Joint Spectral Clustering in Multilayer Networks

Jesús Arroyo

XVI CLAPEM 2023

São Paulo, Brazil

July 11th, 2023

Texas A&M University

Acknowledgements

Joint work with:

Joshua Agterberg (UPenn)

Zachary Lubberts (JHU)

Preprint: J. Agterberg, Z. Lubberts, J. Arroyo, "Joint Spectral Clustering in Multilayer Degree-Corrected Stochastic Blockmodels", *arXiv 2212.05053*.

Community detection

- Networks often exhibit community structure (Girvan and Newman, 2002)
- Many methods to detect communities (Abbe, 2017, Fortunato 2010): modularity maximization, likelihood-based approaches, convex relaxations, random walks... This talk: spectral methods

Comment | Published: 29 July 2022

20 years of network community detection

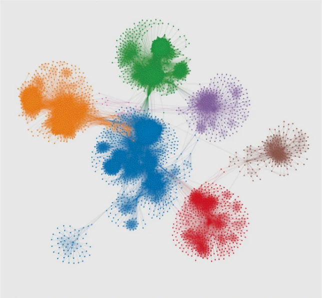
Santo Fortunato & Mark E. J. Newman

Nature Physics **18**, 848–850 (2022) | Cite this article

2653 Accesses | **1** Citations | **104** Altmetric | Metrics

A fundamental technical challenge in the analysis of network data is the automated discovery of communities – groups of nodes that are strongly connected or that share similar features or roles. In this Comment we review progress in the field over the past 20 years.

Fig. 1: Community structure of a social network.

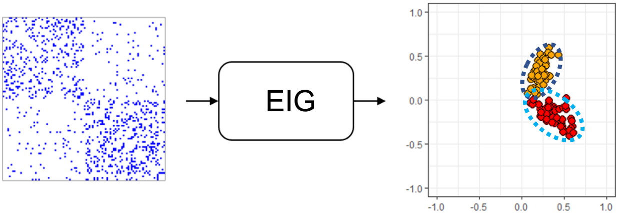


Nodes are Facebook users and edges represent Facebook friendships. Communities, represented by different colours, were found using the InfoMap algorithm[11].

Spectral methods for community detection

- Cluster *embeddings* obtained from top leading eigenvectors of appropriate matrix.
- Example: adjacecy spectral embedding (Sussman et al., 2011):

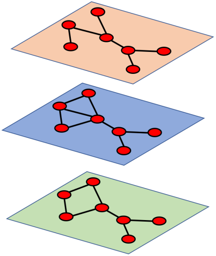
$$
\begin{aligned}
\textit{Eigendecomposition:} \quad & \mathbf{A} = \widehat{\mathbf{V}}\widehat{\mathbf{\Lambda}}\widehat{\mathbf{V}}^\top + \widehat{\mathbf{V}}_\perp\widehat{\mathbf{\Lambda}}_\perp\widehat{\mathbf{V}}_\perp^\top \\
\textit{Embedding:} \quad & \widehat{\mathbf{X}} = \widehat{\mathbf{V}}|\widehat{\mathbf{\Lambda}}|^{1/2} \\
\textit{Clustering}: \quad & \text{kmeans}(\widehat{\mathbf{X}}, K).
\end{aligned}
$$

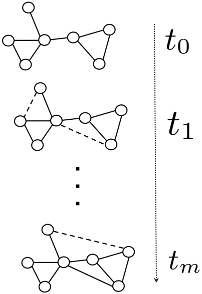


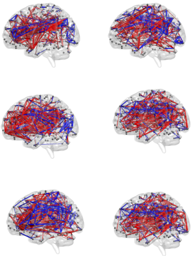
- Practically accurate and computationally efficient with well-developed theory.
- Extensions for degree heterogeneity: SCORE (Jin, 2015), spherical clustering (Lyzinski et al., 2014; Lei and Rinaldo, 2015).

Multilayer networks

Multiple networks over the same set of vertices: multi-view data, time series of networks, independent samples, etc.

