Bias-Variance Tradeoffs in Joint Spectral Embeddings

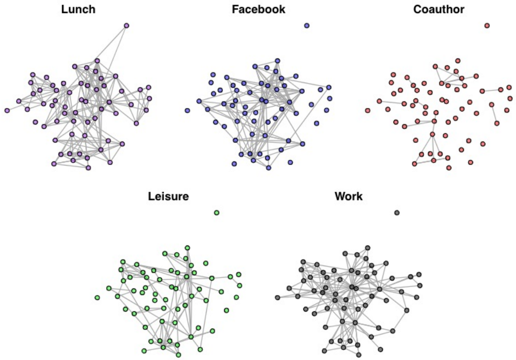
Benjamin Draves   Daniel Sussman

Boston University

Joint Statistical Meetings, 3 August 2020

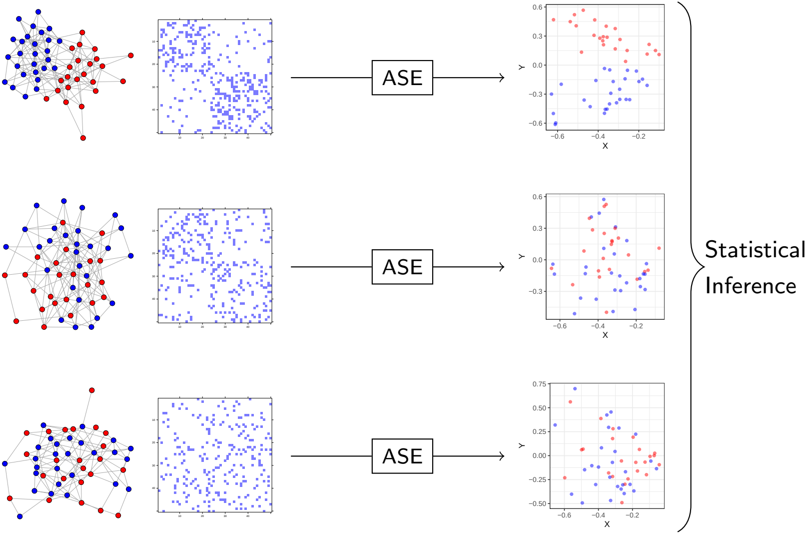
Multiplex Networks

- Multiplex networks encode multiple relationships between entities as a collection of networks (Magnani, Micenkova, and Rossi 2013).

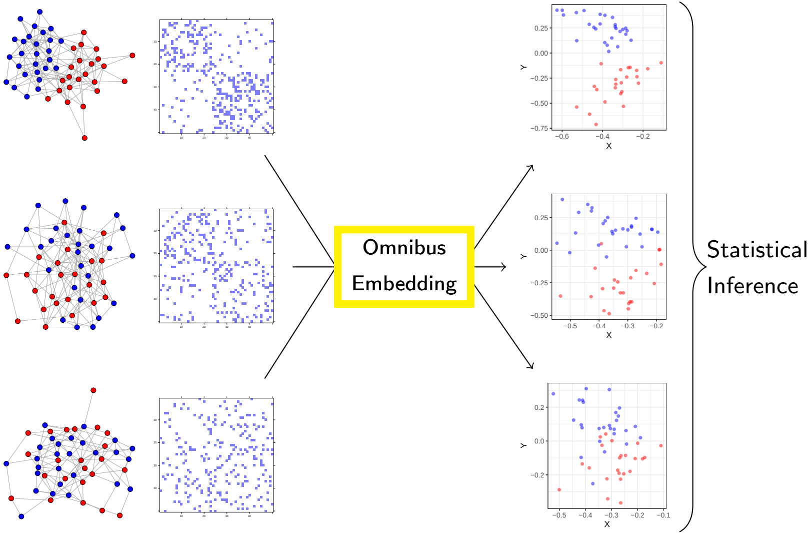


- Application areas; International Trade, Transportation Systems, Terrorist Groups, Neuroscience (Kivelä et al. 2014).

Individual Spectral Embeddings



Joint Spectral Embeddings



Analysis Framework

- Consider m graphs over a common vertex set $\mathcal{V}$ of size n
- Associate $v \in \mathcal{V}$ with a *latent position* $\mathbf{X}_v \in \mathbb{R}^d$

Inner Product Distribution

Let F be a probability distribution over $\mathbb{R}^d$. We say F is a *d-dimensional inner product distribution* if all $\mathbf{x}, \mathbf{y} \in \text{supp}(F)$ has the property $\mathbf{x}^T\mathbf{y} \in [0, 1]$.

