Inference for Multiple Heterogeneous Networks with a Common Invariant Subspace

Jesús Arroyo

September 25th, 2020

UNIVERSITY OF MARYLAND

JOHNS HOPKINS UNIVERSITY

Inference for Multiple Heterogeneous Networks with a Common Invariant Subspace

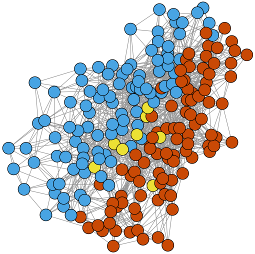
- Joint work with Avanti Athreya, Joshua Cape, Guodong Chen, Carey E. Priebe and Joshua T. Vogelstein
- Talk based on this manuscript: arXiv:1906.10026 (Journal of Machine Learning Research, to appear)

Outline

Outline

Networks

- **Graphs** are a popular structure to represent relational data
  - Vertices/nodes represent the units of a system.
  - Edges/links encode interactions between the units.



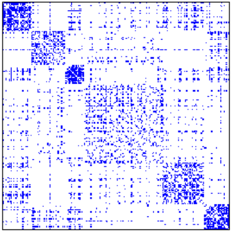
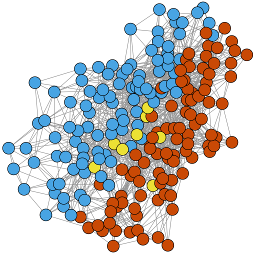


Figure: MRI brain network and its adjacency matrix

Networks

- **Graphs** are a popular structure to represent relational data
  - Vertices/nodes represent the units of a system.
  - Edges/links encode interactions between the units.



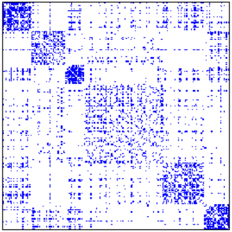
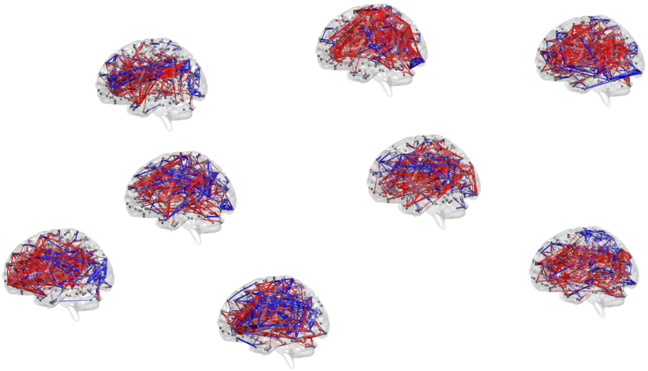


Figure: MRI brain network and its adjacency matrix

- Inference of vertex properties:

Inference for multiple networks

- In many applications, we observe a collection of graphs with matched vertices:
  - Multilayer networks
  - Time-varying networks
  - Multiple samples of networks (e.g. brain networks)



Inference for multiple networks

- Inferential questions about vertex properties:
  - community detection
  - vertex classification / prediction of vertex attributes
  - link prediction
  - ...
- Network heterogeneity. How to leverage information across all the networks?

Inference for multiple networks

- Inferential questions about vertex properties:
  - community detection
  - vertex classification / prediction of vertex attributes
  - link prediction
  - ...
- Network heterogeneity. How to leverage information across all the networks?
- Inference about network properties:
  - hypothesis testing
  - classification / prediction of network attributes
  - clustering / dimensionality reduction

In this talk:

- A model for multiple heterogeneous networks
- A joint spectral embedding method with statistical guarantees
- Applications to subsequent inference tasks

Random graph model

- **Setting:** sample of graphs with matched vertices
  - n nodes
  - m graphs
  - $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(m)}$ adjacency matrices (size $n \times n$)
    - Binary entries
    - Symmetric
    - Zero-diagonal

Random graph model

- **Setting:** sample of graphs with matched vertices
  - n nodes
  - m graphs
  - $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(m)}$ adjacency matrices (size $n \times n$)
    - Binary entries
    - Symmetric
    - Zero-diagonal
- **Bernoulli graph / inhomogeneous Erdős-Rényi model**: edges are independent random variables

$$\mathbb{P}(\mathbf{A}_{uv}^{(i)} = 1) = \mathbf{P}_{uv}^{(i)}, \qquad u < v.$$

  - $\mathbf{P}^{(i)} \in [0,1]^{n \times n}$ is the matrix of edge probabilities.

Random graph model

- **Setting:** sample of graphs with matched vertices
  - n nodes
  - m graphs
  - $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(m)}$ adjacency matrices (size $n \times n$)
    - Binary entries
    - Symmetric
    - Zero-diagonal
- **Bernoulli graph / inhomogeneous Erdős-Rényi model**: edges are independent random variables

$$\mathbb{P}(\mathbf{A}_{uv}^{(i)} = 1) = \mathbf{P}_{uv}^{(i)}, \qquad u < v.$$

  - $\mathbf{P}^{(i)} \in [0,1]^{n \times n}$ is the matrix of edge probabilities.
  - $\mathbf{P}^{(i)} = \mathbb{E}\left[\mathbf{A}^{(i)}\right]$.

Random graph model

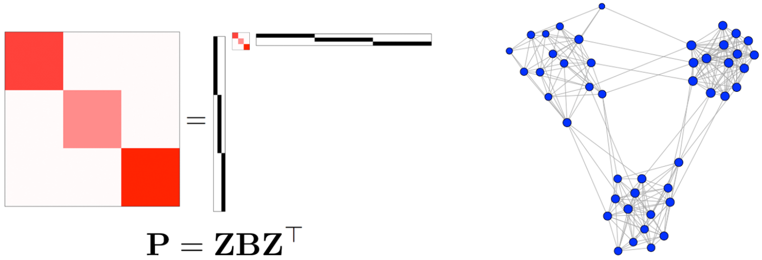
- **Setting:** sample of graphs with matched vertices
  - n nodes
  - m graphs
  - $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(m)}$ adjacency matrices (size $n \times n$)
    - Binary entries
    - Symmetric
    - Zero-diagonal
- **Bernoulli graph / inhomogeneous Erdős-Rényi model**: edges are independent random variables

$$\mathbb{P}(\mathbf{A}^{(i)}_{uv} = 1) = \mathbf{P}^{(i)}_{uv}, \qquad u < v.$$

  - $\mathbf{P}^{(i)} \in [0,1]^{n\times n}$ is the matrix of edge probabilities.
  - $\mathbf{P}^{(i)} = \mathbb{E}\left[\mathbf{A}^{(i)}\right]$.
  - Many popular models are included in this framework: Erdős-Rényi, stochastic block model, latent position graphs, etc.