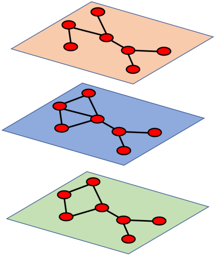


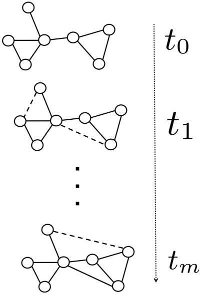


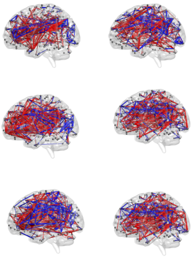


Multilayer networks

Multiple networks over the same set of vertices: multi-view data, time series of networks, independent samples, etc.

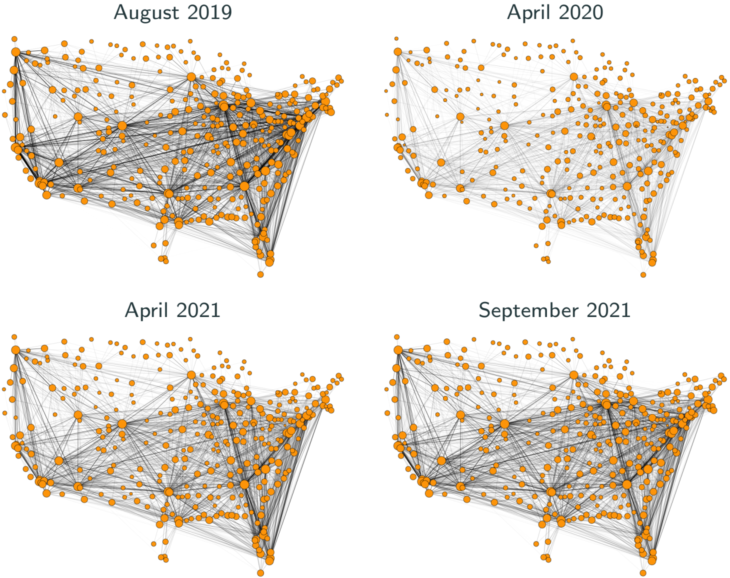






- Common structure across the networks (communities)
- Local and global variability within and between networks

Example: US time series of flights



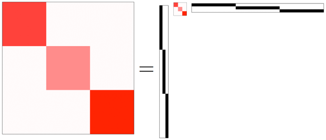
Number of monthly flights between US airports
(data: Bureau of Transportation Statistics)

Model-based community detection

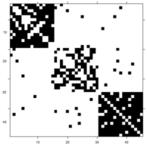
Stochastic block model (SBM) (Holland et al., 1983)

- $\mathbf{A}$ is a $n \times n$ binary symmetric adjacency matrix.
- Nodes are partitioned into K **communities** $\mathcal{C}_1, \ldots, \mathcal{C}_K$
- Edge probabilities only depend on node community labels.

$$\mathbb{E}[\mathbf{A}_{ij}] = \mathbf{P}_{ij} = \mathbf{B}_{rs} \qquad \text{if } i \in \mathcal{C}_r, j \in \mathcal{C}_s.$$



$\mathbf{P} = \mathbf{Z}\mathbf{B}\mathbf{Z}^\top$



$\mathbf{A}$

- $\mathbf{Z} \in \{0,1\}^{n \times K}$ *community membership* indicator matrix.
- $\mathbf{B} \in [0,1]^{K \times K}$ *connection probabilities*.

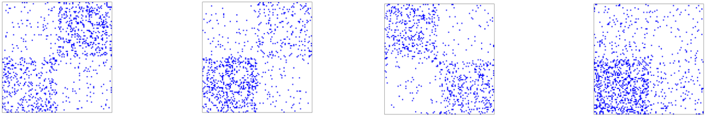
Stochastic block model for multiple networks

Multilayer SBM (Holland et al., 1983)

- L observed adjacency matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)}$.
- Common community structure but different connection probabilities

$$\mathbf{P}^{(l)} = \mathbf{Z}\mathbf{B}^{(l)}\mathbf{Z}^T, \qquad l = 1, \ldots, L.$$

- Connection probabilities can be different on each network.



