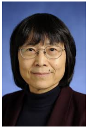
Spectral inference for large stochastic blockmodes with nodal covariates

Angelo Mele
Johns Hopkins Carey Business School
http://meleangelo.github.io
angelo.mele@jhu.edu



Lingxin Hao
Sociology JHU

Joshua Cape
Statistics UMich

Carey E. Priebe
Applied Math JHU

Network Econometrics Juniors' Conference
October, 5 2019

Motivation

- Stochastic blockmodels (SBM) are a workhorse in statistical modeling of network formation (Holland et al 1983)
- Used for community detection (Abbe 2018, Karrer-Newman 2011)
- Used to cluster nodes in applications (Nimczik 2018)
- Inference and estimation are computationally burdensome
- If we include observed covariates in the model, computations are even more problematic
- We want a method that scales well with network size

In this paper

1. Reformulate problem as estimation of latent positions in random dot product graphs (Athreya 2018)
2. Focus on discrete observed covariates
3. **Estimation**: Combine
   - spectral methods
   - dimension reduction tools (ASE)
   - standard clustering methods (GMM, K-means)

   to estimate the block structure and the effect of covariates
4. **Asymptotics**: prove asymptotic results (CLT)
5. **Computation**: Our estimator works well and it is superfast

Stochastic blockmodels (SBM)

K-block stochastic blockmodel (Holland et al 1983)

$$\tau_i \overset{iid}{\sim} Multinomial(1; \pi_1, ..., \pi_K) \quad (1)$$

$$\boldsymbol{A}_{ij}|\tau_i, \tau_j \overset{ind}{\sim} Bernoulli(\boldsymbol{P}_{ij}) \quad (2)$$

$$\boldsymbol{P}_{ij} = \boldsymbol{\theta}_{\tau_i \tau_j} \quad (3)$$

The block-specific probability matrix is

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_{11} & \theta_{12} & \dots & \theta_{1K} \\ \theta_{21} & \theta_{22} & \dots & \theta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{K1} & \theta_{K2} & \dots & \theta_{KK} \end{bmatrix} \quad (4)$$

Random dot product graphs (RDPG)

Random dot product graphs (Young and Scheinerman 2007, Athreya et al 2018)

- Nodes are characterized by a vector of latent positions

$$\boldsymbol{X}_i \in \mathcal{X}_d \subseteq \mathbb{R}^d$$

- Conditional on latent positions $\boldsymbol{X}$ links are independent

$$\boldsymbol{X}_i \overset{iid}{\sim} F \tag{5}$$

$$\boldsymbol{A}_{ij}|\boldsymbol{X}_i, \boldsymbol{X}_j \overset{ind}{\sim} Bernoulli(\boldsymbol{P}_{ij}) \tag{6}$$

$$\boldsymbol{P}_{ij} = \boldsymbol{X}_i^T \boldsymbol{X}_j \tag{7}$$

SBM is a RDPG

SBMs are RDPGs with K fixed latent positions

$$\boldsymbol{\nu} = [\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_K] \tag{8}$$

$$F = \pi_1 \delta_{\boldsymbol{\nu}_1} + \cdots + \pi_K \delta_{\boldsymbol{\nu}_K} \tag{9}$$

and

$$\boldsymbol{P}_{ij} = \boldsymbol{X}_i^T \boldsymbol{X}_j \tag{10}$$

and i and j belong to block k if $\boldsymbol{X}_i = \boldsymbol{X}_j = \boldsymbol{\nu}_k$

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \ldots & \theta_{1K} \\ \theta_{21} & \theta_{22} & \ldots & \theta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{K1} & \theta_{K2} & \ldots & \theta_{KK} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\nu}_1^T \boldsymbol{\nu}_1 & \boldsymbol{\nu}_1^T \boldsymbol{\nu}_2 & \ldots & \boldsymbol{\nu}_1^T \boldsymbol{\nu}_K \\ \boldsymbol{\nu}_2^T \boldsymbol{\nu}_1 & \boldsymbol{\nu}_2^T \boldsymbol{\nu}_2 & \ldots & \boldsymbol{\nu}_2^T \boldsymbol{\nu}_K \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\nu}_K^T \boldsymbol{\nu}_1 & \boldsymbol{\nu}_K^T \boldsymbol{\nu}_2 & \ldots & \boldsymbol{\nu}_K^T \boldsymbol{\nu}_K \end{bmatrix} \tag{11}$$

