Learning Joint and Individual Structure in Network Data with Covariates

Carson James

April 15, 2024

Joint work with:

(a) Dongbang Yuan (Meta)

(b) Irina Gaynanova (UMich)

(c) Jesús Arroyo (Texas A&M)

General Goal

Given multiple datasets, we want to

General Goal

Given multiple datasets, we want to

- isolate information shared across the datasets,

General Goal

Given multiple datasets, we want to

- isolate information shared across the datasets,
- isolate information unique to each dataset,

General Goal

Given multiple datasets, we want to

- isolate information shared across the datasets,
- isolate information unique to each dataset,
- use the above info to better understand the datasets (community structure, node influence, etc).

General Goal

Given multiple datasets, we want to

- isolate information shared across the datasets,
- isolate information unique to each dataset,
- use the above info to better understand the datasets (community structure, node influence, etc).

Here, we specifically observe a network with n nodes in terms of the following data:

General Goal

Given multiple datasets, we want to

- isolate information shared across the datasets,
- isolate information unique to each dataset,
- use the above info to better understand the datasets (community structure, node influence, etc).

Here, we specifically observe a network with n nodes in terms of the following data:

- (connectivity data) adjacency matrix $A \in \mathbb{R}^{n \times n}$ where A_{ij} is the connection strength between nodes i and j,

General Goal

Given multiple datasets, we want to

- isolate information shared across the datasets,
- isolate information unique to each dataset,
- use the above info to better understand the datasets (community structure, node influence, etc).

Here, we specifically observe a network with n nodes in terms of the following data:

- (connectivity data) adjacency matrix $A \in \mathbb{R}^{n \times n}$ where A_{ij} is the connection strength between nodes i and j,
- (covariate data) node covariates $X \in \mathbb{R}^{n \times p}$ so that row i of X are the p covariates observed at row i.

General Goal

Given multiple datasets, we want to

- isolate information shared across the datasets,
- isolate information unique to each dataset,
- use the above info to better understand the datasets (community structure, node influence, etc).

Here, we specifically observe a network with n nodes in terms of the following data:

- (connectivity data) adjacency matrix $A \in \mathbb{R}^{n \times n}$ where A_{ij} is the connection strength between nodes i and j,
- (covariate data) node covariates $X \in \mathbb{R}^{n \times p}$ so that row i of X are the p covariates observed at row i.

The adjacency matrix and node covariates contain information about which nodes are important, if nodes form groups, etc.

Real data example:

Real data example:

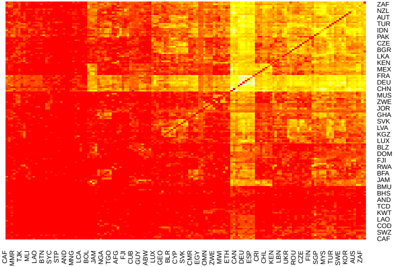
- total food commodity trade volumes between 146 countries

Real data example:

- total food commodity trade volumes between 146 countries
- 10 economic and geographic covariates at each country including GDP, education expenditure, region, etc.

Real data example:

- total food commodity trade volumes between 146 countries
- 10 economic and geographic covariates at each country including GDP, education expenditure, region, etc.

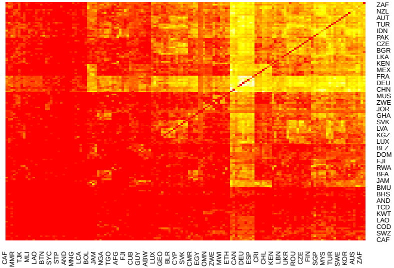


(a) trade adjacency matrix

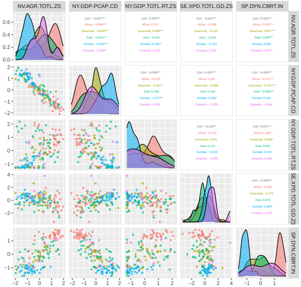
(b) node covariates pairs plot

Real data example:

- total food commodity trade volumes between 146 countries
- 10 economic and geographic covariates at each country including GDP, education expenditure, region, etc.