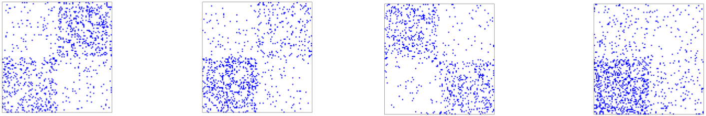
Stochastic block model for multiple networks

Multilayer SBM (Holland et al., 1983)

- L observed adjacency matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(L)}$.
- Common community structure but different connection probabilities

$$\mathbf{P}^{(l)} = \mathbf{Z}\mathbf{B}^{(l)}\mathbf{Z}^T, \qquad l = 1, \ldots, L.$$

- Connection probabilities can be different on each network.



- **Problem:** no hub vertices, expected degrees are the same within community.

Degree corrected SBM

Multilayer degre-corrected SBM (Peixoto, 2016; Bazzi et al., 2020):

- Introduce degree-correction parameters (Karrer and Newman, 2011).

$$\mathbf{P}^{(l)} = \mathbf{\Theta}^{(l)}\mathbf{Z}\mathbf{B}^{(l)}\mathbf{Z}^\top\mathbf{\Theta}^{(l)}, \qquad l = 1, \ldots, L.$$

$\mathbf{\Theta}^{(l)} = \text{diag}(\theta_1^{(l)}, \ldots, \theta_n^{(l)})$ proportional to node degrees.

Degree corrected SBM

Multilayer degre-corrected SBM (Peixoto, 2016; Bazzi et al., 2020):

- Introduce degree-correction parameters (Karrer and Newman, 2011).

$$\mathbf{P}^{(l)} = \boldsymbol{\Theta}^{(l)} \mathbf{Z} \mathbf{B}^{(l)} \mathbf{Z}^\top \boldsymbol{\Theta}^{(l)}, \qquad l = 1, \ldots, L.$$

  $\boldsymbol{\Theta}^{(l)} = \text{diag}(\theta_1^{(l)}, \ldots, \theta_n^{(l)})$ proportional to node degrees.

  For network l and vertices $i \in \mathcal{C}_r$, $j \in \mathcal{C}_s$, the model satisfies

$$\log(\mathbf{P}_{ij}^{(l)}) = \underbrace{\log(\theta_i^{(l)}) + \log(\theta_j^{(l)})}_{\text{vertex effects}} + \underbrace{\log(\mathbf{B}_{rs}^{(l)})}_{\text{community effect}} .$$

- **Remark:** degrees and connection probabilities can vary across networks.

Degree corrected SBM

Multilayer degre-corrected SBM (Peixoto, 2016; Bazzi et al., 2020):

- Introduce degree-correction parameters (Karrer and Newman, 2011).

$$\mathbf{P}^{(l)} = \mathbf{\Theta}^{(l)}\mathbf{Z}\mathbf{B}^{(l)}\mathbf{Z}^\top\mathbf{\Theta}^{(l)}, \qquad l = 1, \ldots, L.$$

  $\mathbf{\Theta}^{(l)} = \text{diag}(\theta_1^{(l)}, \ldots, \theta_n^{(l)})$ proportional to node degrees.

  For network l and vertices $i \in \mathcal{C}_r$, $j \in \mathcal{C}_s$, the model satisfies

$$\log(\mathbf{P}_{ij}^{(l)}) = \underbrace{\log(\theta_i^{(l)}) + \log(\theta_j^{(l)})}_{\text{vertex effects}} + \underbrace{\log(\mathbf{B}_{rs}^{(l)})}_{\text{community effect}}.$$

- **Remark:** degrees and connection probabilities can vary across networks.
- The model has $O(nL + K^2L)$, needs constraints for identifiability.

Community detection identifiability

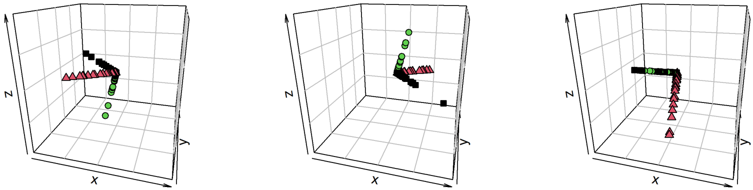
Assumption: K is the smallest value that can fit the model.