Example

- Let $K = 2$
- Let the multinomial probabilities be $\pi = (\pi_1, \pi_2)$
- The 2×2 SBM probability matrix is

$$\boldsymbol{\theta} = \begin{bmatrix} p^2 & pq \\ pq & q^2 \end{bmatrix} \tag{12}$$

Example

- Let $K = 2$
- Let the multinomial probabilities be $\pi = (\pi_1, \pi_2)$
- The 2×2 SBM probability matrix is

$$\boldsymbol{\theta} = \begin{bmatrix} p^2 & pq \\ pq & q^2 \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix} \begin{bmatrix} p & q \end{bmatrix} \tag{12}$$

Example

- Let $K = 2$
- Let the multinomial probabilities be $\pi = (\pi_1, \pi_2)$
- The 2×2 SBM probability matrix is

$$\boldsymbol{\theta} = \begin{bmatrix} p^2 & pq \\ pq & q^2 \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix} \begin{bmatrix} p & q \end{bmatrix} = \boldsymbol{\nu}\boldsymbol{\nu}^T \tag{12}$$

Example

- Let $K = 2$
- Let the multinomial probabilities be $\pi = (\pi_1, \pi_2)$
- The 2×2 SBM probability matrix is

$$\boldsymbol{\theta} = \begin{bmatrix} p^2 & pq \\ pq & q^2 \end{bmatrix} = \begin{bmatrix} p \\ q \end{bmatrix} \begin{bmatrix} p & q \end{bmatrix} = \boldsymbol{\nu}\boldsymbol{\nu}^T \tag{12}$$

- Latent position of RDPG are

$$\boldsymbol{X}_i = \begin{cases} p & \text{w.p. } \pi_1 \\ q & \text{w.p. } \pi_2 \end{cases} \tag{13}$$

Generalized Random Dot Product Graphs (GRDPG)

Generalized Random Dot Product Graphs

Same as RDPG, but allows for disassortative graphs

$$\begin{aligned} \boldsymbol{X}_i &\overset{iid}{\sim} F & (14) \\ \boldsymbol{A}_{ij}|\boldsymbol{X}_i, \boldsymbol{X}_j &\overset{ind}{\sim} Bernoulli(\boldsymbol{P}_{ij}) & (15) \\ \boldsymbol{P}_{ij} &= \boldsymbol{X}_i^T \boldsymbol{I}_{d_1,d_2} \boldsymbol{X}_j & (16) \end{aligned}$$

where the matrix $\boldsymbol{I}_{d_1,d_2}$

$$\boldsymbol{I}_{d_1,d_2} = diag\left[\underbrace{+1,+1,+1,+1,+1,+1}_{d_1},\underbrace{-1,-1,-1,-1}_{d_2}\right] \quad (17)$$

RDPG Estimation intuition

If we observe $\boldsymbol{P}$, easy:

$$\boldsymbol{P} = \boldsymbol{X}\boldsymbol{X}^T \Rightarrow \boldsymbol{P} = \boldsymbol{U_P}\boldsymbol{S_P}\boldsymbol{U_P}^T \Rightarrow \widehat{\boldsymbol{X}} = \boldsymbol{U_P}|\boldsymbol{S_P}|^{1/2} \tag{18}$$

Where $\boldsymbol{S_P}$ is $d \times d$ diagonal matrix with d largest eigenvalues
However, we only observe $\boldsymbol{A}$, a noisy version of $\boldsymbol{P}$

$$\boldsymbol{A} = \boldsymbol{P} + \boldsymbol{E} \Rightarrow \boldsymbol{A}_{ij} = \boldsymbol{P}_{ij} + \boldsymbol{E}_{ij} \text{ or } \mathbb{E}(\boldsymbol{A}_{ij}) = \boldsymbol{P}_{ij} \tag{19}$$

Eigenvalues of $\boldsymbol{A}$ are ‘’close” to eigenvalues of $\boldsymbol{X}\boldsymbol{X}^T$

