

# Probabilistic models for (ecological) networks

---

Sophie Donnet. MIA Paris-Saclay, UMR INRAE - AgroParisTech,  
**APPLIBUGS Meeting.** Dec. 2021

## Introduction

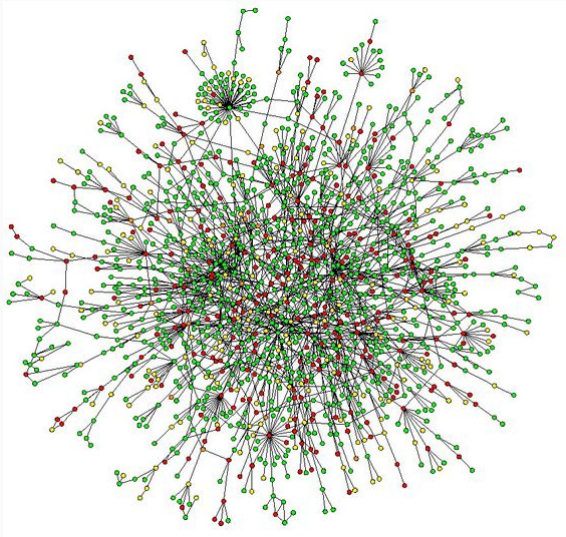
Descriptive statistics

Probabilistic models

Frequentist and Bayesian inference

Conclusion

# Network data



yeastProteinInteractionNetwork



Networks can account for

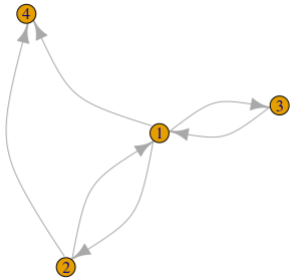
- **Ecological networks** : Food web , Co-existence networks, Host-parasite interactions, Plant-pollinator interactions,
- **Social networks**
- **Inventory datasets** : who cultivates what?
- ...

## A network

- **nodes/vertices** which represent individuals / species /ships which may interact or not,
- **links/edges/connections** which stand for an interaction between a pair of nodes / dyads.

# Directed network and adjacency matrix

For a **directed relation**: advice networks, foodwebs,



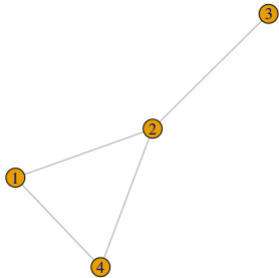
In general, no self-loop

$$Y = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

- $n$  rows and  $n$  columns,
- non symmetric matrix

# Network representation and adjacency matrix

For a **non-directed relation**: co-occurrence, co-publication, ...



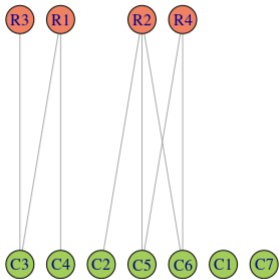
In general, no self-loop

$$Y = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

- $n$  rows and  $n$  columns,
- symmetric matrix

# Bipartite network and incidence matrix

Two types of nodes : Plant-pollinators, farmer-crop species, clients-films on netflix



$$Y = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

- n rows and m columns
- rectangular matrix

- Unraveling / describing / modeling the network topology.
- Discovering particular structure of interaction between some subsets of nodes.
- Understanding network heterogeneity.
- Not inferring the network !



## Additional available data



- the network provided as:
  - an adjacency matrix (for simple network) or an incidence matrix (for bipartite network),
  - a list of pair of nodes / dyads which are linked.
- some additional covariates on nodes, dyads which can account for sampling effort.

Introduction

**Descriptive statistics**

Probabilistic models

Frequentist and Bayesian inference

Conclusion

## Some common features studied on networks

- **Degree of a node:** number of connexion for each node.
- **Degree distribution** can be viewed as a measure of heterogeneity,
- **Nestedness:** a network is said to be nested when its nodes that have the smallest degree, are connected to nodes with the highest degree, [Rodríguez-Gironés and Santamaría, 2006]
- **Betweenness centrality:** for a node, numbers of shortest paths between any pair of nodes passing through this node. [Freeman, 1978]
- **Modularity:** is a measure for a given partition of its tendency of favoring intra-connection over inter-connection.  $\Rightarrow$  Finding the best partition with respect to modularity criterion. [Clauset et al., 2008]

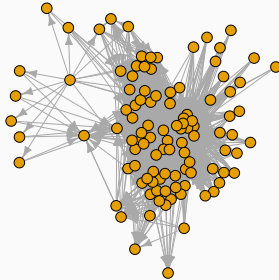
# Some common features studied on networks

All this criterion shall be adapted to:

- directed network,
- bipartite network.

