Learning common structures in a collection of networks

Do the networks share commmon structures?

Saint-Clair Chabert-Liddell
Joint work with S. Donnet and P. Barbillon

07 September 2021
EUSN2021

Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA-Paris

Summary

Modeling a Collection of Networks

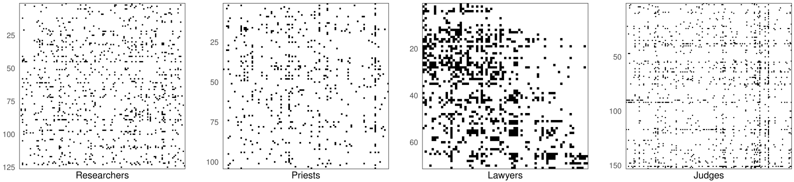
Inference, Model Selection and Partition of Networks

Application to a Collection of Advice Networks

Motivation

Data

- Collection $\mathbf{X} = \{\ldots, X^m, \ldots\}$, $m \in \mathcal{M}$ of $M = |\mathcal{M}|$ networks
- Same type:
  - Simple, Bipartite...
  - Undirected, Directed: *Advice networks*
- Same value type:
  - Binary (Bernoulli), Count (Poisson)...



Motivation

Data

- Collection $\mathbf{X} = \{\ldots, X^m, \ldots\}$, $m \in \mathcal{M}$ of $M = |\mathcal{M}|$ networks
- Same type:
  - Simple, Bipartite...
  - Undirected, Directed: *Advice networks*
- Same value type:
  - Binary (Bernoulli), Count (Poisson)...

Objective Find a common connectivity structure

Question Is the common structure relevant?

Objective Partition networks by connectivity structures

Method Joint modeling with *Stochastic Block Model (SBM)*

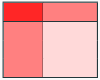
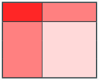
Modeling a Collection of Networks

Model: *iidcolSBM*

SBM for a collection of networks (*iidcolSBM*)

- Network X^m, $m \in \mathcal{M}$
- n_m individuals into common set of blocks $\mathcal{Q}$
- Same blocks proportions: $\mathbb{P}(Z_{iq}^m = 1) = \pi_q$, $q \in \mathcal{Q}$
- Same connectivity structure: $\mathbb{P}(X_{ij}^m = 1 | Z_{iq}^m Z_{jr}^m = 1) = \alpha_{qr}$

Core-Periphery

-  $\alpha = \begin{bmatrix} .9 & .5 \\ .5 & .1 \end{bmatrix}$ $\pi = [.25, .75]$ 
- *iidcolSBM*: 4 parameters Vs. 2 *SBM*s: 8 free parameters (undirected networks)

Model: *iidcolSBM*

SBM for a collection of networks (*iidcolSBM*)

- Network X^m, $m \in \mathcal{M}$
- n_m individuals into common set of blocks $\mathcal{Q}$
- Same blocks proportions: $\mathbb{P}(Z_{iq}^m = 1) = \pi_q$, $q \in \mathcal{Q}$
- Same connectivity structure: $\mathbb{P}(X_{ij}^m = 1 | Z_{iq}^m Z_{jr}^m = 1) = \alpha_{qr}$

i.i.d. assumption too restrictive, 2 new mechanisms:

- Free proportion of blocks between networks
- Density varies between networks

Model with free size of blocks: $\pi colSBM$

SBM for a collection of networks ($\pi colSBM$)

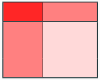
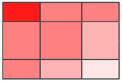
- Network X^m,$m \in \mathcal{M}$
- n_m individuals into set of blocks $\mathcal{Q}_m \subset \mathcal{Q}$
- Network specific proportion of blocks: $\mathbb{P}(Z^m_{iq} = 1) = \pi^m_q$, $q \in \mathcal{Q}_m$
- Same connectivity structure: $\mathbb{P}(X^m_{ij} = 1 | Z^m_{iq} Z^m_{jr} = 1) = \alpha_{qr}$

Model with free size of blocks: $\pi colSBM$

SBM for a collection of networks ($\pi colSBM$)

- Network $X^m, m \in \mathcal{M}$
- n_m individuals into set of blocks $\mathcal{Q}_m \subset \mathcal{Q}$
- Network specific proportion of blocks: $\mathbb{P}(Z^m_{iq} = 1) = \pi^m_q$, $q \in \mathcal{Q}_m$
- Same connectivity structure: $\mathbb{P}(X^m_{ij} = 1 | Z^m_{iq} Z^m_{jr} = 1) = \alpha_{qr}$