Adjacency Spectral Embeddings (ASE) is

$$ASE(\boldsymbol{A}) = \boldsymbol{U_A S_A U_A^T} \Rightarrow \widehat{\boldsymbol{X}} = \boldsymbol{U_A}|\boldsymbol{S_A}|^{1/2} \tag{20}$$

Where $\boldsymbol{S_A}$ is $d \times d$ diagonal matrix

- with d largest eigenvalues in absolute value
- with d_1 positive eigenvalues
- and d_2 negative eigenvalues

And $\boldsymbol{U_A}$ the corresponding eigenvectors
Dimension reduction tool

SBM with covariates

Researcher has data on observed attributes $\boldsymbol{Z}_i$
The model is a SBM with covariates (Airoldi et al 2011, Sweet 2015, Roy et al 2019)

$$\boldsymbol{P}_{ij} = \boldsymbol{B}_{\tau_i\tau_j} + \beta\mathbf{1}_{\{\boldsymbol{Z}_i=\boldsymbol{Z}_j\}} \tag{21}$$

We focus on $\boldsymbol{Z}_i$ scalar and binary (male/female).
We also consider nonlinear models

$$\boldsymbol{P}_{ij} = h\left(\boldsymbol{B}_{\tau_i\tau_j} + \beta\mathbf{1}_{\{\boldsymbol{Z}_i=\boldsymbol{Z}_j\}}\right) \tag{22}$$

where h is a well-behaved function (e.g. logistic)
We rewrite it as

$$\boldsymbol{P}_{ij} = h\left(\boldsymbol{X}_i^T\boldsymbol{X}_j + \beta\mathbf{1}_{\{\boldsymbol{Z}_i=\boldsymbol{Z}_j\}}\right) \tag{23}$$

Go back to example with $K = 2$ blocks...

$$\boldsymbol{B} = \begin{array}{c} \\ block_1 \\ block_2 \end{array} \begin{array}{c} \begin{array}{cc} block_1 & block_2 \end{array} \\ \begin{pmatrix} p^2 & pq \\ pq & q^2 \end{pmatrix} \end{array} \tag{24}$$

Adding binary covariate (male/female) implies a 4−block SBM

$$\boldsymbol{B}_Z = \begin{array}{c} \\ male_1 \\ female_1 \\ male_2 \\ female_2 \end{array} \begin{array}{c} \begin{array}{cccc} male_1 & female_1 & male_2 & female_2 \end{array} \\ \begin{pmatrix} p^2+\beta & p^2 & pq+\beta & pq \\ p^2 & p^2+\beta & pq & pq+\beta \\ pq+\beta & pq & q^2+\beta & q^2 \\ pq & pq+\beta & q^2 & q^2+\beta \end{pmatrix} \end{array}. \tag{25}$$

- $h(p^2 + \beta)$ = prob of two males in block 1 form a link
- $h(q^2 + \beta)$ = prob of two males in block 2 form a link, etc

Algorithm

We have a 4−block SBM ⇒ ∃ GRDPG with latent positions $\boldsymbol{Y}$

$$\boldsymbol{P} = \boldsymbol{Y}\boldsymbol{I}_{d_1,d_2}\boldsymbol{Y}^T \tag{26}$$

1 Perform ASE on adjacency $\boldsymbol{A}$

$$ASE(\boldsymbol{A}) = \boldsymbol{U_A}\boldsymbol{S_A}\boldsymbol{U_A}^T \Rightarrow \widehat{\boldsymbol{Y}} = \boldsymbol{U_A}|\boldsymbol{S_A}|^{1/2} \tag{27}$$

2 Cluster the estimated latent positions $\widehat{\boldsymbol{Y}}$ using Gaussian Mixture Model (GMM) or K-Means clustering. Obtain cluster centers

$$\boldsymbol{\mu} = [\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3, \boldsymbol{\mu}_4] \tag{28}$$