Theorem (Agterberg, Lubberts, A., 2023+)

The matrices $\mathbf{P}^{(1)}, \ldots \mathbf{P}^{(L)}$ *have* K *identifiable communities if and only if the matrices* $\mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(L)}$ *have* K *jointly distinguishable rows.*

Implication: the vertex latent positions of lie on K jointly distinguishable rays

$$\mathbf{P}^{(l)} = \mathbf{U}^{(l)} \mathbf{\Lambda}^{(l)} (\mathbf{U}^{(l)})^\top$$
$$\mathbf{X}^{(l)} = \mathbf{U}^{(l)} |\Lambda^{(l)}|^{1/2}.$$



Spectral clustering for multilayer networks

Multilayer spectral clustering methods: aggregate layers, then perform spectral clustering

- Sum of adjacencies (Tang et al., 2009; Bhattacharyya and Chatterjee, 2022):

$$\mathbf{A}^{\mathsf{Sum}} = \sum_l \mathbf{A}^{(l)}.$$

- Sum of (bias-adjusted) squared adjacencies (Lei and Lin, 2022):

$$\mathbf{A}^{\mathsf{SoS}} = \sum_l (\mathbf{A}^{(l)})^2$$

- SVD on concatenated embeddings (Paul and Chen, 2020; A. et al, 2021):

$$\widehat{\mathbf{V}}^{(l)} = \texttt{eigs}(\mathbf{A}^{(l)}, K)\texttt{\$vectors}$$
$$\widehat{\mathbf{V}}^{\mathsf{MASE}} = \texttt{svds}([\widehat{\mathbf{V}}^{(1)}, \cdots, \widehat{\mathbf{V}}^{(L)}], K)\texttt{\$u}$$

Spectral clustering for multilayer networks

Multilayer spectral clustering methods: aggregate layers, then perform spectral clustering

- Sum of adjacencies (Tang et al., 2009; Bhattacharyya and Chatterjee, 2022):

$$\mathbf{A}^{\mathsf{Sum}} = \sum_l \mathbf{A}^{(l)}.$$

- Sum of (bias-adjusted) squared adjacencies (Lei and Lin, 2022):

$$\mathbf{A}^{\mathsf{SoS}} = \sum_l (\mathbf{A}^{(l)})^2$$

- SVD on concatenated embeddings (Paul and Chen, 2020; A. et al, 2021):

$$\widehat{\mathbf{V}}^{(l)} = \texttt{eigs}(\mathbf{A}^{(l)}, K)\texttt{\$vectors}$$
$$\widehat{\mathbf{V}}^{\mathsf{MASE}} = \texttt{svds}([\widehat{\mathbf{V}}^{(1)}, \cdots, \widehat{\mathbf{V}}^{(L)}], K)\texttt{\$u}$$

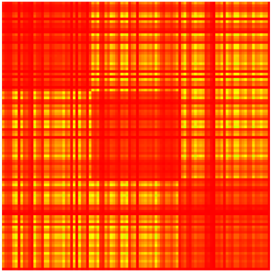
Problem: different degree parameters in DCSBM not considered in aggregation

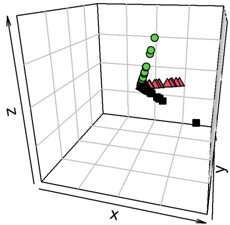
Multilayer DCSBM: spectral geometry

Observation 1

The scaled rows of the top leading eigenvectors of each $\mathbf{P}^{(l)}$ are supported on at most K different rays

$$\mathbf{P}^{(l)} = \mathbf{U}^{(l)}\Lambda^{(l)}(\mathbf{U}^{(l)})^\top, \qquad \mathbf{X}^{(l)} = \mathbf{U}^{(l)}|\Lambda^{(l)}|^{1/2}.$$



