The Importance of Being Correlated: Implications of Dependence in Joint Spectral Inference across Multiple Networks

Konstantinos Pantazis

kpantaz1@jhu.edu

Applied Mathematics & Statistics (AMS)

Johns Hopkins University

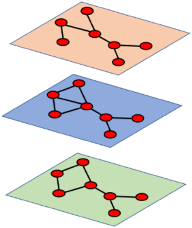
December 17, 2022

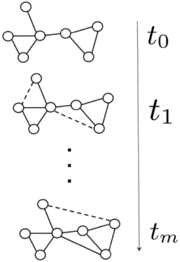
*Joint work with A. Athreya, J. Arroyo, W. Frost, E. Hill, V. Lyzinski

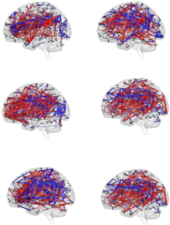
Inference across Multiple Networks

Often, data consist of a collection of networks with aligned vertices:

- Multilayer networks
- Time-varying networks
- Multiple samples of networks







Implications of Dependence in Joint Spectral Inference across Multiple Networks

- Why spectral inference on multiple networks?

Implications of Dependence in Joint Spectral Inference across Multiple Networks

- Why spectral inference on multiple networks?

> To build statistical analogues of classical multisample methods for network-valued data

Implications of Dependence in Joint Spectral Inference across Multiple Networks

- Why spectral inference on multiple networks?

> To build statistical analogues of classical multisample methods for network-valued data

- Many spectral methods simultaneously embed all the networks into a *common low-dimensional Euclidean space* & assume independence across the multiple network realizations.

Implications of Dependence in Joint Spectral Inference across Multiple Networks

- Why spectral inference on multiple networks?

> To build statistical analogues of classical multisample methods for network-valued data

- Many spectral methods simultaneously embed all the networks into a *common low-dimensional Euclidean space* & assume independence across the multiple network realizations.

> 1. Embedding networks into a Euclidean space allows us to utilize tools from traditional statistics.
> 2. The independence assumption rarely holds in real-data applications.

Implications of Dependence in Joint Spectral Inference across Multiple Networks

- Why spectral inference on multiple networks?

To build statistical analogues of classical multisample methods for network-valued data

- Many spectral methods simultaneously embed all the networks into a *common low-dimensional Euclidean space* & assume independence across the multiple network realizations.

1. Embedding networks into a Euclidean space allows us to utilize tools from traditional statistics.
2. The independence assumption rarely holds in real-data applications.

- The **main goal** is to bring awareness to the **induced correlation** that may arise in such joint network embeddings.

Statistical network inference

What are some inference tasks on multiple networks with aligned vertices?

Statistical network inference

What are some inference tasks on multiple networks with aligned vertices?

- *Consistent estimation* of the underlying model parameters
- *Asymptotic normality* results
- *Subsequent inference tasks:*
  1. 2-graph or M-graph hypothesis testing
  2. Change point detection on vertices, vertex clouds or whole-graphs
  3. Clustering, Classification, etc.

Statistical network inference

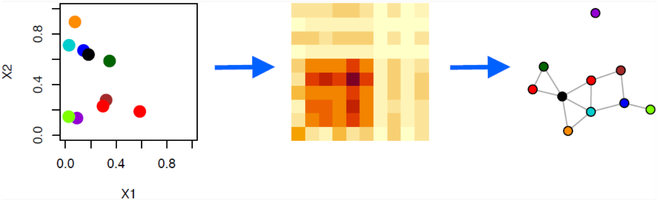
What are some inference tasks on multiple networks with aligned vertices?

- *Consistent estimation* of the underlying model parameters
- *Asymptotic normality* results
- *Subsequent inference tasks:*
  1. 2-graph or M-graph hypothesis testing
  2. Change point detection on vertices, vertex clouds or whole-graphs
  3. Clustering, Classification, etc.

What are the moving parts in joint network embedding procedures?

