Bias-Variance Tradeoffs in Joint Spectral Embeddings

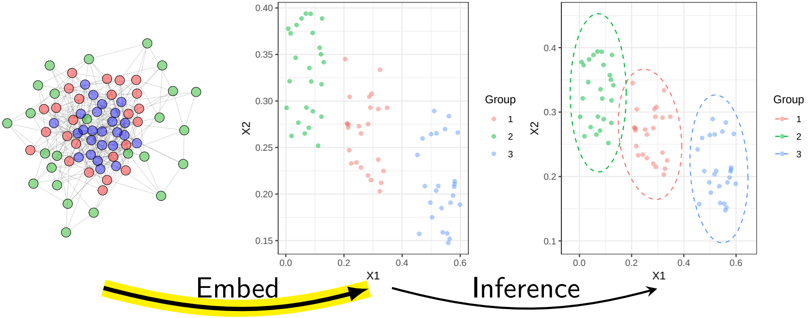
Benjamin Draves   Daniel Sussman

Boston University

Joint Statistical Meetings, 1 August 2019

Network Embeddings

- Goal: Represent vertices of a network in a low dimensional space
- Reason: Use machine learning techniques to answer network based questions



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 $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n \overset{i.i.d.}{\sim} F$. Organize in the rows of a matrix $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \dots \mathbf{X}_n]^T$.

Technical Assumption

For $\mathbf{y} \sim F$ assume $\Delta = \mathbb{E}[\mathbf{y}\mathbf{y}^T]$ is diagonal with $\min_{i \in [d]} \Delta_{ii} > 0$

Multiplex Colinear Random Dot Product Graph

Multiplex Colinear Random Dot Product Graph (MCRDPG)

- Let $\mathbf{C}^{(1)}, \dots, \mathbf{C}^{(m)} \in \mathbb{R}^{d \times d}$ be a diagonal matrices with non-negative entries so that $\mathbf{X}_i^T \mathbf{C}^{(g)} \mathbf{X}_j \in [0, 1]$ for all $i \in [n]$ and $g \in [m]$.
- The random adjacency matrices $\{\mathbf{A}^{(g)}\}_{g=1}^m$ are distributed according to the *MCRDPG* with *latent positions* $\mathbf{X}$ iff $\{\mathbf{A}_{ij}^{(g)}\}$ are conditionally independent with

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

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

Adjacency Spectral Embedding

- Idea 1: Ignore joint structure and estimate $\mathbf{X}\sqrt{\mathbf{C}^{(g)}}$ individually

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 $\text{ASE}(\mathbf{A}^{(g)}, d) = \mathbf{U}_{\mathbf{A}^{(g)}}\mathbf{S}_{\mathbf{A}^{(g)}}^{1/2}$.

- Estimate $\mathbf{X}\sqrt{\mathbf{C}^{(g)}}$ by $\hat{\mathbf{X}}_{\text{ASE}}^{(g)} = \text{ASE}(\mathbf{A}^{(g)}, d)$
- Idea 2: Assume i.i.d. ($\mathbf{C}^{(g)} = \mathbf{C}^*$) and estimate a global $\hat{\mathbf{X}}$
- Estimate $\mathbf{X}\sqrt{\mathbf{C}^{(g)}}$ by $\hat{\mathbf{X}}_{\text{Abar}} = \text{ASE}(\bar{\mathbf{A}}, d)$ where $\bar{\mathbf{A}} = m^{-1}\sum_{g=1}^{m}\mathbf{A}^{(g)}$

Omnibus Embedding

- Idea 3: Incorporate the joint structure in the embedding

Omnibus Embedding (Levin et al. 2017)

