Presentation

Carson James

July 11, 2022

Definition

We define
$\Lambda_+^{n\times r} = \{\Sigma \in \mathbb{R}^{n\times r} : \Sigma \text{ is diagonal and positive semi-definite}\}$ and
$O_n = \{U \in \mathbb{R}^{n\times n} : U \text{ is orthogonal}\}$.

Model 1:

1. Fix $M \in \mathbb{R}^{n \times n_M}$columns of M are orthogonal and set $P_M = M(M^T M)^{-1} M^T$?
2. Choose $\Sigma_Z \in \Lambda_+^{n_M \times r}$, $\Sigma_X \in \Lambda_+^{n \times p}$, $U_Z \in O_r$ and $U_X \in O_p$.
3. Set $V_Z^T = \Sigma_Z U_Z$ and $V_X^T = \Sigma_X U_X$.
4. Set $J_Z = M V_Z^T$ and $J_X = M V_X^T$.
5. Choose $I_Z \in \mathbb{R}^{n \times r}$, $I_X \in \mathbb{R}^{n \times p}$ such that $\mathcal{C}(I_Z) \cup \mathcal{C}(I_X) \subset \mathcal{C}(I - P_M)$.
6. Choose $E_X \in \mathbb{R}_{n \times p}$ with $(E_X)_{i,j} \sim N(0, \sigma^2)$
7. Set $Z = J_Z + I_Z$ and $X = J_X + I_X + E_X$

Then $\mathcal{C}(M) \perp \mathcal{C}(I_Z), \mathcal{C}(I_X)$.

Model 2: We consider a modification of the planted partition model which is a submodel of the stochastic block model with n nodes and r blocks (for now $r = 2$).

- Choose $U \in \mathbb{R}^{n \times 2}$ such that for each $i \in \{1, \dots, n\}$,

$$U_{i,j} = \begin{cases} 1 & \text{node } i \text{ is in block } j \\ 0 & \text{else} \end{cases}$$

  as in the stochastic block model. We choose

$$U = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

- Choose $W \in \mathbb{R}^{n\times 1}$, with $W = (1, -1, \ldots, 1, -1)^T$ and $\mathcal{C}(W) \perp \mathcal{C}(U)$.
- Choose $0 < b < a < 1$ and set

$$Q = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$$

  to be the block probability matrix from the planted partition model.
- Set $B = Q^{1/2}$.
- Choose $\alpha \in (0, 1)$ and set $Z = (1 - \alpha)(0, W) + \alpha(UB, 0)$
- Set $X = (0, W)$

Then

- $U^T W = 0$
- $BB^T \in [0,1]^{2\times 2}$ and $(UB)(UB)^T \in [0,1]^{n\times n}$
- if α is close enough to 1, then $Z \in [0,1]^{n\times n}$
- here $J_Z = J_X = (0, W)$ and $I_Z = (UB, 0)$, $I_X = (0, 0)$.

Model 3: We consider another modification of the planted partition model which is a submodel of the stochastic block model with n nodes and r blocks (for now $r = 2$).

- Choose $U \in \mathbb{R}^{n\times 2}$ such that for each $i \in \{1, \dots, n\}$,

$$U_{i,j} = \begin{cases} 1 & \text{node } i \text{ is in block } j \\ 0 & \text{else} \end{cases}$$

  as in the stochastic block model. We choose

$$U = \begin{pmatrix} 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \\ 0 & 1 \\ \vdots & \vdots \\ 0 & 1 \end{pmatrix}$$

- Choose $W \in \mathbb{R}^{n \times 1}$, with $W = (1, -1, \ldots, 1, -1)^T$ and $\mathcal{C}(W) \perp \mathcal{C}(U)$.
- Choose $0 < b < a < 1$ and set
$$Q = \begin{pmatrix} a & b \\ b & a \end{pmatrix}$$
to be the block probability matrix from the planted partition model.
- Set $B = Q^{1/2}$.
- Choose $K^T \in O(p)$ and $I_X \in \mathbb{R}^{n \times p}$ such that $\mathcal{C}(I_X) \perp \mathcal{C}(W)$.
- Choose $E_X \in \mathbb{R}^{n \times p}$ with $(E_X)_{i,j} \sim N(0, \sigma^2)$
- Define $J_Z = W \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}$, $I_Z = UB$, $J_X = \begin{pmatrix} W & 0 \end{pmatrix} K^T$
- Choose $\alpha \in (0, 1)$ and set $Z = (1 - \alpha) J_Z + \alpha I_Z$
- Set $X = J_X + I_X + E_X$