(a) trade adjacency matrix

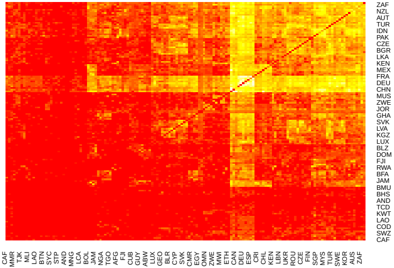


(b) node covariates pairs plot

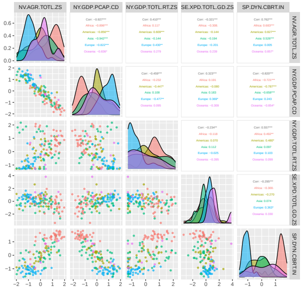
Questions:

Real data example:

- total food commodity trade volumes between 146 countries
- 10 economic and geographic covariates at each country including GDP, education expenditure, region, etc.



(a) trade adjacency matrix



(b) node covariates pairs plot

Questions: What info about the nodes can we extract from both datasets? What can each dataset tell us about the other?

Modeling joint and individual structure

Related work:

- Joint and individual for covariate data: JIVE (Lock et al., 2013), AJIVE (Feng et al., 2018), DMMD (Yuan and Gaynanova, 2022)
- Joint and individual for network data: MASE (Arroyo et al., 2021), MultiNeSS (MacDonald et al., 2022)
- Network and covariate data: CASC (Binkiewicz et al., 2017)

Modeling joint and individual structure

Model:

- Signal + noise model

$$A = P + E^A, \qquad X = W + E^X$$

- $\mathbb{E}(E^A) = 0, \quad \mathbb{E}(E^X) = 0$

Modeling joint and individual structure

Model:

- Signal + noise model

$$A = P + E^A, \qquad X = W + E^X$$

- $\mathbb{E}(E^A) = 0, \quad \mathbb{E}(E^X) = 0$

Assumption:

- data matrices are low rank for model nontrviality and feasible computation
- observed values are close to their means

Included models:

- stochastic block model
- random dot-product graphs,
- inhomogeneous bernoulli

Modeling joint and individual structure

- Given $P = \mathbb{E}(A)$ and $W = \mathbb{E}(X)$, define the joint and individual subspaces as

Modeling joint and individual structure

- Given $P = \mathbb{E}(A)$ and $W = \mathbb{E}(X)$, define the joint and individual subspaces as

Joint:	$\mathcal{M} = \mathcal{C}(P) \cap \mathcal{C}(W)$,
Network individual:	$\mathcal{R}^{(1)} = \mathcal{P}_{\mathcal{M}^\perp}\mathcal{C}(P)$,
Covariate individual:	$\mathcal{R}^{(2)} = \mathcal{P}_{\mathcal{M}^\perp}\mathcal{C}(W)$,

where $\mathcal{C}(\cdot)$ indicates the column space and $\mathcal{P}_{\mathcal{M}^\perp}$ is the orthogonal projection onto $\mathcal{M}^\perp$.

Modeling joint and individual structure

- Given $P = \mathbb{E}(A)$ and $W = \mathbb{E}(X)$, define the joint and individual subspaces as

Joint:	$\mathcal{M} = \mathcal{C}(P) \cap \mathcal{C}(W)$,
Network individual:	$\mathcal{R}^{(1)} = \mathcal{P}_{\mathcal{M}^\perp}\mathcal{C}(P)$,
Covariate individual:	$\mathcal{R}^{(2)} = \mathcal{P}_{\mathcal{M}^\perp}\mathcal{C}(W)$,

  where $\mathcal{C}(\cdot)$ indicates the column space and $\mathcal{P}_{\mathcal{M}^\perp}$ is the orthogonal projection onto $\mathcal{M}^\perp$.
- Set $r_M = \dim \mathcal{M}$, $r_k = \dim \mathcal{R}^{(k)}$. Then

$$\operatorname{rank}(P) = r_M + r_1, \qquad \operatorname{rank}(W) = r_M + r_2$$

Modeling joint and individual structure

Structure and identifiability:

Modeling joint and individual structure

Structure and identifiability:

- There exist matrices $M \in \mathbb{O}_{n,r_M}$ and $R^{(k)} \in \mathbb{O}_{n,r_k}$ such that

$$\mathcal{C}(M) = \mathcal{M}, \qquad \mathcal{C}(R^{(k)}) = \mathcal{R}^{(k)}$$

  and P and W factor as

$$P = \begin{pmatrix} M & R^{(1)} \end{pmatrix} \Gamma^{(1)} \qquad W = \begin{pmatrix} M & R^{(2)} \end{pmatrix} \Gamma^{(2)}$$

  where

Modeling joint and individual structure

Structure and identifiability:

- There exist matrices $M \in \mathbb{O}_{n,r_M}$ and $R^{(k)} \in \mathbb{O}_{n,r_k}$ such that

$$\mathcal{C}(M) = \mathcal{M}, \qquad \mathcal{C}(R^{(k)}) = \mathcal{R}^{(k)}$$

and P and W factor as

$$P = \begin{pmatrix} M & R^{(1)} \end{pmatrix} \Gamma^{(1)} \qquad W = \begin{pmatrix} M & R^{(2)} \end{pmatrix} \Gamma^{(2)}$$

where

- $\Gamma^{(k)}$ is full rank

Modeling joint and individual structure

Structure and identifiability:

- There exist matrices $M \in \mathbb{O}_{n,r_M}$ and $R^{(k)} \in \mathbb{O}_{n,r_k}$ such that

$$\mathcal{C}(M) = \mathcal{M}, \qquad \mathcal{C}(R^{(k)}) = \mathcal{R}^{(k)}$$

and P and W factor as

$$P = \begin{pmatrix} M & R^{(1)} \end{pmatrix} \Gamma^{(1)} \qquad W = \begin{pmatrix} M & R^{(2)} \end{pmatrix} \Gamma^{(2)}$$

where

- $\Gamma^{(k)}$ is full rank
- $M \perp R^{(k)}$ are orthogonal

Modeling joint and individual structure

Structure and identifiability:

- There exist matrices $M \in \mathbb{O}_{n,r_M}$ and $R^{(k)} \in \mathbb{O}_{n,r_k}$ such that

$$\mathcal{C}(M) = \mathcal{M}, \qquad \mathcal{C}(R^{(k)}) = \mathcal{R}^{(k)}$$

  and P and W factor as

$$P = \begin{pmatrix} M & R^{(1)} \end{pmatrix} \Gamma^{(1)} \qquad W = \begin{pmatrix} M & R^{(2)} \end{pmatrix} \Gamma^{(2)}$$

  where

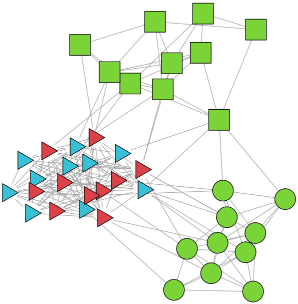
- $\Gamma^{(k)}$ is full rank
- $M \perp R^{(k)}$ are orthogonal
- These matrices are unique up to orthogonal transformation

Illustrative Example

Data: 40 nodes each belonging to one of 4 groups, at each node we observe 3 covariates.

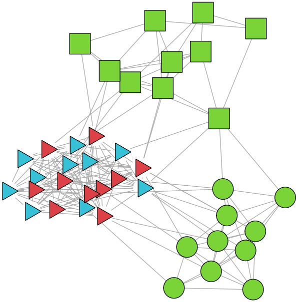
Illustrative Example

Data: 40 nodes each belonging to one of 4 groups, at each node we observe 3 covariates.



Illustrative Example

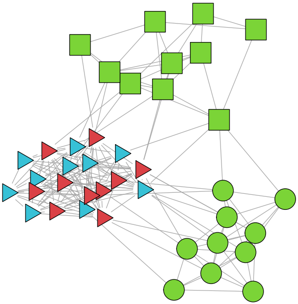
Data: 40 nodes each belonging to one of 4 groups, at each node we observe 3 covariates.



Network:

Illustrative Example

Data: 40 nodes each belonging to one of 4 groups, at each node we observe 3 covariates.

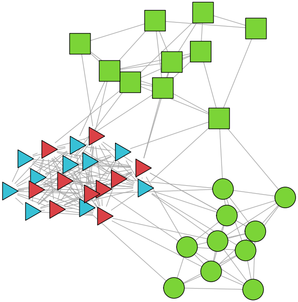


Network:

- 3 communities (shape).

Illustrative Example

Data: 40 nodes each belonging to one of 4 groups, at each node we observe 3 covariates.

