A Random Dot Product Graph Model for Weighted and Directed Networks

Bernardo Marenco, Paola Bermolen, Marcelo Fiori, **Federico Larroca** and Gonzalo Mateos

Asilomar Conference on Signals, Systems, and Computers
October 2024

Random dot product graphs

- Consider a latent space $\mathcal{X}_d \subset \mathbb{R}^d$ such that for all

$$\mathbf{x}, \mathbf{y} \in \mathcal{X}_d \quad \Rightarrow \quad \mathbf{x}^\top \mathbf{y} \in [0, 1]$$

  $\Rightarrow$ Inner-product distribution $F : \mathcal{X}_d \mapsto [0, 1]$

- Random dot product graphs (RDPGs) are defined as follows:

$$\mathbf{x}_1, \ldots, \mathbf{x}_{N_v} \overset{\text{i.i.d.}}{\sim} F,$$
$$A_{ij} \,\big|\, \mathbf{x}_i, \mathbf{x}_j \sim \text{Bernoulli}(\mathbf{x}_i^\top \mathbf{x}_j)$$

  for $1 \leq i, j \leq N_v$, where $A_{ij} = A_{ji}$ and $A_{ii} \equiv 0$

- A particularly tractable latent position random graph model

  $\Rightarrow$ Vertex positions $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top \in \mathbb{R}^{N_v \times d}$

S. J. Young and E. R. Scheinerman, "Random dot product graph models for social networks," *WAW*, 2007

Estimation of latent positions

- **Q:** Given $G = (V, E)$ from an RDPG, find the 'best' $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top$?
- MLE is well motivated but it is intractable for large N_v

$$\hat{\mathbf{X}}_{ML} = \operatorname*{argmax}_{\mathbf{X}} \prod_{i<j} (\mathbf{x}_i^\top \mathbf{x}_j)^{A_{ij}} (1 - \mathbf{x}_i^\top \mathbf{x}_j)^{1-A_{ij}}$$

Estimation of latent positions

- **Q:** Given $G = (V, E)$ from an RDPG, find the 'best' $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top$?
- MLE is well motivated but it is intractable for large N_v

$$\hat{\mathbf{X}}_{ML} = \underset{\mathbf{X}}{\operatorname{argmax}} \prod_{i<j} (\mathbf{x}_i^\top \mathbf{x}_j)^{A_{ij}} (1 - \mathbf{x}_i^\top \mathbf{x}_j)^{1-A_{ij}}$$

- Instead, let $P_{ij} = \mathrm{P}\left((i,j) \in \mathcal{E}\right)$ and define $\mathbf{P} = [P_{ij}] \in [0,1]^{N_v \times N_v}$
  - $\Rightarrow$ RDPG model specifies that $\mathbf{P} = \mathbf{X}\mathbf{X}^\top$
  - $\Rightarrow$ **Key:** Observed $\mathbf{A}$ is a noisy realization of $\mathbf{P}$ $(\mathbb{E}\{\mathbf{A}\} = \mathbf{P})$
- Suggests a LS regression approach to find $\mathbf{X}$ s.t. $\mathbf{X}\mathbf{X}^\top \approx \mathbf{A}$

$$\hat{\mathbf{X}}_{LS} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X}\mathbf{X}^\top - \mathbf{A}\|_F^2$$

A. Athreya et al, "Statistical inference on random dot product graphs: A survey," *JMLR*, 2018

Adjacency spectral embedding

- Since $\mathbf{A}$ is real and symmetric, can decompose it as $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$
  - $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_v}]$ is the orthogonal matrix of eigenvectors
  - $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{N_v})$, with eigvenvalues $\lambda_1 \geq \ldots \geq \lambda_{N_v}$

Adjacency spectral embedding

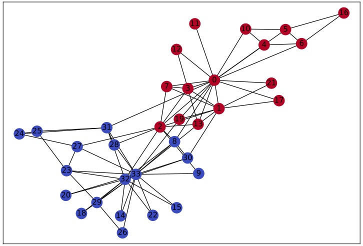
- Since $\mathbf{A}$ is real and symmetric, can decompose it as $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$
  - $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_v}]$ is the orthogonal matrix of eigenvectors
  - $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_{N_v})$, with eigvenvalues $\lambda_1 \geq \ldots \geq \lambda_{N_v}$
- Define $\hat{\mathbf{\Lambda}} = \text{diag}(\lambda_1^+, \ldots, \lambda_d^+)$ and $\hat{\mathbf{U}} = [\mathbf{u}_1, \ldots, \mathbf{u}_d]$ $(\lambda^+ := \max(0, \lambda))$
- Best rank-d, positive semi-definite (PSD) approximation of $\mathbf{A}$ is $\hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^\top$

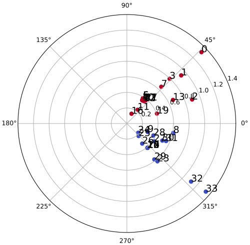
$\Rightarrow$ Ajacency spectral embedding (ASE) is $\hat{\mathbf{X}}_{LS} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}^{1/2}$ since

$$\mathbf{A} \approx \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^\top = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}^{1/2}\hat{\mathbf{\Lambda}}^{1/2}\hat{\mathbf{U}}^\top = \hat{\mathbf{X}}_{LS}\hat{\mathbf{X}}_{LS}^\top$$

Interpretability of the embeddings

- Ex: Zachary's karate club graph with $N_v = 34$, $N_e = 78$ (left)





- Node embeddings (rows of $\hat{\mathbf{X}}_{LS}$) for $d = 2$ (right)
  - Club's administrator ($i = 0$) and instructor ($j = 33$) are orthogonal
- Interpretability of embeddings a valuable asset for RDPGs

  $\Rightarrow$ Vector magnitudes indicate how well connected nodes are

  $\Rightarrow$ Vector angles indicate nodes' affinity

Weighted graphs

- **Q:** Can we extend the RDPG model to the weighted case?
- **Idea** latent positions related to the moment generating function (MGF) of weights ω_{ij}

$\Rightarrow$ **Weighted RDPG**: Each node now has a sequence of vectors $(\mathbf{x}_i[k] \in \mathbb{R}^{d_k})_{k\in\mathbb{N}}$ where

$$\mathbb{E}[\omega_{ij}^k] = \mathbf{x}_i[k]^\top \mathbf{x}_j[k]$$

$\Rightarrow$ Weights are independently drawn from distributions with MGF

$$\mathbb{E}\{e^{t\omega_{ij}}\} = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}\{\omega_{ij}^k\}}{k!} = \sum_{k=0}^{\infty} \frac{t^k \mathbf{x}_i[k]^\top \mathbf{x}_j[k]}{k!}$$

Weighted graphs

- **Q:** Can we extend the RDPG model to the weighted case?
- **Idea** latent positions related to the moment generating function (MGF) of weights ω_{ij}