Then

- $U^T W = 0$
- $BB^T \in [0,1]^{2\times 2}$ and $(UB)(UB)^T \in [0,1]^{n\times n}$
- if α is close enough to 1, then $Z \in [0,1]^{n\times n}$

Analysis of Initial Solution:

Consider the following model for the data:

$$A_{ij} \sim \mathrm{Ber}(ZZ^T)_{ij}), \quad i > j, i, j \in [n],$$

$$X_{uv} = (W)_{uv} + \epsilon_{uv}, \quad u \in [n], v \in [p].$$

Assume $\epsilon_{uv} \overset{iid}{\sim} N(0, \sigma^2)$. Write

$$Z = [M, R_Z]\Gamma,$$

$$W = [M, R_W]U^T,$$

where $M \in \mathbb{R}^{n \times r_M}$, $R_Z \in \mathbb{R}^{n \times r_Z}$ and $R_W \in \mathbb{R}^{n \times r_w}$ are matrices with orthogonal columns, and $\Gamma \in \mathbb{R}^{(r_M + r_Z) \times (r_M + r_Z)}$ and $U \in \mathbb{R}^{p \times (r_M + r_W)}$ some other matrices.

Define $\hat{V}^{(1)}$ as the matrix of $(r_M + r_Z)$ leading eigenvectors of A, and $\hat{V}^{(2)}$ as the matrix of $(r_M + r_W)$ left leading singular vectors of X. Then define $\hat{M}$ as the matrix of r_M left leading singular vectors of $[\hat{V}^{(1)}, \hat{V}^{(2)}]$. Set

$$\epsilon = \sqrt{\frac{1}{2}\left(\frac{\delta(ZZ^T)}{\lambda_{\min}^2(\Gamma\Gamma^T)} + \frac{n}{\lambda_{\min}^2(U^T U)}\right)}$$

Conjecture

With the assumptions as above, and some regularity conditions (TBD) (maybe if $\epsilon = o(1)$ as $n \to \infty$), there exists some orthogonal matrix U such that

$$\mathbb{E}\|\hat{M} - MU\|_F = O\left(\frac{r_M^{1/2}}{\sqrt{n}}\right).$$

idea

$$U^{(1)} = A \qquad U^{(2)} = X$$

$$\hat{\Pi}^{(i)} = \hat{V}^{(i)}(\hat{V}^{(i)})^T \qquad \hat{\Pi} = \frac{1}{2}(\hat{\Pi}^{(1)} + \hat{\Pi}^{(2)})$$

$$\tilde{\Pi}^{(i)} = \mathbb{E}\hat{\Pi}^{(i)} \qquad \tilde{\Pi} = \frac{1}{2}(\tilde{\Pi}^{(1)} + \tilde{\Pi}^{(2)})$$

$$\Pi = MM^T$$

Define $\tilde{M}$ to be the matrix consisting of the r_M left leading singular vectors of $\tilde{\Pi}$. Then

$$\begin{aligned}\min_{W \in O_{r_M}} \|\hat{M} - MW\|_F &\leq \|\hat{M}\hat{M}^T - MM^T\|_F \\ &\leq \|\hat{M}\hat{M}^T - \tilde{M}\tilde{M}^T\|_F + \|\tilde{M}\tilde{M}^T - MM^T\|_F\end{aligned}$$

Then we control both of these errors.