1. Network model
2. Technique to aggregate networks
3. Embedding method

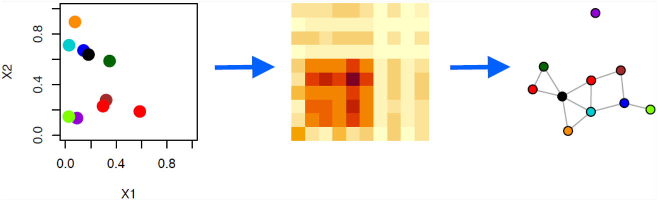
Random Dot Product Graph (RDPG)



Latent positions $X \to P = XX^T \to$ Sampled graph G

- Each vertex i is associated with a **latent position** $X_i \in \mathbb{R}^d$ drawn from a distribution F, with support supp$(F) \subseteq B_d(1)$

Random Dot Product Graph (RDPG)



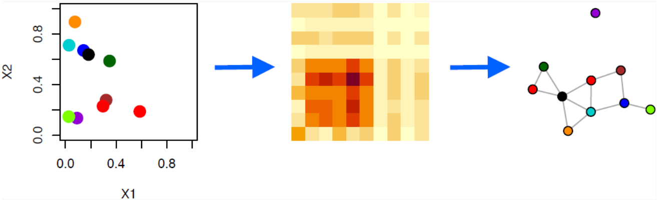
Latent positions $X \rightarrow P = XX^T \rightarrow$ Sampled graph G

- Each vertex i is associated with a **latent position** $X_i \in \mathbb{R}^d$ drawn from a distribution F, with support supp$(F) \subseteq B_d(1)$
- Given X_i, X_j, the adjacency matrix A of a sampled graph G is given by:

$$A_{ij} = \begin{cases} 1 & \text{,with probability } \langle X_i, X_j \rangle \\ 0 & \text{,otherwise.} \end{cases}$$

- **Note:** The $\mathrm{d\text{-}RDPG}$ model is specified only up to *orthogonal transformation* of its latent positions.

Random Dot Product Graph (RDPG)



Latent positions $X \rightarrow P = XX^T \rightarrow$ Sampled graph G

- Each vertex i is associated with a **latent position** X_i 2 R^d drawn from a distribution F, with support supp(F) $B_d(1)$
- Given X_i, X_j, the adjacency matrix A of a sampled graph G is given by:

$$A_{ij} = \begin{cases} 1 & \text{,with probability } hX_i, X_j i \\ 0 & \text{,otherwise.} \end{cases}$$

- **Note:** The $\mathrm{d\text{-}RDPG}$ model is specified only up to *orthogonal transformation* of its latent positions.
- Why $\mathrm{d\text{-}RDPG}$?
  - $\mathrm{d\text{-}RDPG}$ is an analytically tractable model.
  - Yet, encompasses a broad range of random graph models such as positive semidefinite SBM and Erdos-Renyi.

Adjacency Spectral Embedding (ASE)

The d-*dimensional adjacency spectral embedding* (ASE) of A is obtained by

$$\widehat{X}_A = U_A S_A^{1/2} \in \mathbb{R}^{n\times d}$$

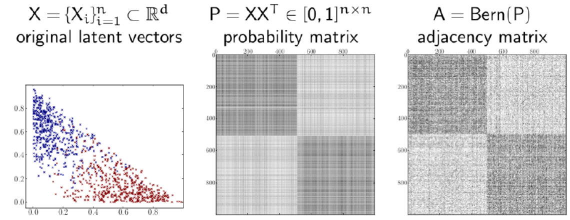
- $S_A \in \mathbb{R}^{d\times d}$:= diagonal matrix whose entries are the top d eigenvalues of $|A| = (A^T A)^{1/2}$
- $U_A \in \mathbb{R}^{n\times d}$:= $n \times d$ matrix whose columns are the orthonormal eigenvectors corresponding to the eigenvalues in S_A.

Adjacency Spectral Embedding (ASE)

The d*-dimensional adjacency spectral embedding (*ASE*)* of A is obtained by

$$\widehat{X}_A = U_A S_A^{1/2} \; 2 \; \mathrm{R}^{n\times d}$$

- S_A 2 R$^{d\times d}$:= diagonal matrix whose entries are the top d eigenvalues of jAj $= (A^T A)^{1/2}$
- U_A 2 R$^{n\times d}$:= n d matrix whose columns are the orthonormal eigenvectors corresponding to the eigenvalues in S_A.