⇒ **Weighted RDPG**: Each node now has a sequence of vectors $(\mathbf{x}_i[k] \in \mathbb{R}^{d_k})_{k\in\mathbb{N}}$ where

$$\mathbb{E}[\omega_{ij}^k] = \mathbf{x}_i[k]^\top \mathbf{x}_j[k]$$

⇒ Weights are independently drawn from distributions with MGF

$$\mathbb{E}\{e^{t\omega_{ij}}\} = \sum_{k=0}^{\infty} \frac{t^k \mathbb{E}\{\omega_{ij}^k\}}{k!} = \sum_{k=0}^{\infty} \frac{t^k \mathbf{x}_i[k]^\top \mathbf{x}_j[k]}{k!}$$

- We now have a sequence of matrices $\mathbf{X}[k] = [\mathbf{x}_1[k], \ldots, \mathbf{x}_{N_v}[k]]^\top$ such that

$$\mathbb{E}\{\underbrace{\mathbf{A} \circ \mathbf{A} \circ \cdots \circ \mathbf{A}}_{k \text{ times}}\} = \mathbb{E}\left\{\mathbf{A}^{(k)}\right\} := \mathbf{M}[k] = \mathbf{X}[k]\mathbf{X}[k]^\top$$

Weighted RDPG

Advantages

- ✓ Backwards compatibility: vanilla RDPG is recovered by setting $\mathbf{x}_i[k] = \mathbf{x}_i$ for all $k > 0$
- ✓ Flexible way of specifying a distribution per edge
  - Acommodates discrete and/or continuous distribution
  - Prior art relied on fixed, known, parametric distribution $F\left(A_{ij}; \boldsymbol{\theta} = \{\mathbf{x}_i^\top[k]\mathbf{x}_j[k]\}_{k=1}^{K}\right)$

R. Tang et al, "Robust estimation from multiple graphs under gross error contamination", arXiv:1707.03487, 2017

D. DeFord et al, "A Random Dot Product Model for Weighted Networks", arXiv:1611.02530, 2016

Weighted RDPG

Advantages

- ✓ Backwards compatibility: vanilla RDPG is recovered by setting $\mathbf{x}_i[k] = \mathbf{x}_i$ for all $k > 0$
- ✓ Flexible way of specifying a distribution per edge
  - Acommodates discrete and/or continuous distribution
  - Prior art relied on fixed, known, parametric distribution $F\left(A_{ij}; \boldsymbol{\theta} = \{\mathbf{x}_i^\top[k]\mathbf{x}_j[k]\}_{k=1}^K\right)$
- ✓ Sparsity pattern of $\mathbf{A}^{(k)}$ is maintained for all k
- ✓ Observation $\mathbf{A}^{(k)}$ is a noisy realization of $\mathbf{M}[k]$

$\Rightarrow$ **Inference** of the embedding sequence $(\hat{\mathbf{X}}[k])$ via the ASE of $\mathbf{A}^{(k)}$

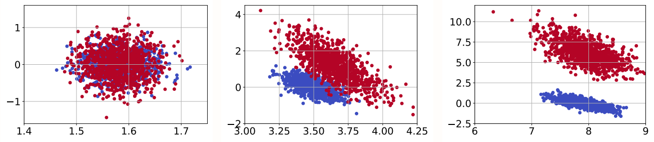
R. Tang et al, "Robust estimation from multiple graphs under gross error contamination", arXiv:1707.03487, 2017

D. DeFord et al, "A Random Dot Product Model for Weighted Networks", arXiv:1611.02530, 2016

Weighted RDPG: Discriminative power

- Ex: $Q = 2$ block weighted SBM graph G with $N_v = 2000$, edges present w.p. $p = 0.5$

  $\Rightarrow$ Weights $A_{ij} \sim \mathcal{N}(5, 0.1)$ except among nodes $i > 1000$, where $A_{ij} \sim \text{Poisson}(5)$



- ASE estimates $\hat{\mathbf{x}}_i[k]$ for $k = 1$ (left), $k = 2$ (center), $k = 3$ (right), where $d = 2$
  - Indistinguishable for $k = 1$, since $\hat{\mathbf{x}}_i[1]$ are centered around $(\sqrt{\mu p}, 0) = (\sqrt{\lambda p}, 0) \approx (1.58, 0)$
  - Noise hinders discriminability for $k = 2$, even though

$$\mathbf{x}_i[2] = \begin{cases} (\sqrt{p(\mu^2 + \sigma^2)}, 0) \approx (3.55, 0) & i \leq 1000, \\ (\sqrt{p(\mu^2 + \sigma^2)}, \sqrt{p(\lambda^2 + \lambda - (\mu^2 + \sigma^2)}) \approx (3.55, 1.58) & i > 1000 \end{cases}$$

  - Skewness kicks in for $k = 3$ and group separation is apparent

Weighted and Directed Graphs

- So far, matrix $\mathbf{M}[k]$ is restricted to be
  - ✗ Positive semi-definitive: what about heterophilous behaviour?
  - ✗ Symmetric: what about directed graphs?

Weighted and Directed Graphs

- So far, matrix $\mathbf{M}[k]$ is restricted to be
  - ✗ Positive semi-definitive: what about heterophilous behaviour?
  - ✗ Symmetric: what about directed graphs?
- **Extension to digraphs**
  - Each node has an associated sequence $\mathbf{x}_i[k]$ – now in $\mathbb{R}^{2d}$
  - Or two vectors: $\mathbf{x}_i^l[k]$ and $\mathbf{x}_i^r[k]$ (first and last d entries of $\mathbf{x}_i$)
- Model:

$$\mathbb{E}\left\{\mathbf{A}^{(k)}\right\} := \mathbf{M}[k] = \mathbf{X}^l[k]\mathbf{X}^r[k]^\top \quad (1)$$

Weighted and Directed RDPG

- **Inference:**
  - $\mathbf{M}[k] = \mathbb{E}\left\{\mathbf{A}^{(k)}\right\}$ still holds

$\Rightarrow$ Seek $\{\hat{\mathbf{X}}^l[k], \hat{\mathbf{X}}^r[k]\}$ s.t. $\hat{\mathbf{X}}^l[k]\hat{\mathbf{X}}^r[k]^\top$ is the best rank-d approximation of $\mathbf{A}^{(k)}$

Weighted and Directed RDPG

- **Inference:**
  - $\mathbf{M}[k] = \mathbb{E}\left\{\mathbf{A}^{(k)}\right\}$ still holds

  $\Rightarrow$ Seek $\{\hat{\mathbf{X}}^l[k], \hat{\mathbf{X}}^r[k]\}$ s.t. $\hat{\mathbf{X}}^l[k]\hat{\mathbf{X}}^r[k]^\top$ is the best rank-d approximation of $\mathbf{A}^{(k)}$