Algorithm

3 We compute an estimate for $\boldsymbol{\theta}_Z$ as

$$\widehat{\boldsymbol{\theta}}_Z = \begin{array}{c} male_1 \\ female_1 \\ male_2 \\ female_2 \end{array} \overset{\begin{array}{cccc} male_1 & female_1 & male_2 & female_2 \end{array}}{\begin{pmatrix} \widehat{\boldsymbol{\mu}}_1^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_1 & \widehat{\boldsymbol{\mu}}_1^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_2 & \widehat{\boldsymbol{\mu}}_1^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_3 & \widehat{\boldsymbol{\mu}}_1^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_4 \\ \widehat{\boldsymbol{\mu}}_2^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_1 & \widehat{\boldsymbol{\mu}}_2^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_2 & \widehat{\boldsymbol{\mu}}_2^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_3 & \widehat{\boldsymbol{\mu}}_2^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_4 \\ \widehat{\boldsymbol{\mu}}_3^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_1 & \widehat{\boldsymbol{\mu}}_3^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_2 & \widehat{\boldsymbol{\mu}}_3^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_3 & \widehat{\boldsymbol{\mu}}_3^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_4 \\ \widehat{\boldsymbol{\mu}}_4^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_1 & \widehat{\boldsymbol{\mu}}_4^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_2 & \widehat{\boldsymbol{\mu}}_4^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_3 & \widehat{\boldsymbol{\mu}}_4^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_4 \end{pmatrix}}.$$

Compare to theoretical probability matrix

$$\boldsymbol{\theta}_Z = \begin{array}{c} male_1 \\ female_1 \\ male_2 \\ female_2 \end{array} \overset{\begin{array}{cccc} male_1 & female_1 & male_2 & female_2 \end{array}}{\begin{pmatrix} h(p^2+\beta) & h(p^2) & h(pq+\beta) & h(pq) \\ h(p^2) & h(p^2+\beta) & h(pq) & h(pq+\beta) \\ h(pq+\beta) & h(pq) & h(q^2+\beta) & h(q^2) \\ h(pq) & h(pq+\beta) & h(q^2) & h(q^2+\beta) \end{pmatrix}}.$$

and estimate the block assignments $\widehat{\boldsymbol{\tau}}$

Algorithm

4 Estimate β as

$$\widehat{\beta} = h^{-1}(\widehat{\boldsymbol{\mu}}_1^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_1) - h^{-1}(\widehat{\boldsymbol{\mu}}_1^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_2).$$

5 Estimate the latent positions (if you want)

$$\begin{array}{c} \\ male_1 \\ male_2 \end{array} \begin{array}{c} \begin{array}{cc} \quad female_1 & \qquad\qquad female_2 \end{array} \\ \begin{pmatrix} h^{-1}(\widehat{\boldsymbol{\mu}}_1^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_2) & h^{-1}(\widehat{\boldsymbol{\mu}}_1^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_4) \\ h^{-1}(\widehat{\boldsymbol{\mu}}_3^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_2) & h^{-1}(\widehat{\boldsymbol{\mu}}_3^T \boldsymbol{I}_{\hat{d}_1,\hat{d}_2} \widehat{\boldsymbol{\mu}}_4) \end{pmatrix} \end{array} = \begin{pmatrix} \widehat{p^2} & \widehat{pq} \\ \widehat{pq} & \widehat{q^2} \end{pmatrix}.$$

Asymptotics

Our stochastic blockmodel with covariates is

$$\tau_i \overset{iid}{\sim} Multinomial(1; \pi_1, \ldots, \pi_K), \tag{29}$$

$$\boldsymbol{Z}_i | \tau_i \overset{ind}{\sim} Bernoulli(b_{\tau_i}), \tag{30}$$

$$\boldsymbol{A}_{ij} | \tau_i, \tau_j, \boldsymbol{Z}_i, \boldsymbol{Z}_j \overset{ind}{\sim} Bernoulli(\boldsymbol{P}_{ij}), \tag{31}$$

$$\boldsymbol{P}_{ij} = h\left(\boldsymbol{B}_{\tau_i \tau_j} + \beta \mathbf{1}_{\{\boldsymbol{Z}_i = \boldsymbol{Z}_j\}}\right). \tag{32}$$

and we rewrite it as a GRDPG

$$\boldsymbol{X}_i \overset{iid}{\sim} \pi_1 \delta_{\nu_1} + \pi_2 \delta_{\nu_2} + \cdots + \pi_K \delta_{\nu_K}, \tag{33}$$

