# Bias in Joint Spectral Embeddings

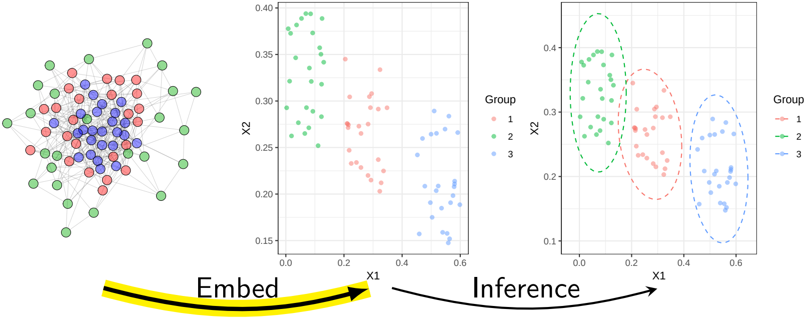
Benjamin Draves

Advisor: D. Sussman

CSGW 6.0 January 2019

# Network Embeddings

- Goal: Obtain a low dimensional representation of the network
- Reason: Use machine learning techniques to answer network based questions



# Latent Position Models

- Assume each vertex $v \in V$ is associated with a random vector $X_v \in \mathbb{R}^d$ that characterize the connectivity structure

**Random Dot Product Graph (RDPG) (Sussman et al. 2012)**

Suppose that $F$ is a probability distribution on $\mathbb{R}^d$ such that for all $x, y \in \text{supp}(F)$, $x^T y \in [0,1]$. Let $\{X_i\}_{i=1}^n \overset{i.i.d.}{\sim} F$ and $\mathbf{X} = [X_1, \dots, X_n]^T$. We say $(\mathbf{A}, \mathbf{X}) \sim RDPG(F, n)$ with *latent positions* $\mathbf{X}$ iff $\{\mathbf{A}_{ij}\}$ are conditionally independent with

$$\mathbb{P}(\mathbf{A}_{ij} = 1 | \mathbf{X}) = X_i^T X_j$$

- In essence, $\mathbf{A}_{ij} | \mathbf{X} \overset{ind.}{\sim} \text{Bern}(X_i^T X_j)$ and $\mathbf{P} \equiv \mathbb{E}(\mathbf{A}|\mathbf{X}) = \mathbf{X}\mathbf{X}^T$
- Task: Given $\{\mathbf{A}^{(k)}\}_{k=1}^m$ with latent positions $\mathbf{X}^{(k)}$, how do we estimate $\mathbf{X}^{(k)}$?

# Omnibus Embedding

## Omnibus Embedding (Levin et al. 2017)

The *Omnibus matrix* is given by

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

The $d$-dimensional *Omnibus embedding* is $\widehat{\mathbf{Z}} \equiv \mathbf{U_M S_M^{1/2}}$ where

$$\mathbf{M} = \mathbf{U_M S_M U_M^T} + \widetilde{\mathbf{U}}_\mathbf{M} \widetilde{\mathbf{S}}_\mathbf{M} \widetilde{\mathbf{U}}_\mathbf{M}^T$$

$\mathbf{S_M} \in \mathbb{R}^{d \times d}$ is a diagonal matrix of the top $d$ eigenvalues and $\mathbf{U_M} \in \mathbb{R}^{nm \times d}$ has columns that are the corresponding eigenvectors

# Omnibus Embedding - Intuition

- Suppose that $\mathbf{A}^{(k)}$ have the same latent positions $\mathbf{X}^{(k)} = \mathbf{X}$
- Then we have

$$\mathbf{M} = \mathbf{U_M S_M U_M^T} + \widetilde{\mathbf{U}}_\mathbf{M}\widetilde{\mathbf{S}}_\mathbf{M}\widetilde{\mathbf{U}}_\mathbf{M}^T = \underbrace{\widehat{\mathbf{Z}}}_{nm\times d}\underbrace{\widehat{\mathbf{Z}}^T}_{d\times nm} + \widetilde{\mathbf{Z}}\widetilde{\mathbf{Z}}^T$$

$$\widetilde{\mathbf{P}} = \mathbb{E}(\mathbf{M}|\mathbf{X}) = \begin{bmatrix} \mathbf{XX}^T & \dots & \mathbf{XX}^T \\ \vdots & \ddots & \vdots \\ \mathbf{XX}^T & \dots & \mathbf{XX}^T \end{bmatrix} = \underbrace{\left(\begin{bmatrix} \mathbf{X} \\ \vdots \\ \mathbf{X} \end{bmatrix}\right)}_{nm\times d}\underbrace{\left(\begin{bmatrix} \mathbf{X} \\ \vdots \\ \mathbf{X} \end{bmatrix}\right)^T}_{d\times nm}$$