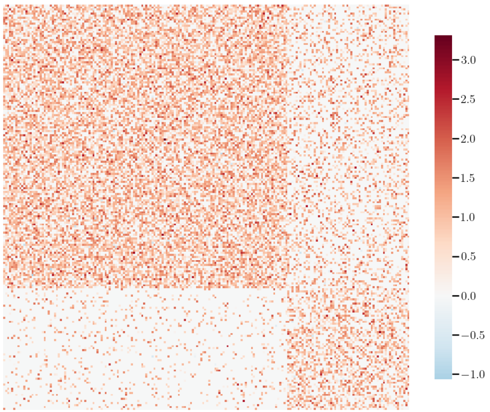
  - SVD: $\mathbf{A}^{(k)} = \mathbf{U}[k]\mathbf{D}[k]\mathbf{V}[k]^\top$

$$\Rightarrow \hat{\mathbf{X}}^l[k] = \hat{\mathbf{U}}[k]\hat{\mathbf{D}}[k]^{1/2} \text{ and } \hat{\mathbf{X}}^r[k] = \hat{\mathbf{V}}[k]\hat{\mathbf{D}}[k]^{1/2}$$

  - What about backwards compatibility? Enforced by the choice of $\hat{\mathbf{D}}[k]^{1/2}$
    - $\mathbf{M}[k]$ is symmetric $\Leftrightarrow \mathbf{X}^l[k] = \mathbf{X}^r[k]$

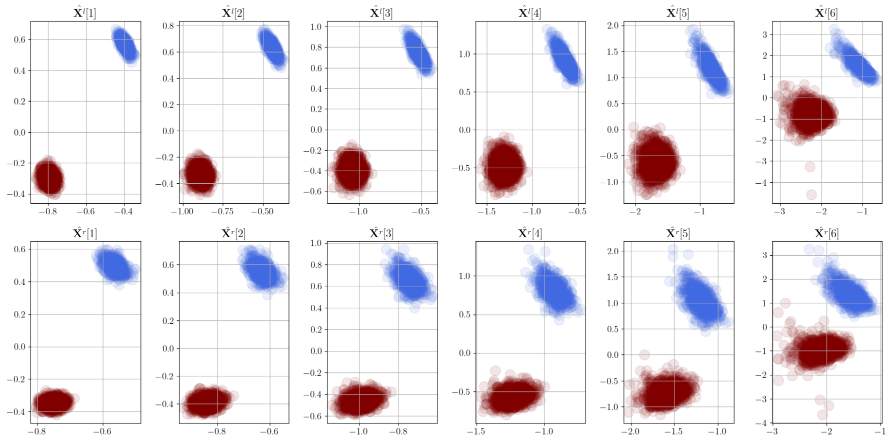
Embedding a Weighted SBM Graph

- A weighted SBM graph G with $N_v = 2000$, number of classes $Q = 2$, inter-class connection probability matrix $\mathbf{\Pi} = \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.5 \end{pmatrix}$ and weights $A_{ij} \sim \mathcal{N}(1, 0.5)$



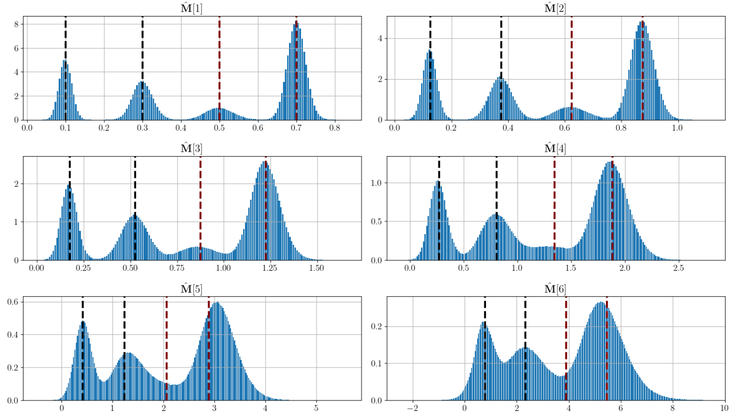
Embedding a Weighted SBM Graph

- A weighted SBM graph G with $N_v = 2000$, number of classes $Q = 2$, inter-class connection probability matrix $\mathbf{\Pi} = \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.5 \end{pmatrix}$ and weights $A_{ij} \sim \mathcal{N}(1, 0.5)$
  - $\hat{\mathbf{X}}^l[k]$ and $\hat{\mathbf{X}}^r[k]$ for $k = 1, \ldots, 6$



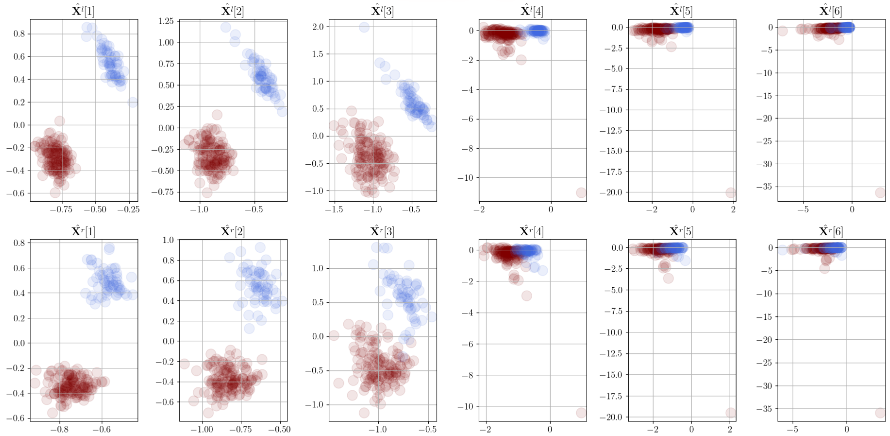
Embedding a Weighted SBM Graph

- A weighted SBM graph G with $N_v = 2000$, number of classes $Q = 2$, inter-class connection probability matrix $\mathbf{\Pi} = \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.5 \end{pmatrix}$ and weights $A_{ij} \sim \mathcal{N}(1, 0.5)$
  - $\hat{\mathbf{X}}^l[k]$ and $\hat{\mathbf{X}}^r[k]$ for $k = 1, \ldots, 6$ can reconstruct accurate values of $\hat{\mathbf{M}}[k]$ up to $k \approx 4$



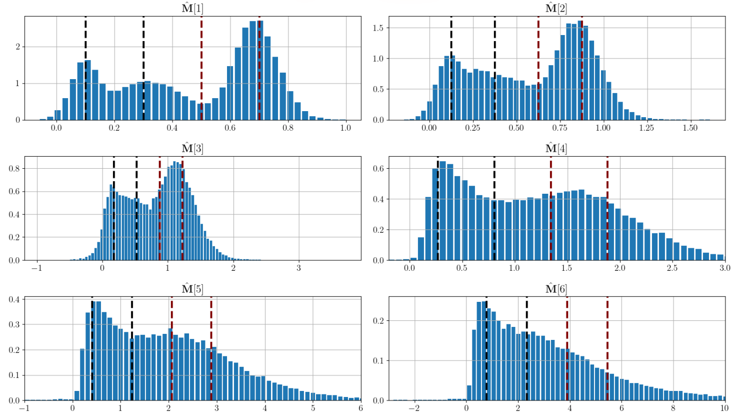
Embedding a Weighted SBM Graph

- A weighted SBM graph G with $N_v = 200$, number of classes $Q = 2$, inter-class connection probability matrix $\mathbf{\Pi} = \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.5 \end{pmatrix}$ and weights $A_{ij} \sim \mathcal{N}(1, 0.5)$
  - $\hat{\mathbf{X}}^l[k]$ and $\hat{\mathbf{X}}^r[k]$ for $k = 1, \ldots, 6$