To control the first error, we need the following lemma:

Lemma

Let $X \in \mathbb{R}^{n\times p}$ with $X_{i,j} \sim N(0,\sigma^2)$ and $a > 1$. Set $C_{n,p} = \frac{a}{a-1}\frac{3}{2}\left[(\sqrt{n}+\sqrt{p}) + \frac{5}{\log(3/2)}\sqrt{\log(n \wedge p)}\right]\sigma$. Then for each $t \geq C_{n,p}$,

$$P(\|X\| \geq t) \leq \exp\left(-\frac{t^2}{2(a\sigma)^2}\right)$$

and for each $q \geq 1$,

$$\mathbb{E}(\|X\|^q) = O(\sigma^q(\sqrt{n}+\sqrt{p})^q)$$

Focusing on the first error, the Davis-Kahan theorem for rectangular matrices tells us that

$$\begin{aligned}\|\hat{M}\hat{M}^T - \tilde{M}\tilde{M}^T\|_F &\leq \frac{2^{3/2}(2\sigma_1(\tilde{\Pi}) + \|\hat{\Pi} - \tilde{\Pi}\|_{op})\|\hat{\Pi} - \tilde{\Pi}\|_F}{\sigma_{r_M}(\tilde{\Pi})^2 - \sigma_{r_M+1}(\tilde{\Pi})^2} \\ &\leq \frac{2^{3/2}(2\sigma_1(\tilde{\Pi}) + \|\hat{\Pi} - \tilde{\Pi}\|_F)\|\hat{\Pi} - \tilde{\Pi}\|_F}{\sigma_{r_M}(\tilde{\Pi})^2 - \sigma_{r_M+1}(\tilde{\Pi})^2}\end{aligned}$$

To bound $\|\hat{\Pi} - \tilde{\Pi}\|_F$, we note that the triangle inequality implies that

$$\begin{aligned}
\|\hat{\Pi} - \tilde{\Pi}\|_F &= \frac{1}{2}\left\| \sum_{i=1}^{2} \hat{\Pi}^{(i)} - \tilde{\Pi}^{(i)} \right\|_F \\
&\leq \frac{1}{2} \sum_{i=1}^{2} \|\hat{\Pi}^{(i)} - \tilde{\Pi}^{(i)}\|_F
\end{aligned}$$

Since $\|\cdot\|_1 \leq \|\cdot\|_p$ on a probability space, we may bound

$$\begin{aligned}
\mathbb{E}\Big(\|\hat{\Pi}^{(i)} - \tilde{\Pi}^{(i)}\|_F^q\Big)^{1/q} &\leq \mathbb{E}\Big(\|\hat{\Pi}^{(i)} - \Pi\|_F^q\Big)^{1/q} + \mathbb{E}\Big(\|\Pi - \tilde{\Pi}^{(i)}\|_F^q\Big)^{1/q} \\
&= \mathbb{E}\Big(\|\hat{\Pi}^{(i)} - \Pi\|_F^q\Big)^{1/q} + \|\mathbb{E}(\Pi - \hat{\Pi}^{(i)})\|_F \\
&\leq \mathbb{E}\Big(\|\hat{\Pi}^{(i)} - \Pi\|_F^q\Big)^{1/q} + \mathbb{E}\|(\Pi - \hat{\Pi}^{(i)})\|_F \\
&\leq \mathbb{E}\Big(\|\hat{\Pi}^{(i)} - \Pi\|_F^q\Big)^{1/q} + \mathbb{E}\Big(\|\Pi - \hat{\Pi}^{(i)}\|_F^q\Big)^{1/q} \\
&= 2\mathbb{E}\Big(\|\hat{\Pi}^{(i)} - \Pi\|_F^q\Big)^{1/q}
\end{aligned}$$

To bound $\|\hat{\Pi}^{(i)} - \Pi\|_F^q$, we again apply the Davis-Kahan theorem to obtain

$$\|\hat{\Pi}^{(i)} - \Pi\|_F \leq \frac{2^{3/2}(2\sigma_1(\Pi) + \|U^{(i)} - \mathbb{E}U^{(i)}\|_{op}) r_M^{1/2} \|U^{(i)} - \mathbb{E}U^{(i)}\|_{op}}{\sigma_{r_M}(\mathbb{E}U^{(i)})^2 - \sigma_{r_M+1}(\mathbb{E}U^{(i)})^2} \tag{1}$$