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

- Vertices are divided into K **communities**
- Edge probabilities only depends on the vertex community assignments



- Vertex memberships: $\mathbf{Z} \in \{0,1\}^{n \times K}$
  - $\mathbf{Z}_{uj} = 1$ if node u is in community j.
- Edge probabilities: $\mathbf{B} \in [0,1]^{K \times K}$
  - $\mathbf{B}_{jh}$ probability of an edge between communities j and h

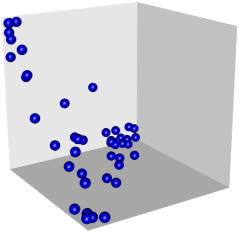
Random dot product graph (RDPG) (Young et al., 2007)

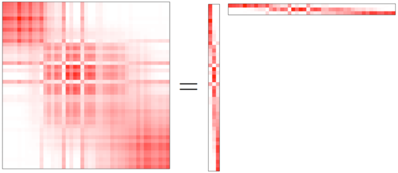
- Latent positions in $\mathbb{R}^d$ represent the vertices

$$\mathbf{X} = [X_1, \ldots, X_n]^\top \in \mathbb{R}^{n \times d}.$$

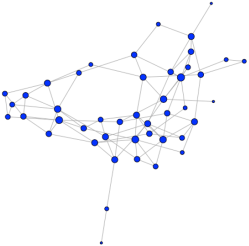
- Edge probabilities are given by the dot product of latent positions

$$\mathbf{P}_{uv} = \langle X_u, X_v \rangle$$





$$\mathbf{P} = \mathbf{X}\mathbf{X}^\top$$



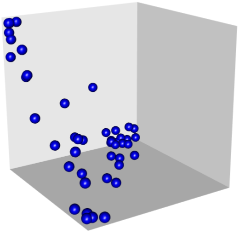
Random dot product graph (RDPG) (Young et al., 2007)

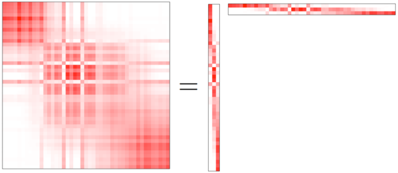
- Latent positions in $\mathbb{R}^d$ represent the vertices

$$\mathbf{X} = [X_1, \ldots, X_n]^\top \in \mathbb{R}^{n \times d}.$$

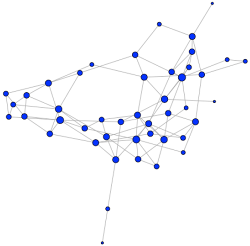
- Edge probabilities are given by the dot product of latent positions

$$\mathbf{P}_{uv} = \langle X_u, X_v \rangle$$





$$\mathbf{P} = \mathbf{X}\mathbf{X}^\top$$



- RDPG and its generalizations are a flexible model
  - SBM $\subset$ Generalized RDPG .
  - Degree corrected SBM, overlapping or hierarchical communities, etc.

Outline

Common subspace independent edge (COSIE) model

Definition

Consider the following parameters

- Matrix with orthonormal columns $\mathbf{V} \in \mathbb{R}^{n \times d}$
- Symmetric matrices $\mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(m)} \in \mathbb{R}^{d \times d}$

The COSIE model with bounded rank d is defined as

$$\mathbf{P}^{(i)} = \mathbb{E}\left[\mathbf{A}^{(i)}\right] = \mathbf{V}\mathbf{R}^{(i)}\mathbf{V}^\top, \qquad i = 1, \ldots, m.$$

Common subspace independent edge (COSIE) model

Definition

Consider the following parameters

- Matrix with orthonormal columns $\mathbf{V} \in \mathbb{R}^{n\times d}$
- Symmetric matrices $\mathbf{R}^{(1)}, \ldots, \mathbf{R}^{(m)} \in \mathbb{R}^{d\times d}$

The COSIE model with bounded rank d is defined as

$$\mathbf{P}^{(i)} = \mathbb{E}\left[\mathbf{A}^{(i)}\right] = \mathbf{V}\mathbf{R}^{(i)}\mathbf{V}^\top, \qquad i = 1, \ldots, m.$$

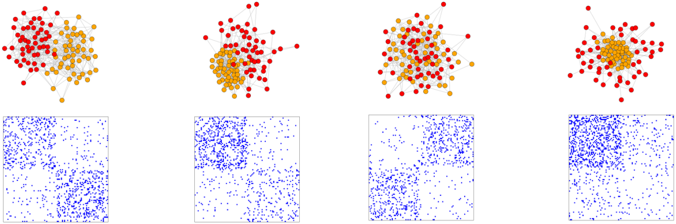
- The columns of $\mathbf{V}$ are a basis of the common invariant subspace for $\mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(m)}$.
- $\mathbf{R}^{(i)}$ are individual score matrices
- The rows of $\mathbf{V} = [V_1, \ldots, V_n]^\top$ can be interpreted as latent positions

$$\mathbf{P}^{(i)}_{uv} = V_u^\top \mathbf{R}^{(i)} V_v.$$

Examples of COSIE graphs

- **Multilayer SBM:** common community structure but different edge probabilities

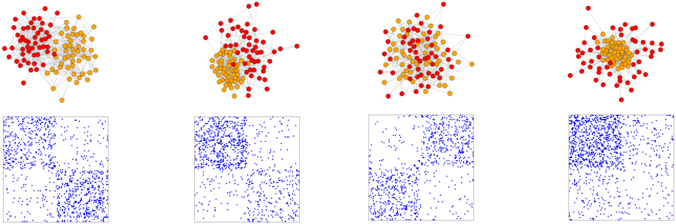
$$\mathbf{P}^{(i)} = \mathbf{Z}\mathbf{B}^{(i)}\mathbf{Z}^T$$



Examples of COSIE graphs

- **Multilayer SBM:** common community structure but different edge probabilities

$$\mathbf{P}^{(i)} = \mathbf{Z}\mathbf{B}^{(i)}\mathbf{Z}^T$$

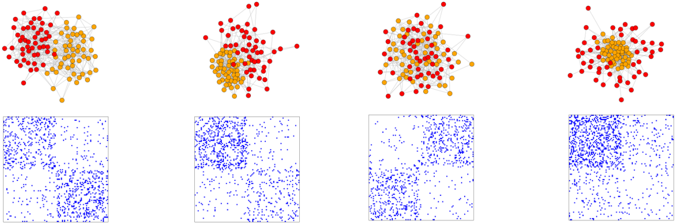


- **RDPG models** (including many SBM extensions) can be translated into a multilayer setting using COSIE:
  - Degree-corrected SBM
  - Mixed and overlapping memberships
  - Hierarchical communities
  - Manifold structure