- Assume latent positions $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_n \overset{i.i.d.}{\sim} F$. Organize in the rows of a matrix $\mathbf{X} = [\mathbf{X}_1\mathbf{X}_2 \ldots \mathbf{X}_n]^T$.
- Associate each network with a diagonal matrix $\mathbf{C}^{(g)} \in \mathbb{R}^{d\times d}_{\geq 0}$ such that for all $\mathbf{x}, \mathbf{y} \in \text{supp}(F)$ and $g \in [m]$, $\mathbf{x}^T\mathbf{C}^{(g)}\mathbf{y} \in [0, 1]$.

Eigen-Scaling Random Dot Product Graph

Eigen-Scaling Random Dot Product Graph

- Suppose that for $\mathbf{y} \sim F$, $\Delta = \mathbb{E}[\mathbf{y}\mathbf{y}^T]$ is diagonal and full rank and the matrices $\{\mathbf{C}^{(g)}\}_{g=1}^m$ satisfy $\min_{i\in[d]} \max_{g\in[m]} \mathbf{C}_{ii}^{(g)} > 0$.
- Then the random adjacency matrices $\{\mathbf{A}^{(g)}\}_{g=1}^m$ are said to be jointly distributed according to the *ESRDPG* with *latent positions* $\mathbf{X}$ iff $\{\mathbf{A}_{ij}^{(g)}\}$ are conditionally independent with

$$\mathbb{P}(\mathbf{A}_{ij}^{(g)} = 1 | \mathbf{X}_i, \mathbf{X}_j) = \mathbf{X}_i^T \mathbf{C}^{(g)} \mathbf{X}_j$$

- In essence, $\mathbf{A}_{ij}^{(g)} | \mathbf{X} \overset{ind.}{\sim} \text{Bern}(\mathbf{X}_i^T \mathbf{C}^{(g)} \mathbf{X}_j)$.
- Goal: Given $\{\mathbf{A}^{(g)}\}_{g=1}^m$, estimate $\{\mathbf{X}\sqrt{\mathbf{C}^{(g)}}\}_{g=1}^m$

Individual Network Embedding Techniques

- First approach: ignore shared structure and individually embedd networks $\mathbf{A}^{(g)}$ for $g \in [m]$.

Adjacency Spectral Embedding (Sussman et al. 2012)

Let $\mathbf{A}^{(g)}$ have eigendecomposition

$$\mathbf{A}^{(g)} = [\mathbf{U}_{\mathbf{A}^{(g)}} | \tilde{\mathbf{U}}_{\mathbf{A}^{(g)}}][\mathbf{S}_{\mathbf{A}^{(g)}} \oplus \tilde{\mathbf{S}}_{\mathbf{A}^{(g)}}][\mathbf{U}_{\mathbf{A}^{(g)}} | \tilde{\mathbf{U}}_{\mathbf{A}^{(g)}}]^T$$

where $\mathbf{U}_{\mathbf{A}^{(g)}} \in \mathbb{R}^{n \times d}$ and $\mathbf{S}_{\mathbf{A}^{(g)}} \in \mathbb{R}^{d \times d}$ contains the top d eigenvalues of $\mathbf{A}^{(g)}$. Then the ASE of $\mathbf{A}^{(g)}$ is defined by $\text{ASE}(\mathbf{A}^{(g)}, d) = \mathbf{U}_{\mathbf{A}^{(g)}} \mathbf{S}_{\mathbf{A}^{(g)}}^{1/2}$.

- Second approach: assume identical structure and embed the sample mean matrix $\bar{\mathbf{A}} = m^{-1} \sum_{g=1}^{m} \mathbf{A}^{(g)}$ by $\text{ASE}(\bar{\mathbf{A}}, d)$.

Joint Network Embedding Techniques

- Third approach: *jointly* embed the networks $\{\mathbf{A}^{(g)}\}_{g=1}^m$.

