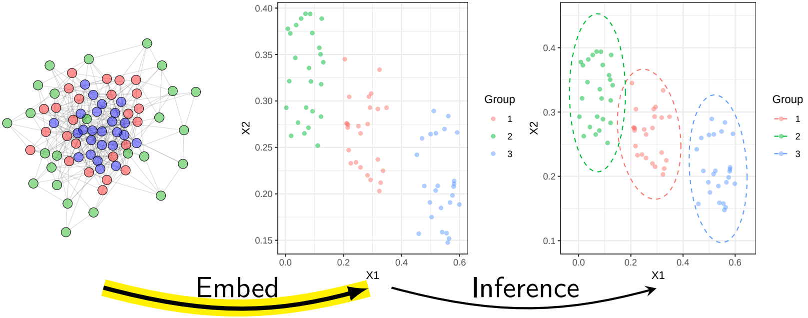
Joint Spectral Embeddings of Random Dot Product Graphs

Benjamin Draves

December 5, 2020

Network Embeddings

- Goal: Obtain a low dimensional representation of the network.
- Reason: Address network analysis questions with methods from multivariate statistics and machine learning.



Motivation - Multiplex Networks

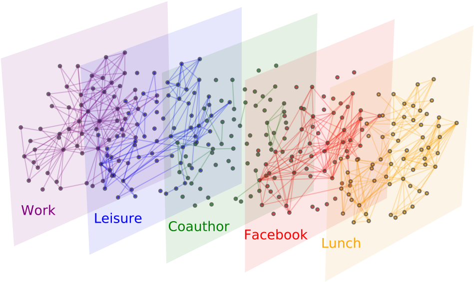
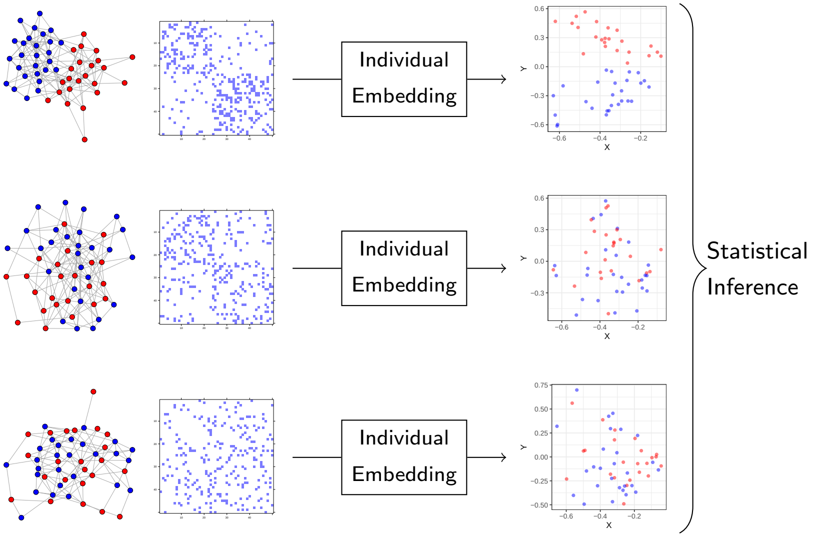


