\|\theta^{(l)}\|^2} \leq \frac{\theta^{(l)}_{\max}}{\|\theta^{(l)}\|^2},$$

and is tightest when $\theta^{(l)}_{\max} = \theta^{(l)}_{\min}$.

Degree Heterogeneity

- $\frac{\|\theta^{(l)}\|_3^3}{\|\theta^{(l)}\|^4}$ and $\frac{\theta^{(l)}_{\max}}{\|\theta^{(l)}\|^2}$ measure *degree heterogeneity relative to sparsity* Why?
  - First term satisfies
$$\frac{\theta^{(l)}_{\min}}{\|\theta^{(l)}\|^2} \leq \frac{\|\theta^{(l)}\|_3^3}{\|\theta^{(l)}\|^2} \leq \frac{\theta^{(l)}_{\max}}{\|\theta^{(l)}\|^2},$$
    and is tightest when $\theta^{(l)}_{\max} = \theta^{(l)}_{\min}$.
  - Second term satisfies
$$\frac{1}{n} \leq \frac{\theta^{(l)}_{\max}}{\|\theta^{(l)}\|^2} \leq \frac{1}{n}\left(\frac{\theta^{(l)}_{\max}}{\theta^{(l)}_{\min}}\right)$$
    and is tightest when $\theta^{(l)}_{\max} = \theta^{(l)}_{\min}$.

Misclassification Error Rate

Loss Function

Define:

$$\ell(\widehat{z}, z) := \min_{\text{Permutations } \mathcal{P}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{\widehat{z}(i) \neq \mathcal{P}(z(i))\}.$$

Misclassification Error Rate

Loss Function

Define:

$$\ell(\widehat{z}, z) := \min_{\text{Permutations } \mathcal{P}} \frac{1}{n} \sum_{i=1}^{n} \mathbb{I}\{\widehat{z}(i) \neq \mathcal{P}(z(i))\}.$$

Provide bounds on *expected* error rate in terms of

- K and L;
- Degree heterogeneity;
- Community separation;
- Relationship between degree heterogeneity and community separation

Main Result

Theorem (Agterberg et al. (2022))

Suppose certain individual network signal-strength conditions hold. Let $\widehat{z}$ denote the output of K-Means with DC-MASE. Define

$$\mathrm{err}_{\mathrm{ave}}^{(i)} := \frac{1}{L} \sum_{l} \frac{\|\theta^{(l)}\|_3^3}{\theta_i^{(l)} \|\theta^{(l)}\|^4 (\lambda_{\min}^{(l)})^2}; \qquad \left(\begin{array}{c}\textit{Average Degree Heterogeneity}\\ \textit{and Community Separation}\end{array}\right)$$

$$\mathrm{err}_{\max}^{(i)} := \max_{l} \frac{\theta_{\max}^{(l)}}{\theta_i^{(l)} \|\theta^{(l)}\|^2 \lambda_{\min}^{(l)}}. \qquad \left(\begin{array}{c}\textit{Worst-Case Degree Heterogeneity}\\ \textit{and Community Separation}\end{array}\right)$$

Then

$$\mathbb{E}\ell(\widehat{z}, z) \leq \frac{2K}{n} \sum_{i=1}^{n} \exp\left(- cL \min\left\{ \frac{\bar{\lambda}^2}{K^2 \mathrm{err}_{\mathrm{ave}}^{(i)}}, \frac{\bar{\lambda}}{K \mathrm{err}_{\max}^{(i)}} \right\}\right) + O(n^{-10}).$$

Network Homogeneity

Corollary (Agterberg et al. (2022))

Suppose that $\lambda_{\min}^{(l)} = \lambda_{\min}$ and $\theta_i^{(l)} = \theta_i$ for all l. Then

$$\mathbb{E}\ell(\widehat{z}, z) \leq \frac{2K}{n} \sum_{i=1}^{n} \exp\left(-cL\theta_i \min\left\{ \frac{\|\theta\|^4 \lambda_{\min}^4}{K^2 \|\theta\|_3^3}, \frac{\|\theta\|^2 \lambda_{\min}^2}{K\theta_{\max}} \right\} \right) + O(n^{-10}).$$

Network Homogeneity

Corollary (Agterberg et al. (2022))

Suppose that $\lambda_{\min}^{(l)} = \lambda_{\min}$ *and* $\theta_i^{(l)} = \theta_i$ *for all* l*. Then*

$$\mathbb{E}\ell(\widehat{z}, z) \leq \frac{2K}{n} \sum_{i=1}^{n} \exp\left(-cL\theta_i \min\left\{ \frac{\|\theta\|^4 \lambda_{\min}^4}{K^2 \|\theta\|_3^3}, \frac{\|\theta\|^2 \lambda_{\min}^2}{K\theta_{\max}} \right\}\right) + O(n^{-10}).$$

Key Takeaway

More networks (bigger L) means better misclustering error!