Set $c_i = \frac{2^{3/2} r_M^{1/2}}{\sigma_{r_M}(\mathbb{E}U^{(i)})^2 - \sigma_{r_M+1}(\mathbb{E}U^{(i)})^2}$. The previous lemma then implies that

$$\begin{aligned}
\mathbb{E}\|\hat{\Pi}^{(2)} - \Pi\|_F^q &\leq c_2^q \mathbb{E}\left[(2\sigma_1(\Pi) + \|U^{(2)} - \mathbb{E}U^{(2)}\|_{op})^q \|U^{(2)} - \mathbb{E}U^{(2)}\|_{op}^q\right] \\
&= c_2^q \mathbb{E}\Big[\text{Binomial}(q, 2\sigma_1(\Pi), \|U^{(2)} - \mathbb{E}U^{(2)}\|_{op}^q)\|U^{(2)} - \mathbb{E}U \\
&= c_2^q O(\sigma^{2q}(\sqrt{n} + \sqrt{r_M})^{2q})
\end{aligned}$$

and

$$(\mathbb{E}\|\hat{\Pi}^{(2)} - \Pi\|_F^q)^{1/q} = c_2 O(\sigma^2(\sqrt{n} + \sqrt{r_M})^2)$$

Now we need to use a lemma from the paper *Distributed estimation of principal eigenspaces* (Fan et al) to combine these bounds as well as to be used in bounding the second error term $\|\tilde{M}\tilde{M}^T - MM^T\|_F$

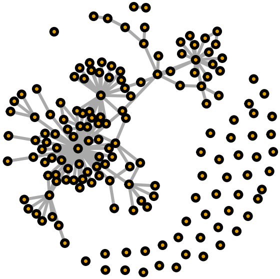
Data

Data was obtained from

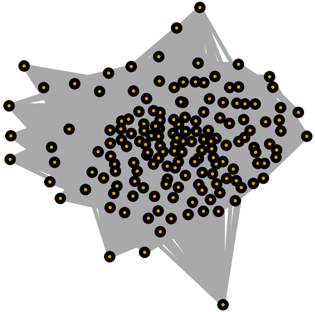
- World Bank API via the `wbstats` package.
- Github: lukes/ISO-3166
- Kaggle: Trade Network (I think I need to pull the data myself from the UN comtrade database using the `comtradr` package)

GDP	regionAfrica	regionAmericas
3127.891	1	0
3952.803	0	0

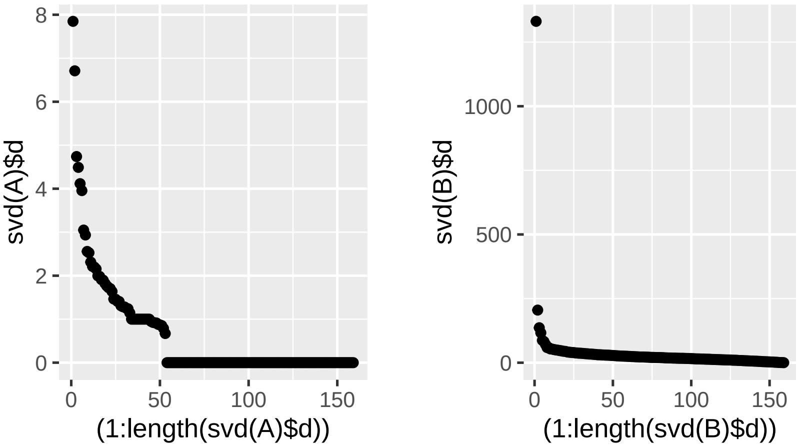
The first network, with adjacency matrix A, is defined so that two countries have a connection if the at least one of the countries receives at least a certain percentage of the other country's exports. Here the threshold is 0.15