Omnibus Embedding (Levin et al. 2017)

Let the *omnibus matrix* be defined as

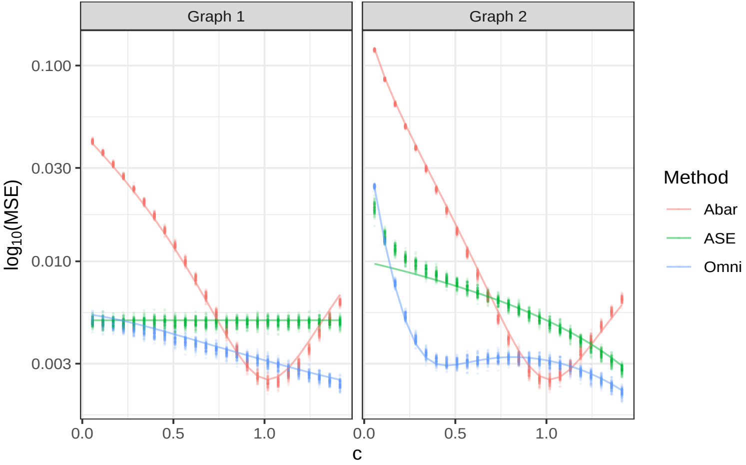
$$\tilde{\mathbf{A}} = \begin{bmatrix} \mathbf{A}^{(1)} & \frac{1}{2}[\mathbf{A}^{(1)} + \mathbf{A}^{(2)}] & \dots & \frac{1}{2}[\mathbf{A}^{(1)} + \mathbf{A}^{(m)}] \\ \frac{1}{2}[\mathbf{A}^{(2)} + \mathbf{A}^{(1)}] & \mathbf{A}^{(2)} & \dots & \frac{1}{2}[\mathbf{A}^{(2)} + \mathbf{A}^{(m)}] \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}[\mathbf{A}^{(m)} + \mathbf{A}^{(1)}] & \frac{1}{2}[\mathbf{A}^{(m)} + \mathbf{A}^{(2)}] & \dots & \mathbf{A}^{(m)} \end{bmatrix}.$$

Then the *omnibus embedding* of $\{\mathbf{A}^{(g)}\}_{g=1}^m$ is given by $\hat{\mathbf{L}} = \text{ASE}(\tilde{\mathbf{A}}, d)$.

- Notice $\hat{\mathbf{L}} \in \mathbb{R}^{nm \times d}$ so each vertex has a latent position estimate for each graph.

Mean Squared Error Comparison

- Suppose $\mathbf{A}^{(1)} \sim \text{ER}(p)$ and $\mathbf{A}^{(2)} \sim \text{ER}(c^2 p)$
- Under ESRDPG: $\mathbf{X} = \sqrt{p}\mathbf{1}_n$, $\mathbf{C}^{(1)} = \mathbf{I}$, and $\mathbf{C}^{(2)} = c^2\mathbf{I}$



Main Results

- Let $\hat{\mathbf{L}} = \text{ASE}(\tilde{\mathbf{A}}, d)$ and $h = n(g-1)+i$ for $i \in [n]$ and $g \in [m]$ so that $\hat{\mathbf{L}}_h \in \mathbb{R}^{d\times 1}$ is some row of $\hat{\mathbf{L}}$ written as a column vector.

Theorem

- *There exists diagonal matrices* $\{\mathbf{S}^{(g)}\}_{g=1}^m$ *that only depend on* $\{\mathbf{C}^{(g)}\}_{g=1}^m$ *and a sequence of orthogonal matrices* $\{\tilde{\mathbf{W}}_n\}_{n=1}^\infty$ *such that*

$$\hat{\mathbf{L}}\tilde{\mathbf{W}}_n - \mathbf{L} = (\mathbf{S}^{(g)} - \sqrt{\mathbf{C}^{(g)}})\mathbf{X}_i + \mathbf{R}_h \tag{1}$$

where $\mathbf{R}_h$ *is a residual.*

- $\mathbf{R}_h$ *satisfies* $\max_{h\in[nm]} \|\mathbf{R}_h\|_2 = O_{\mathbb{P}}\left(m^{3/2}\frac{\log nm}{\sqrt{n}}\right)$ *and has asympoptic distribution*

$$\lim_{n\to\infty} \mathbb{P}\left[\sqrt{n}\mathbf{R}_h \leq \mathbf{x}\right] = \int_{supp(F)} \Phi(\mathbf{x}; \mathbf{0}, \Sigma_g(\mathbf{y}))dF(\mathbf{y}). \tag{2}$$