Examples of COSIE graphs

- **Multilayer SBM:** common community structure but different edge probabilities

$$\mathbf{P}^{(i)} = \mathbf{Z}\mathbf{B}^{(i)}\mathbf{Z}^T$$



- **RDPG models** (including many SBM extensions) can be translated into a multilayer setting using COSIE:
  - Degree-corrected SBM
  - Mixed and overlapping memberships
  - Hierarchical communities
  - Manifold structure
- For any collection of matrices $\mathbf{P}^{(1)}, \ldots, \mathbf{P}^{(m)}$, there exists $d \leq n$ such that the matrices have a common invariant subspace (trivial if $d = n$)

Outline

Adjacency spectral embedding (ASE)

- Write the eigendecomposition of $\mathbf{A}$ as

$$\mathbf{A} = \widetilde{\mathbf{V}}\widetilde{\mathbf{D}}\widetilde{\mathbf{V}}^\top + \widetilde{\mathbf{V}}_\perp \widetilde{\mathbf{D}}_\perp \widetilde{\mathbf{V}}_\perp^\top,$$

  - $(\widetilde{\mathbf{V}}, \widetilde{\mathbf{V}}_\perp)$ orthogonal matrix of eigenvectors
  - $\widetilde{\mathbf{D}}, \widetilde{\mathbf{D}}_\perp$ are diagonal matrices containing the d largest eigenvalues in magnitude.

Adjacency spectral embedding (ASE)

- Write the eigendecomposition of $\mathbf{A}$ as

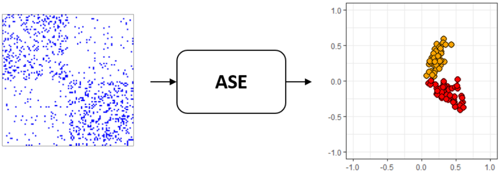
$$\mathbf{A} = \widetilde{\mathbf{V}}\widetilde{\mathbf{D}}\widetilde{\mathbf{V}}^\top + \widetilde{\mathbf{V}}_\perp\widetilde{\mathbf{D}}_\perp\widetilde{\mathbf{V}}_\perp^\top,$$

  - $(\widetilde{\mathbf{V}}, \widetilde{\mathbf{V}}_\perp)$ orthogonal matrix of eigenvectors
  - $\widetilde{\mathbf{D}}, \widetilde{\mathbf{D}}_\perp$ are diagonal matrices containing the d largest eigenvalues in magnitude.

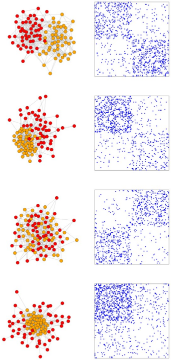
- The *scaled* **adjacency spectral embedding** of $\mathbf{A}$ is defined as

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{V}}|\widetilde{\mathbf{D}}|^{1/2}.$$

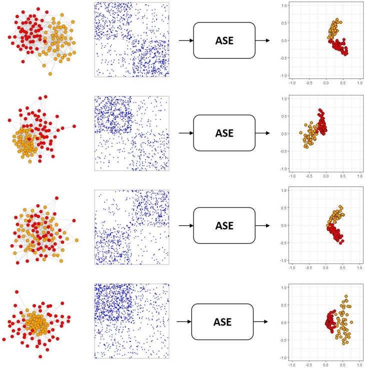
- We call $\widetilde{\mathbf{V}}$ the *unscaled* **adjacency spectral embedding** of $\mathbf{A}$



Multiple adjacency spectral embedding (MASE)

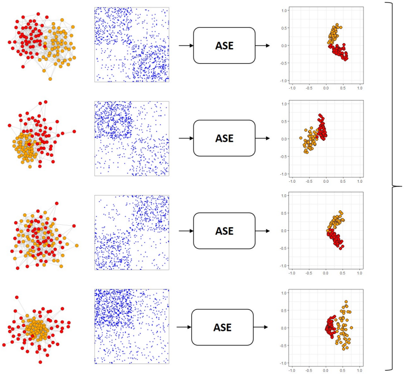


Multiple adjacency spectral embedding (MASE)



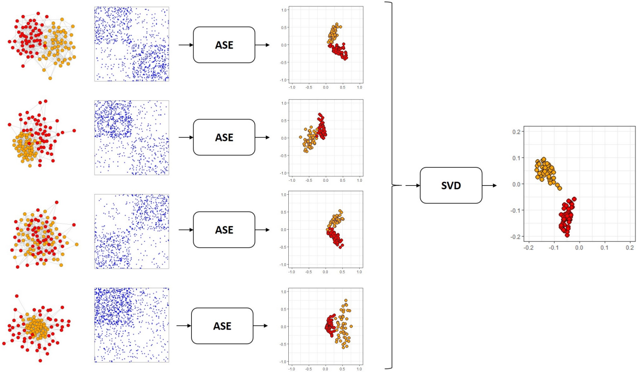
1. For each i, $\widetilde{\mathbf{V}}^{(i)} \in \mathbb{R}^{n \times d_i}$ are the d_i leading eigenvectors of $\widehat{\mathbf{A}}^{(i)}$.

Multiple adjacency spectral embedding (MASE)



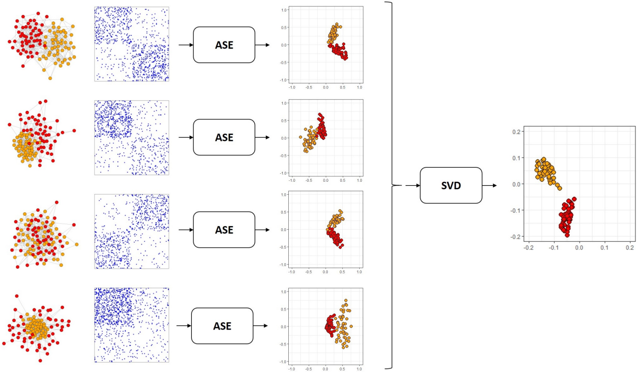
1. For each i, $\widetilde{\mathbf{V}}^{(i)} \in \mathbb{R}^{n \times d_i}$ are the d_i leading eigenvectors of $\widehat{\mathbf{A}}^{(i)}$.
2. $\widehat{\mathbf{U}} = \left[\widetilde{\mathbf{V}}^{(1)} \cdots \widetilde{\mathbf{V}}^{(m)}\right]$ is the $n \times \sum d_i$ matrix of concatenated eigenvectors.