$$\boldsymbol{Z}_i | \boldsymbol{X}_i \overset{ind}{\sim} Bernoulli(b(\boldsymbol{X}_i)), \tag{34}$$

$$\boldsymbol{A}_{ij} | \boldsymbol{X}_i, \boldsymbol{X}_j, \boldsymbol{Z}_i, \boldsymbol{Z}_j \overset{ind}{\sim} Bernoulli(\boldsymbol{P}_{ij}), \tag{35}$$

$$\boldsymbol{P}_{ij} = h\left(\boldsymbol{X}_i^T \boldsymbol{X}_j + \beta \mathbf{1}_{\{\boldsymbol{Z}_i = \boldsymbol{Z}_j\}}\right). \tag{36}$$

CLT for $\widehat{\beta}$

Central limit theorem for β
Let $\widehat{\boldsymbol{\tau}} : [n] \to [K]$ be the estimated function that assigns nodes to blocks, estimated using GMM clustering of $\widehat{\boldsymbol{Y}} = \widehat{\boldsymbol{U}}|\widehat{\boldsymbol{S}}|^{1/2}$'s rows. Let function g be the inverse of h. Let $g'(\boldsymbol{\nu}_1^T \boldsymbol{\nu}_1 + \beta) \neq 0$ and $g'(\boldsymbol{\nu}_1^T \boldsymbol{\nu}_2) \neq 0$. Then there exists a sequence of permutations $\phi \equiv \phi_n$ on $[K]$ such that the estimator $\widehat{\beta} = h^{-1}(\widehat{\boldsymbol{\theta}}_{Z,\phi(1)\phi(1)}) - h^{-1}(\widehat{\boldsymbol{\theta}}_{Z,\phi(1)\phi(2)})$ is asymptotically normal, that is as $n \to \infty$

$$n\left(\widehat{\beta} - \beta - \frac{\widehat{\psi}_\beta}{n}\right) \xrightarrow{d} N(0, \widehat{\sigma}^2_\beta) \tag{37}$$

CLT for $\widehat{\beta}$ with semi-sparse networks

Central limit theorem for semi-sparse networks
Let the SBM model include a sparsity coefficient ρ_n

$$\boldsymbol{P}_{ij} = \rho_n h\left(\boldsymbol{X}_i^T \boldsymbol{X}_j + \beta \mathbf{1}_{\{\boldsymbol{Z}_i = \boldsymbol{Z}_j\}}\right) \tag{38}$$

such that $\rho_n \to 0$ and $n\rho_n = \omega(\sqrt{n})$ as $n \to \infty$. Let $\hat{\boldsymbol{\tau}}$ be assignment of each node to a block, estimated using ASE and GMM (or K-means) clustering. Then there exists a sequence of permutations $\phi \equiv \phi_n$ on $[K]$ such that the estimator $\widehat{\beta} = h^{-1}(\widehat{\boldsymbol{\theta}}_{Z,\phi(1)\phi(1)}) - h^{-1}(\widehat{\boldsymbol{\theta}}_{Z,\phi(1)\phi(2)})$ is asymptotically normal, that is as $n \to \infty$

$$n\rho_n^{1/2}\left(\hat{\beta} - \beta - \frac{\ddot{\psi}_\beta}{n\rho_n}\right) \xrightarrow{d} N\left(0, \ddot{\sigma}_\beta^2\right) \tag{39}$$

Computational Advantage

Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	Time (s)	ARI
GRDPG	2000	2	0.1	0.09993	0.7	0.70065	1.513	1

Computational Advantage

Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	Time (s)	ARI
GRDPG	2000	2	0.1	0.09993	0.7	0.70065	1.513	1
VEM	2000	2	0.1	0.10008	0.7	0.70061	39.679	1

Computational Advantage

Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	Time (s)	ARI
GRDPG	2000	2	0.1	0.09993	0.7	0.70065	1.513	1
VEM	2000	2	0.1	0.10008	0.7	0.70061	39.679	1
GRDPG	5000	2	0.1	0.10004	0.7	0.69977	8.548	1