Nested core-periphery

-  $\pi^1 = [.25, 0, .75]$ $\alpha = \begin{bmatrix} .9 & .5 & .5 \\ .5 & .5 & .3 \\ .5 & .3 & .1 \end{bmatrix}$
-  $\pi^2 = [.25, .50, .25]$
- $\pi colSBM$: 9 parameters Vs. 2 *SBM*s: 12 free parameters (undirected networks)

Model with density factor $(\delta\text{–}\delta\pi)$*colSBMs*

SBM for a collection of networks (δ*colSBM*)

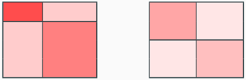
- Network X^m, $m \in \mathcal{M}$
- n_m individuals into common set of blocks $\mathcal{Q}$
  - δ*colSBM*: $\mathbb{P}(Z_{iq}^m = 1) = \pi_q$, $q \in \mathcal{Q}$ OR
  - $\delta\pi$*colSBM*: $\mathbb{P}(Z_{iq}^m = 1) = \pi_q^m$, $q \in \mathcal{Q}_m \subset \mathcal{Q}$
- Common connectivity structure up to a density parameter: $\mathbb{P}(X_{ij}^m = 1 | Z_{iq}^m Z_{jr}^m = 1) = \delta_m \alpha_{qr}$ with $\delta_1 = 1$ (identifiability)

Model with density factor $(\delta$–$\delta\pi)$*colSBMs*

SBM for a collection of networks (δ*colSBM*)

- Network X^m, $m \in \mathcal{M}$
- n_m individuals into common set of blocks $\mathcal{Q}$
  - δ*colSBM*: $\mathbb{P}(Z^m_{iq} = 1) = \pi_q$, $q \in \mathcal{Q}$ OR
  - $\delta\pi$*colSBM*: $\mathbb{P}(Z^m_{iq} = 1) = \pi^m_q$, $q \in \mathcal{Q}_m \subset \mathcal{Q}$
- Common connectivity structure up to a density parameter: $\mathbb{P}(X^m_{ij} = 1 | Z^m_{iq} Z^m_{jr} = 1) = \delta_m \alpha_{qr}$ with $\delta_1 = 1$ (identifiability)

Community structure

- 
- $\pi^1 = (.25, .75)$ $\pi^2 = (.50, .50)$ $\alpha = \begin{pmatrix} .7 & .2 \\ .2 & .5 \end{pmatrix}$ $\delta = (1, 0.5)$
- $\delta\pi$*colSBM*: 7 parameters Vs. 2 *SBM*s: 10 free parameters (undirected networks)

Identifiability

Identifiability of all *colSBM*s for both parameters and block matching

For $|\mathcal{Q}_m|$ known with:

- Classical assumptions for SBM on n_m, $|\mathcal{Q}_m|$ ratio and $\{\alpha, \pi\}$
- Assumption on block support: $S = \bigotimes_{m \in \mathcal{M}} \mathcal{Q}_m$ for $(\pi\text{–}\delta\pi)$*colSBM*s

Inference, Model Selection and Partition of Networks

Maximum Likelihood Inference

For fixed support S, $\boldsymbol{\theta} = \{\alpha, \pi, \delta\}$:

Objective Joint clustering of $\mathbf{Z} = \{Z^1, \dots, Z^{|\mathcal{M}|}\}$ and estimates of $\boldsymbol{\theta}$

Method Maximum likelihood of the observed data

Idea Compute complete likelihood and integrate on $\mathbf{Z}$

Problem Intractable, sum of $\prod_{m \in \mathcal{M}} |\mathcal{Q}_m|^{n_m}$ terms

Solution EM algorithm

Problem $\mathcal{L}(\mathbf{Z}|\mathbf{X})$ also intractable

Solution Variational approach of the EM algorithm

Daudin et al. (2008)

Variational EM

$$\begin{aligned}\ell(\mathbf{X};\boldsymbol{\theta}) &\geq \sum_{m\in\mathcal{M}} \ell(X^m;\boldsymbol{\theta}) - D_{KL}(\mathcal{R}(Z^m)\|p(Z^m|X^m))\\ &= \sum_{m\in\mathcal{M}} \left(\mathbb{E}_{\mathcal{R}}[\ell(X^m, Z^m;\boldsymbol{\theta})] + \mathcal{H}(\mathcal{R}(Z^m))\right) =: \mathcal{J}(\mathcal{R}(\mathbf{Z}),\boldsymbol{\theta}).\end{aligned}$$