Multiple adjacency spectral embedding (MASE)



1. For each i, $\widetilde{\mathbf{V}}^{(i)} \in \mathbb{R}^{n \times d_i}$ are the d_i leading eigenvectors of $\widehat{\mathbf{A}}^{(i)}$.
2. $\widehat{\mathbf{U}} = \left[\widetilde{\mathbf{V}}^{(1)} \cdots \widetilde{\mathbf{V}}^{(m)} \right]$ is the $n \times \sum d_i$ matrix of concatenated eigenvectors.
3. $\widehat{\mathbf{V}} \in \mathbb{R}^{n \times d}$ are the d leading left singular vectors of $\widehat{\mathbf{U}}$.

Multiple adjacency spectral embedding (MASE)



1. For each i, $\widetilde{\mathbf{V}}^{(i)} \in \mathbb{R}^{n \times d_i}$ are the d_i leading eigenvectors of $\widehat{\mathbf{A}}^{(i)}$.
2. $\widehat{\mathbf{U}} = \left[\widetilde{\mathbf{V}}^{(1)} \cdots \widetilde{\mathbf{V}}^{(m)}\right]$ is the $n \times \sum d_i$ matrix of concatenated eigenvectors.
3. $\widehat{\mathbf{V}} \in \mathbb{R}^{n \times d}$ are the d leading left singular vectors of $\widehat{\mathbf{U}}$.
4. For each i, set $\widehat{\mathbf{R}}^{(i)} = \widehat{\mathbf{V}}^\top \mathbf{A}^{(i)} \widehat{\mathbf{V}}$.

Why MASE?

- The method is computationally efficient and robust
- The SVD of $\hat{\mathbf{U}}$ (step 3) leverages the information in all the graphs to estimate $\mathbf{V}$.

Why MASE?

- The method is computationally efficient and robust
- The SVD of $\hat{\mathrm{U}}$ (step 3) leverages the information in all the graphs to estimate $\mathbf{V}$.
- Alternative approach: estimate $\mathbf{V}$ by using the leading eigenvectors of

$$\bar{\mathbf{A}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{A}^{(i)}.$$

  Similar estimation error rate to MASE, but requires stronger assumptions.
- Optimization approaches (such as maximum likelihood) require to solve a non-convex problem, which is only guaranteed to converge to a local minimum
- MASE is related to other matrix factorization methods: population SVD (Crainiceanu et al., 2011), distributed PCA (Fan et al., 2019).

Outline

Common invariant subspace estimation error

- $\delta(\mathbf{P}) = \max_u \sum_{v=1}^n \mathbf{P}_{uv}$ is the largest expected degree of a graph
- $\lambda_d(\mathbf{P})$ is the dth-eigenvalue (in magnitude) of $\mathbf{P}$.
- Define $\varepsilon = \left(\frac{1}{m}\sum_{i=1}^n \frac{\delta(\mathbf{P}^{(i)})}{\lambda_d^2(\mathbf{P}^{(i)})}\right)^{1/2}$.
- $\mathcal{O}_d$ is the set of $d \times d$ orthogonal matrices

Common invariant subspace estimation error

- $\delta(\mathbf{P}) = \max_u \sum_{v=1}^{n} \mathbf{P}_{uv}$ is the largest expected degree of a graph
- $\lambda_d(\mathbf{P})$ is the dth-eigenvalue (in magnitude) of $\mathbf{P}$.
- Define $\varepsilon = \left(\frac{1}{m}\sum_{i=1}^{n} \frac{\delta(\mathbf{P}^{(i)})}{\lambda_d^2(\mathbf{P}^{(i)})}\right)^{1/2}$.
- $\mathcal{O}_d$ is the set of $d \times d$ orthogonal matrices

Theorem (A. et al, 2020)

Suppose that $\varepsilon = o(1)$ *and* $\delta(\mathbf{P}^{(i)}) = \omega(\log n)$. *Then, the matrix* $\widehat{\mathbf{V}}$ *estimated by MASE satisfies*

$$\mathbb{E}\left[\min_{\mathbf{W}\in\mathcal{O}_d} \|\widehat{\mathbf{V}} - \mathbf{V}\mathbf{W}\|_F\right] \lesssim \sqrt{\frac{d}{m}}\varepsilon + \sqrt{d}\varepsilon^2.$$

Common invariant subspace estimation error

- $\delta(\mathbf{P}) = \max_u \sum_{v=1}^{n} \mathbf{P}_{uv}$ is the largest expected degree of a graph
- $\lambda_d(\mathbf{P})$ is the dth-eigenvalue (in magnitude) of $\mathbf{P}$.
- Define $\varepsilon = \left(\frac{1}{m}\sum_{i=1}^{n} \frac{\delta(\mathbf{P}^{(i)})}{\lambda_d^2(\mathbf{P}^{(i)})}\right)^{1/2}$.
- $\mathcal{O}_d$ is the set of $d \times d$ orthogonal matrices

Theorem (A. et al, 2020)

Suppose that $\varepsilon = o(1)$ and $\delta(\mathbf{P}^{(i)}) = \omega(\log n)$. Then, the matrix $\widehat{\mathbf{V}}$ estimated by MASE satisfies

$$\mathbb{E}\left[\min_{\mathbf{W}\in\mathcal{O}_d} \|\widehat{\mathbf{V}} - \mathbf{V}\mathbf{W}\|_F\right] \lesssim \sqrt{\frac{d}{m}}\varepsilon + \sqrt{d}\varepsilon^2.$$

- In dense graphs (average degree is $\Omega(n)$),

$$\mathbb{E}\left[\min_{\mathbf{W}\in\mathcal{O}_d} \|\widehat{\mathbf{V}} - \mathbf{V}\mathbf{W}\|_F\right] \lesssim \sqrt{\frac{d}{nm}} + \frac{\sqrt{d}}{n}.$$