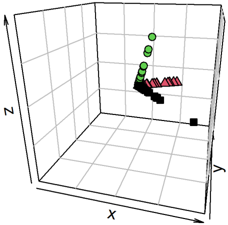


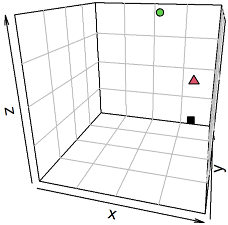
Observation 2

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

$$\mathbf{X} = \mathbf{U}|\Lambda|^{1/2}$$

$$\mathbf{Y}_{i\cdot} = \frac{1}{\|\mathbf{X}_{i\cdot}\|}\mathbf{X}_{i\cdot}$$

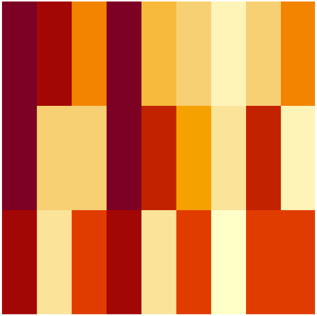




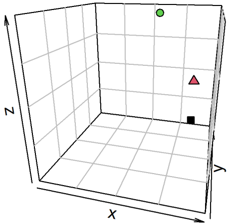
Observation 3

The matrix of concatenated row-normalized embeddings has left singular subspace that reveals community memberships for all networks.

$$\mathcal{Y} = [\mathbf{Y}^{(1)}, \cdots, \mathbf{Y}^{(L)}] = \mathbf{U}\Sigma\mathbf{V}^\top$$

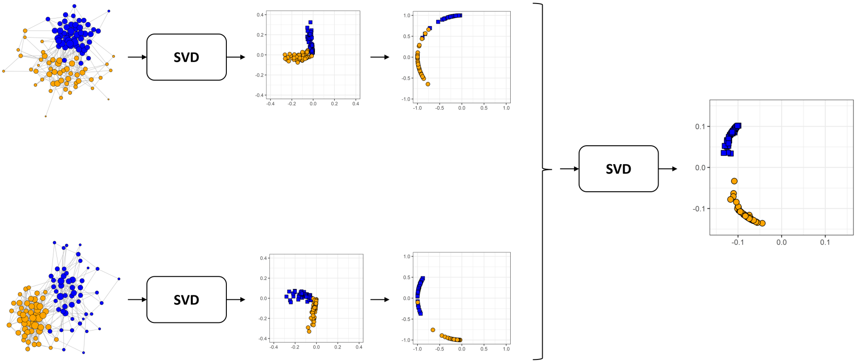


$\mathbf{U}$



Degree Corrected Multiple Adjacency Spectral Embedding (DC-MASE)

1. For each graph $l \in [L]$
   - Compute K scaled leading eigenvectors of $\mathbf{A}^{(l)}$.
   - Row-normalization.
2. Concatenate embeddings and compute the K left leading singular vectors.
3. Cluster the rows via `kmeans`.



Theoretical performance

- Theoretical performance measured using the *misclustering error rate*:

$$\ell(\widehat{z}, z) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(\widehat{z}(i) \neq z(i)).$$

  $\widehat{z}$ and z are the estimated and true memberships (up to a permutation).

- Assume that the number of communities K is known.

Assumptions: each network $l = 1, \ldots, L$ satisfies the following.

- **Balanced community sizes:**

$$|\mathcal{C}(r)| \asymp |\mathcal{C}(s)|, \qquad \|\theta^{(l)}_{\mathcal{C}(r))}\| \asymp \|\theta^{(l)}_{\mathcal{C}(s))}\|, \qquad \forall s, r \in [K], l \in [L].$$

- **Eigenvalues of $\mathbf{B}^{(l)}$ are bounded**: for all $k \in [K]$

$$\lambda^{(l)}_{\min} \leq |\lambda_k(\mathbf{B}^{(l)})| \leq C, \qquad \lambda^{(l)}_{\min} < 1.$$

- **Signal strength:**

$$\left(\frac{\theta^{(l)}_{\max}}{\theta^{(l)}_{\min}}\right) \frac{K^8 \theta_{\max} \|\theta\| \log(n)}{(\lambda_{\min})^2 \|\theta^{(l)}\|^4} \leq \bar{\lambda} := \frac{1}{L} \sum \lambda^{(l)}_{\min}.$$

- **Degree heterogeneity:**

$$\frac{\theta^{(l)}_{\min}}{\theta^{(l)}_{\max}} \geq \sqrt{\frac{\log(n)}{n}}$$

- **Minimum degree growth:** $\theta^{(l)}_{\min} \|\theta^{(l)}\| \geq c \log(n)$.