$\mathcal{R}(\mathbf{Z})$ is a mean-field approximation of $\mathbf{Z}|\mathbf{X}$

$\mathcal{H}$ is the entropy

V-EM algorithm
2 steps iterative algorithm, for each $m \in \mathcal{M}$:

VE Maximize $\mathcal{J}(\mathcal{R}(Z^m),\boldsymbol{\theta})$ w.r.t. $\mathcal{R}(\mathbf{Z})$

M Maximize $\mathcal{J}(\mathcal{R}(\mathbf{Z}),\boldsymbol{\theta})$ w.r.t. $\boldsymbol{\theta}$

- Introduce stochasticy in the V-EM algorithm
- $(\delta\text{–}\delta\pi)$*colSBM*: no closed form for M–Step for Bernoulli model (Can use Poisson)

Model selection

Penalized model-based criterion

- To choose $S = \bigotimes_{m \in \mathcal{M}} \mathcal{Q}_m$
- To determine if common structure is relevant
- Based on Integrated Classification Likelihood (ICL)
- Adapted to allow for empty blocks

$$ICL(\mathcal{M}|S) = \mathcal{J}(\hat{\mathcal{R}}(\mathbf{Z}), \hat{\boldsymbol{\theta}}) - pen_{colSBM}(\mathcal{M}|S)$$

Structure relevant if:

$$\sum_{m \in \mathcal{M}} \max_{Q_m} ICL_{SBM}(m, Q_m) < \max_{S} ICL(\mathcal{M}, S)$$

Biernacki et al. (2000)

Partition of networks

Groups of networks may have different connectivity structures.

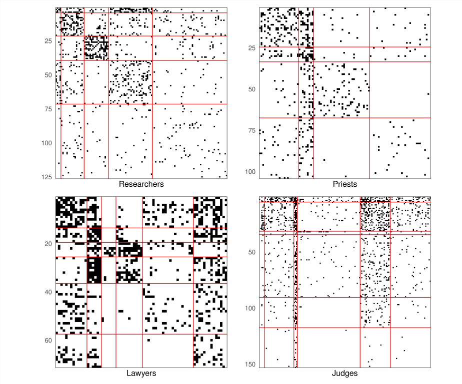
Find the partition with the highest *ICL*

$$\mathcal{G}^* = \arg \max_{\mathcal{G} \in \mathcal{P}(\mathcal{M})} \sum_{g \in \mathcal{G}} \max_{S \in \bigotimes_{m \in \mathcal{M}_g} \mathcal{Q}_m} ICL(\mathcal{M}_g | S)$$

Application to a Collection of Advice Networks

Application to advice networks (1)

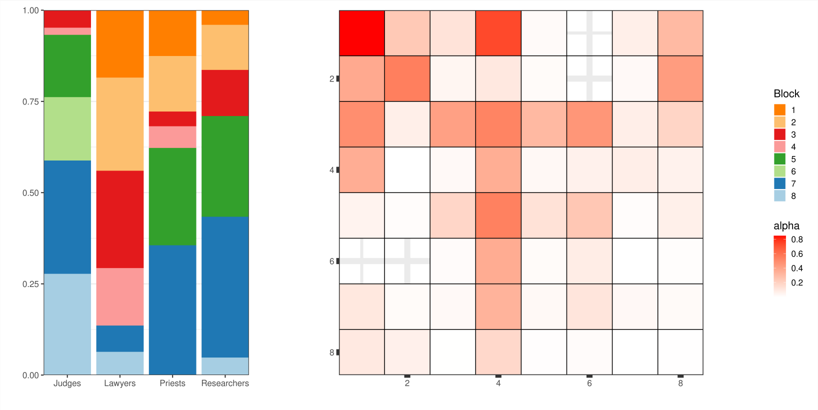
- 4 advice networks [3]
- $(126, 104, 71, 153)$ individuals in $(5, 4, 6, 6)$ SBM Blocks.
- Density: $(.061, .049, .18, .053)$



[3] Courtesy of E. Lazega

Application to advice networks (2)