Embedding a Weighted SBM Graph

- A weighted SBM graph G with $N_v = 200$, number of classes $Q = 2$, inter-class connection probability matrix $\mathbf{\Pi} = \begin{pmatrix} 0.7 & 0.3 \\ 0.1 & 0.5 \end{pmatrix}$ and weights $A_{ij} \sim \mathcal{N}(1, 0.5)$
  - $\hat{\mathbf{X}}^l[k]$ and $\hat{\mathbf{X}}^r[k]$ for $k = 1, \ldots, 6$ reconstructs accurate values of $\hat{\mathbf{M}}[k]$ up to $k \approx 2$ (somewhat)



Consistency of ASE for WD-RDPG

Theorem (Consistency): Let $\mathbf{B} \in \mathbb{R}^{N\times N}$ be a random matrix such that $\{B_{ii}\} = 0$, and $\{B_{ij}\}_{i\neq j}$ are bounded and independent with $\mathbb{E}[B_{ij}] = E_{ij}$, $\mathbf{E} = \mathbf{X}\mathbf{Y}^\top$ for fixed $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N\times d}$.
Assume $\text{rank}(\mathbf{E}) = d$ and that the singular values of $\mathbf{E}$ $\sigma_1 > \sigma_2 > \ldots > \sigma_d > 0$ are such that $\min_{i\neq j} |\sigma_i - \sigma_j| > \delta N$ and $\sigma_d > \delta N$ for some $\delta > 0$. Let $\hat{\mathbf{X}}, \hat{\mathbf{Y}} \in \mathbb{R}^{N\times d}$ be the ASE of $\mathbf{B}$. Then, there almost always exist an invertible matrix $\mathbf{W} \in \mathbb{R}^{d\times d}$ such that, for all $i \in \{1, \ldots, N\}$ and all $\gamma < 1$,

$$\mathbb{P}\left[||(\hat{\mathbf{X}}\mathbf{W} - \mathbf{X})_i||_2^2 > N^{-\gamma}\right] = o\left(N^{\gamma-1}\log N\right)$$

$$\mathbb{P}\left[||(\hat{\mathbf{Y}}\mathbf{W}^{-\top} - \mathbf{Y})_i||_2^2 > N^{-\gamma}\right] = o\left(N^{\gamma-1}\log N\right)$$

where $\mathbf{C}_i$ is the i-th row of matrix $\mathbf{C}$.

Consistency of ASE for WD-RDPG

Theorem (Consistency): Let $\mathbf{B} \in \mathbb{R}^{N \times N}$ be a random matrix such that $\{B_{ii}\} = 0$, and $\{B_{ij}\}_{i \neq j}$ are bounded and independent with $\mathbb{E}[B_{ij}] = E_{ij}$, $\mathbf{E} = \mathbf{X}\mathbf{Y}^\top$ for fixed $\mathbf{X}, \mathbf{Y} \in \mathbb{R}^{N \times d}$.
Assume $\text{rank}(\mathbf{E}) = d$ and that the singular values of $\mathbf{E}$ $\sigma_1 > \sigma_2 > \ldots > \sigma_d > 0$ are such that $\min_{i \neq j} |\sigma_i - \sigma_j| > \delta N$ and $\sigma_d > \delta N$ for some $\delta > 0$. Let $\hat{\mathbf{X}}, \hat{\mathbf{Y}} \in \mathbb{R}^{N \times d}$ be the ASE of $\mathbf{B}$. Then, there almost always exist an invertible matrix $\mathbf{W} \in \mathbb{R}^{d \times d}$ such that, for all $i \in \{1, \ldots, N\}$ and all $\gamma < 1$,

$$\mathbb{P}\left[||(\hat{\mathbf{X}}\mathbf{W} - \mathbf{X})_i||_2^2 > N^{-\gamma} \right] = o\left(N^{\gamma - 1} \log N\right)$$

$$\mathbb{P}\left[||(\hat{\mathbf{Y}}\mathbf{W}^{-\top} - \mathbf{Y})_i||_2^2 > N^{-\gamma} \right] = o\left(N^{\gamma - 1} \log N\right)$$

where $\mathbf{C}_i$ is the i-th row of matrix $\mathbf{C}$.

- For each fixed k, we let $\mathbf{B} = \mathbf{A}^{(k)}$ to ensure consistency of the ASE to $\mathbf{X}^l[k], \mathbf{X}^r[k]$

Asymptotic Normality of ASE for WD-RDPG

Theorem (Central Limit Theorem): Let F be a weighted, directed, inner-product distribution. Assume $\mathbf{B}$ and $\mathbf{E}$ as before, only now $\mathbb{E}\left[B_{ij}|\mathbf{X},\mathbf{Y}\right] = E_{ij}$, and $\mathbf{X},\mathbf{Y} \sim F$. Then there almost always exist a sequence of invertible matrices $\mathbf{W}_N \in \mathbb{R}^{d\times d}$ such that, for all $i \in \{1,\dots,N\}$ and all $\mathbf{x} \in \mathbb{R}^d$:

$$\lim_{N\to\infty} \mathbb{P}\left[N^{1/2}(\hat{\mathbf{X}}\mathbf{W}_N - \mathbf{X})_i \leq \mathbf{z}\right] = \int_{\text{supp } F} \Phi(\mathbf{z},\mathbf{\Sigma}_{\mathbf{X}}(\mathbf{x}))dF(\mathbf{x})$$

$$\lim_{N\to\infty} \mathbb{P}\left[N^{1/2}(\hat{\mathbf{Y}}\mathbf{W}_N^{-\top} - \mathbf{Y})_i \leq \mathbf{z}\right] = \int_{\text{supp } F} \Phi(\mathbf{z},\mathbf{\Sigma}_{\mathbf{Y}}(\mathbf{y}))dF(\mathbf{y})$$

where $\Phi(\mathbf{z},\mathbf{\Sigma})$ is zero-mean multivariate normal with covariance matrix $\mathbf{\Sigma}$, and

$$\mathbf{\Sigma}_{\mathbf{X}}(\mathbf{x}) = \mathbf{\Delta}_{\mathbf{X}}^{-1}\mathbb{E}\left[\left(\mathbf{x}^\top\mathbf{X}_1 - (\mathbf{x}^\top\mathbf{X}_1)^2\right)\mathbf{X}_1\mathbf{X}_1^\top\right]\mathbf{\Delta}_{\mathbf{X}}^{-1}\ ,\ \mathbf{\Delta}_{\mathbf{X}} = \mathbb{E}\left[\mathbf{X}_1\mathbf{X}_1^\top\right]$$

$$\mathbf{\Sigma}_{\mathbf{Y}}(\mathbf{y}) = \mathbf{\Delta}_{\mathbf{Y}}^{-1}\mathbb{E}\left[\left(\mathbf{y}^\top\mathbf{Y}_1 - (\mathbf{y}^\top\mathbf{Y}_1)^2\right)\mathbf{Y}_1\mathbf{Y}_1^\top\right]\mathbf{\Delta}_{\mathbf{Y}}^{-1}\ ,\ \mathbf{\Delta}_{\mathbf{Y}} = \mathbb{E}\left[\mathbf{Y}_1\mathbf{Y}_1^\top\right]$$