- Therefore, $\widehat{\mathbf{Z}}$ serves as a natural estimator of the latent positions

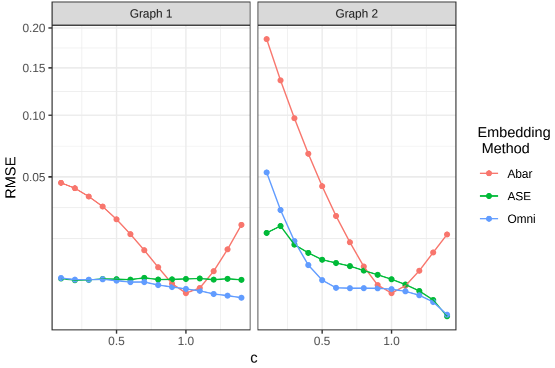
$$\mathbf{L} = [\mathbf{X}^{(1)T}\mathbf{X}^{(2)T}\dots\mathbf{X}^{(m)T}]^T$$

up to an orthogonal rotation.

# Trivia Time! What's the name of this building?

# Simulation: (Root) Mean Squared Error

- Motivating Question: What if the latent positions are different for each network?
- Suppose $\mathbf{A}_{ij}^{(1)}$ and $\mathbf{A}_{ij}^{(2)}$ have latent positions $\sqrt{p}$ and $c\sqrt{p}$ so that $\mathbf{P}_{ij}^{(1)} = p$ and $\mathbf{P}_{ij}^{(2)} = c^2 p$



# Analysis Setup

- What are some properties of the Omnibus embedding, $\widehat{\mathbf{Z}}$?
- First, consider
$$\widetilde{\mathbf{P}} = \mathbf{U}_{\widetilde{\mathbf{P}}}\mathbf{S}_{\widetilde{\mathbf{P}}}\mathbf{U}_{\widetilde{\mathbf{P}}}^T + \widetilde{\mathbf{U}}_{\widetilde{\mathbf{P}}}\widetilde{\mathbf{S}}_{\widetilde{\mathbf{P}}}\widetilde{\mathbf{U}}_{\widetilde{\mathbf{P}}}^T$$
and define $\mathbf{Z} \equiv \mathbf{U}_{\widetilde{\mathbf{P}}}\mathbf{S}_{\widetilde{\mathbf{P}}}^{1/2} \in \mathbb{R}^{nm\times d}$
- For some orthogonal matrix $\mathbf{W}$ consider
$$\left(\widehat{\mathbf{Z}} - \mathbf{L}\mathbf{W}\right)_i = \left(\mathbf{Z} - \mathbf{L}\mathbf{W}\right)_i + \left(\widehat{\mathbf{Z}} - \mathbf{Z}\right)_i$$

- Bias Term
- Variance Term

# Exploiting Kronecker Structure

- In order to analyze the bias, we need a closed form of $\mathbf{Z}$
- First consider the following decomposition

$$\begin{bmatrix} 1 & \frac{1+c^2}{2} \\ \frac{1+c^2}{2} & c^2 \end{bmatrix} \equiv \left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right)\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}\right)^T + \left(\begin{bmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{bmatrix}\right)\left(\begin{bmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{bmatrix}\right)^T$$

- Then using the Kronecker Structure of $\widetilde{\mathbf{P}} = \mathbb{E}(\mathbf{M}|\mathbf{X})$

$$\widetilde{\mathbf{P}} = \begin{bmatrix} \mathbf{P}^{(1)} & \frac{1}{2}\left(\mathbf{P}^{(1)} + \mathbf{P}^{(2)}\right) \\ \frac{1}{2}\left(\mathbf{P}^{(1)} + \mathbf{P}^{(2)}\right) & \mathbf{P}^{(2)} \end{bmatrix} = \begin{bmatrix} 1 & \frac{1+c^2}{2} \\ \frac{1+c^2}{2} & c^2 \end{bmatrix} \otimes \mathbf{X}\mathbf{X}^T$$

$$= \left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \otimes \mathbf{X}\right)\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \otimes \mathbf{X}\right)^T + \left(\begin{bmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{bmatrix} \otimes \mathbf{X}\right)\left(\begin{bmatrix} \tilde{\alpha} \\ \tilde{\beta} \end{bmatrix} \otimes \mathbf{X}\right)^T$$