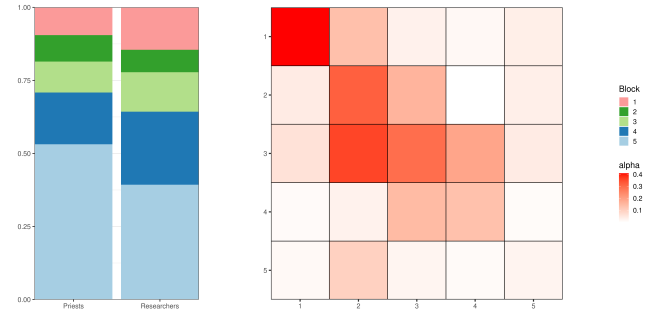
- Modeling 4 networks with $\delta\pi colSBM$
- $ICL_{\delta\pi colSBM} \approx -11147 > -11209 \approx ICL_{SBM}$
- No good common structure for the other models



$\hat{\delta} = (1, 0.7, 0.45, .79)$

Application to advice networks (3)

- $\delta\pi colSBM$ difficult to analyze
- Other *colSBMs*: structure of network with judges is different
- Best partition for $\pi colSBM$: Priests-Researchers, Lawyers, Judges ($ICL_{\pi colSBM} \approx -11177$)



Predicting missing advices

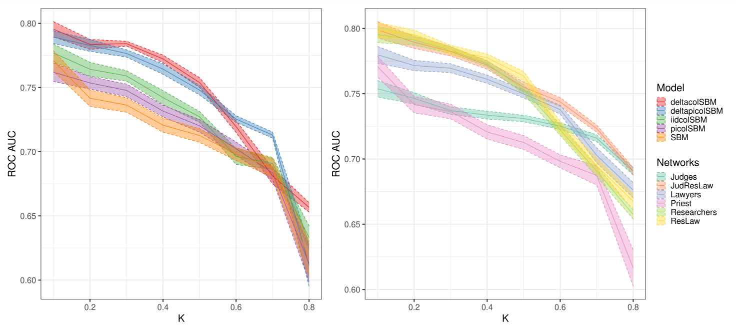
Can we better predict advices between priests thanks to other advice networks?

- Encoding proportion K of entries as `NA`
- Fit *colSBM*s (using Poisson model instead of (δ–$\delta\pi$)*colSBM*s for inference purpose)
- Using information from Researchers networks with all *colSBM*s
- Using information from different networks with δ*colSBM*
- $\hat{p}_{ij}^{priest} = \sum_{q,r \in \hat{\mathcal{Q}}_{priest}} \hat{\mathbb{P}}_{\mathcal{R}}(Z_{iq}^{priest} = 1)\hat{\mathbb{P}}_{\mathcal{R}}(Z_{jr}^{priest} = 1)\hat{\delta}^{priest}\hat{\alpha}_{qr}$

Predicting missing advices

Can we better predict advices between priests thanks to other advice networks?

- $(\delta$–$\delta\pi)$*colSBM*s better at prediction
- Researchers, Lawyers information very insightful when K small
- Judges good for large K



Left: with Researchers for *colSBM*s, Right: for δ*colSBM* with different networks

Take Home Message

- Joint modeling of a collection of networks with *colSBMs*
  - Find a common structure between the different networks
  - Identify blocks between networks
  - Model selection criterion:
    - Determine the relevance of the joint modeling
    - Classify networks from their connectivity patterns
- Extension to other types of networks: bipartite, multipartite...
- Dealing with covariates on nodes, edges and networks

Any questions? `saint-clair.chabert-liddell@inrae.fr`

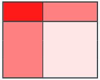
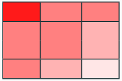
References

Biernacki, C., G. Celeux, and G. Govaert (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence 22*(7), 719–725.

Daudin, J.-J., F. Picard, and S. Robin (2008). A mixture model for random graphs. *Statistics and computing 18*(2), 173–183.

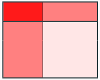
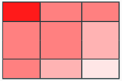
Examples (1): Nested structures

2 separated SBM: 16 parameters

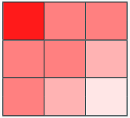
-  $\pi^{(1)} = [.25, .75]$ $\alpha = \begin{bmatrix} .9 & .5 \\ .5 & .1 \end{bmatrix}$
-  $\pi^{(2)} = [.25, .50, .25]$ $\alpha = \begin{bmatrix} .9 & .5 & .5 \\ .5 & .5 & .3 \\ .5 & .3 & .1 \end{bmatrix}$

