Intérêt des modèles à blocs stochastiques pour les réseaux écologiques Partie 2

Sophie Donnet. INRA@

Ecole de printemps, Chaire MMB, Aussois, juin 2022





Introduction

Finding a consensus in a collection of networks

Partition of networks according to their mesoscale structure

By infering the SBM on my network I get:

- Z: A clustering of the nodes/ species playing the same role in the network
- $\hat{lpha}, \hat{\pi}$: a mesoscopic view ot the network

What can I do with that?

- Clustering of the nodes: studying the species, linking the clustering with other traits
- Resume of the structure:
 - Using the influence of this mescopic structure to its robustness to species extinction for instance? [Chabert-Liddell et al., 2022]
 - Comparing networks? Clustering networks?

Introduction

Finding a consensus in a collection of networks Block models for a collection of networks Statistical inference Application on our 3 foodwebs

Partition of networks according to their mesoscale structure

Objectives

Looking for commun patterns in networks involving non-common sets of nodes



Applications

- Compare the structure of ecological networks
- Compare sociological networks : advices between lawyers, researchers, judges, or priests

Three foodwebs

- Pine-forest stream food webs issued from Maine, North-Caroline and New-Zealand [Thompson and Townsend, 2003]
- Involve respectively 105, 58 and 71 species.
- $Y_{ij} = 1$ if *i* is eaten by *j*. Directed relation



Look for similarities and differences between network structures.



- Fitted SBM on each separately
- Reordered the matrices following the blocks
- Label the blocks following the average out-degrees order



- Two bottom groups in each matrix are basal species : eaten by many species and not eating anybody.
 - Martins: has a separation into 5 blocks, the third one is a medium trophic level, which preys on basal species and is highly preyed by species of the 1st block.
 - **Cooper**. Higher trophic levels grouped together in the same block (lack of statistical power).
 - **Herlzier**: higher trophic level is separated into 2 blocks determined on how much they prey on the less preyed basal block.

Introduction

Finding a consensus in a collection of networks Block models for a collection of networks Several models Statistical inference

Application on our 3 foodwebs

Partition of networks according to their mesoscale structure

Introduction

Finding a consensus in a collection of networks Block models for a collection of networks Several models Statistical inference

Partition of networks according to their mesoscale structure

Towards a joint modeling of the networks

- Need to model jointly the networks
- Identify the groups playing the same role through out the networks, with an unsupervised strategy.
- Let (Y^m)_{m=1,...,M} denote the collection of networks each involving n_m nodes.
- (**Y**^{*m*}) independent.

$$\mathbf{Y}^m \sim \mathsf{SBM}_{n_m}(K_m, \pi^m, \alpha^m)$$

- Conditions on the parameters $(\pi^m)_{m=1,...,M}$ and $(lpha^m)_{m=1,...,M}$

iid-coISBM

$$\mathbf{Y}^m \sim \text{SBM}_{n_m}(K, \pi, \alpha)$$

with $\pi_k > 0 \ \forall k \in \{1, \dots, K\}$ and $\sum_{k=1}^{K} \pi_k = 1$.

- (K 1) + K² unknown parameters, M clusterings (one for each network)
- Too strict to be applied to the Thomson's dataset

Same structure of connection $\boldsymbol{\alpha},$ specific proportions of blocks in each network

 $\pi\text{-colSBM}$

$$\mathbf{Y}^m \sim \mathsf{SBM}_{n_m}(K, \pi^m, \alpha)$$

On the block proportions

- $\pi_k^m \ge 0$
- If $\pi_k^m = 0$ then block k is not represented in network m

M = 2 networks

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{12} & \alpha_{22} & \alpha_{23} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{pmatrix} \qquad \pi^1 = [.25, .25, .50] \\ \pi^2 = [.20, .50, .30]$$

- Same connection structure between blocks
- Different block proportions
- $2 \times (3-1) + 3^2 = 15$ parameters.