Computational Advantage

Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	Time (s)	ARI
GRDPG	2000	2	0.1	0.09993	0.7	0.70065	1.513	1
VEM	2000	2	0.1	0.10008	0.7	0.70061	39.679	1
GRDPG	5000	2	0.1	0.10004	0.7	0.69977	8.548	1
VEM	5000	2	0.1	0.10008	0.7	0.69975	593.203	1

Computational Advantage

Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	Time (s)	ARI
GRDPG	2000	2	0.1	0.09993	0.7	0.70065	1.513	1
VEM	2000	2	0.1	0.10008	0.7	0.70061	39.679	1
GRDPG	5000	2	0.1	0.10004	0.7	0.69977	8.548	1
VEM	5000	2	0.1	0.10008	0.7	0.69975	593.203	1
GRDPG	10000	2	0.1	0.09994	0.7	0.69988	32.169	1

Computational Advantage

Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	Time (s)	ARI
GRDPG	2000	2	0.1	0.09993	0.7	0.70065	1.513	1
VEM	2000	2	0.1	0.10008	0.7	0.70061	39.679	1
GRDPG	5000	2	0.1	0.10004	0.7	0.69977	8.548	1
VEM	5000	2	0.1	0.10008	0.7	0.69975	593.203	1
GRDPG	10000	2	0.1	0.09994	0.7	0.69988	32.169	1
VEM	10000	2	0.1	0.09996	0.7	0.69987	4171.218	1

Computational Advantage

Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	Time (s)	ARI
GRDPG	2000	2	0.1	0.09993	0.7	0.70065	1.513	1
VEM	2000	2	0.1	0.10008	0.7	0.70061	39.679	1
GRDPG	5000	2	0.1	0.10004	0.7	0.69977	8.548	1
VEM	5000	2	0.1	0.10008	0.7	0.69975	593.203	1
GRDPG	10000	2	0.1	0.09994	0.7	0.69988	32.169	1
VEM	10000	2	0.1	0.09996	0.7	0.69987	4171.218	1
GRDPG	20000	2	0.1	0.09998	0.7	0.70005	128.633	1
VEM	20000	2	NA					
GRDPG	30000	2	0.1	0.09998	0.7	0.69995	386.210	1

Computational Advantage

Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	β	$\hat{\beta}$	Time (s)	ARI
GRDPG	2000	2	-1.5	-1.49744	1	1.00077	no covariates		4.672	1
VEM	2000	2	-1.5	-1.49712	1	1.00067	no covariates		48.619	1

Computational Advantage

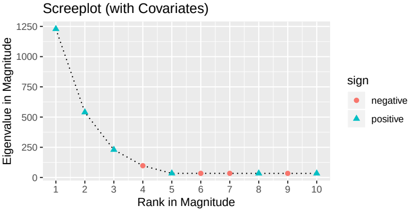
Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	β	$\hat{\beta}$	Time (s)	ARI
GRDPG	2000	2	-1.5	-1.49744	1	1.00077	no covariates		4.672	1
VEM	2000	2	-1.5	-1.49712	1	1.00067	no covariates		48.619	1
GRDPG	2000	2	-1.5	-1.49454	1	0.99926	1.5	1.51201	7.557	1
VEM	2000	2	-1.5	-1.49712	1	1.00067	1.5	1.50335	6903.673	1

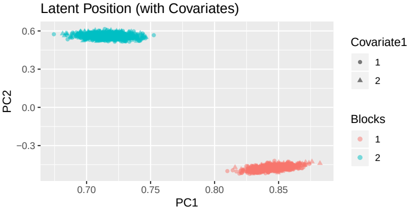
Computational Advantage

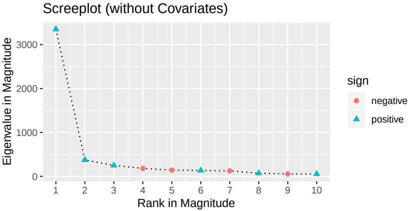
Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	β	$\hat{\beta}$	Time (s)	ARI
GRDPG	2000	2	-1.5	-1.49744	1	1.00077	no covariates		4.672	1
VEM	2000	2	-1.5	-1.49712	1	1.00067	no covariates		48.619	1
GRDPG	2000	2	-1.5	-1.49454	1	0.99926	1.5	1.51201	7.557	1
VEM	2000	2	-1.5	-1.49712	1	1.00067	1.5	1.50335	6903.673	1
GRDPG	5000	2	-1.5	-1.50029	1	1.00030	no covariates		17.539	1
VEM	5000	2	-1.5	-1.50019	1	1.00024	no covariates		537.831	1
GRDPG	5000	2	-1.5	-1.49995	1	1.00064	1.5	1.49981	27.312	1
VEM	5000	2	-1.5	-1.50019	1	1.00024	1.5	1.49955	35331.012	1