Real-life dataset (I): UN roll calls

- For each roll call in the UN General Assembly, members vote 'Yes', 'No' or 'Abstain'.
- 'Abstain' is frequently used as another level of agreement with the roll call

⇒ Consider the bipartite digraph for 2003, where

  - Nodes correspond to member countries and roll calls,
  - Edge weight is either 1 (affirmative vote), -1 (negative) or 0 (abstain or absent).

Real-life dataset (I): UN roll calls

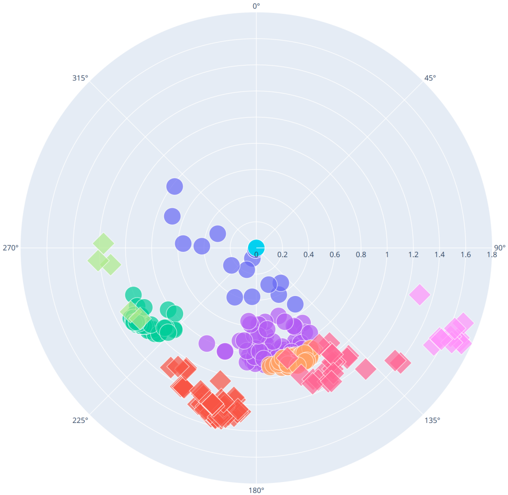
- For each roll call in the UN General Assembly, members vote ‘Yes’, ‘No’ or ‘Abstain’.
- ‘Abstain’ is frequently used as another level of agreement with the roll call

⇒ Consider the bipartite digraph for 2003, where
  - Nodes correspond to member countries and roll calls,
  - Edge weight is either 1 (affirmative vote), -1 (negative) or 0 (abstain or absent).
  - Each country has a probability distribution (p_{-1}, p_0, p_1) for each roll call

- **Q:** What can we learn by visualizing the embeddings? Note that
  - $X_l[k] = 0$ for roll calls (roll calls do not vote)
  - $X_r[k] = 0$ for countries (countries are not voted)

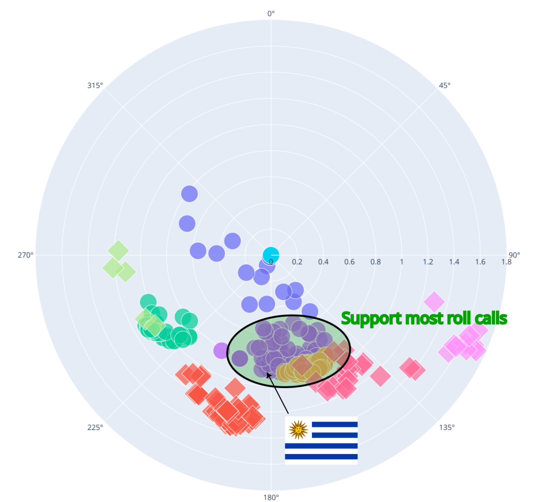
Real-life dataset (I): UN roll calls - Interpretability

- $\mathbf{k} = \mathbf{1}$ and $d = 2$. Countries (●) and roll calls (◆) are colored using a GMM clustering



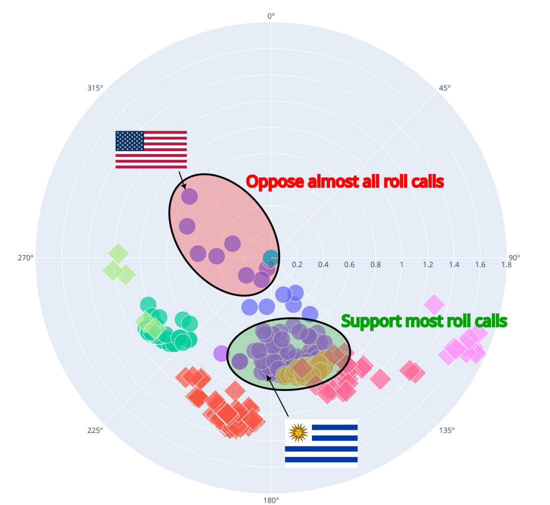
Real-life dataset (I): UN roll calls - Interpretability

- $\mathbf{k} = \mathbf{1}$ and $d = 2$. Countries (●) and roll calls (◆) are colored using a GMM clustering



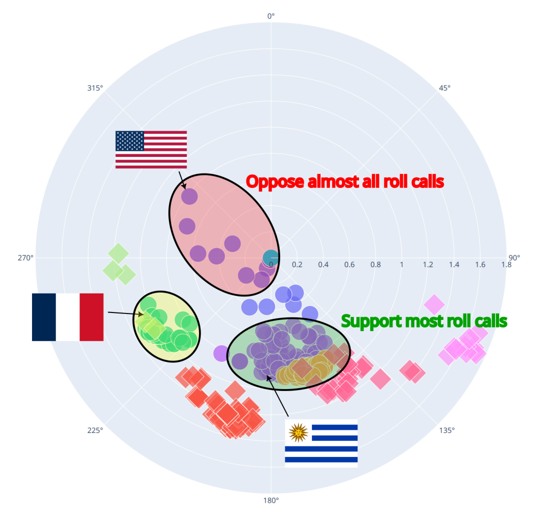
Real-life dataset (I): UN roll calls - Interpretability

- $\mathbf{k} = \mathbf{1}$ and $d = 2$. Countries (●) and roll calls (◆) are colored using a GMM clustering



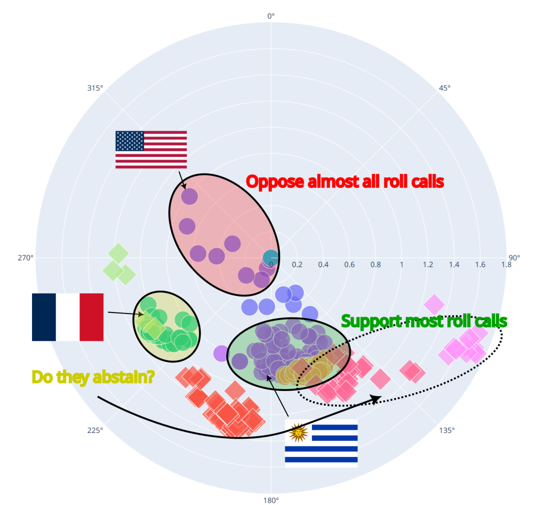
Real-life dataset (I): UN roll calls - Interpretability