 $\pi_k^m \geq 0$

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{12} & \alpha_{22} & \alpha_{23} \\ \alpha_{13} & \alpha_{23} & \alpha_{33} \end{pmatrix} \qquad \begin{array}{c} \pi^1 = [.25, .25, .50] \\ \pi^2 = [.40, \ 0, .60] \end{array}$$

- Blocks 1 and 3 are represented in the two networks while block 2 only exists in network 1.
- $3 1 + 3 2 + 3^2 = 14$ parameters

π -colSBM: partially nested structures

$$\alpha = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \cdot \\ \alpha_{31} & \cdot & \alpha_{33} \end{pmatrix} \qquad \pi^1 = [.25, .75, 0] \\ \pi^2 = [.40, 0, .60].$$

- The two networks share block 1 (for instance super predators or basal species)
- The remaining nodes of each network not equivalent in terms of connectivity.
- Blocks 2 and 3 never interact because their elements do not belong to the same network and so α₂₃ and α₃₂ are not required to define the model.
- (2-1) + (2-1) + 7 = 11 parameters.

Let S be the support $M \times K$ matrix such that

$$S_{mk} = egin{cases} 1 & ext{ if } \pi^m_k > 0 \ 0 & ext{ otherwise }. \end{cases}$$

Then,

$$Nb(\pi\text{-colSBM}) = \sum_{m=1}^{M} \left(\sum_{k=1}^{K} S_{km} - 1 \right) + \sum_{k,\ell=1}^{K} \mathbf{1}_{(S'S)_{k\ell} > 0}$$

$\delta\text{-colSBM}$

$$\mathbf{Y}^m \sim \mathsf{SBM}_{n_m}(K, \pi, \delta^m \alpha)$$

with $\pi_k > 0$,

- *M* networks exhibit similar intra- and inter blocks connectivity patterns but with proper densities.
- δ^m be a density parameter, specific to each network. $\delta^1 = 1$.
- Mimics differences of effort sampling or abundances
- $(K-1) + K^2 + (M-1)$ parameters.

$\delta\pi\text{-colSBM}$

$$\mathbf{Y}^m \sim \mathsf{SBM}_{n_m}(K, \pi^m, \delta^m \alpha)$$

with $\pi_k^m \ge 0$

- Most flexible model
- $Nb(\pi$ -colSBM) + (M 1) parameters.

M independent networks.

$$\mathbf{Y}^m \sim \mathsf{SBM}(\mathcal{K}_m, \pi^m, \alpha^m)$$

Model name	Block prop.	Connexion param.	Nb of param.
iid-coISBM	$\pi_k^m = \pi_k, \ \pi_k > 0$	$\alpha_{k\ell}^m = \alpha_{k\ell}$	$(K-1)+K^2$
π-coISBM	$\pi_k^m, \pi_k^m \ge 0$	$\alpha_{k\ell}^m = \alpha_{k\ell}$	$\leq M(K-1)+K^2$
δ -colSBM	$\pi_k^m = \pi_k, \ \pi_k > 0$	$\alpha_{k\ell}^m = \delta^m \alpha_{k\ell}$	$(K-1)+K^2+(M-1)$
$\delta\pi$ -coISBM	$\pi_k^m, \ \pi_k^m \ge 0$	$\alpha_{k\ell}^m = \delta^m \alpha_{k\ell}$	$\leq M(K-1)+K^2+M-1$
sep-SBM	$\pi_k^m, \ \pi_k^m > 0$	$\alpha^m_{k\ell}$	$\sum_{m=1}^{M}(K_m-1)+K_m^2$

Introduction

Finding a consensus in a collection of networks

Block models for a collection of networks

Statistical inference

Application on our 3 foodwebs

Partition of networks according to their mesoscale structure

Demonstrated for the most complex SBM, upto label switching of the blocks under light conditions.