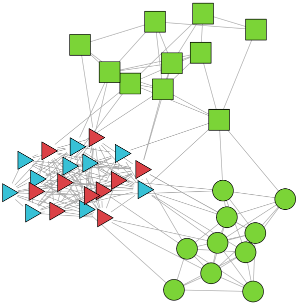


Network:

- 3 communities (shape).
- nodes in the same community are more connected.

Illustrative Example

Data: 40 nodes each belonging to one of 4 groups, at each node we observe 3 covariates.



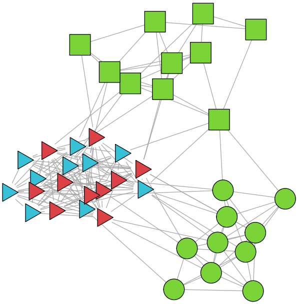
Network:

- 3 communities (shape).
- nodes in the same community are more connected.

Covariates:

Illustrative Example

Data: 40 nodes each belonging to one of 4 groups, at each node we observe 3 covariates.



Network:

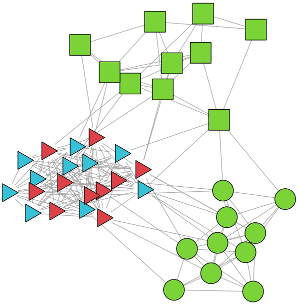
- 3 communities (shape).
- nodes in the same community are more connected.

Covariates:

- 3 clusters (color).

Illustrative Example

Data: 40 nodes each belonging to one of 4 groups, at each node we observe 3 covariates.



Network:

- 3 communities (shape).
- nodes in the same community are more connected.

Covariates:

- 3 clusters (color).
- nodes in the same cluster have similar covariates.

Estimating joint and individual structure (spectal method)

- Step 1: Extract singular subspaces

Estimating joint and individual structure (spectal method)

- Step 1: Extract singular subspaces

$$\widehat{V}^{(1)} = \mathrm{SV}(A, r_M + r_1), \qquad \text{(top left singular vectors)}$$
$$\widehat{V}^{(2)} = \mathrm{SV}(X, r_M + r_2), \qquad \text{(top left singular vectors)}$$

Estimating joint and individual structure (spectal method)

- Step 1: Extract singular subspaces

$$\widehat{V}^{(1)} = \mathrm{SV}(A, r_M + r_1), \qquad \text{(top left singular vectors)}$$
$$\widehat{V}^{(2)} = \mathrm{SV}(X, r_M + r_2), \qquad \text{(top left singular vectors)}$$

- Step 2: Extract joint singular subspace

Estimating joint and individual structure (spectal method)

- Step 1: Extract singular subspaces

$$\widehat{V}^{(1)} = \mathrm{SV}(A, r_M + r_1), \qquad \text{(top left singular vectors)}$$
$$\widehat{V}^{(2)} = \mathrm{SV}(X, r_M + r_2), \qquad \text{(top left singular vectors)}$$

- Step 2: Extract joint singular subspace

$$\widehat{U} = \begin{pmatrix} \widehat{V}^{(1)} & \widehat{V}^{(2)} \end{pmatrix}$$
$$\widehat{M} = \mathrm{SV}(\widehat{U}, r_M)$$

Estimating joint and individual structure (spectal method)

- Step 1: Extract singular subspaces

$$\widehat{V}^{(1)} = \mathrm{SV}(A, r_M + r_1), \qquad \text{(top left singular vectors)}$$
$$\widehat{V}^{(2)} = \mathrm{SV}(X, r_M + r_2), \qquad \text{(top left singular vectors)}$$

- Step 2: Extract joint singular subspace

$$\widehat{U} = \begin{pmatrix} \widehat{V}^{(1)} & \widehat{V}^{(2)} \end{pmatrix}$$
$$\widehat{M} = \mathrm{SV}(\widehat{U}, r_M)$$

- Step 3: Extract individual singular subspaces

Estimating joint and individual structure (spectal method)

- Step 1: Extract singular subspaces

$$\widehat{V}^{(1)} = \mathrm{SV}(A, r_M + r_1), \qquad \text{(top left singular vectors)}$$
$$\widehat{V}^{(2)} = \mathrm{SV}(X, r_M + r_2), \qquad \text{(top left singular vectors)}$$