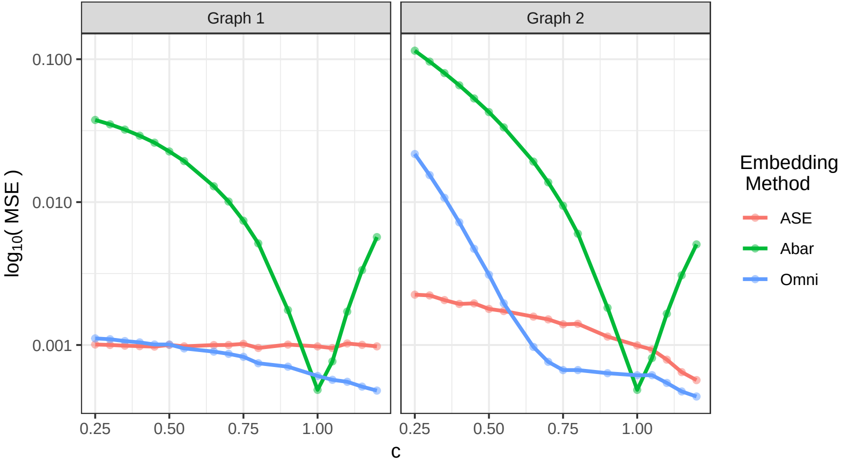
The *Omnibus matrix* is given by

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

- Notice $\text{ASE}(\tilde{\mathbf{A}}, d) \in \mathbb{R}^{nm \times d}$
- Estimate $\mathbf{X}\sqrt{\mathbf{C}^{(g)}}$ by $\hat{\mathbf{X}}^{(g)}_{\text{Omni}} = [\text{ASE}(\tilde{\mathbf{A}}, d)]_g$ where $[\cdot]_g$ denotes the g-th n row block.

Simulation: Mean Squared Error

- Suppose $\mathbf{A}^{(1)} \sim \mathrm{ER}(p)$ and $\mathbf{A}^{(2)} \sim \mathrm{ER}(c^2 p)$
- Then the latent positions are $\sqrt{p}$ for $\mathbf{A}^{(1)}$ and $c\sqrt{p}$ for $\mathbf{A}^{(2)}$



Bias in Joint Spectral Embeddings

- Preliminary notation:
  - Let $h = i + n(g-1)$ for $i \in [n]$ and $g \in [m]$ and define $(\mathbf{M})_i = \mathbf{M}_{i\cdot}^T$
  - Let $\mathbf{L} = [\sqrt{\mathbf{C}^{(1)}}\mathbf{X}^T \ \sqrt{\mathbf{C}^{(2)}}\mathbf{X}^T \dots \sqrt{\mathbf{C}^{(m)}}\mathbf{X}^T]^T$.

Theorem

- *Let* $(\{\mathbf{A}^{(g)}\}_{g=1}^m, \mathbf{X}) \sim MCRDPG(F, n, \{\mathbf{C}^{(g)}\}_{g=1}^m)$. *Let* $\hat{\mathbf{L}} = ASE(\tilde{\mathbf{A}}, d)$ *be the Omnibus embedding of* $\{\mathbf{A}^{(g)}\}_{g=1}^m$.
- *There exists diagonal matrices* $\{\mathbf{S}^{(g)}\}_{g=1}^m$ *that only depend on* $\{\mathbf{C}^{(g)}\}_{g=1}^m$ *and an orthogonal matrix* $\mathbf{W}$ *such that*

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

Moreover, with high probability

$$\max_{h \in [nm]} \|\mathbf{R}_h\|_2 \leq O\left(\frac{m^{3/2}\log nm}{\sqrt{n}}\right) \tag{2}$$

Simulation Design

- Base model: Balanced two group SBM, $\mathbf{B} = \begin{bmatrix} 1/4 & 1/20 \\ 1/20 & 1/4 \end{bmatrix}$

Network	**Weighting**	**Block Probabilities**	**Model**
$\mathbf{A}^{(1)}$	$\mathbf{C}^{(1)} = \begin{bmatrix} 3/4 & 0 \\ 0 & 1/2 \end{bmatrix}$	$\mathbf{B}_1 = \begin{bmatrix} 13/80 & 1/16 \\ 1/16 & 13/80 \end{bmatrix}$	Connected SBM
$\mathbf{A}^{(2)}$	$\mathbf{C}^{(2)} = \begin{bmatrix} 1/2 & 0 \\ 0 & 3/4 \end{bmatrix}$	$\mathbf{B}_2 = \begin{bmatrix} 3/20 & 0 \\ 0 & 3/20 \end{bmatrix}$	Disconnected SBM
$\mathbf{A}^{(3)}$	$\mathbf{C}^{(3)} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$	$\mathbf{B}_3 = \begin{bmatrix} 3/20 & 3/20 \\ 3/20 & 3/20 \end{bmatrix}$	Erdös-Réyni