Common invariant subspace estimation error

- $\delta(\mathbf{P}) = \max_u \sum_{v=1}^{n} \mathbf{P}_{uv}$ is the largest expected degree of a graph
- $\lambda_d(\mathbf{P})$ is the dth-eigenvalue (in magnitude) of $\mathbf{P}$.
- Define $\varepsilon = \left(\frac{1}{m} \sum_{i=1}^{n} \frac{\delta(\mathbf{P}^{(i)})}{\lambda_d^2(\mathbf{P}^{(i)})} \right)^{1/2}$.
- $\mathcal{O}_d$ is the set of $d \times d$ orthogonal matrices

Theorem (A. et al, 2020)

Suppose that $\varepsilon = o(1)$ and $\delta(\mathbf{P}^{(i)}) = \omega(\log n)$. Then, the matrix $\widehat{\mathbf{V}}$ estimated by MASE satisfies

$$\mathbb{E}\left[\min_{\mathbf{W} \in \mathcal{O}_d} \|\widehat{\mathbf{V}} - \mathbf{V}\mathbf{W}\|_F \right] \lesssim \sqrt{\frac{d}{m}}\varepsilon + \sqrt{d}\varepsilon^2.$$

- In dense graphs (average degree is $\Omega(n)$),

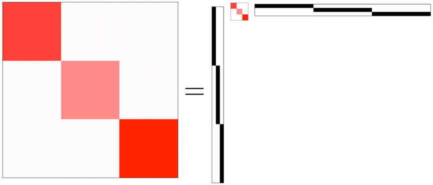
$$\mathbb{E}\left[\min_{\mathbf{W} \in \mathcal{O}_d} \|\widehat{\mathbf{V}} - \mathbf{V}\mathbf{W}\|_F \right] \lesssim \sqrt{\frac{d}{nm}} + \frac{\sqrt{d}}{n}.$$

- The proof relies on recent distributed PCA results Fan et al., 2019.

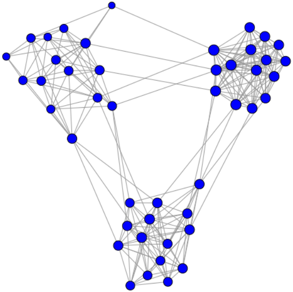
Community detection in multilayer networks

Consider m networks from the multilayer SBM with K communities

$$(\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(m)}) \sim \mathrm{SBM}\left(\mathbf{Z}; \mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(m)}\right).$$



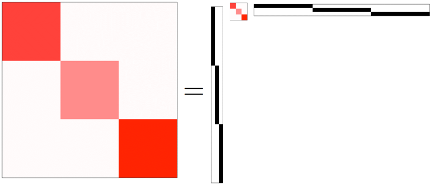
$$\mathbf{P}^{(i)} = \mathbf{V}\mathbf{R}^{(i)}\mathbf{V}^\top$$



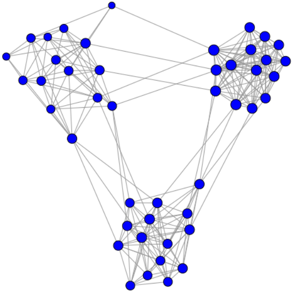
Community detection in multilayer networks

Consider m networks from the multilayer SBM with K communities

$$(\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(m)}) \sim \mathrm{SBM}\left(\mathbf{Z}; \mathbf{B}^{(1)}, \ldots, \mathbf{B}^{(m)}\right).$$



$$\mathbf{P}^{(i)} = \mathbf{V}\mathbf{R}^{(i)}\mathbf{V}^\top$$



Community detection algorithm:

1. Compute $\widehat{\mathbf{V}}$ using MASE
2. Cluster the rows of $\widehat{\mathbf{V}}$ using K-means

Community detection in multilayer networks

Theorem

Suppose that there exist some absolute constants $\kappa \in (0, 1], \gamma > 0$ *such that*

$$\underbrace{\frac{\sqrt{K}\left(\sum_{j=1}^{K} n_j^2\right)^{1/2}}{n_{\min}^{2-\kappa}}}_{\textit{community sizes}} \underbrace{\left(\frac{1}{m}\sum_{i=1}^{m}\frac{\lambda_1(\mathbf{B}^{(i)})}{\lambda_{\min}^2(\mathbf{B}^{(i)})}\right)}_{\textit{block connectivity}} \leq \gamma.$$

and that $n_{\min} = \omega(1)$, $\delta(\mathbf{P}^{(i)}) = \omega(\log n)$. *Then*

$$\mathbb{E}\left[\|\widehat{\mathbf{Z}} - \mathbf{Z}\|_F\right] \lesssim \sqrt{\frac{\gamma K n_{\max}}{m n_{\min}^{\kappa}}} + \frac{\gamma\sqrt{K n_{\max}}}{n_{\min}^{\kappa}}.$$

When all the communities have a proportional number of nodes $(n_{\min} \asymp n_{\max})$,

$$\#\ \text{misclustered nodes} = O_P\left(\frac{\gamma K}{m} + \frac{\gamma^2 K}{n}\right).$$

Asymptotic normality of the score matrices

- $\mathrm{vec}(\mathbf{R}^{(i)}) \in \mathbb{R}^r$ is the vectorized upper triangular part of $\mathbf{R}^{(i)}$ ($r = \frac{d(d+1)}{2}$).

Theorem (A. et al, 2020)

Under appropriate regularity conditions, there exists orthogonal matrices $\mathbf{W}_n$ *such that*

$$\left(\mathbf{\Sigma}^{(i)}\right)^{-1/2} \mathrm{vec}\left(\mathbf{W}_n^\top \widehat{\mathbf{R}}^{(i)} \mathbf{W}_n - \mathbf{R}^{(i)} + \mathbf{H}_m^{(i)}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}_r, \mathbf{I}_r),$$

- $\mathbf{\Sigma}^{(i)} \in \mathbb{R}^{r \times r}$ *is a covariance matrix that depends on* $\mathbf{V}$ *and* $\mathbf{P}^{(i)}$*.*
- $\mathbb{E}\left[\|\mathbf{H}_m^{(i)}\|_F\right] = O(\frac{d}{\sqrt{m}})$*.*

Asymptotic normality of the score matrices

- $\text{vec}(\mathbf{R}^{(i)}) \in \mathbb{R}^r$ is the vectorized upper triangular part of $\mathbf{R}^{(i)}$ $(r = \frac{d(d+1)}{2})$.