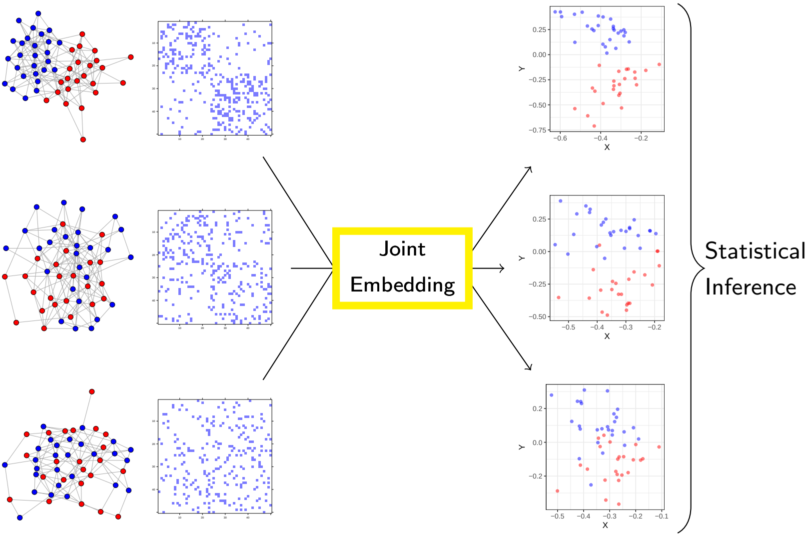
Figure: Multiplex Network of Aarhus Computer Science Department. Vertices are members of the department and each layer encodes a different type of interaction.(Magnani, Micenkova, and Rossi 2013).

- Multiplex networks - set of networks over a common vertex set.
- Adjacency matrices $A^{(g)} \in [0,1]^{n\times n}$ often share structure.

Multiple Network Embeddings



Multiple Network Embeddings



Multiplex Random Dot Product Graph

- Consider m graphs over a common vertex set $\mathcal{V}$ of size n.
- We extend the Random Dot Product Graph (Young and Scheinerman 2007) to multiplex networks.

Multiplex Random Dot Product Graph

- For $g \in [m]$, suppose that F_g is a probability distribution on $\mathbb{R}^{d_g}$ such that for all $x, y \in \text{supp}(F_g)$, $x^T y \in [0, 1]$.
- Let $\{X_i^{(g)}\}_{i=1}^n \overset{i.i.d.}{\sim} F_g$ and $\mathrm{X}^{(g)} = [X_1^{(g)}, \dots, X_n^{(g)}]^T \in \mathbb{R}^{n \times d_g}$ be the *latent positions* for graph $g \in [m]$.
- We say $\{(\mathrm{A}^{(g)}, \mathrm{X}^{(g)})\}_{g=1}^m \sim MRDPG(F, n)$ iff $\{\mathrm{A}_{ij}^{(g)}\}_{i<j, g \in [m]}$ are conditionally independent with

$$\mathbb{P}(\mathrm{A}_{ij}^{(g)} = 1 | \mathrm{X}^{(g)}) = \langle X_i^{(g)}, X_j^{(g)} \rangle$$

Submodels & Past Work

- Each vertex in each layer is described by a latent position $X_i^{(g)}$.
- In essence, $A^{(g)}|X^{(g)} \sim \text{Bern}(P^{(g)})$ where $P^{(g)} \equiv X^{(g)}X^{(g)T}$.

Model	Latent Positions $X^{(g)}$	Source
Joint RDPG	X	Levin et al. (2017)
Multiple RDPG	$U\sqrt{D^{(g)}}$	Nielsen and Witten (2018)
Eigen-Scaling RDPG	$X\sqrt{D^{(g)}}$	Draves and Sussman (2020+)
MREG	$V\sqrt{D^{(g)}}$	Wang et al. (2017)
COSIE	$U\sqrt{R^{(g)}}$	Arroyo et al. (2019)
Multilayer RDPG	$X\sqrt{R^{(g)}}$	Jones and Rubin-Delanchy (2020)

Table: Global structure: U is matrix with orthogonal columns, V has columns with unit Euclidean norm, and X is a latent position matrix. Layer variation: matrix square root of diagonal matrix $D^{(g)}$ or symmetric matrix $R^{(g)}$.

- **Key Point**: *Network Embeddings are latent position estimators.*

Adjacency Spectral Embedding

- Ignore joint structure and embed $\mathsf{A}^{(g)}$ individually.

Adjacency Spectral Embedding (Sussman et al. 2012)

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

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

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

- The i-th row of $\mathrm{ASE}(\mathsf{A}^{(g)}, d_g)$ is consistent for $\mathsf{X}_i^{(g)}$ up to Gaussian error for each $g \in [m]$ and $i \in [n]$ (Athreya et al. 2017).