- $\mathbf{k} = \mathbf{1}$ and $d = 2$. Countries (●) and roll calls (◆) are colored using a GMM clustering



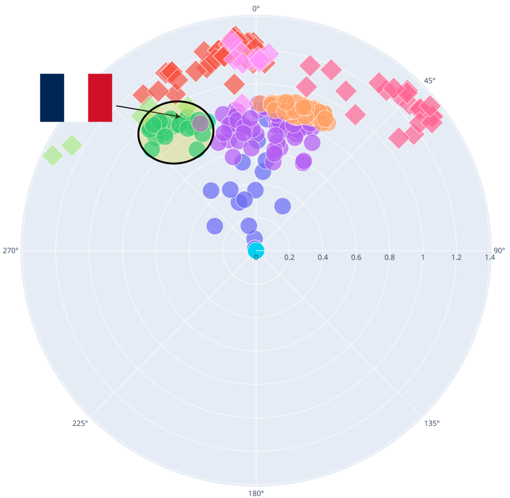
Real-life dataset (I): UN roll calls - Interpretability

- $\mathbf{k} = \mathbf{1}$ and $d = 2$. Countries (●) and roll calls (◆) are colored using a GMM clustering



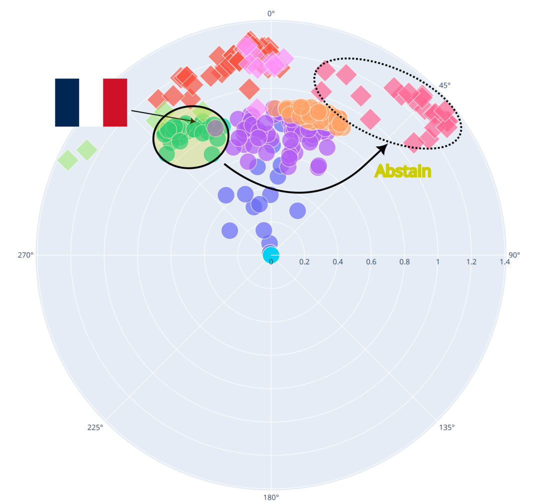
Real-life dataset (I): UN roll calls - Interpretability

- $\mathbf{k} = \mathbf{2}$ and $d = 2$. Countries (●) and roll calls (◆) are colored using a GMM clustering



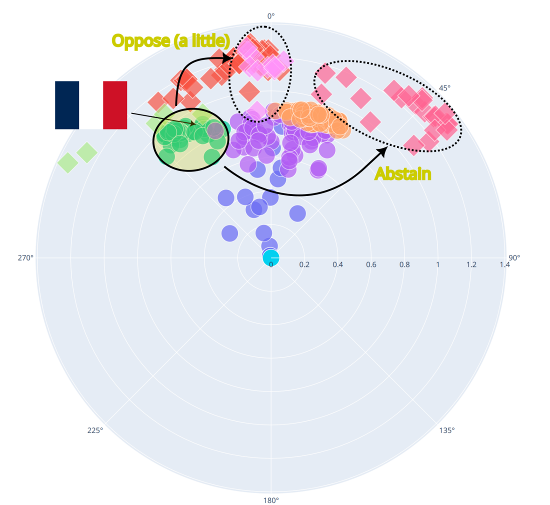
Real-life dataset (I): UN roll calls - Interpretability

- $\mathbf{k} = \mathbf{2}$ and $d = 2$. Countries (●) and roll calls (◆) are colored using a GMM clustering



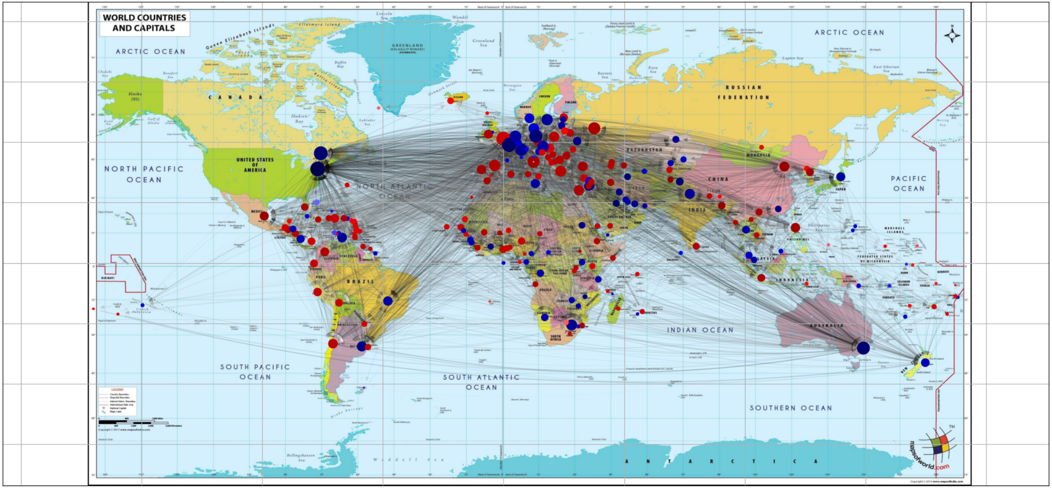
Real-life dataset (I): UN roll calls - Interpretability

- $\mathbf{k}=\mathbf{2}$ and $d=2$. Countries (●) and roll calls (◆) are colored using a GMM clustering



Real-life dataset (II): UN migration data

- Migration between countries in 1990 (based on UN data)
  - Nodes: size indicative of total degree, net balance is positive or negative
  - Edge thickness indicative of total flow



Real-life dataset (II): UN migration data - Graph Generation

- How can we generate similar graphs?

$\Rightarrow$ WD-RDPG for **graph generation**

- Problem statement: Generate $\mathbf{A}$ such that, for $1 \leq i, j \leq N$, A_{ij} follows a distribution whose first $K+1$ moments are $\hat{\mathbf{x}}_i^l[k]^\top \hat{\mathbf{x}}_j^r[k] = \mu_k$ for $k = 0, 1, \ldots, K$.

Real-life dataset (II): UN migration data - Graph Generation

- How can we generate similar graphs?