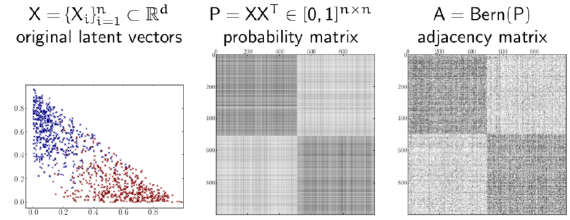


Adjacency Spectral Embedding (ASE)

The d-*dimensional adjacency spectral embedding* (ASE) of A is obtained by

$$\widehat{X}_A = U_A S_A^{1/2} \in \mathbb{R}^{n \times d}$$

- $S_A \in \mathbb{R}^{d \times d}$:= diagonal matrix whose entries are the top d eigenvalues of $|A| = (A^T A)^{1/2}$
- $U_A \in \mathbb{R}^{n \times d}$:= $n \times d$ matrix whose columns are the orthonormal eigenvectors corresponding to the eigenvalues in S_A.



Under RDPG, ASE **consistently estimates** (up to orthogonal transformation) the data matrix $\mathbf{X} \in \mathbb{R}^{\mathbf{n} \times \mathbf{d}}$ (Sussman et al, 2014).

OMNIBUS Embedding (Levin et al. 2017)

- Joint-RDPG **model:** Given data matrix $\mathbf{X}$, for each $k \in [m]$,

$$A^{(k)} \sim \text{RDPG}(\mathbf{X})$$

Note: $A^{(k)}$'s are assumed independent, and the same data matrix $\mathbf{X}$ is used to generate all m graphs.

OMNIBUS Embedding (Levin et al. 2017)

- Joint-RDPG **model:** Given data matrix $\mathbf{X}$, for each $k \in [m]$,

$$A^{(k)} \sim \text{RDPG}(\mathbf{X})$$

Note: $A^{(k)}$'s are assumed independent, and the same data matrix $\mathbf{X}$ is used to generate all m graphs.

- **Omnibus matrix:**

$$M = \begin{bmatrix} A^{(1)} & \frac{A^{(1)}+A^{(2)}}{2} & \frac{A^{(1)}+A^{(3)}}{2} & \cdots & \frac{A^{(1)}+A^{(m)}}{2} \\ \frac{A^{(2)}+A^{(1)}}{2} & A^{(2)} & \frac{A^{(2)}+A^{(3)}}{2} & \cdots & \frac{A^{(2)}+A^{(m)}}{2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{A^{(m)}+A^{(1)}}{2} & \frac{A^{(m)}+A^{(2)}}{2} & \frac{A^{(m)}+A^{(3)}}{2} & \cdots & A^{(m)} \end{bmatrix} \in \mathbb{R}^{mn \times mn}$$

OMNIBUS Embedding (Levin et al. 2017)

- Joint-RDPG **model:** Given data matrix $\mathbf{X}$, for each $k \in [m]$,

$$A^{(k)} \sim \mathrm{RDPG}(\mathbf{X})$$

Note: $A^{(k)}$'s are assumed independent, and the same data matrix $\mathbf{X}$ is used to generate all m graphs.

- **Omnibus matrix:**

$$M = \begin{bmatrix} A^{(1)} & \frac{A^{(1)}+A^{(2)}}{2} & \frac{A^{(1)}+A^{(3)}}{2} & \cdots & \frac{A^{(1)}+A^{(m)}}{2} \\ \frac{A^{(2)}+A^{(1)}}{2} & A^{(2)} & \frac{A^{(2)}+A^{(3)}}{2} & \cdots & \frac{A^{(2)}+A^{(m)}}{2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{A^{(m)}+A^{(1)}}{2} & \frac{A^{(m)}+A^{(2)}}{2} & \frac{A^{(m)}+A^{(3)}}{2} & \cdots & A^{(m)} \end{bmatrix} \in \mathbb{R}^{mn \times mn}$$

- OMNI **Embedding:**

$$\mathrm{OMNI}(A^{(1)}, A^{(2)}, \cdots, A^{(m)}, d) = \mathrm{ASE}(M, d) = U_M S_M^{1/2} \in \mathbb{R}^{mn \times d}$$

Advantages of OMNI embedding

1. *Consistency* and *asymptotic normality* results when the graphs are sampled from the same distribution
2. The omnibus embedding produces m *distinct estimates for each vertex*
3. The omnibus matrix incorporates *a priori an alignment*.