For π -colSBM, let us define $\mathcal{K}_m = \{k \in \{1, \dots, K\} | \pi_k^m > 0\}.$

- 1. $\forall m : n_m \ge 2|\mathcal{K}_m|$ 2. $\forall k = 1, \dots, K, \quad \exists m : K \in \mathcal{K}_m$ 3. $(\alpha \cdot \pi^m)_k \neq (\alpha \cdot \pi^m)_\ell$ for all $(k \neq \ell) \in \mathcal{K}_m^2$
- 4. Each diagonal entry of α is unique

VEM algorithm

- Direct extension of VEM previously described for $\it iid\mbox{-}colSBM$ and $\ensuremath{\pi\mbox{-}colSBM}$
- Less obvious with $\delta_m \alpha$: M step not explicit.

ICL can be directly extended for *iid*-colSBM and the δ -colSBM

$$ICL(K) = \mathcal{I}(\hat{\tau}, \hat{\theta}) - \frac{K-1}{2} \log\left(\sum_{m=1}^{M} n_m\right) \\ -\frac{1}{2} \left(K^2 + \nu(\delta)\right) \log\left(\sum_{m=1}^{M} n_m(n_m - 1)\right), \qquad (1)$$

where $\nu(\delta) = M - 1$ for $\delta colSBM$ and 0 otherwise.

- For *iid*-colSBM and the δ -colSBM
- π_k^m possibly null. Asymptotic approximation do not hold
- Each couple (K, S) defines a model.

$$eBICL(\mathbf{Y}, K) = \max_{S} \left[\max_{\theta_{S} \in \Theta_{S}} \mathcal{I}(\widehat{\mathcal{R}}, \theta_{S}) - \frac{1}{2} \left(\operatorname{pen}_{\pi}(K, S) + \operatorname{pen}_{\alpha}(K, S) + \operatorname{pen}_{\delta}(K, S) + \operatorname{pen}_{\delta}(K) \right) \right],$$

Model selection ii

where

$$\begin{aligned} \operatorname{pen}_{\pi}(K,S) &= \sum_{m=1}^{M} (K_m - 1) \log(n_m) \\ \operatorname{pen}_{\alpha}(K,S) &= \left(\sum_{q,r=1}^{K} \mathbf{1}_{(S'S)_{qr} > 0} \right) \log(N_M) , \\ \operatorname{pen}_{\delta}(K,S) &= \begin{cases} 0 & \text{for } \pi\text{-}colSBM \\ (M-1) \log(N_M) & \text{for } \delta\pi\text{-}colSBM \end{cases} , \\ \operatorname{pen}_{S}(K) &= -2 \log p_K(S). \end{aligned}$$

Finally, K is chosen such that:

$$\widehat{K} = \underset{K \in \{1, \dots, K_{\max}\}}{\operatorname{arg max}} \operatorname{eBIC-L}(\mathbf{Y}, K)$$

Introduction

Finding a consensus in a collection of networks

Block models for a collection of networks

Statistical inference

Application on our 3 foodwebs

Partition of networks according to their mesoscale structure



Separate sbm

Model	ICL
sepSBM	-2080
iid-colSBM	-1966
π -colSBM	-1982
δ -colSBM	-1969
$\delta\pi ext{-colSBM}$	-1989

• Reject sepSBM : commun structure in the networks

Our 4 consensus models



Top left : iid (-1966). Top right: π -colSBM (-1982) Bottom-left: δ -colSBM (-1969). Bottom-right: $\delta\pi$ -colSBM (-1989)

iid-colSBM: the prefered model



- Makes 5 blocks
- Block 3 (light green) is a small block of intermediate trophic level species with some within block predation.
- The higher trophic level is divided into 2 more blocks,
 - block 2 (dark green) only preys on the 2 basal blocks
 - block 1 (pink) preys on the intermediate block 3 level but only on the most connected basal species block.