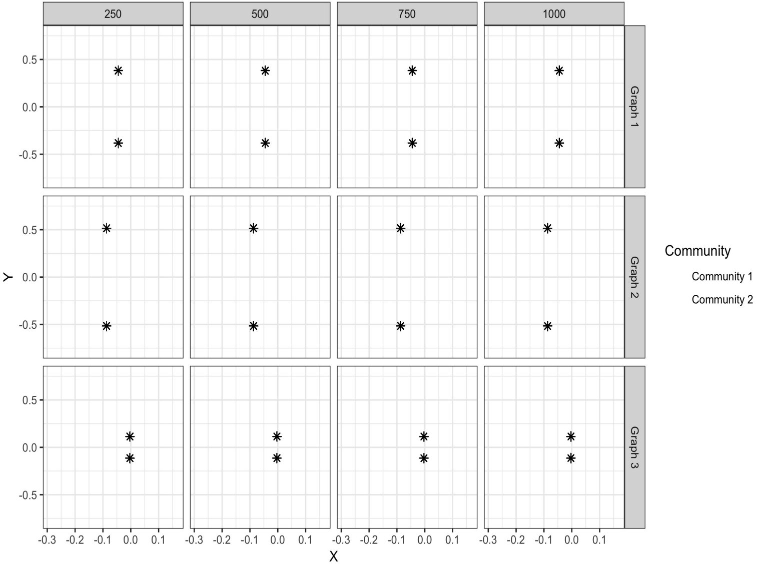
Simulation Design (continued)

Simulation Design

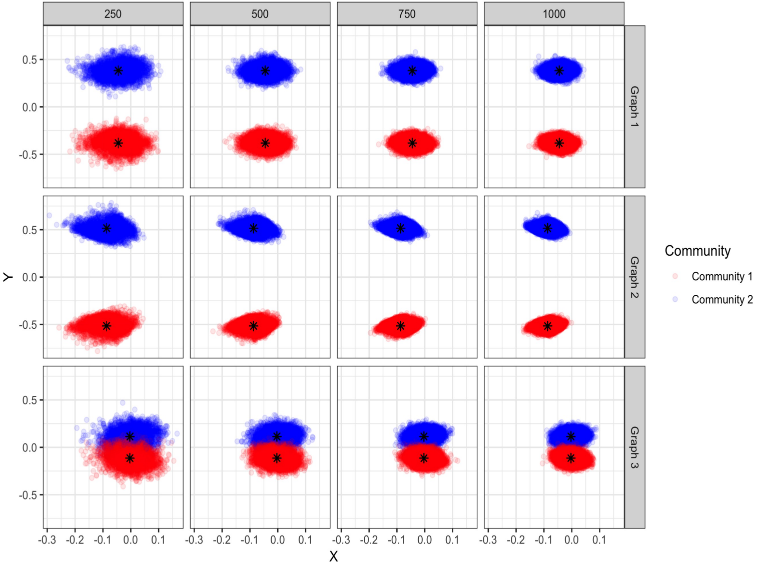
1. Assign vertices to each group with probability $\pi = .5$
2. Sample adjacency matrices
   - $\mathbf{A}^{(1)} \sim \mathrm{SBM}(\mathbf{B}_1, \pi)$
   - $\mathbf{A}^{(2)} \sim \mathrm{SBM}(\mathbf{B}_2, \pi)$
   - $\mathbf{A}^{(3)} \sim \mathrm{ER}(p = 0.15)$
3. Embed $\tilde{\mathbf{A}}$ to obtain $\hat{\mathbf{L}} = [\hat{\mathbf{X}}_{\mathrm{Omni}}^{(1)T} \ \hat{\mathbf{X}}_{\mathrm{Omni}}^{(2)T} \ \hat{\mathbf{X}}_{\mathrm{Omni}}^{(3)T}]^T$