Advantages of OMNI embedding

1. *Consistency* and *asymptotic normality* results when the graphs are sampled from the same distribution
2. The omnibus embedding produces m *distinct estimates for each vertex*
3. The omnibus matrix incorporates *a priori an alignment*.

Multi-scale inference:

$$\mathrm{ASE}(M,d) = \begin{bmatrix} \hat{X} \\ \hat{Y} \end{bmatrix} = \begin{bmatrix} \text{Estimated latent position of } vertex\ 1 \text{ in } G_1 \\ \text{Estimated latent position of } vertex\ 2 \text{ in } G_1 \\ \vdots \\ \text{Estimated latent position of } vertex\ n \text{ in } G_1 \\ \text{Estimated latent position of } vertex\ 1 \text{ in } G_2 \\ \text{Estimated latent position of } vertex\ 2 \text{ in } G_2 \\ \vdots \\ \text{Estimated latent position of } vertex\ n \text{ in } G_2 \end{bmatrix} \in \mathbb{R}^{2n \times d}$$

Understanding induced correlation

OMNI embedding induces correlation across estimates.

Understanding induced correlation

OMNI embedding induces correlation across estimates.

- **CLT (Avanti et al., 2016):** Consider $A \sim \text{d-RDPG}(X)$ and $\widehat{X}_A = U_A S_A^{1/2}$. Under mild assumptions, there exists an orthogonal matrix Q such that

$$\sqrt{n}(\widehat{X}_A Q - X)_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(X_i))$$

.

Understanding induced correlation

OMNI embedding induces correlation across estimates.

- **CLT (Avanti et al., 2016):** Consider $A \sim \text{d-RDPG}(X)$ and $\widehat{X}_A = U_A S_A^{1/2}$. Under mild assumptions, there exists an orthogonal matrix Q such that

$$\sqrt{n}(\widehat{X}_A Q - X)_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(X_i))$$

.

- **CLT for correlated networks:** Consider two correlated networks $A, B \sim \text{d-RDPG}(X)$ with edge-correlation ρ, i.e.,

$$\text{correlation}(A_{ij}, B_{ij}) = \rho \text{ for all } i, j.$$

Understanding induced correlation

OMNI embedding induces correlation across estimates.

- **CLT (Avanti et al., 2016):** Consider $A \sim \text{d-RDPG}(X)$ and $\widehat{X}_A = U_A S_A^{1/2}$. Under mild assumptions, there exists an orthogonal matrix Q such that

$$\sqrt{n}(\widehat{X}_A Q - X)_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(X_i))$$

.

- **CLT for correlated networks:** Consider two correlated networks $A, B \sim \text{d-RDPG}(X)$ with edge-correlation ρ, i.e.,

$$\text{correlation}(A_{ij}, B_{ij}) = \rho \text{ for all } i, j.$$

  - Let $\widehat{X}_A = U_A S_A^{1/2}$ and $\widehat{X}_B = U_B S_B^{1/2}$ their embeddings. Then, there exist orthogonal matrices Q_1, Q_2

$$\sqrt{n}(\widehat{X}_A Q_1 - \widehat{X}_B Q_2)_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2(1-\rho)\Sigma(X_i))$$

OMNI induces "flat" correlation between estimates

- **CLT for OMNI estimates:** Let M denote the omnibus matrix. Let $\widehat{\mathbf{X}}_M = U_M S_M^{1/2}$ and denote the estimates from network $A^{(s)}$ as $\widehat{\mathbf{X}}_M^{(s)}$. For fixed indices s_1, s_2, there exist orthogonal matrix W

$$\sqrt{n}\Big((\widehat{X}_M^{(s_1)} - \widehat{X}_M^{(s_2)})W\Big)_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{2}\Sigma(X_i))$$

OMNI induces “flat" correlation between estimates

- **CLT for OMNI estimates:** Let M denote the omnibus matrix. Let $\widehat{\mathbf{X}}_M = U_M S_M^{1/2}$ and denote the estimates from network $A^{(s)}$ as $\widehat{\mathbf{X}}_M^{(s)}$. For fixed indices s_1, s_2, there exist orthogonal matrix W