- Also 5 blocks.
- There are no empty blocks
- the block proportions are roughly corresponding to the ones of iid-colSBM .
- Flexibility of the *π*-colSBM of little use compared to the iid-colSBM on this collection.

Introduction

Finding a consensus in a collection of networks

Partition of networks according to their mesoscale structure

- If the networks in a collection do not have the same connectivity structure, we aim to partition them accordingly.
- Finding a partition \$\mathcal{G} = (\mathcal{M}_g)_{g=1,...,G}\$ of \$\{1,...,M\}\$. such that

$$\forall g \in \{1, \ldots, G\}, \quad \forall m \in \mathcal{M}_g, \quad \mathbf{Y}^m \sim \mathsf{SBM}(\mathcal{K}^g, \pi^m, \alpha^g)$$

networks belonging to the subcollection \mathcal{M}_g share the same mesoscale structure given by π -colSBM.

• To any partition ${\mathcal G}$ we associate the following score:

$$\mathsf{Sc}(\mathcal{G}) = \sum_{g=1}^{\mathsf{G}} \mathsf{BIC-L}((\mathbf{Y}^m)_{m \in \mathcal{M}_g}, \widehat{K^g}).$$

Best partition G is chosen as follows:

$$\mathcal{G}^* = rg\max_{\mathcal{G}} \mathsf{Sc}(\mathcal{G}).$$

- 67 networks issued from the Mangal database belonging to 33 datasets. [Vissault et al., 2020]
- predation networks which are all directed networks with more than 30 species,
- number of species ranges from 31 to 106 (3395 in total) by network
- Density ranging from .01 to .32 (14934 total predation links).

Aim use our model to propose partition of the networks into group of networks with common mesoscale structure.





Groupe A

- 7 networks and 12 blocks are required to describe this group of networks
- 5 networks are issued from the same dataset (id: 80).
- These 5 networks populate the 12 blocks, while the other 2 networks only populate parts of them.
- Average density is about 0.18
- Blocks 1 to 3 represent the higher trophic levels, blocks 4 to 8 the intermediate ones and block 9 to 12 the lower ones.



Group B : structure with 8 blocks

- 26 networks with heterogeneous size and density.
- Issued from various datasets
- Most networks populate only parts of the 8 blocks
- Block 4 is represented in only 5 networks where it is either an intermediate or a bottom trophic level.
- It introduces some symmetry in the connectivity matrix rendering it difficult to order the blocks by trophic order.
- Species from top trophic levels prey on basal species.



Group C : structure with 7 blocks

- 6 networks with density ranging from .06 to .11.
- All networks are represented in 5 or 6 of the 7 blocks, including the first three blocks.
- 3 of the 5 networks of dataset 48 (diff. collecting sites).
- Top trophic level divided into 2 blocks, species from those blocks preying only on intermediate trophic level species.
 - Species from block 2 prey on species from block 4, which prey more on basal species (block 7) than on others intermediate trophic species (block 6),



Group D : structure with 7 blocks

- 23 networks.
- The 10 networks from dataset 157 (stream food webs from New Zealand) are divided between groups B and D based on the type of ecosystem. The data from group B were collected in creeks, while the one from group D were collected on streams.



Group E : structure with 7 blocks

- Paper submitted to AoAS Arxiv + package
- Still need some work to convince ecologists
- Much needed for bipartite and tripartite networks
- On going work...

Merci



Chabert-Liddell, S., Barbillon, P., and Donnet, S. (2022).

Impact of the mesoscale structure of a bipartite ecological interaction network on its robustness through a probabilistic modeling.

Environmetrics, 33(2).



Thompson, R. M. and Townsend, C. R. (2003).

Impacts on stream food webs of native and exotic forest: an intercontinental comparison.

Ecology, 84(1):145-161.



Vissault, S., Cazelles, K., Bergeron, G., Mercier, B., Violet, C., Gravel, D., and Poisot, T. (2020).

rmangal: An R package to interact with Mangal database.

R package version 2.0.2.