Theorem (A. et al, 2020)

Under appropriate regularity conditions, there exists orthogonal matrices $\mathbf{W}_n$ such that

$$\left(\boldsymbol{\Sigma}^{(i)}\right)^{-1/2} \text{vec}\left(\mathbf{W}_n^\top \widehat{\mathbf{R}}^{(i)} \mathbf{W}_n - \mathbf{R}^{(i)} + \mathbf{H}_m^{(i)}\right) \xrightarrow{d} \mathcal{N}(\mathbf{0}_r, \mathbf{I}_r),$$

- $\boldsymbol{\Sigma}^{(i)} \in \mathbb{R}^{r \times r}$ *is a covariance matrix that depends on* $\mathbf{V}$ *and* $\mathbf{P}^{(i)}$.
- $\mathbb{E}\left[\|\mathbf{H}_m^{(i)}\|_F\right] = O(\frac{d}{\sqrt{m}})$.

- After a proper alignment between $\widehat{\mathbf{R}}^{(i)}$ and $\mathbf{R}^{(i)}$, and if n and m are sufficiently large, then

$$\text{vec}\left(\widehat{\mathbf{R}}^{(i)} - \mathbf{R}^{(i)}\right) \approx \mathcal{N}(\mathbf{0}_r, \boldsymbol{\Sigma}^{(i)}).$$

Graph hypothesis testing

- Given a pair of observed graphs $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, are their underlying distributions the same?

$$\mathcal{H}_0 : \mathbf{P}^{(1)} = \mathbf{P}^{(2)} \qquad \text{vs.} \qquad \mathcal{H}_1 : \mathbf{P}^{(1)} \neq \mathbf{P}^{(2)}.$$

Graph hypothesis testing

- Given a pair of observed graphs $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, are their underlying distributions the same?

$$\mathcal{H}_0 : \mathbf{P}^{(1)} = \mathbf{P}^{(2)} \qquad \text{vs.} \qquad \mathcal{H}_1 : \mathbf{P}^{(1)} \neq \mathbf{P}^{(2)}.$$

- Under the COSIE model, the hypothesis testing problem reduces to

$$\mathcal{H}_0 : \mathbf{R}^{(1)} = \mathbf{R}^{(2)} \qquad \text{vs.} \qquad \mathcal{H}_1 : \mathbf{R}^{(1)} \neq \mathbf{R}^{(2)}$$

Graph hypothesis testing

- Given a pair of observed graphs $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, are their underlying distributions the same?

$$\mathcal{H}_0 : \mathbf{P}^{(1)} = \mathbf{P}^{(2)} \qquad \text{vs.} \qquad \mathcal{H}_1 : \mathbf{P}^{(1)} \neq \mathbf{P}^{(2)}.$$

- Under the COSIE model, the hypothesis testing problem reduces to

$$\mathcal{H}_0 : \mathbf{R}^{(1)} = \mathbf{R}^{(2)} \qquad \text{vs.} \qquad \mathcal{H}_1 : \mathbf{R}^{(1)} \neq \mathbf{R}^{(2)}$$

Hypothesis tesing with MASE:

- **Test statistic**: reject $\mathcal{H}_0$ if $\|\hat{\mathbf{R}}^{(1)} - \hat{\mathbf{R}}^{(2)}\|_F^2$ is large
- **Null distribution**:
  - *Monte Carlo simulations* if $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ are known
  - *Parametric boostrap* using $\hat{\mathbf{P}}^{(1)}$ and $\hat{\mathbf{P}}^{(2)}$
  - *Asymptotic distribution* of $\|\hat{\mathbf{R}}^{(1)} - \hat{\mathbf{R}}^{(2)}\|_F^2$ is a generalized chi-square

Graph hypothesis testing

- Given a pair of observed graphs $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, are their underlying distributions the same?

$$\mathcal{H}_0 : \mathbf{P}^{(1)} = \mathbf{P}^{(2)} \qquad \text{vs.} \qquad \mathcal{H}_1 : \mathbf{P}^{(1)} \neq \mathbf{P}^{(2)}.$$

- Under the COSIE model, the hypothesis testing problem reduces to

$$\mathcal{H}_0 : \mathbf{R}^{(1)} = \mathbf{R}^{(2)} \qquad \text{vs.} \qquad \mathcal{H}_1 : \mathbf{R}^{(1)} \neq \mathbf{R}^{(2)}$$

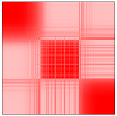
Hypothesis tesing with MASE:

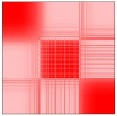
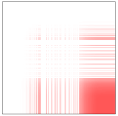
- **Test statistic**: reject $\mathcal{H}_0$ if $\|\hat{\mathbf{R}}^{(1)} - \hat{\mathbf{R}}^{(2)}\|_F^2$ is large
- **Null distribution**:
  - *Monte Carlo simulations* if $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$ are known
  - *Parametric boostrap* using $\hat{\mathbf{P}}^{(1)}$ and $\hat{\mathbf{P}}^{(2)}$
  - *Asymptotic distribution* of $\|\hat{\mathbf{R}}^{(1)} - \hat{\mathbf{R}}^{(2)}\|_F^2$ is a generalized chi-square

- The **omnibus embedding** (Levin et al., 2017) is an alternative approach for this task under the RDPG model.

Simulation 1

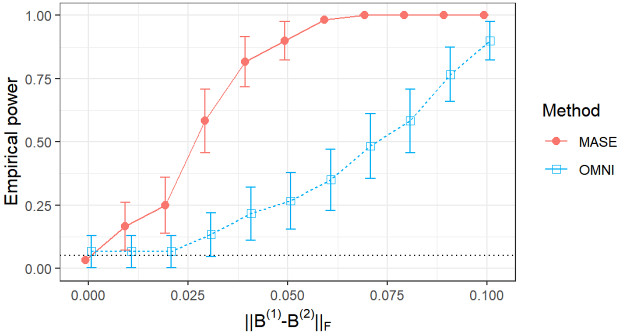
- Generate graphs from a mixed-membership SBM with $d = 3$ communities
  - $\mathbf{P}^{(i)} = \mathbf{Z}\mathbf{B}^{(i)}\mathbf{Z}^\top$
- Connection probabilities are perturbed: $\mathbf{B}_{11}^{(2)} = \mathbf{B}_{11}^{(1)} + \varepsilon$.