⇒ WD-RDPG for **graph generation**

- Problem statement: Generate $\mathbf{A}$ such that, for $1 \leq i, j \leq N$, A_{ij} follows a distribution whose first $K+1$ moments are $\hat{\mathbf{x}}_i^l[k]^\top \hat{\mathbf{x}}_j^r[k] = \mu_k$ for $k = 0, 1, \ldots, K$.
- Today I'll discuss the discrete case
  - ⇒ Each migration flow will be converted to $(Q+1)$-quantiles
  - ⇒ $A_{ij} = 0$ indicates relatively low number of migrants, $A_{ij} = Q$ indicates high migration
  - We assume $Q = K$ (i.e., as many moments as symbols)

Real-life dataset (II): UN migration data - Graph Generation

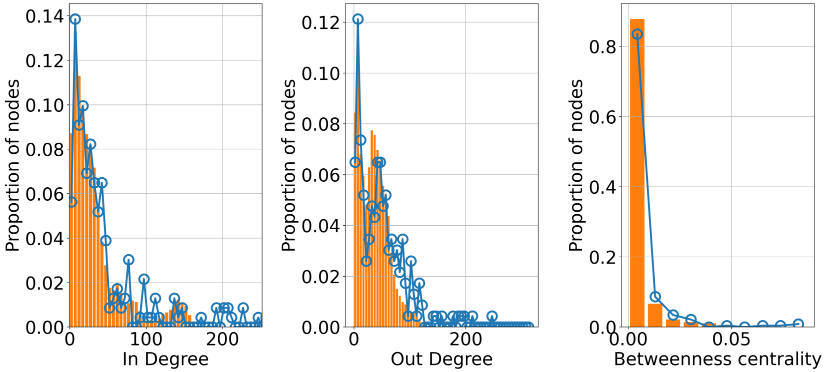
- Consider a single link $(i,j) \Rightarrow$ How can we estimate $p_l = \mathrm{P}\left[A_{ij} = l\right]$?
- We have the following system of equations:

$$\left\{\begin{array}{rcl} p_0 + \ p_1 + \cdots + \ \ p_K & = & \mu_0 \\ 0p_0 + 1p_1 + \cdots + Kp_K & = & \mu_1 \\ 0^2 p_0 + 1^2 p_1 + \cdots + K^2 p_K & = & \mu_2 \\ \vdots & & \vdots \\ 0^K p_0 + 1^K p_1 + \cdots + K^K p_K & = & \mu_K \end{array}\right. \Leftrightarrow \mathbf{V}\mathbf{p} = \boldsymbol{\mu} \tag{2}$$

- $\mathbf{V}$ is a Vandermonde matrix of the possible symbols (in this case $0, 1, \ldots, Q = K$)

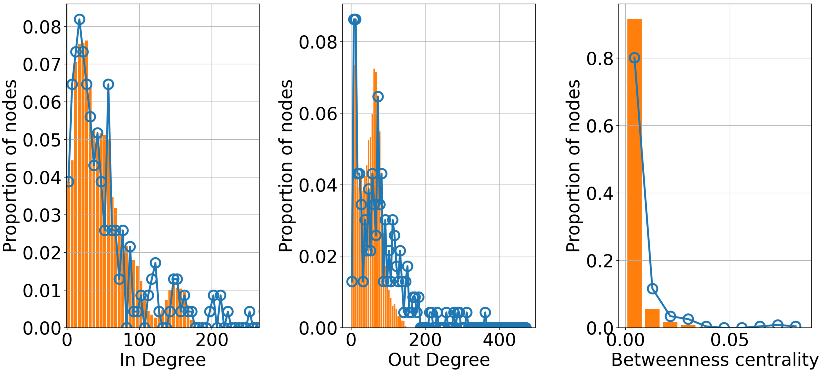
Real-life dataset (II): UN migration data - Graph Generation

- Simulations: For each pair of nodes estimate **p**, generate 100 graphs and compute:
  - Degree distribution
  - Nodes' betweenness centrality
- Great fit for $Q = 2$



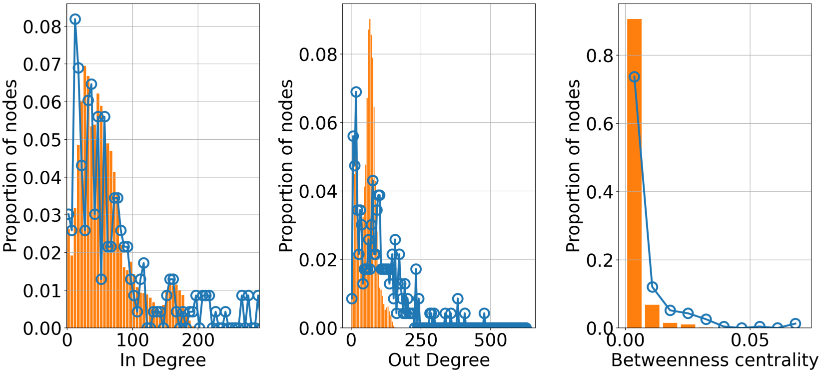
Real-life dataset (II): UN migration data - Graph Generation

- Simulations: For each pair of nodes estimate **p**, generate 100 graphs and compute:
  - Degree distribution
  - Nodes' betweenness centrality
- Great fit for $Q = 2, 3$.



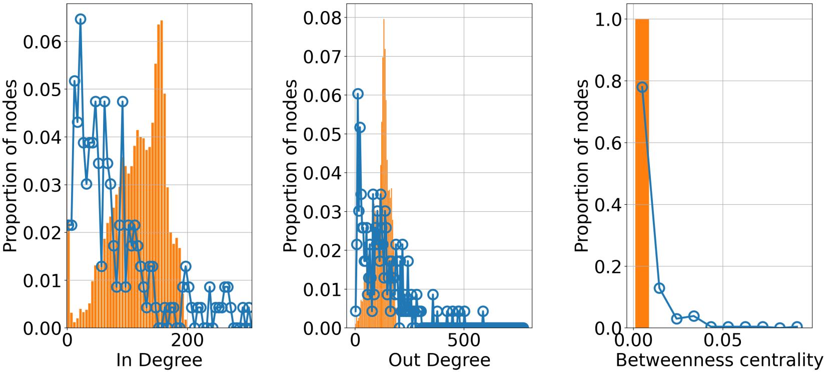
Real-life dataset (II): UN migration data - Graph Generation

- Simulations: For each pair of nodes estimate $\mathbf{p}$, generate 100 graphs and compute:
  - Degree distribution
  - Nodes' betweenness centrality
- Great fit for $Q = 2, 3$. Not so good for $Q = 4$.



Real-life dataset (II): UN migration data - Graph Generation

- Simulations: For each pair of nodes estimate **p**, generate 100 graphs and compute:
  - Degree distribution
  - Nodes' betweenness centrality
- Great fit for $Q = 2, 3$. Not so good for $Q = 4$. And as we add more symbols...



Limitations and Future Work

- Estimating several moments is challenging (unless we have a large graph)

$\Rightarrow$ How can we estimate a distribution from just a few moments?

Limitations and Future Work

- Estimating several moments is challenging (unless we have a large graph)

$\Rightarrow$ How can we estimate a distribution from just a few moments?

- Estimating a distribution per-link does not scale

$\Rightarrow$ Grouping nodes should help. What's the impact on the estimation?

Thanks!

Questions?