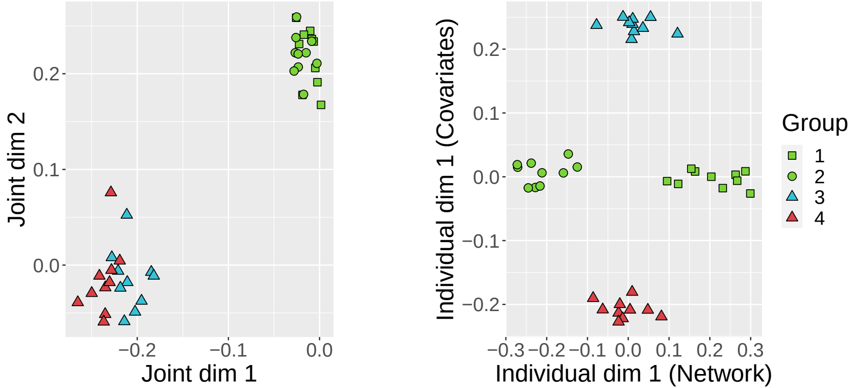
- Step 2: Extract joint singular subspace

$$\widehat{U} = \begin{pmatrix} \widehat{V}^{(1)} & \widehat{V}^{(2)} \end{pmatrix}$$
$$\widehat{M} = \mathrm{SV}(\widehat{U}, r_M)$$

- Step 3: Extract individual singular subspaces

$$\widehat{R}^{(k)} = \mathrm{SV}(\mathcal{P}_{\mathcal{C}(\widehat{M})^\perp} \widehat{V}^{(k)}, r_k)$$

Example



(a) Joint components

(b) Individual components

Figure: network communities: (triangle, circle square),
covariate clusters: (red, green, cyan)

Theory: expected error (spectral)

Notion of distance:

Theory: expected error (spectral)

Notion of distance: For matrices $A, B \in \mathbb{R}^{n \times r}$, define the *procrustes distance*,

$$d(A, B) = \inf_{Q \in \mathbb{O}_r} \|A - BQ\|_F$$

where $\mathbb{O}_r$ is the set of orthogonal matrices

Theory: expected error (spectral)

Notion of distance: For matrices $A, B \in \mathbb{R}^{n \times r}$, define the *procrustes distance*,

$$d(A, B) = \inf_{Q \in \mathbb{O}_r} \|A - BQ\|_F$$

where $\mathbb{O}_r$ is the set of orthogonal matrices

Important parameters:

Theory: expected error (spectral)

Notion of distance: For matrices $A, B \in \mathbb{R}^{n\times r}$, define the *procrustes distance*,

$$d(A,B) = \inf_{Q\in\mathbb{O}_r} \|A - BQ\|_F$$

where $\mathbb{O}_r$ is the set of orthogonal matrices

Important parameters:

- eigen/singular values: $\lambda_{r_M+r_1}(P)$, $\sigma_{r_M+r_2}(W)$

Theory: expected error (spectral)

Notion of distance: For matrices $A, B \in \mathbb{R}^{n \times r}$, define the *procrustes distance*,

$$d(A, B) = \inf_{Q \in \mathbb{O}_r} \|A - BQ\|_F$$

where $\mathbb{O}_r$ is the set of orthogonal matrices

Important parameters:

- eigen/singular values: $\lambda_{r_M + r_1}(P)$, $\sigma_{r_M + r_2}(W)$
- individual subspace separation: $\delta = 1 - \sigma_1((R^{(1)})^\top R^{(2)})$

Theory: expected error (spectral)

Notion of distance: For matrices $A, B \in \mathbb{R}^{n \times r}$, define the *procrustes distance*,

$$d(A, B) = \inf_{Q \in \mathbb{O}_r} \|A - BQ\|_F$$

where $\mathbb{O}_r$ is the set of orthogonal matrices

Important parameters:

- eigen/singular values: $\lambda_{r_M + r_1}(P)$, $\sigma_{r_M + r_2}(W)$
- individual subspace separation: $\delta = 1 - \sigma_1((R^{(1)})^\top R^{(2)})$
- standard deviation of covariate entries: τ

Theory: expected error (spectral)

Notion of distance: For matrices $A, B \in \mathbb{R}^{n \times r}$, define the *procrustes distance*,

$$d(A, B) = \inf_{Q \in \mathbb{O}_r} \|A - BQ\|_F$$

where $\mathbb{O}_r$ is the set of orthogonal matrices

Important parameters:

- eigen/singular values: $\lambda_{r_M + r_1}(P)$, $\sigma_{r_M + r_2}(W)$
- individual subspace separation: $\delta = 1 - \sigma_1((R^{(1)})^\top R^{(2)})$
- standard deviation of covariate entries: τ
- variance-type term of network: κ such that $P(\|E^A\| \geq t) \leq Ce^{-\frac{t}{\kappa}}$

Theory: expected error (spectral)

- Set

$$\epsilon^{(1)} = \frac{\kappa\sqrt{r_M + r_1}}{\lambda_{r_M+r_1}(P)}.$$

Theory: expected error (spectral)

- Set
$$\epsilon^{(1)} = \frac{\kappa\sqrt{r_M + r_1}}{\lambda_{r_M+r_1}(P)}.$$
Here $\epsilon^{(1)}$ measures how noisy the network is.

Theory: expected error (spectral)

- Set
$$\epsilon^{(1)} = \frac{\kappa\sqrt{r_M + r_1}}{\lambda_{r_M+r_1}(P)}.$$
Here $\epsilon^{(1)}$ measures how noisy the network is.

- Similarly, define the noise level in the covariates as
$$\epsilon^{(2)} = \frac{\tau\sqrt{n(r_M + r_2)(\sigma^2_{r_M+r_2}(W) + p)}}{\sigma^2_{r_M+r_2}(W)} \wedge \sqrt{r_M + r_2}.$$

Theory: expected error (spectral)

Theorem

If $\|E^A\|$ is subexponential and entries of E^X_{ij} are iid subgaussian, then

$$\text{Joint:} \qquad \mathbb{E}[d(\widehat{M}, M)] = O\left(\frac{\sqrt{r_M}}{\delta}[\epsilon^{(1)} + \epsilon^{(2)}]\right),$$

$$\text{Individual:} \qquad \mathbb{E}[d(\widehat{R}^{(k)}, R^{(k)})] = O\left(\frac{\sqrt{r_M r_k}}{\delta}[\epsilon^{(1)} + \epsilon^{(2)}]\right)$$

Estimating joint and individual structure (optimization)

- Pulling top singular vectors may discard important information
- Refine the spectral estimate by minimizing an associated loss function:

$$
\begin{aligned}
\min_{P', M, W} \quad & \|A' - P'\|_F^2 + \|X - W\|_F^2 \\
\text{s.t.} \quad & \mathcal{C}(M) \subset \mathcal{C}(P') \cap \mathcal{C}(W) \\
& \text{rank}(P') = r_M + r_1, \\
& \text{rank}(W) = r_M + r_2, \\
& \text{rank}(M) = r_M
\end{aligned}
$$

  where $A' = |A|^{1/2}$.
- Can be solved locally by iteratively optimizing a pair of related losses analogously to block cooridnate descent.
- Can initialize at spectral estimate

Data Exploration

Network: International food commodity trade where nodes are countries and edges are trade volumes

Covariates: We observe economic/geographic information at each nation like GDP, education expenditure, and geographic region

Possible Questions:

- Can the covariates identify groupings of countries based on how they trade?
- What information about trade relationships is explained by economic and regional information?

Data Exploration

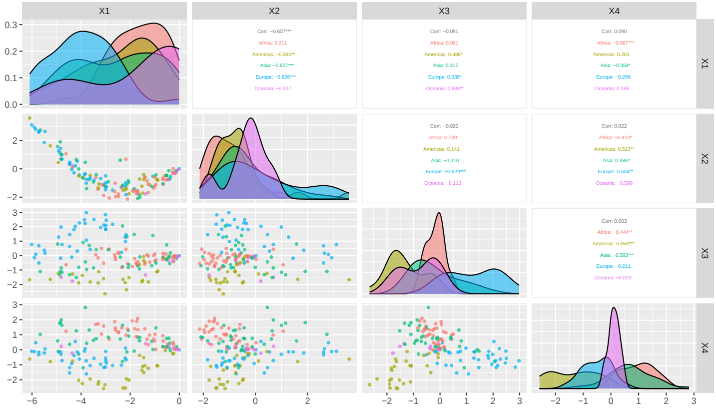


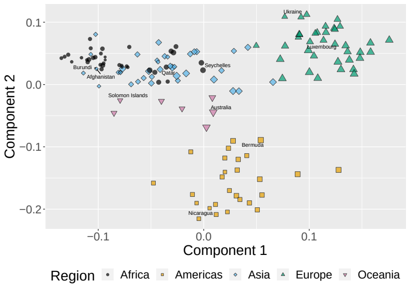
Figure: PCA (network)

Upon inspection, regional trade structure is not obvious.