R packages: `igraph`, `sna`, `vegan`.

## Example Chilean food web

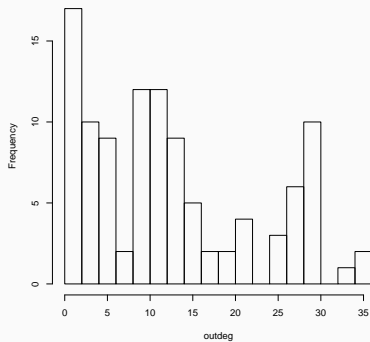


- $n = 106$  species / nodes,
- density of edges: 12.1%.

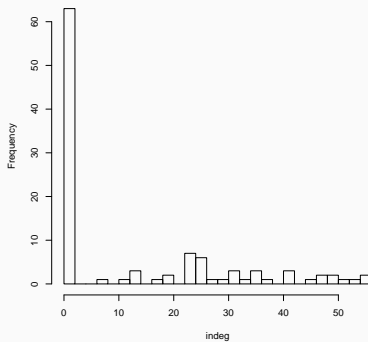
[Kéfi et al., 2016]

# Degree distribution

Histogram of outdeg



Histogram of indeg



Introduction

Descriptive statistics

Probabilistic models

- The stochastic Block Model

- Latent block models

- Some possible extensions

Frequentist and Bayesian inference

Conclusion

# Probabilistic approach

- **Context:** our matrix  $Y$  is the realization of a stochastic process.
- **Aim:** Propose a stochastic process is able to mimic heterogeneity in the connections.
- **Advantage:** benefit from the statistical tools (tests, model selection, etc...)

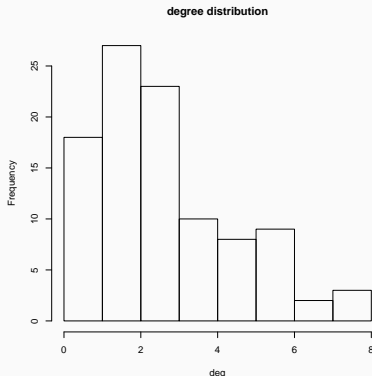
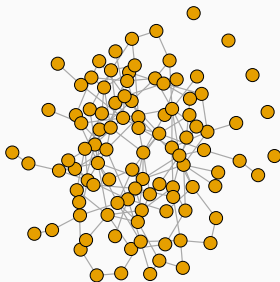


# A first random graph model for network: null model

Erdős-Rényi (1959) Model for  $n$  nodes

$$\forall 1 \leq i, j \leq n, \quad Y_{ij} \stackrel{i.i.d.}{\sim} \text{Bern}(p),$$

where  $p \in [0, 1]$  a probability for a link to exist.



# Limitations of an ER graph to describe real networks

- Homogeneity of the connections
- Degree distribution too concentrated, no high degree nodes,
- All nodes are equivalent (no nestedness...),
- No modularity, no hubs

Introduction

Descriptive statistics

Probabilistic models

The stochastic Block Model

Latent block models

Some possible extensions

Frequentist and Bayesian inference

Conclusion

# Stochastic Block Model

[Nowicki and Snijders, 2001] Let  $(Y_{ij})$  be an adjacency matrix

## Latent variables

- The nodes  $i = 1, \dots, n$  are partitionned into  $K$  clusters
- $Z_i = k$  if node  $i$  belongs to cluster (block)  $k$
- $Z_i$  independant variables

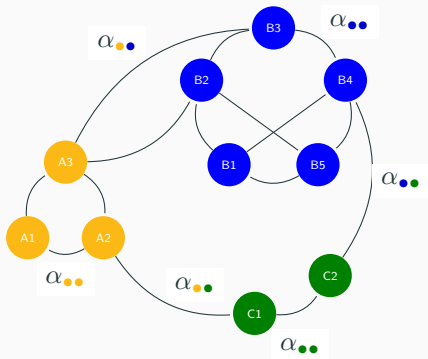
$$\mathbb{P}(Z_i = k) = \pi_k$$

## Conditionally to $(Z_i)_{i=1, \dots, n} \dots$

$(Y_{ij})$  independant and

$$Y_{ij} | Z_i, Z_j \sim \text{Bern}(\alpha_{Z_i, Z_j}) \quad \Leftrightarrow \quad P(Y_{ij} = 1 | Z_i = k, Z_j = \ell) = \alpha_{k\ell}$$