Multiple Adjacency Spectral Embedding

- Embed $A^{(g)}$ individually then map to a common space.

Multiple Adjacency Spectral Embedding (Arroyo et al. 2019)

Let $\hat{X}_{ASE}^{(g)} = ASE(A^{(g)}, d_g)$ and define $\hat{X}_{ASE} = [\hat{X}_{ASE}^{(1)}\hat{X}_{ASE}^{(2)} \dots \hat{X}_{ASE}^{(m)}]$. Let $\hat{X}_{ASE}$ has singular value decomposition

$$\hat{X}_{ASE} = \hat{U}\hat{\Sigma}\hat{V}^T + \hat{U}_\perp\hat{\Sigma}_\perp\hat{V}_\perp^T$$

where $\hat{\Sigma} \in \mathbb{R}^{d\times d}$. Partition $\hat{V} = [\hat{V}^{(1)T}\hat{V}^{(2)T} \dots \hat{V}^{(m)T}]^T$. Then for all $g \in [m]$, $\hat{X}_{MASE}^{(g)} = [MASE(\{A^{(g)}\}_{g=1}^m, d)]_g = \hat{U}\hat{\Sigma}\hat{V}^{(g)T}$.

- Under a submodel of the MRDPG, $\hat{U}$ is consistent estimator for global model parameter (Arroyo et al. 2019).

Omnibus Embedding

- Simultaneously embed $\{\mathrm{A}^{(g)}\}_{g=1}^{m}$ into a common space.

Omnibus Embedding (Levin et al. 2017)

The *Omnibus matrix* is given by

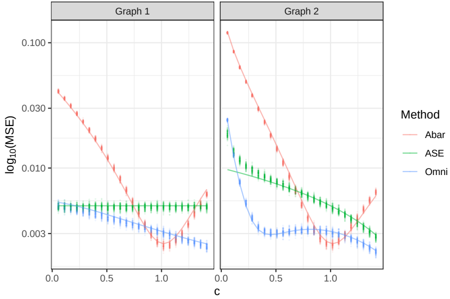
$$\tilde{\mathrm{A}} = \begin{bmatrix} \mathrm{A}^{(1)} & \frac{1}{2}[\mathrm{A}^{(1)} + \mathrm{A}^{(2)}] & \dots & \frac{1}{2}[\mathrm{A}^{(1)} + \mathrm{A}^{(m)}] \\ \frac{1}{2}[\mathrm{A}^{(2)} + \mathrm{A}^{(1)}] & \mathrm{A}^{(2)} & \dots & \frac{1}{2}[\mathrm{A}^{(2)} + \mathrm{A}^{(m)}] \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{2}[\mathrm{A}^{(m)} + \mathrm{A}^{(1)}] & \frac{1}{2}[\mathrm{A}^{(m)} + \mathrm{A}^{(2)}] & \dots & \mathrm{A}^{(m)} \end{bmatrix} \in \mathbb{R}^{nm \times nm}.$$

The omnibus embedding of graph $g \in [m]$ is $\hat{\mathrm{X}}^{(g)}_{\mathrm{OMNI}} = [\mathrm{ASE}(\tilde{\mathrm{A}}, d)]_g$ where $[\cdot]_g$ denotes the g-th, $n \times d$ block.

- Under a submodel of the MRDPG, the i-th row of $\hat{\mathrm{X}}^{(g)}_{\mathrm{OMNI}}$ is a biased estimator of $\mathrm{X}_i^{(g)}$ up to Gaussian error (Draves and Sussman 2020+).

Bias Variance Tradeoffs in Joint Spectral Embeddings

- Assess embedding techniques as latent positions estimators.
- Draves and Sussman 2020+ analyzed bias-variance tradeoff and ramifications of utilizing biased latent position estimates in subsequent inference.



- Unbiased latent position estimation *not necessarily* needed to achieve accurate statistical inference.