Theorem (Agterberg, Lubberts, A., 2023+)

Under assumptions, if $L \lesssim n^5$, then the expected misclustering rate satisfies

$$\mathbb{E}\left[\ell(\widehat{z}, z)\right] \lesssim \frac{K}{n} \sum_{i=1}^{n} \exp\left(- cL \min\left\{ \underbrace{\frac{\bar{\lambda}^2}{K^4 \mathrm{err}_{\mathsf{ave}}^{(i)}}}_{\text{average layer}}, \underbrace{\frac{\bar{\lambda}}{K^2 \mathrm{err}_{\max}^{(i)}}}_{\text{worst layer}} \right\} \right) + O\left(n^{-10}\right).$$

- Rates depend on average smallest eigenvalue $\bar{\lambda}$ and

$$\mathrm{err}_{\mathsf{ave}}^{(i)} := \frac{1}{L} \sum_{l} \frac{\|\theta^{(l)}\|_3^3}{\theta_i^{(l)} \|\theta^{(l)}\|^4 \lambda_{\min}^{(l)}}; \qquad \mathrm{err}_{\max}^{(i)} := \max_{l} \frac{\theta_{\max}^{(l)}}{\theta_i^{(l)} \|\theta^{(l)}\|^2 (\lambda_{\min}^{(l)})^{1/2}}.$$

- **Remarks:**
  - Error rate improves with L: effective for layer aggregation
  - No conditions on average layer: flexible for heterogeneous networks.

Corollary (Homogeneous Degrees)

Under the conditions of the theorem, if all the networks have the same parameters and all degrees are proportional to $\sqrt{\rho_n}$*, then*

$$\mathbb{E}\left[\ell(\widehat{z}, z)\right] \lesssim K \exp\left(- cLn\rho_n\lambda_{\min}^3\right) + O\left(n^{-10}\right).$$

- For $L = 1$, the misclustering error of SCORE (Jin et al., 2021) is

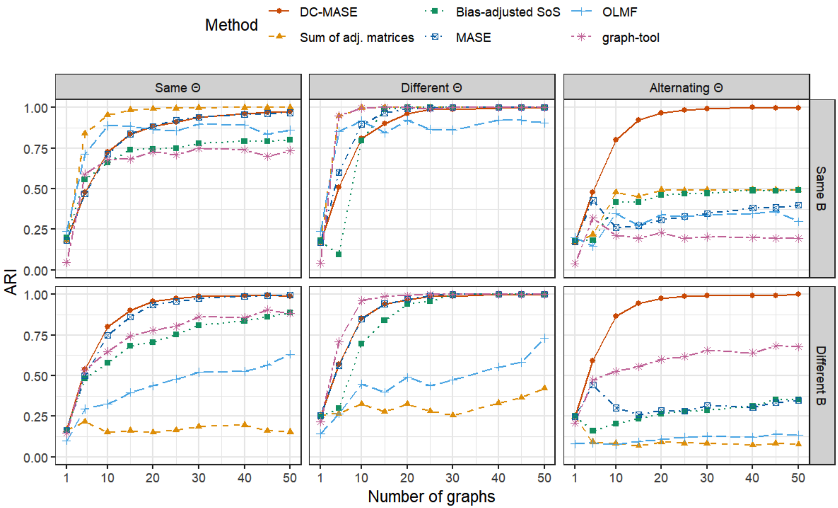
$$\mathbb{E}\left[\ell(\widehat{z}, z)\right] \lesssim K \exp(-cn\rho_n\lambda_{\min}^2) + o(n^{-3}).$$