$\mathbf{P}^{(1)} =$ 

$\mathbf{P}^{(2)} =$  $+\ \varepsilon$ 

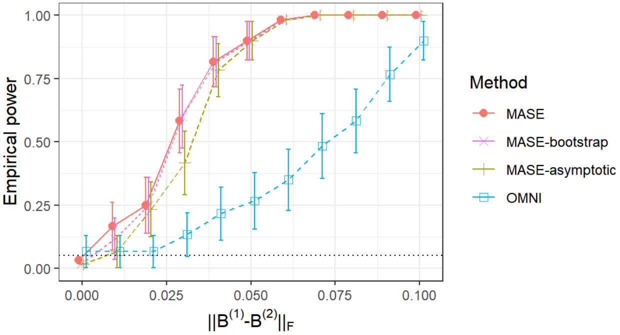
Simulation 1

- Generate graphs from a mixed-membership SBM with $d = 3$ communities
  - $\mathbf{P}^{(i)} = \mathbf{Z}\mathbf{B}^{(i)}\mathbf{Z}^\top$
- Connection probabilities are perturbed: $\mathbf{B}_{11}^{(2)} = \mathbf{B}_{11}^{(1)} + \varepsilon$.



Simulation 1

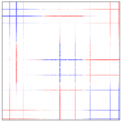
- Generate graphs from a mixed-membership SBM with $d = 3$ communities
  - $\mathbf{P}^{(i)} = \mathbf{Z}\mathbf{B}^{(i)}\mathbf{Z}^\top$
- Connection probabilities are perturbed: $\mathbf{B}_{11}^{(2)} = \mathbf{B}_{11}^{(1)} + \varepsilon$.



Simulation 2

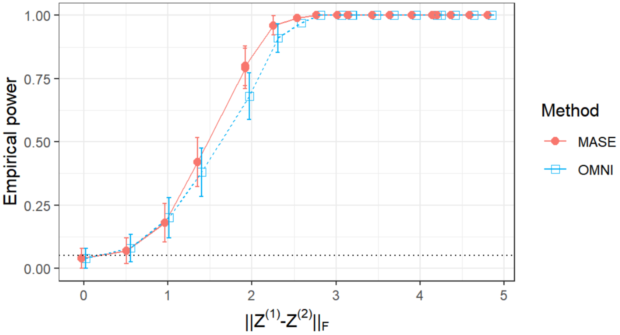
- Community connectivities are the same but some vertex memberships are perturbed
  - $\mathbf{P}^{(i)} = \mathbf{Z}^{(i)} \mathbf{B} \mathbf{Z}^{(i)\top}$
- Each $\mathbf{P}^{(i)}$ is rank 3, but the common invariant subspace has dimension $d = 5$.

$\mathbf{P}^{(1)} =$ 

$\mathbf{P}^{(2)} =$  $+$ 

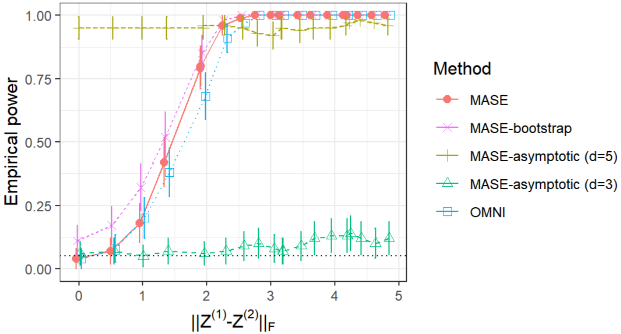
Simulation 2

- Community connectivities are the same but some vertex memberships are perturbed
  - $\mathbf{P}^{(i)} = \mathbf{Z}^{(i)}\mathbf{B}\mathbf{Z}^{(i)^{\top}}$
- Each $\mathbf{P}^{(i)}$ is rank 3, but the common invariant subspace has dimension $d = 5$.



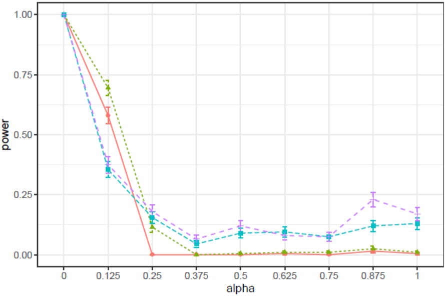
Simulation 2

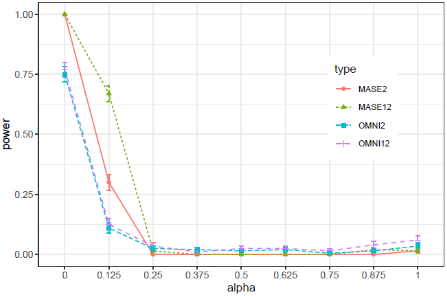
- Community connectivities are the same but some vertex memberships are perturbed
  - $\mathbf{P}^{(i)} = \mathbf{Z}^{(i)}\mathbf{B}\mathbf{Z}^{(i)\top}$
- Each $\mathbf{P}^{(i)}$ is rank 3, but the common invariant subspace has dimension $d = 5$.



Anomaly detection

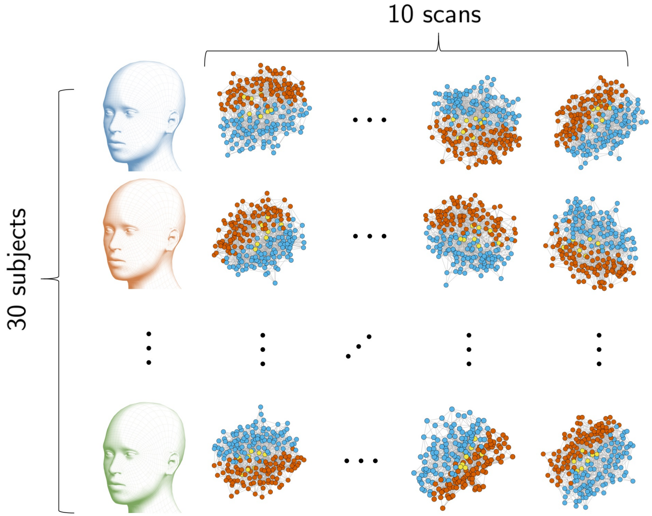
- Using hypothesis testing to detect anomalies in time series of graphs (Chen et al., 2020) https://arxiv.org/abs/2008.10055
- A parameter `alpha` controls the dissimilarity between the invariant subspaces of $\mathbf{P}^{(1)}$ and $\mathbf{P}^{(2)}$