Simulation Experiment

Simulation Design

1. Draw $\mathbf{X}_1, \ldots \mathbf{X}_n \overset{i.i.d}{\sim} F$ where F corresponds to a two-group SBM with parameters $(a = 0.25, b = 0.05)$.
2. For $t \in [0, 1]$, draw $(\{\mathbf{A}^{(g)}\}_{g=1}^2, \mathbf{X}) \sim \text{ESRDPG}(F, n, \{\mathbf{C}^{(g)}\}_{g=1}^m)$ with

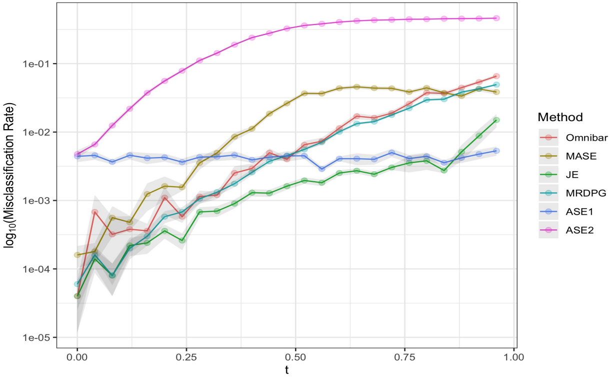
$$\mathbf{C}^{(1)} = \mathbf{I} \qquad \mathbf{C}^{(2)} := \mathbf{C}(t) = \begin{bmatrix} 1+t & 0 \\ 0 & 1-t \end{bmatrix}$$

3. Jointly embed $\hat{\mathbf{L}} = \text{ASE}(\tilde{\mathbf{A}}, d)$

- At $t = 0$, $\mathbf{A}^{(2)}$ is a SBM and at $t = 1$, $\mathbf{A}^{(2)}$ is an Erdös-Réyni graph with parameter $p = 0.3$.
- Goal: Analyze techniques that utilize $\hat{\mathbf{L}}$ for statistical inference.

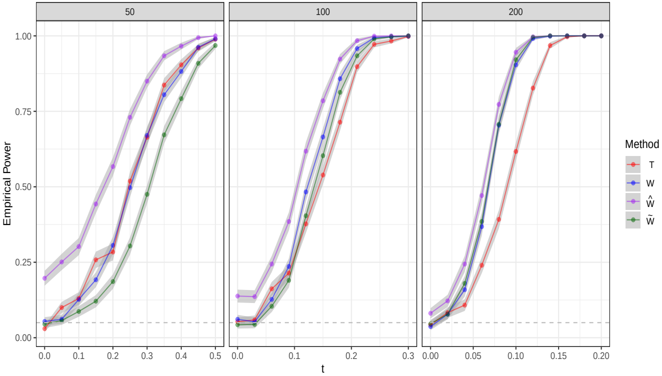
Community Detection

- Task: Recover community labels using $\hat{\mathbf{L}} = [\hat{\mathbf{X}}^{(1)T} \hat{\mathbf{X}}^{(2)T}]^T$.
- Apply Gaussian Mixture Models to the rows of $\bar{\mathbf{X}} = 2^{-1}(\hat{\mathbf{X}}^{(1)} + \hat{\mathbf{X}}^{(2)})$



Two Graph Hypothesis Testing

- Task: Test the hypothesis $H_0 : \mathbf{C}^{(1)} = \mathbf{C}^{(2)}$.
- Construct pivotal test statistic based on the Mahalanobis distance between rows of $\hat{\mathbf{X}}^{(1)}$ and $\hat{\mathbf{X}}^{(2)}$.



Conclusion & Future Work

Preprint available: `https://arxiv.org/abs/2005.02511`

- Extend this analysis to more general graph models
- Power Analysis for proposed test statistic
- Refine finite sample testing procedures

References I

- Mikko Kivelä et al. "Multilayer networks". In: *Journal of Complex Networks* 2.3 (2014), pp. 203–271.
- Keith Levin et al. "A Central Limit Theorem for an Omnibus Embedding of Multiple Random Dot Product Graphs". In: Nov. 2017, pp. 964–967.
- Matteo Magnani, Barbora Micenkova, and Luca Rossi. *Combinatorial Analysis of Multiple Networks*. 2013. arXiv: 1303.4986 [cs.SI].
- Daniel L. Sussman et al. "A Consistent Adjacency Spectral Embedding for Stochastic Blockmodel Graphs". In: *Journal of the American Statistical Association* 107.499 (2012), pp. 1119–1128.