- Compare analytic bias $(\mathbf{S}^{(g)} - \sqrt{\mathbf{C}^{(g)}})\mathbf{X}_i$ to observed bias $(\hat{\mathbf{L}}\mathbf{W} - \mathbf{L})_h$.
- Compare residual bound $O\left(\frac{m^{3/2}\log nm}{\sqrt{n}}\right)$ to observed residuals $\hat{\mathbf{R}}_h$.

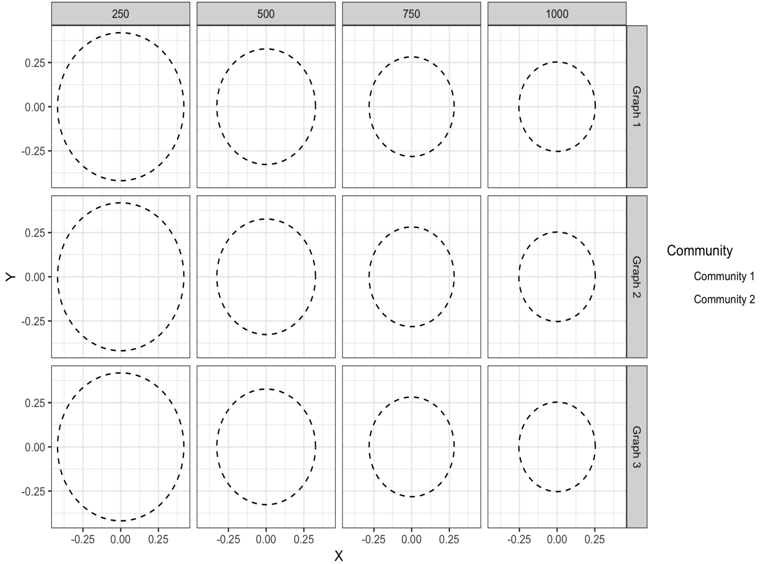
Simulation Results: Bias



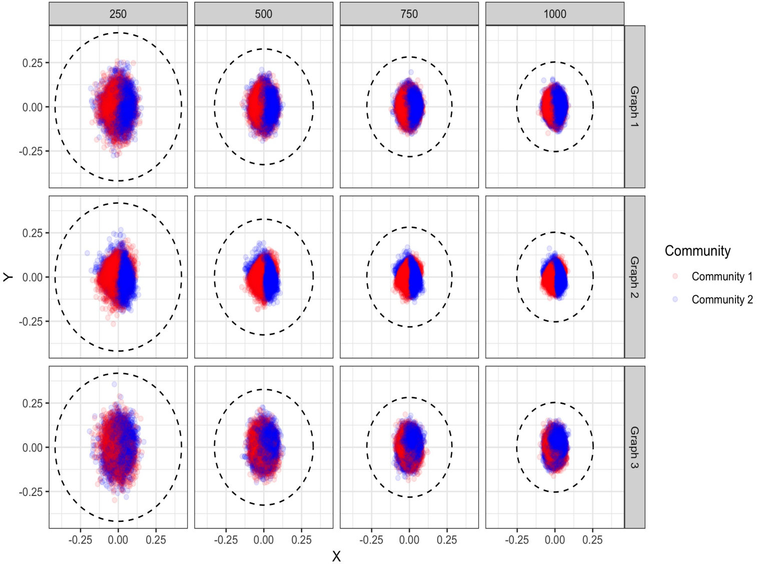
Simulation Results: Bias



Simulation Results: Residuals



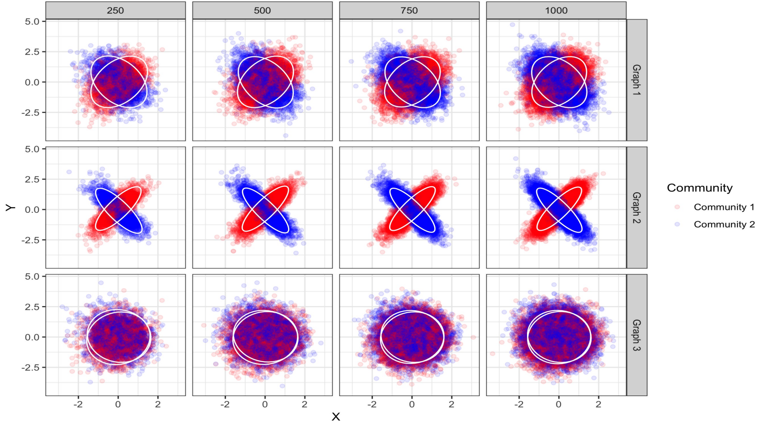
Simulation Results: Residuals



Variance in Joint Spectral Embeddings

Residual Decomposition

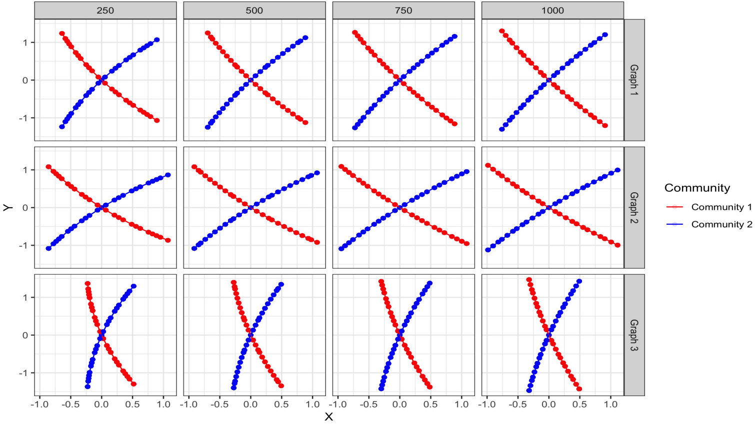
Decompose the residual term $\sqrt{n}\mathbf{R}_i = \sqrt{n}\mathbf{R}_i^{(1)} + \sqrt{n}\mathbf{R}_i^{(2)}$



Variance in Joint Spectral Embeddings

Residual Decomposition

Decompose the residual term $\sqrt{n}\mathbf{R}_i = \sqrt{n}\mathbf{R}_i^{(1)} + \sqrt{n}\mathbf{R}_i^{(2)}$



Conclusion

- Introduced the *MCRDPG* probability model
- Highlighted an advantageous bias-variance tradeoff given by the Omnibus Embedding
- Established
  1. Bias of the Omnibus Estimator under the *MCRDPG*
  2. Uniform bound on the residual term at a $O(m^{3/2}\log nm/\sqrt{n})$ rate
- Highlighted second moment properties of the Omnibus Embedding

Questions?

References I

- Avanti Athreya et al. "Statistical inference on random dot product graphs: A survey". In: *Journal of Machine Learning Research* 18 (Sept. 2017).
- A. Athreya et al. "A Limit Theorem for Scaled Eigenvectors of Random Dot Product Graphs". In: *Sankhya A* 78.1 (Feb. 2016), pp. 1–18.
- Peter D Hoff, Adrian E Raftery, and Mark S Handcock. "Latent Space Approaches to Social Network Analysis". In: *Journal of the American Statistical Association* 97.460 (2002), pp. 1090–1098.
- Keith Levin et al. "A Central Limit Theorem for an Omnibus Embedding of Multiple Random Dot Product Graphs". In: Nov. 2017, pp. 964–967.

References II

- 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.
- Shangsi Wang et al. *Joint Embedding of Graphs*. 2017.
- Stephen J. Young and Edward R. Scheinerman. "Random Dot Product Graph Models for Social Networks". In: *Algorithms and Models for the Web-Graph*. Ed. by Anthony Bonato and Fan R. K. Chung. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 138–149.