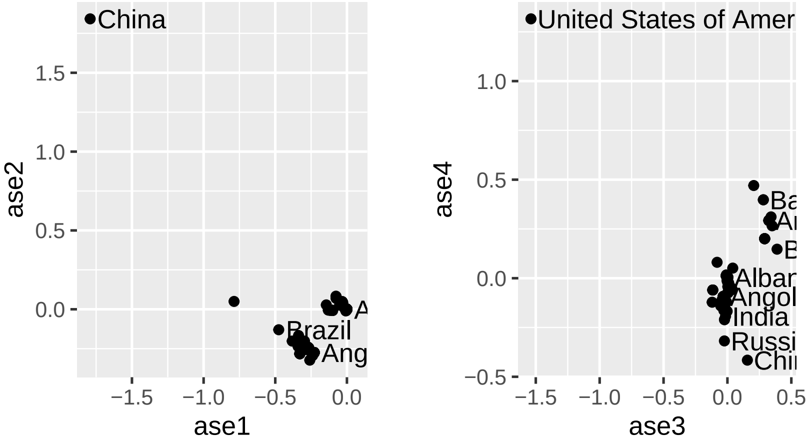
The second network, with weighted adjacency matrix B, is defined so that two countries have a connection if they had some nonnegligible trade. The weight of a connection is the log of the total trade between the two countries.



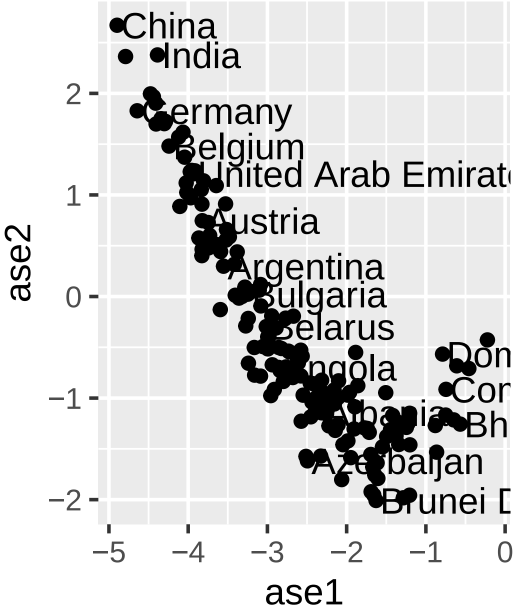
Scree plots for A and B:



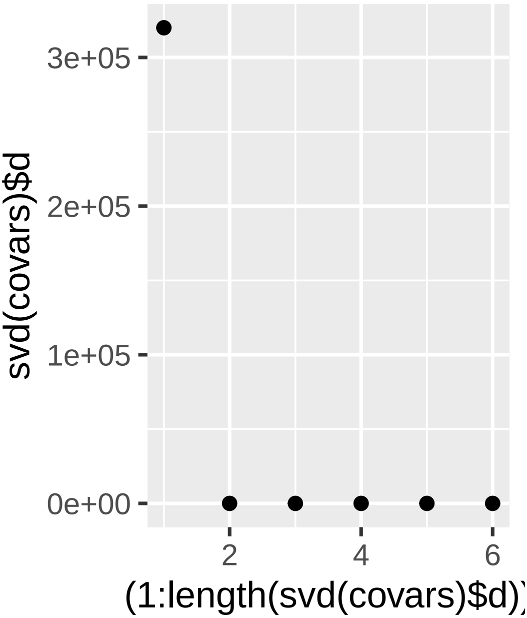
Some ASE plots for A:



ASE plots for B:



Scree plot for the covariates:



I added a snippet of code to Dongbangs algorithm to let me initialize M as the COSIE estimate,

$$\hat{M} = r_M\text{-lsv}(\hat{V}^{(1)}, \hat{V}^{(2)})$$

where

- $\hat{V}^{(1)} = (r_M + r_Z)\text{-lsv}(A)$
- $\hat{V}^{(2)} = (r_M + r_W)\text{-lsv}(X)$

Initializing Dongbangs algorithm for A by averaging principal vectors and by the COSIE estimate yields a loss of 7718508 and 221727943 respectively.
Initializing the algorithm for B by averaging principal vectors and by the COSIE estimate yields a loss of 140245606 and 1743031541 respectively. Maybe I made a mistake in my code since we were expecting the COSIE estimate to yield a better starting estimate.