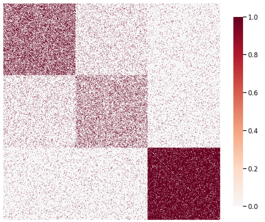
Federico "Larroca" La Rocca

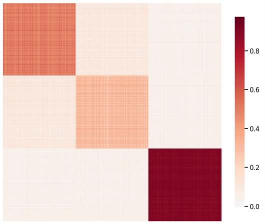
flarroca@fing.edu.uy
@fedelarrocca
https://github.com/git-artes/

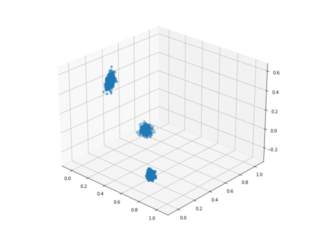
Embedding an SBM graph

- Ex: SBM with $N_v = 1500$, $Q = 3$ and mixing parameters

$$\boldsymbol{\alpha} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \quad \boldsymbol{\Pi} = \begin{bmatrix} 0.5 & 0.1 & 0.05 \\ 0.1 & 0.3 & 0.05 \\ 0.05 & 0.05 & 0.9 \end{bmatrix}$$







- Sample adjacency $\mathbf{A}$ (left), $\hat{\mathbf{X}}_{LS}\hat{\mathbf{X}}_{LS}^\top$ (center), rows of $\hat{\mathbf{X}}_{LS}$ (right)
- Use embeddings to bring to bear geometric methods of analysis

Ambiguity

- **Q:** Is the solution to ASE unique? Nope, inner-products are rotation invariant

$$\mathbf{P} = \mathbf{XW}(\mathbf{XW})^\top = \mathbf{XX}^\top, \quad \mathbf{WW}^\top = \mathbf{I}_d$$

⇒ RDPG embedding problem is identifiable modulo rotations

Ambiguity

- **Q:** Is the solution to ASE unique? Nope, inner-products are rotation invariant

$$\mathbf{P} = \mathbf{X}\mathbf{W}(\mathbf{X}\mathbf{W})^\top = \mathbf{X}\mathbf{X}^\top, \quad \mathbf{W}\mathbf{W}^\top = \mathbf{I}_d$$

$\Rightarrow$ RDPG embedding problem is identifiable modulo rotations

- Ambiguity in DRDPG got worse:
  - a transformation with *any invertible* matrix $\mathbf{W} \in \mathbb{R}^{d\times d}$ will result in the same $\mathbf{M}[k]$.
    undirected (W)RDPG: $\mathbf{W}$ orthonormal matrix
  - $\hat{\mathbf{X}}^l = \mathbf{X}^l\mathbf{W}$ and $\hat{\mathbf{X}}^r = \mathbf{X}^r\mathbf{W}^{-\top}$ (omitting k for clarity)

Ambiguity

- **Q:** Is the solution to ASE unique? Nope, inner-products are rotation invariant

$$\mathbf{P} = \mathbf{X}\mathbf{W}(\mathbf{X}\mathbf{W})^\top = \mathbf{X}\mathbf{X}^\top, \quad \mathbf{W}\mathbf{W}^\top = \mathbf{I}_d$$

$\Rightarrow$ RDPG embedding problem is identifiable modulo rotations

- Ambiguity in DRDPG got worse:
  - a transformation with *any invertible* matrix $\mathbf{W} \in \mathbb{R}^{d\times d}$ will result in the same $\mathbf{M}[k]$. undirected (W)RDPG: $\mathbf{W}$ orthonormal matrix
  - $\hat{\mathbf{X}}^l = \mathbf{X}^l\mathbf{W}$ and $\hat{\mathbf{X}}^r = \mathbf{X}^r\mathbf{W}^{-\top}$ (omitting k for clarity)

$\Rightarrow \hat{\mathbf{X}}^l(\hat{\mathbf{X}}^l)^\top = \mathbf{X}^l\mathbf{W}(\mathbf{X}^r\mathbf{W}^{-\top})^\top = \mathbf{X}^l\mathbf{W}\mathbf{W}^{-1}(\mathbf{X}^r)^\top = \mathbf{X}^l(\mathbf{X}^r)^\top = \mathbf{M}$

Ambiguity

- **Q:** Is the solution to ASE unique? Nope, inner-products are rotation invariant

$$\mathbf{P} = \mathbf{X}\mathbf{W}(\mathbf{X}\mathbf{W})^\top = \mathbf{X}\mathbf{X}^\top, \quad \mathbf{W}\mathbf{W}^\top = \mathbf{I}_d$$

$\Rightarrow$ RDPG embedding problem is identifiable modulo rotations

- Ambiguity in DRDPG got worse:
  - a transformation with *any invertible* matrix $\mathbf{W} \in \mathbb{R}^{d\times d}$ will result in the same $\mathbf{M}[k]$.
    undirected (W)RDPG: $\mathbf{W}$ orthonormal matrix
  - $\hat{\mathbf{X}}^l = \mathbf{X}^l\mathbf{W}$ and $\hat{\mathbf{X}}^r = \mathbf{X}^r\mathbf{W}^{-\top}$ (omitting k for clarity)

  $\Rightarrow$ $\hat{\mathbf{X}}^l(\hat{\mathbf{X}}^l)^\top = \mathbf{X}^l\mathbf{W}(\mathbf{X}^r\mathbf{W}^{-\top})^\top = \mathbf{X}^l\mathbf{W}\mathbf{W}^{-1}(\mathbf{X}^r)^\top = \mathbf{X}^l(\mathbf{X}^r)^\top = \mathbf{M}$
  - $\mathbf{W}$ is necessarily orthonormal by enforcing $\mathbf{X}^l[k]$ and $\mathbf{X}^r[k]$ with orthogonal columns

Weighted RDPG

- Vanilla RDPG require $0 \leq \mathbf{x}_i \mathbf{x}_i^\top \leq 1$. Is any sequence $\mathbf{M}[k]$ valid in WRDPG?
  - Certainly not! E.g. $M_{ij}[k] = -1$ for all k cannot be correct

Weighted RDPG

- Vanilla RDPG require $0 \leq \mathbf{x}_i \mathbf{x}_i^\top \leq 1$. Is any sequence $\mathbf{M}[k]$ valid in WRDPG?
  - Certainly not! E.g. $M_{ij}[k] = -1$ for all k cannot be correct
- A sequence $\{m[k]\}_{k \geq 0}$ is an admisible moment sequence if $m[0] = 1$ and the matrix

$$\mathbf{B} = \begin{pmatrix} m[0] & m[1] & m[2] & \dots & m[p] \\ m[1] & m[2] & m[3] & \dots & m[p+1] \\ m[2] & m[3] & m[4] & \dots & m[p+3] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m[p] & m[p+1] & m[p+2] & \dots & m[2p] \end{pmatrix}$$

is positive-semidefinite for all $p \geq 0$

$\Rightarrow$ $M_{ij}[k]$ has to be an admisible moment sequence for all i, j