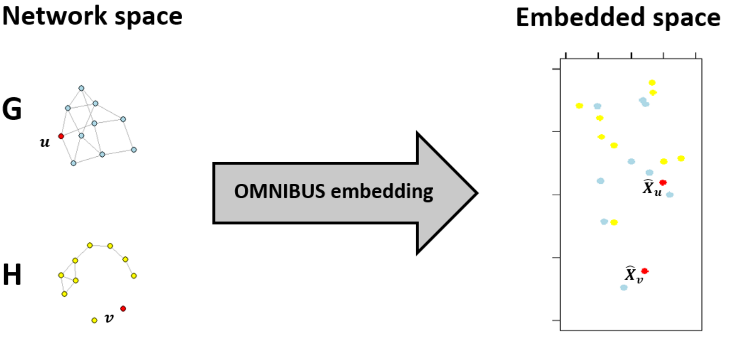
$$\sqrt{n}\Big((\widehat{X}_M^{(s_1)} - \widehat{X}_M^{(s_2)})W\Big)_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \frac{1}{2}\Sigma(X_i))$$

- **CLT for correlated networks:**

$$\sqrt{n}(\widehat{X}_A Q_1 - \widehat{X}_B Q_2)_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2(1-\rho)\Sigma(X_i))$$

OMNI embedding induces correlation equal to $\rho = 0.75$ between estimates across all networks.

Induced correlation in OMNI embedding



$corr(G, H) = 0$

$corr(\hat{X}_u, \hat{X}_v) = 0.75$

Real-data experiment

"Flat" correlation can mask the signal present in a time-series of networks application

Application: Analysis of Aplysia californica escape motor program of Hill et al. (2020)

- 20 min recording of action for 82 neurons
- One minute into the recording, stimulus was applied to nerve 9.

The stimulus results to initial rapid galloping followed by a slower rhythmic crawling

Real-data experiment

“Flat" correlation can mask the signal present in a time-series of networks application

Application: Analysis of Aplysia californica escape motor program of Hill et al. (2020)

- 20 min recording of action for 82 neurons
- One minute into the recording, stimulus was applied to nerve 9.

The stimulus results to initial rapid galloping followed by a slower rhythmic crawling

Real-data experiment

"Flat" correlation can mask the signal present in a time-series of networks application

Application: Analysis of Aplysia californica escape motor program of Hill et al. (2020)

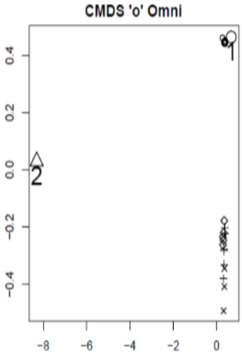
- 20 min recording of action for 82 neurons
- One minute into the recording, stimulus was applied to nerve 9.

The stimulus results to initial rapid galloping followed by a slower rhythmic crawling

To extract a **network time series** from the recording

- Bin the motor program into 24 bins, each 50 second long
- Convert each bin into a weighted matrix
- Embed these 24 matrices using OMNI procedure.

Real-data experiment



First graph $\Longrightarrow$ relaxing state.
Second graph $\Longrightarrow$ firing state.
Rest of the graphs $\Longrightarrow$ galloping and crawling states.

OMNI:

- Successfully detects the stimulus in the second graph
- However, the induced flat correlation masks the transition from galloping to crawling and creates an artificial similarity between graphs 1 and some of the graphs $k > 2$ in the embedded space

Generalized OMNI (genOMNI)

- $\mathrm{R\text{-}RDPG}$ **model:** Extend $\mathrm{Joint\text{-}RDPG}$ model to incorporate latent correlation across networks
- **Generalized omnibus matrix:** The block entries of the generalized matrix `M` are *convex combinations of* $A^{(i)}$*'s*.

Generalized OMNI (genOMNI)

- $\mathrm{R\text{-}RDPG}$ **model:** Extend $\mathrm{Joint\text{-}RDPG}$ model to incorporate latent correlation across networks
- **Generalized omnibus matrix:** The block entries of the generalized matrix `M` are *convex combinations of* $A^{(i)}$*'s*.