Network Embeddings as Dim. Reduction Techniques

- If biased latent position estimates can still lead to accurate inference, how should we assess embedding techniques?
- Goal: Study network embeddings *as their own statistical method*.
- **Key Point**: *Network Embeddings are latent position estimators and dimensionality reduction techniques*.

- Observation: ASE implicitly imitates *uncentered* principal component analysis (UPCA) on the matrix $A^{(g)}$.
- Question: Can we interpret MASE and OMNI as dimensionality reduction techniques?

Multiple Population Dimensionality Reduction

- Focus on dimensionality reduction of the matrix

$$\mathsf{X} = [\mathsf{X}^{(1)}\mathsf{X}^{(2)} \dots \mathsf{X}^{(m)}] \in \mathbb{R}^{n \times (\sum_{g=1}^{m} d_g)}.$$

Uncentered PCA of X

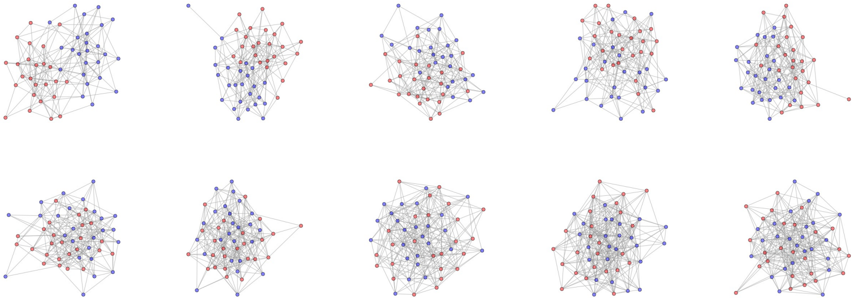
- Suppose $r = \text{rank}(\mathsf{X})$ and let X have singular value decomposition $\mathsf{X} = \mathsf{U}\Sigma\mathsf{V}^T + \mathsf{U}_\perp\Sigma_\perp\mathsf{V}_\perp^T$ where $\Sigma \in \mathbb{R}^{d\times d}$ for $d \in [r]$.
- The *uncentered principal components* are the rows of $\tilde{\mathsf{X}} = \mathsf{U}\Sigma$.
- Partition $\mathsf{V} = [\mathsf{V}^{(1)T} \dots \mathsf{V}^{(m)T}]^T$ so that $\mathsf{V}^{(g)} \in \mathbb{R}^{d_g \times d}$. Then $\tilde{\mathsf{X}}^{(g)} = \mathsf{U}\Sigma\mathsf{V}^{(g)T}$ is the total least squares fit of the latent positions for graph $g \in [m]$.

- UPCA of X is identical to performing total least squares (TLS) regression on the latent positions $\{\mathsf{X}^{(g)}\}_{g=1}^m$.

Example: Hierarchical Stochastic Block Model

- Suppose that $\{A^{(g)}\}_{g=1}^{25}$ are SBMs with 2 groups of even size, $n = 25$.
- Community assignment fixed and block probabilities vary.
- For $b_g \in [0, 0.25]$, model parameters and latent positions are

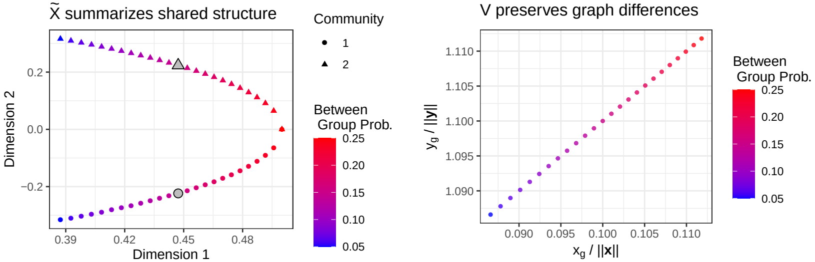
$$Z = \begin{bmatrix} 1_n & 0_n \\ 0_n & 1_n \end{bmatrix} \qquad B^{(g)} = \begin{bmatrix} 0.25 & b_g \\ b_g & 0.25 \end{bmatrix} \qquad L^{(g)} = Z \begin{bmatrix} x_g & -y_g \\ x_g & y_g \end{bmatrix}.$$