Comparison to One Network

Error rate in Jin et al. (2021) shows that for small K,

$$\mathbb{E}\ell(\widehat{z}, z) \lesssim \frac{1}{n}\sum_{i=1}^{n} \exp\left(- c\theta_i \min\left\{ \frac{\|\theta\|^4 \lambda_{\min}^2}{\|\theta\|_3^3}, \frac{\|\theta\|^2 \lambda_{\min}}{\theta_{\max}} \right\}\right).$$

Our rate:

$$\mathbb{E}\ell(\widehat{z}, z) \lesssim \frac{1}{n}\sum_{i=1}^{n} \exp\left(- cL\theta_i \min\left\{ \frac{\|\theta\|^4 \lambda_{\min}^4}{\|\theta\|_3^3}, \frac{\|\theta\|^2 \lambda_{\min}^2}{\theta_{\max}} \right\}\right).$$

Comparison to One Network

Error rate in Jin et al. (2021) shows that for small K,

$$\mathbb{E}\ell(\widehat{z}, z) \lesssim \frac{1}{n}\sum_{i=1}^{n} \exp\left(- c\theta_i \min\left\{ \frac{\|\theta\|^4 \lambda_{\min}^2}{\|\theta\|_3^3}, \frac{\|\theta\|^2 \lambda_{\min}}{\theta_{\max}} \right\}\right).$$

Our rate:

$$\mathbb{E}\ell(\widehat{z}, z) \lesssim \frac{1}{n}\sum_{i=1}^{n} \exp\left(- cL\theta_i \min\left\{ \frac{\|\theta\|^4 \lambda_{\min}^4}{\|\theta\|_3^3}, \frac{\|\theta\|^2 \lambda_{\min}^2}{\theta_{\max}} \right\}\right).$$

Key Takeaway

More networks (bigger L) means better misclustering error!

Comparison to One Network

Error rate in Jin et al. (2021) shows that for small K,

$$\mathbb{E}\ell(\widehat{z}, z) \lesssim \frac{1}{n}\sum_{i=1}^{n} \exp\left(- c\theta_i \min\left\{ \frac{\|\theta\|^4 \lambda_{\min}^2}{\|\theta\|_3^3}, \frac{\|\theta\|^2 \lambda_{\min}}{\theta_{\max}} \right\}\right).$$

Our rate:

$$\mathbb{E}\ell(\widehat{z}, z) \lesssim \frac{1}{n}\sum_{i=1}^{n} \exp\left(- cL\theta_i \min\left\{ \frac{\|\theta\|^4 \lambda_{\min}^4}{\|\theta\|_3^3}, \frac{\|\theta\|^2 \lambda_{\min}^2}{\theta_{\max}} \right\}\right).$$

Key Takeaway

More networks (bigger L) means better misclustering error! *And well-separated communities (larger $\lambda_{\min}$) means better misclustering error.*

Theory: Sparse Networks

Corollary (Agterberg et al. (2022))

Suppose that $\theta_i \asymp \sqrt{\rho_n}$ for all i and l. Then

$$\mathbb{E}\ell(\widehat{z}, z) \leq 2K \exp\left(- cL\lambda_{\min}^4 n\rho_n \right) + O(n^{-10}).$$

Theory: Sparse Networks

Corollary (Agterberg et al. (2022))

Suppose that $\theta_i \asymp \sqrt{\rho_n}$ for all i and l. Then

$$\mathbb{E}\ell(\widehat{z}, z) \leq 2K \exp\left(- cL\lambda_{\min}^4 n\rho_n \right) + O(n^{-10}).$$

Key Takeaway

More networks (bigger L) means better misclustering error! And well-separated communities (larger $\lambda_{\min}$) means better misclustering error

Theory: Sparse Networks

Corollary (Agterberg et al. (2022))

Suppose that $\theta_i \asymp \sqrt{\rho_n}$ for all i and l. Then

$$\mathbb{E}\ell(\widehat{z}, z) \leq 2K \exp\Big(- cL\lambda_{\min}^4 n\rho_n \Big) + O(n^{-10}).$$

Key Takeaway

More networks (bigger L) means better misclustering error! And well-separated communities (larger $\lambda_{\min}$) means better misclustering error *even when each network is quite sparse (e.g. $n\rho_n \asymp \log(n)$).*