Examples (1): Nested structures

2 separated SBM: 16 parameters

-  $\pi^{(1)} = [.25, .75] \qquad \alpha = \begin{bmatrix} .9 & .5 \\ .5 & .1 \end{bmatrix}$
-  $\pi^{(2)} = [.25, .50, .25] \qquad \alpha = \begin{bmatrix} .9 & .5 & .5 \\ .5 & .5 & .3 \\ .5 & .3 & .1 \end{bmatrix}$

π*colSBM*: 9 paramètres

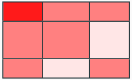
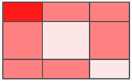
-  $\pi^{(1)} = [.25, 0, .75]$, $\pi^{(2)} = [.25, .50, .25] \qquad \alpha = \begin{bmatrix} .9 & .5 & .5 \\ .5 & .5 & .3 \\ .5 & .3 & .1 \end{bmatrix}$
- $pen_{SBM}(2) + pen_{SBM}(3) \approx 45 > 39 \approx pen_{\pi colSBM}(3)$ for $n_1 = n_2 = 100$

Common structure relevant

Examples (2): Partially nested structures

Undirected networks:

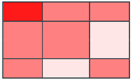
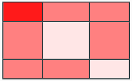
2 separated SBM: 16 parameters

-  $\pi^{(1)} = [.25, .50, .25]$ $\alpha = \begin{bmatrix} .9 & .5 & .5 \\ .5 & .5 & .1 \\ .5 & .1 & .5 \end{bmatrix}$
-  $\pi^{(2)} = [.25, .50, .25]$ $\alpha = \begin{bmatrix} .9 & .5 & .5 \\ .5 & .1 & .5 \\ .5 & .5 & .1 \end{bmatrix}$

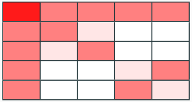
Examples (2): Partially nested structures

Undirected networks:

2 separated SBM: 16 parameters

-  $\pi^{(1)} = [.25, .50, .25]$ $\alpha = \begin{bmatrix} .9 & .5 & .5 \\ .5 & .5 & .1 \\ .5 & .1 & .5 \end{bmatrix}$
-  $\pi^{(2)} = [.25, .50, .25]$ $\alpha = \begin{bmatrix} .9 & .5 & .5 \\ .5 & .1 & .5 \\ .5 & .5 & .1 \end{bmatrix}$

π*colSBM*: 15 parameters

-  $\pi^{(1)} = [.25, 50, .25, 0, 0]$ $\pi^{(2)} = [.25, 0, 0, .50, .25]$ $\alpha = \begin{bmatrix} .9 & .5 & .5 & .5 & .5 \\ .5 & .5 & .1 & \cdot & \cdot \\ .5 & .1 & .5 & \cdot & \cdot \\ .5 & \cdot & \cdot & .1 & .5 \\ .5 & \cdot & \cdot & .5 & .1 \end{bmatrix}$
- $pen_{SBM}(3) + pen_{SBM}(3) \approx 60 < 67 \approx pen_{\pi colSBM}(5)$ for $n_1 = n_2 = 100$

Common structure not relevant

Partition of networks

All the networks in the collection may not have the same structure.

$$\mathcal{G}^* = \arg \max_{\mathcal{G} \in \mathcal{P}(\mathcal{M})} \sum_{g \in \mathcal{G}} \max_{S \in \bigotimes_{m \in \mathcal{M}_g} \mathcal{Q}_m} ICL(\mathcal{M}_g | S).$$

Need 2^M partitions to compute all partitions. Too costly if M large.

Dissimilarity

- *colSBM*s allow to match Z^ms
- Compute dissimilarity matrix using MLE of SBM on *colSBM*s block:

$$D(m, m') = \sum_{q,r \in \mathcal{Q}} \max\left(\hat{\pi}_q^m, \hat{\pi}_q^{m'}\right) \max\left(\hat{\pi}_r^m, \hat{\pi}_r^{m'}\right) \left(\frac{\hat{\alpha}_{qr}^m}{\hat{\delta}^m} - \frac{\hat{\alpha}_{qr}^{m'}}{\hat{\delta}^{m'}}\right)^2$$

- Use clustering algorithm on D (hierarchical clustering, k-medoids...)
- Compute ICL_{colSBM} on obtained partition

Extension: Partition of Predation Networks

- $|\mathcal{M}| = 67$ networks from Mangal database
- 31 to 106 species nodes
- Density range in $[.01, .32]$
- Modeling the collection with $\pi colSBM$