Example: Hierarchical Stochastic Block Model

- Let $\mathrm{x} = (x_1 x_2 \dots x_m)^T \in \mathbb{R}^{m\times 1}$ and $\mathrm{y} = (y_1 y_2 \dots y_m)^T \in \mathbb{R}^{m\times 1}$.
- Let $\bar{\mathrm{B}} = m^{-1}\sum_{g=1}^{m} \mathrm{B}^{(g)}$ and let $\bar{\mathrm{L}} = \mathrm{ASE}(\bar{\mathrm{B}}, 2)$
- Then the UPCA estimates are $\tilde{\mathrm{X}}$ and the $\{\mathrm{V}^{(g)}\}_{g=1}^{m}$ as

$$\tilde{\mathrm{X}} = \sqrt{m}\mathrm{Z}\bar{\mathrm{L}} \qquad \mathrm{V}^{(g)} = \mathrm{diag}\left(\frac{x_g}{\|\mathrm{x}\|}, \frac{y_g}{\|\mathrm{y}\|}\right).$$



Preliminary Results

- $\tilde{\mathrm{X}}$ captures layer similarities by implicitly embedding $\bar{\mathrm{P}}$ into $\mathbb{R}^d$.

Lemma - $\tilde{\mathrm{X}}$ Summarizes Shared Structure

There exists $\mathrm{W} \in \mathcal{O}_d$ such that $\tilde{\mathrm{X}}\mathrm{W} = m^{1/2}\mathrm{ASE}(\bar{\mathrm{P}}, d)$.

- $\mathrm{V}^{(g)} : \mathbb{R}^d \to \mathbb{R}^{d_g}$ maps the rows of $\tilde{\mathrm{X}}$ to the TLS fit of the latent positions $\{\mathrm{X}^{(g)}\}_{g=1}^m$.

Lemma - $\mathrm{V}^{(g)}$ describes graph differences

Let $\{\tilde{\mathrm{X}}^{(g)}\}_{g=1}^m$ be the TLS fit of the latent positions$\{\mathrm{X}^{(g)}\}_{g=1}^m$. Then $\mathrm{V}^{(g)}\tilde{\mathrm{x}}_i = \tilde{\mathrm{x}}_i^{(g)}$.

- UPCA of X succinctly segregates the common structure shared across layers and layer specific variation.

Preliminary Results: MASE & OMNI

- Let $X_{MASE}^{(g)}$ and $X_{OMNI}^{(g)}$ be the population equivalent of $\hat{X}_{MASE}^{(g)}$ and $\hat{X}_{OMNI}^{(g)}$ by replacing $\{A^{(g)}\}_{g=1}^{m}$ with $\{P^{(g)}\}_{g=1}^{m}$.

Lemma - MASE

There exists $W \in \mathcal{O}_{d_g}$ such that $X_{MASE}^{(g)}W = \tilde{X}^{(g)} = \tilde{X}V^{(g)T}$.

Lemma - OMNI

Let $\tilde{Y} = m^{1/2}\text{ASE}(\bar{P}, r)$ and $\tilde{V}$ be the omnibus matrix of $\{V^{(g)T}V^{(g)}\}_{g=1}^{m}$. Let $B = \text{OMNI}(\tilde{V}, d)$ and partition $B^T = [B^{(1)}B^{(2)} \dots B^{(m)}]$ so that $B_g \in \mathbb{R}^{r \times d}$. Then there exists $W \in \mathcal{O}_d$ such that $X_{OMNI}^{(g)}W = \tilde{Y}B^{(g)T}$.

- MASE implicitly imitates rank-d UPCA of X.
- OMNI implicitly imitates full rank UPCA of X then maps the rows into $\mathbb{R}^d$.