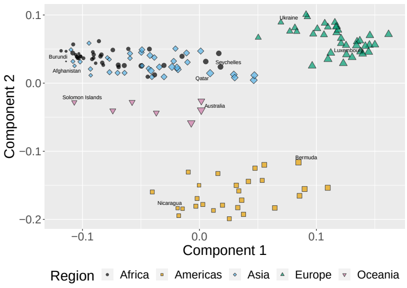
Figure: PCA (covariates)

Clear regional structure in the covariates. Note that the covariates can separate nations in Africa and Asia.

Data Exploration



(a) $\widehat{M}$ spectral



(b) $\widehat{M}$ optimization

- Optimization improved group separation.
- Since the covariates separate Asia and Africa while the joint does not, the trade relation data cannot distinguish between Asia and Africa.

Data Exploration

Variation Explained: Identify variation explained by the joint and individual structure in each dataset.

- Partition the data as

$$A = \mathcal{P}_{\mathcal{C}(\widehat{M})}A + \mathcal{P}_{\mathcal{C}(\widehat{R}^{(1)})}A + \mathcal{P}_{\mathcal{C}(\widehat{V}^{(1)})^{\perp}}A,$$

- Define

$$\mathsf{Var}_{\widehat{M}}(A) = \|\mathcal{P}_{\mathcal{C}(\widehat{M})}A\|_F^2/\|A\|_F^2,$$

$$\mathsf{Var}_{\widehat{R}^{(1)}}(A) = \|\mathcal{P}_{\mathcal{C}(\widehat{R}^{(1)})}A\|_F^2/\|A\|_F^2$$

- Define $\mathsf{Var}_{\widehat{M}}(X)$ and $\mathsf{Var}_{\widehat{R}^{(2)}}(X)$ similarly.

Data Exploration

- $\text{Var}_{\widehat{M}}(A)$ close to one means that the covariates can explain most of the variation in the network.
- Similar interpretations for $\text{Var}_{\widehat{R}^{(1)}}(A)$, $\text{Var}_{\widehat{M}}(X)$ and $\text{Var}_{\widehat{R}^{(2)}}(X)$.

Dataset	Component	Variation
Network	Joint	12.1%
Network	Individual	79.12%
Network	Residual	8.78%
Covariates	Joint	50.56%
Covariates	Individual	29.13%
Covariates	Residual	20.31%

Table: Proportion of variation explained by component for network and covariate datasets

Take home:

- We can partition the information in multiple datasets using shared and unique structure.
- Each dataset helps to inform about the other, the partition gives holistic view

Take home:

- We can partition the information in multiple datasets using shared and unique structure.
- Each dataset helps to inform about the other, the partition gives holistic view

Future extensions:

- more than two datasets
- group structure recovery

Take home:

- We can partition the information in multiple datasets using shared and unique structure.
- Each dataset helps to inform about the other, the partition gives holistic view

Future extensions:

- more than two datasets
- group structure recovery

Thank You!

Arroyo, J., A. Athreya, J. Cape, G. Chen, C. E. Priebe, and J. T. Vogelstein (2021). Inference for multiple heterogeneous networks with a common invariant subspace. *Journal of Machine Learning Research 22*(142), 1–49.

Binkiewicz, N., J. T. Vogelstein, and K. Rohe (2017). Covariate-assisted spectral clustering. *Biometrika 104*(2), 361–377.

Feng, Q., M. Jiang, J. Hannig, and J. Marron (2018). Angle-based joint and individual variation explained. *Journal of multivariate analysis 166*, 241–265.

Lock, E. F., K. A. Hoadley, J. S. Marron, and A. B. Nobel (2013). Joint and individual variation explained (jive) for integrated analysis of multiple data types. *The Annals of Applied Statistics 7*(1), 523.

MacDonald, P. W., E. Levina, and J. Zhu (2022). Latent space models for multiplex networks with shared structure. *Biometrika 109*(3), 683–706.

Yuan, D. and I. Gaynanova (2022). Double-matched matrix decomposition for multi-view data. *Journal of Computational and Graphical Statistics 31*(4), 1114–1126.