- Therefore $\mathbf{Z} = \mathbf{U}_{\widetilde{\mathbf{P}}}\mathbf{S}_{\widetilde{\mathbf{P}}}^{1/2} = \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \otimes \mathbf{X}$

# Theorem: Bias in Joint Spectral Embeddings

## Bias in Joint Spectral Embeddings

Suppose $\mathbf{X} = \mathbf{U\Sigma V}^T$ and $\widetilde{\mathbf{X}} = \mathbf{U\Sigma}$. Then with the notation as above, there exists a matrix $\mathbf{W} \in \mathcal{O}^{d\times d}$ such that

$$\left(\mathbf{Z} - \mathbf{LW}\right)_i = \begin{cases} (\alpha - 1)\widetilde{\mathbf{X}}_i & 1 \leq i \leq n \\ (\beta - c)\widetilde{\mathbf{X}}_i & n+1 \leq i \leq 2n \end{cases} \tag{1}$$

Moreover, for all $i \in [2n]$ and $\gamma < 1$

$$\mathbb{P}\left( \| \left(\widehat{\mathbf{Z}} - \mathbf{Z}\right)_i \|^2 \geq n^{-\gamma} \right) = O\left(\frac{\log 2n}{n^{1-\gamma}}\right) \tag{2}$$

# Central Limit ~~Theorem~~ Conjecture

- Conjecture: scaled rows of $\left(\widehat{\mathbf{Z}} - \mathbf{Z}\right)_i$ are asymptotically normal

**Central Limit ~~Theorem~~ Conjecture**

With notation as above we have the following limit theorem

$$\mathbb{P}\left(\sqrt{n}\left(\widehat{\mathbf{Z}} - \mathbf{Z}\right)_i \leq z\right) \xrightarrow{D} \int_{\text{supp}(F)} \Phi\left(z, \Sigma_g(c, x_i)\right) dF(x_i)$$

where $\Phi$ is a the Normal cumulative distribution function.

- Have analytic expressions for $\Sigma_g(c, x_i)$
- Weighted sum of variances derived in Athreya et al. 2016
- Weights in terms of $\alpha$ and $\beta$
- Several other extensions & improvements of the results presented here

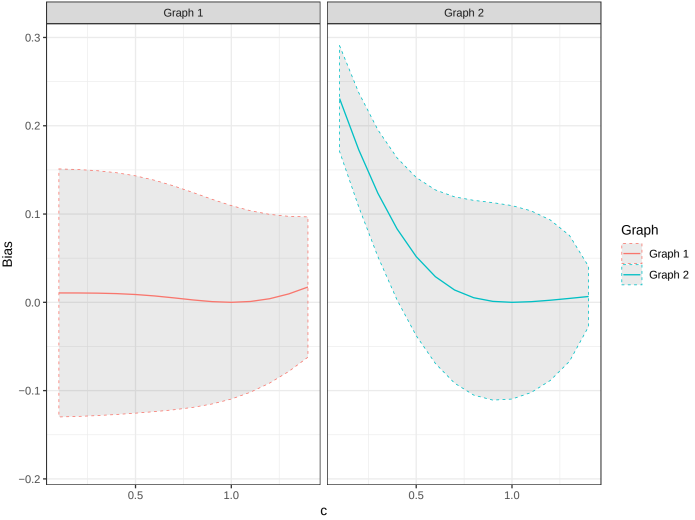
# Simulation Design

- Suppose that $\mathbf{A}^{(1)} \sim \text{ER}(p)$ and $\mathbf{A}^{(2)} \sim \text{ER}(c^2 p)$
- Then the latent positions are $\mathbf{X}_i = \sqrt{p}$ and $c\mathbf{X}_i = c\sqrt{p}$
- Simulation layout
  1. Sample $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}$ with $n \in \{100, 500\}$ vertices from each model with $p = 1/2$
  2. Jointly embed $\mathbf{A}^{(1)}, \mathbf{A}^{(2)}$ to obtain estimates $\widehat{\mathbf{Z}}_i$
  3. Compare estimates to theoretical quantile intervals