Simulation Design

Consider SBM with $K = 3$ groups of size $n_k = 100$ so that $\mathrm{Z} = \mathrm{I}_{3\times 3} \otimes 1_{n_k}$.

$$\mathrm{B} = \begin{bmatrix} 0.25 & 0.05 & 0.05 \\ 0.05 & 0.25 & 0.05 \\ 0.05 & 0.05 & 0.25 \end{bmatrix} \quad \mathrm{L_B} = \begin{bmatrix} 0.34 & -0.37 & 0 \\ 0.34 & 0.18 & -0.32 \\ 0.34 & 0.18 & 0.32 \end{bmatrix} \quad \Delta = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0.2 \\ 0 & 0.2 & 0 \end{bmatrix}$$

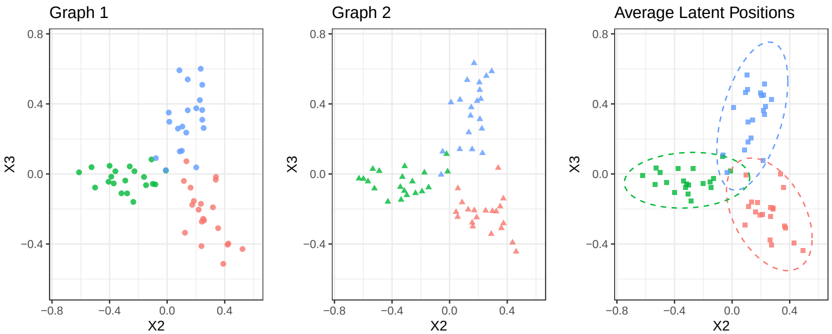
Anomalous Vertices

1. Set $\mathrm{X}^{(1)} = \mathrm{X}^{(2)} = \mathrm{ZL_B}$
2. $\mathrm{A}^{(1)} \sim \mathrm{Bern}(\mathrm{X}^{(1)}\mathrm{X}^{(1)T})$
3. For $\alpha \in [0, 0.1]$, replace $(100 \times \alpha)\%$ rows of $\mathrm{X}^{(2)}$ sample from $\mathrm{Dir}(1, 1, 1)$
4. $\mathrm{A}_\alpha^{(2)} \sim \mathrm{Bern}(\mathrm{X}_\alpha^{(2)}\mathrm{X}_\alpha^{(2)T})$

Hierarchical SBM

1. Set $\mathrm{X}^{(1)} = \mathrm{X}^{(2)} = \mathrm{ZL_B}$
2. $\mathrm{A}^{(1)} \sim \mathrm{Bern}(\mathrm{X}^{(1)}\mathrm{X}^{(1)T})$
3. For $\lambda \in [0, 0.25]$, let $\mathrm{L}_\lambda = \mathrm{ASE}(\mathrm{B} + \lambda\Delta, 3)$ and set $\mathrm{X}_\lambda^{(2)} = \mathrm{ZL}_\lambda$
4. $\mathrm{A}_\lambda^{(2)} \sim \mathrm{Bern}(\mathrm{X}_\lambda^{(2)}\mathrm{X}_\lambda^{(2)T})$

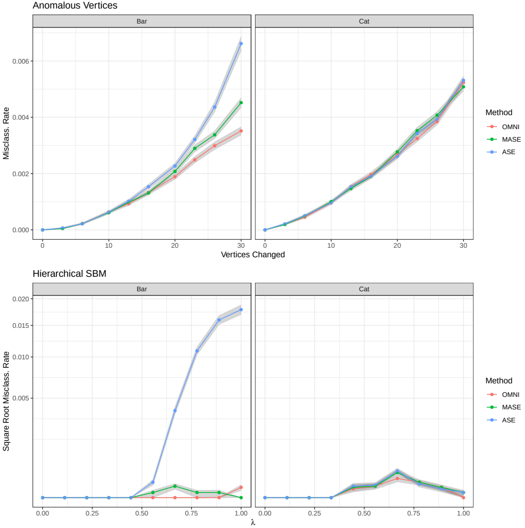
Multiplex Network Analysis: Community Detection