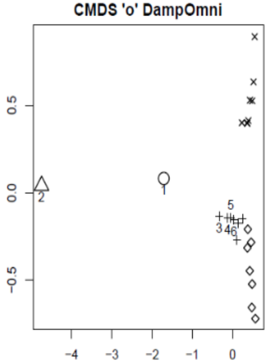
CLT for genOMNI estimates:

$\sqrt{n}\Big((\widehat{X}_{\mathtt{M}}^{(s_1)} - \widehat{X}_{\mathtt{M}}^{(s_2)})W\Big)_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, 2(1-\rho(s_1,s_2))\Sigma(X_i))$, where

$$\rho(s_1,s_2) = \underbrace{1 - \frac{\sum_{q=1}^{m}(\alpha(s_1,q)-\alpha(s_2,q))^2}{2m^2}}_{\text{method-induced correlation}} + \underbrace{\frac{\sum_{q<l}\Big(\alpha(s_1,q)-\alpha(s_2,q))(\alpha(s_2,l)-\alpha(s_1,l)\Big)\rho_{q,l}}{m^2}}_{\text{model-inherent correlation}}$$

and $\alpha(k,q)$ is the total weight put on $A^{(q)}$ in the k-th block-row of `M`.

Real data experiment cont'd



Dampened OMNI:

$$\mathtt{M}_{damp}^{(k,\ell)} = \begin{cases} \frac{A^{(k)} + \ell A^{(\ell)}}{\ell+1} & \text{if } k < \ell, \\ A^{(k)} & \text{if } k = \ell \end{cases}$$

- Detects the stimulus in the second graph
- Distinguishes the relaxing state from the other states.
- Captures (imperfectly) the transition from galloping (graphs $3, 4$) to crawling (graphs 5-24).
- Picks out an unstable dynamic (graphs 13-24) not apparent from simple visual inspection of the firing traces.

Summary

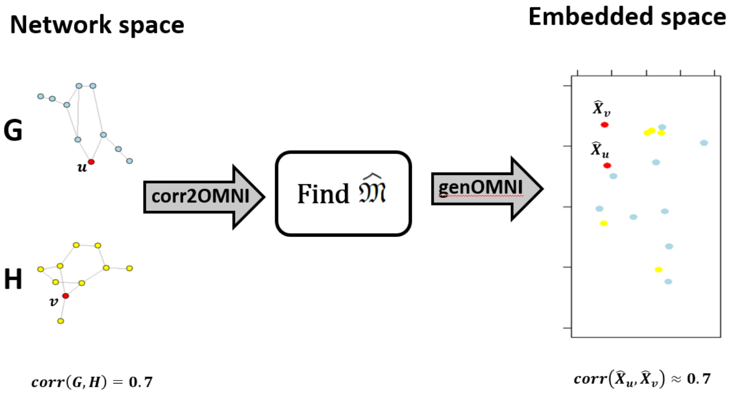
1. Identify and analyze the *phenomenon of induced correlation*, which is an artifice in joint network embeddings.

2. First to show theoretical guarantees (*consistency, asymptotic normality*) under a correlated multiple network model.

3. Extend previous methodology to a *family of models* making genOMNI suitable for **meaningful subsequent inference**, especially for *time series of networks* applications.

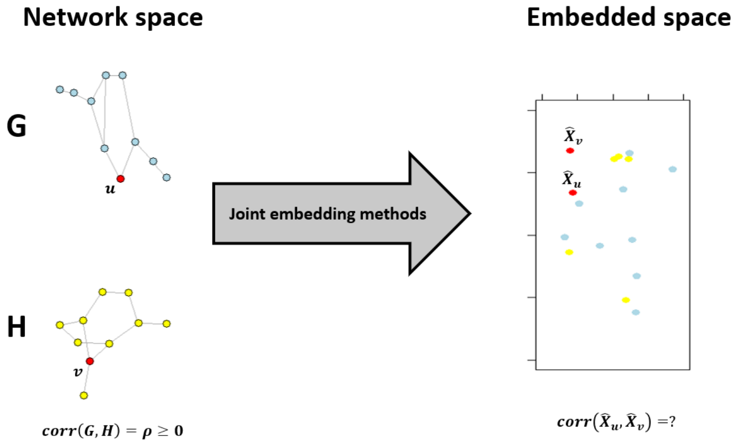
Future work: corr2omni algorithm

Given a correlation structure for a collection of networks, choose weights in the genOMNI setting that would reproduce (approximately) this structure in the embedded space.



Future work

Identify and analyze the induced correlation in other joint embedding procedures (e.g., COSIE-MASE, Arroyo et al. 2020).



Thank You !

Paper: `https://www.jmlr.org/papers/v23/20-944.html`

Contact: `kpantaz1@jhu.edu`