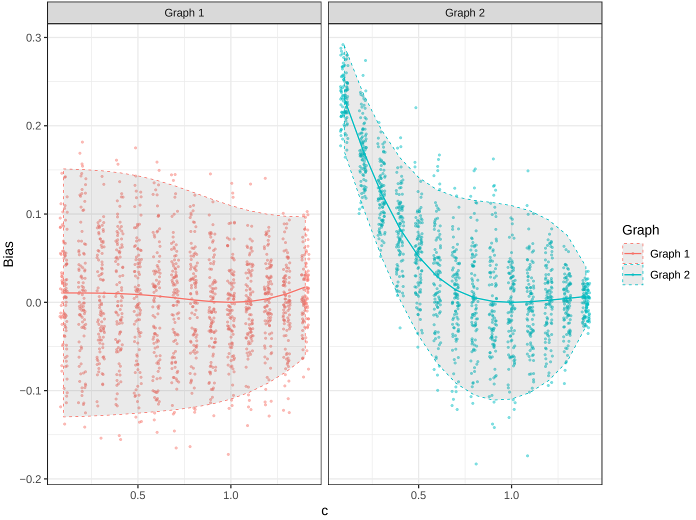
$$(\alpha - 1)\sqrt{p} \pm z_{\alpha/2} n^{-1/2} \sqrt{\Sigma_1(c, \sqrt{p})}$$

$$(\beta - c)\sqrt{p} \pm z_{\alpha/2} n^{-1/2} \sqrt{\Sigma_2(c, \sqrt{p})}$$

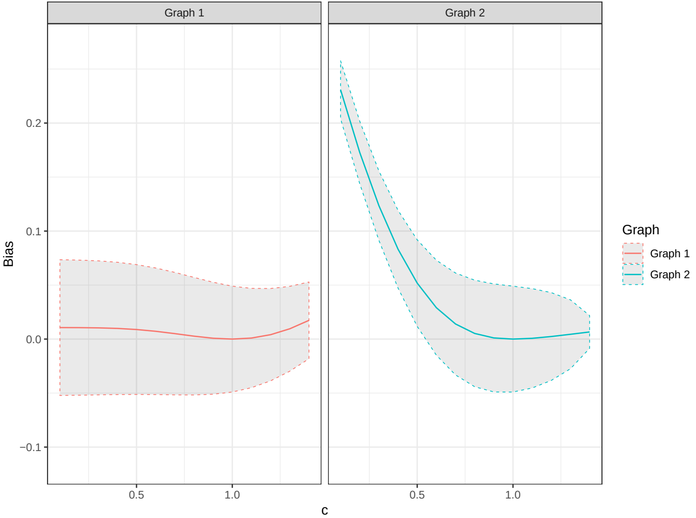
# Results: $n = 100$



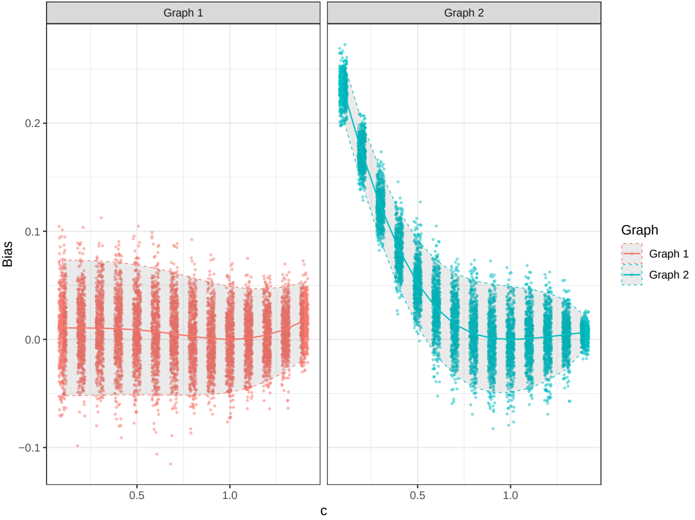
# Results: $n = 100$



# Results: $n = 500$



# Results: $n = 500$



# Conclusion & Future Work

- Discussed the graph embedding approach to network inference
- Uncovered the bias when jointly embedding networks with different connectivity structure
- Discussed the asymptotic properties of the embedding

- Extend this analysis to a more general set of latent positions
- Work with $m > 2$ networks

Questions? Comments?

# References


- 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. ISSN: 0976-8378. DOI: 10.1007/s13171-015-0071-x. URL: https://doi.org/10.1007/s13171-015-0071-x.
- Keith Levin et al. *A central limit theorem for an omnibus embedding of random dot product graphs*. 2017. eprint: arXiv:1705.09355.
- 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. DOI: 10.1080/01621459.2012.699795.
- Shangsi Wang et al. *Joint Embedding of Graphs*. 2017. eprint: arXiv:1703.03862.