Simulations

- Comparison with other multilayer community detection methods:
    - Aggregated sum of adjacency matrices (Han et al., 2015)
    - Bias-adjusted sum-of-squared (Lei and Lin, 2022)
    - Multiple adjacency spectral embedding (A. et al, 2021)
    - Orthogonal linked matrix factorization (Paul and Chen, 2020)
    - MCMC via graph-tool (Peixoto, 2014)
- Adjusted rand index (ARI): values close to 1 indicate perfect clustering.
- Different types of heterogeneity across networks:
    - Degree corrections: same, different, or alternating degrees.
    - Connection probabilities: same or different at random.

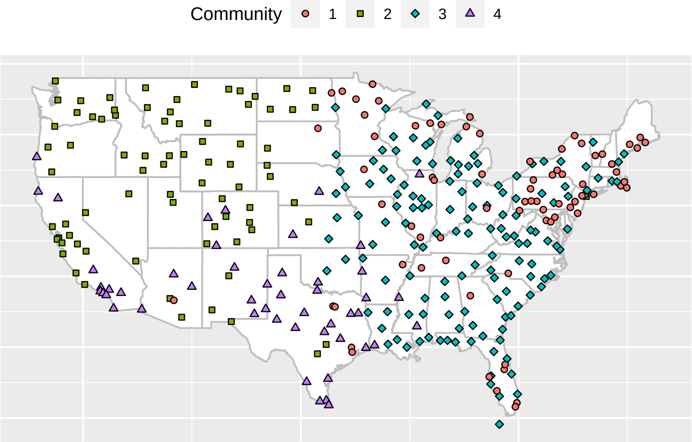
Simulations

- Columns: same, random, or alternating degrees $\Theta^{(l)}$ across graphs.
- Rows: same or different connectivity $\mathbf{B}^{(l)}$ across graphs.



US airport network

- Monthly data of US commercial flights (January 2016 - September 2021)
- Vertices: 343 airports from 48 states.
- Edges: number of flights between airports.



Parameter estimation

Given community memberships, parameters are estimated as follows:

- **Global probability matrices:** compute average connectivity within a block

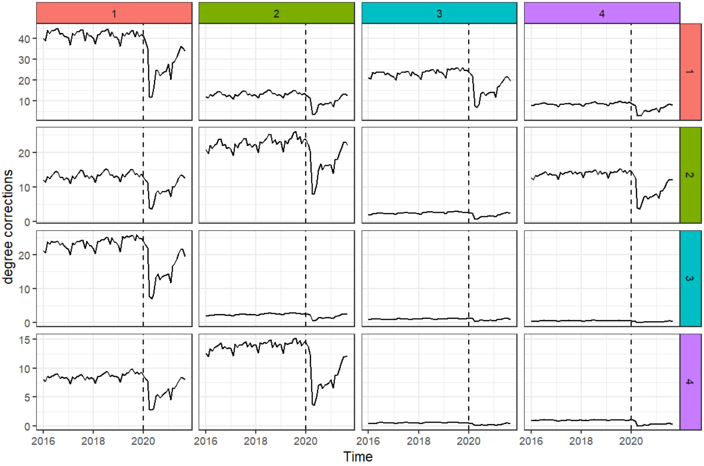
$$\widehat{\mathbf{B}}_{rs}^{(l)} = \frac{1}{|\mathcal{C}(r)|\ |\mathcal{C}(s)|} \sum_{i\in\mathcal{C}(r), j\in\mathcal{C}(s)} \mathbf{A}_{ij}^{(l)}.$$

- **Degree corrections:** given degrees $d_i^{(l)} = \sum_j \mathbf{A}_{ij}^{(l)}$, estimate

$$\widehat{\theta}_i^{(l)} = \frac{d_i^{(l)}}{\frac{1}{|\mathcal{C}(r)|}\sum_{j\in\mathcal{C}(r)} d_j^{(l)}}.$$