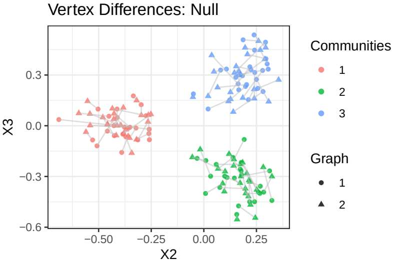
Community Detection

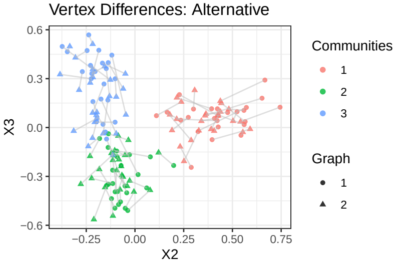
1. Embed $(\hat{X}^{(1)}, \hat{X}^{(2)}) \leftarrow \text{EMBED}(A^{(1)}, A^{(2)}, d)$
2. Apply clustering algorithms to rows of $[\hat{X}^{(1)}\hat{X}^{(2)}]$ or $\bar{X} = 2^{-1}(\hat{X}^{(1)} + \hat{X}^{(2)})$.

Inference Ramifications: Community Detection



Multiplex Network Analysis: Hypothesis Testing

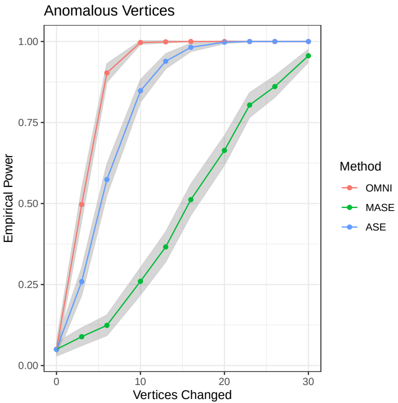




Hypothesis Testing

1. Embed $(\hat{\mathrm{X}}^{(1)}, \hat{\mathrm{X}}^{(2)}) \leftarrow \mathrm{EMBED}(\mathrm{A}^{(1)}, \mathrm{A}^{(2)}, d)$
2. Set test statistic $T = \|\hat{\mathrm{X}}^{(1)} - \hat{\mathrm{X}}^{(2)}\|_F$
3. Bootstrap reference distribution $\{T_b : b \in [B]\}$ (Levin et al. 2017).
4. Return $p \leftarrow |\{b \in [B] : T_b \geq T\}|/B$

Inference Ramifications: Hypothesis Testing



Conclusion

- By pooling information across networks, joint spectral embeddings can learn more representative node embeddings.
- MASE and OMNI complete multiple population dimensionality reduction on the latent positions.
- Differences between these dimensionality reduction techniques make MASE and OMNI node embeddings more/less useful in certain statistical applications.

Questions?

References I

- Jesús Arroyo et al. "Inference for multiple heterogeneous networks with a common invariant subspace". In: (2019). arXiv preprint: 1906.10026.
- Avanti Athreya et al. "Statistical inference on random dot product graphs: A survey". In: *Journal of Machine Learning Research* 18 (Sept. 2017).
- Benjamin Draves and Daniel L. Sussman. *Bias-Variance Tradeoffs in Joint Spectral Embeddings*. 2020+. arXiv: 2005.02511 [math.ST].
- Andrew Jones and Patrick Rubin-Delanchy. *The multilayer random dot product graph*. 2020. arXiv: 2007.10455 [stat.ML].
- 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

- Matteo Magnani, Barbora Micenkova, and Luca Rossi. *Combinatorial Analysis of Multiple Networks*. 2013. arXiv: 1303.4986 [cs.SI].
- Agnes Martine Nielsen and Daniela Witten. "The Multiple Random Dot Product Graph Model". In: (2018). arXiv preprint: 1811.12172.
- 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". In: (2017). arXiv preprint: 1703.03862.

References III

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.