Outline

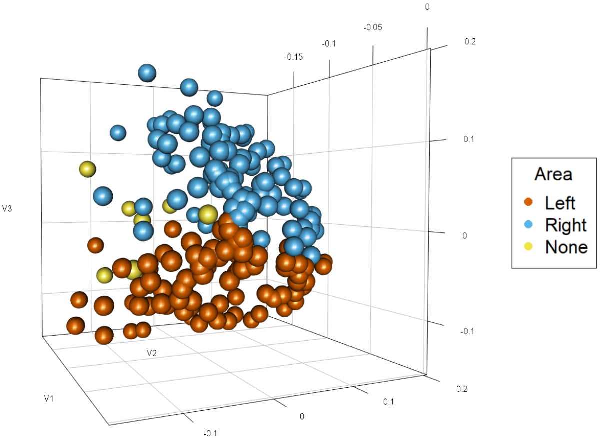
HNU1 study (Zuo et al., 2014)

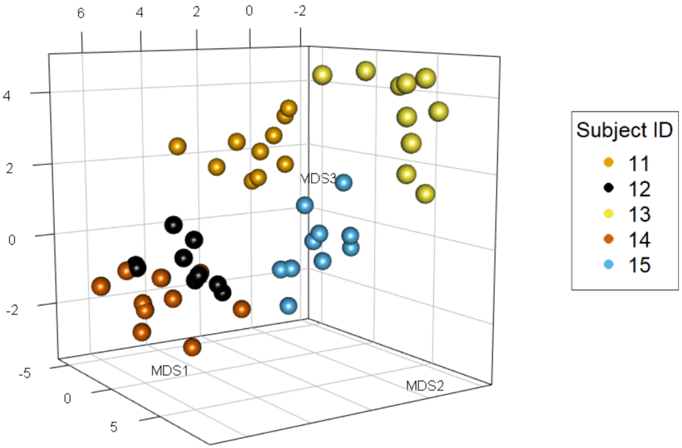


Structural brain graphs from diffusion MRI ($m = 300$, 30 subjects $\times$ 10 scans) encoding the existence of nerve tracts between $n = 200$ brain regions

COSIE model of brain networks

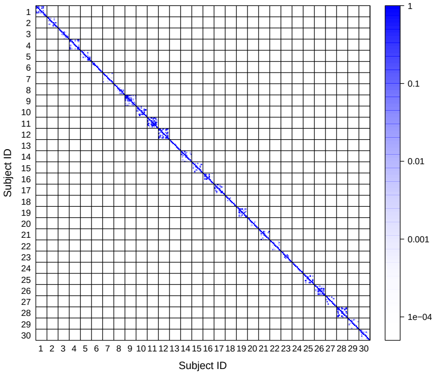
- Fit COSIE model with MASE using $d = 15$
- The plot shows the first three dimensions of $\widehat{\mathbf{V}}$
- Vertices are colored according to their location in the brain (left and right hemispheres)



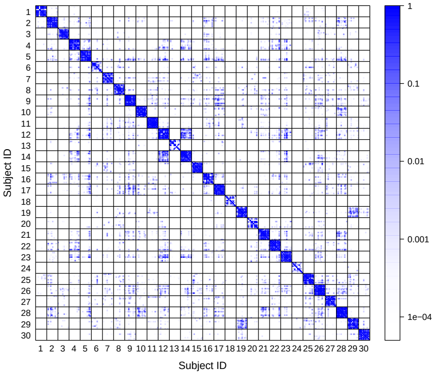


Estimated score matrices $\{\widehat{\mathbf{R}}^{(i)}\}$, embedded into $\mathbb{R}^3$ using MDS

Matrix of p-values for the hypothesis test $\mathcal{H}_0 : \mathbf{P}^{(i)} = \mathbf{P}^{(j)}$. Diagonal blocks correspond to same-subject graphs.

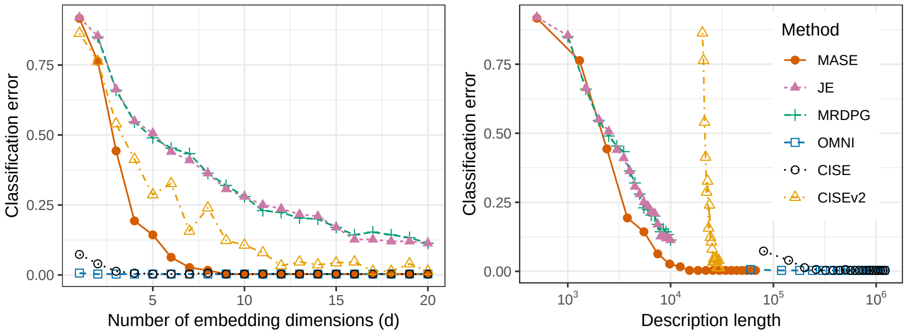


Asymptotic null distribution



Parametric bootstrap

- Classification error of a 1-nearest neighbor classifier, comparing different graph embedding methods as a function of the number of embedding dimensions.



Summary

- COSIE is a flexible and tractable model for multiple random graphs that retains enough homogeneity and permits sufficient heterogeneity.
- MASE provides consistent estimates for the COSIE parameters.
- Applications to subsequent graph inference tasks: classification, community detection, dimensionality reduction, hypothesis testing.
- COSIE and MASE provide domain-relevant insights in real data.

Summary

- COSIE is a flexible and tractable model for multiple random graphs that retains enough homogeneity and permits sufficient heterogeneity.
- MASE provides consistent estimates for the COSIE parameters.
- Applications to subsequent graph inference tasks: classification, community detection, dimensionality reduction, hypothesis testing.
- COSIE and MASE provide domain-relevant insights in real data.

Thank you!

Arroyo, J., Athreya, A., Cape, J., Chen, G., Priebe, C. E., Vogelstein, J. T. (2020). *"Inference for multiple heterogeneous networks with a common invariant subspace"*, Journal of Machine Learning Research (to appear), arXiv:1906.10026.

References I

Crainiceanu, Ciprian M et al. (2011). "Population value decomposition, a framework for the analysis of image populations". In: *Journal of the American Statistical Association* 106.495, pp. 775–790.

Fan, Jianqing et al. (2019). "Distributed estimation of principal eigenspaces". In: *Annals of statistics* 47.6, p. 3009.

Holland, Paul W., Kathryn Blackmond Laskey, and Samuel Leinhardt (June 1983). "Stochastic blockmodels: First steps". In: *Social Networks* 5.2, pp. 109–137.

Levin, Keith et al. (May 2017). "A central limit theorem for an omnibus embedding of multiple random dot product graphs". In: *arXiv preprint arXiv:1705.09355*. arXiv: 1705.09355.

Young, Stephen J and Edward R Scheinerman (2007). "Random dot product graph models for social networks". In: *International Workshop on Algorithms and Models for the Web-Graph*. Springer, pp. 138–149.