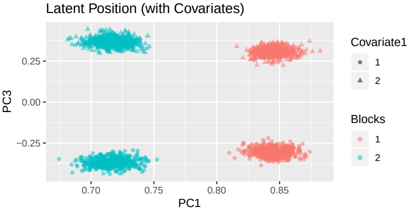
Computational Advantage

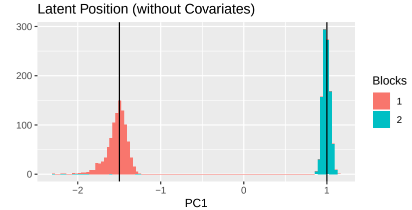
Estimator	n	K	p	$\hat{p}$	q	$\hat{q}$	β	$\hat{\beta}$	Time (s)	ARI
GRDPG	2000	2	-1.5	-1.49744	1	1.00077	no covariates		4.672	1
VEM	2000	2	-1.5	-1.49712	1	1.00067	no covariates		48.619	1
GRDPG	2000	2	-1.5	-1.49454	1	0.99926	1.5	1.51201	7.557	1
VEM	2000	2	-1.5	-1.49712	1	1.00067	1.5	1.50335	6903.673	1
GRDPG	5000	2	-1.5	-1.50029	1	1.00030	no covariates		17.539	1
VEM	5000	2	-1.5	-1.50019	1	1.00024	no covariates		537.831	1
GRDPG	5000	2	-1.5	-1.49995	1	1.00064	1.5	1.49981	27.312	1
VEM	5000	2	-1.5	-1.50019	1	1.00024	1.5	1.49955	35331.012	1
GRDPG	10000	2	-1.5	-1.49989	1	1.00029	no covariates		55.428	1
GRDPG	10000	2	-1.5	1.49992	1	0.99992	1.5	1.50190	91.067	1

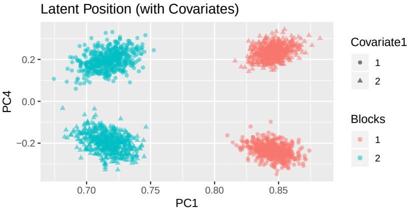

Screeplot (with Covariates)
Eigenvalue in Magnitude
1250
1000
750
500
250
0
1 2 3 4 5 6 7 8 9 10
Rank in Magnitude
sign
negative
positive


Latent Position (with Covariates)
PC2
0.6
0.3
0.0
−0.3
0.70 0.75 0.80 0.85
PC1
Covariate1
1
2
Blocks
1
2


Screeplot (without Covariates)
Eigenvalue in Magnitude
3000
2000
1000
0
1 2 3 4 5 6 7 8 9 10
Rank in Magnitude
sign
negative
positive


Latent Position (with Covariates)
PC3
0.25
0.00
−0.25
0.70 0.75 0.80 0.85
PC1
Covariate1
1
2
Blocks
1
2


Latent Position (without Covariates)
300
200
100
0
−2 −1 0 1
PC1
Blocks
1
2


Latent Position (with Covariates)
PC4
0.2
0.0
−0.2
0.70 0.75 0.80 0.85
PC1
Covariate1
1
2
Blocks
1
2

Next steps...

1. Estimate GRDPG with binary (discrete) observed covariates
2. Estimate GRDPG with continuous covariates (harder)
3. Use GRDPGs to approximate structural models of network formation
4. Extend to directed and bipartite graphs (non-trivial)
5. Apply GRDPGs to correct for endogeneity of (large) network in social interactions models (Johnson-Moon 2018, Shalizi-MacFowland 2018, Imbens et al 2013, and many others)