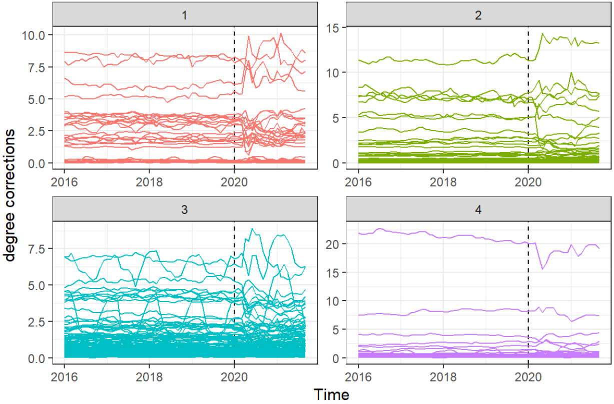
  Values larger than 1 indicate relative importance within the community at time l.

Tracking community-level dynamics



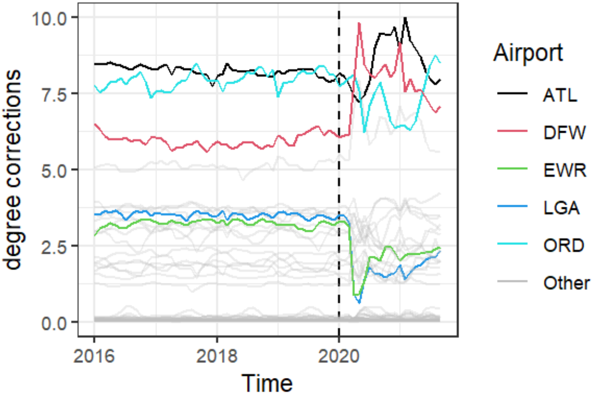
Estimated connectivity matrices $\widehat{\mathbf{B}}^{(1)}, \ldots, \widehat{\mathbf{B}}^{(L)}$.

Tracking airport-level dynamics



Estimated degree-correction parameters from each community.

Tracking airport-level dynamics: community 1



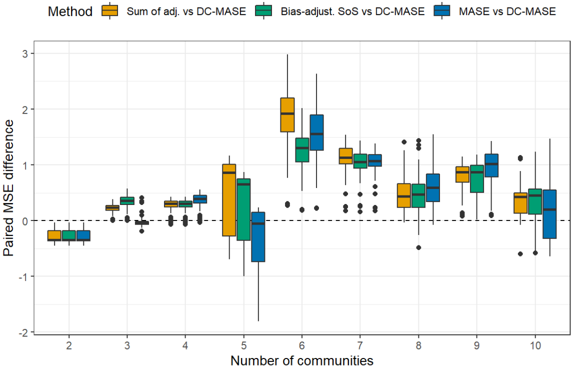
Estimated degree-correction parameters from community 1.

Flight data: comparison with other methods

- Comparison with other methods via out-of-sample mean squared error (MSE):

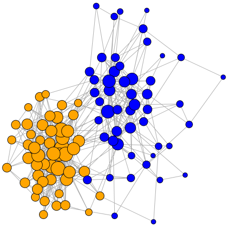
$$\text{MSE}(K, l) = \frac{1}{n^2}\|\mathbf{A}^{(l)} - \widehat{\mathbf{P}}^{(l)}_{\widehat{\mathbf{Z}}(-l,K)}\|_F^2.$$

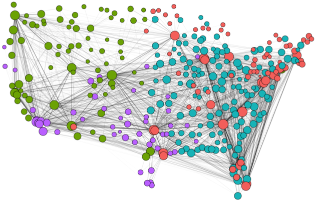
- Paired out-of-sample MSE difference between other methods and DC-MASE: positive values indicate better community quality for DC-MASE



Discussion

- Multilayer DCSBM: flexible and interpretable model.
- DC-MASE: efficient method for multilayer community detection.
- Multilayer spectral methods in other models? Mixed memberships, popularity-adjusted, networks with covariates, time series model, etc.





Thank you!

Questions?

✉jarroyo@tamu.edu

Preprint: J. Agterberg, Z. Lubberts, J. Arroyo, “Joint Spectral Clustering in Multilayer Degree-Corrected Stochastic Blockmodels”, *arXiv 2212.05053*.