# Stochastic Block Model : illustration



## Parameters

Let  $n$  nodes divided into 3 clusters

- $\mathcal{K} = \{\bullet, \bullet, \bullet\}$  clusters
- $\pi_{\bullet} = \mathbb{P}(i \in \bullet), \bullet \in \mathcal{K}, i = 1, \dots, n$
- $\alpha_{\bullet, \bullet} = \mathbb{P}(i \leftrightarrow j | i \in \bullet, j \in \bullet)$

$$Z_i = \mathbf{1}_{\{i \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \pi), \quad \forall \bullet \in \mathcal{K},$$
$$Y_{ij} | \{i \in \bullet, j \in \bullet\} \sim^{\text{ind}} \mathcal{B}(\alpha_{\bullet, \bullet})$$

# SBM : A great generative model

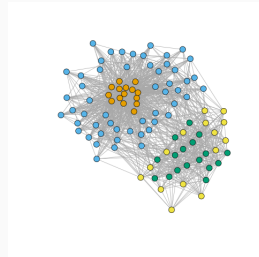
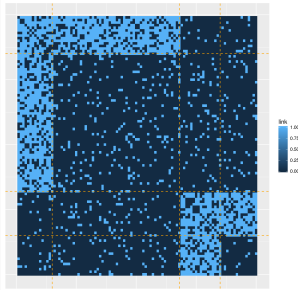
- Generative model : easy to simulate
- Very flexible
- Combination of modularity, nestedness, etc...

# Networks with hubs generated by SBM

- $\pi = c(.15, .35, .15, .35)$

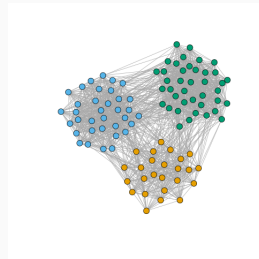
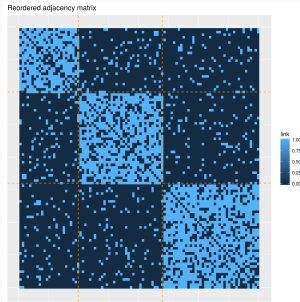
- $\alpha = \begin{pmatrix} 0.80 & 0.80 & 0.20 & 0.20 \\ 0.80 & 0.20 & 0.20 & 0.20 \\ 0.20 & 0.20 & 0.80 & 0.80 \\ 0.20 & 0.20 & 0.80 & 0.20 \end{pmatrix}$

Reordered adjacency matrix



# Community network generated by SBM

- $\pi = c(0.25, 0.35, 0.40)$
- $\alpha = \begin{pmatrix} 0.80 & 0.20 & 0.20 \\ 0.20 & 0.80 & 0.20 \\ 0.20 & 0.20 & 0.80 \end{pmatrix}$



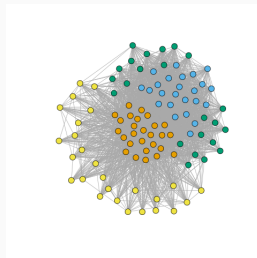
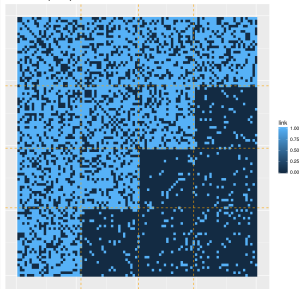


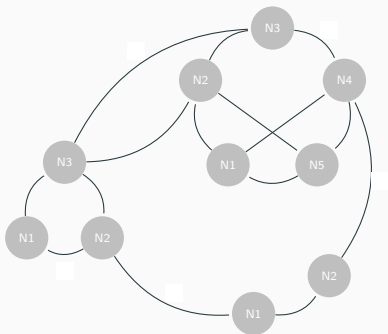
# Nestedness generated by SBM

- $\pi = c(.15, .35, .15, .35)$

- $\alpha = \begin{pmatrix} 0.80 & 0.80 & 0.80 & 0.80 \\ 0.80 & 0.80 & 0.80 & 0.20 \\ 0.20 & 0.80 & 0.20 & 0.80 \\ 0.80 & 0.20 & 0.20 & 0.20 \end{pmatrix}$

Reordered adjacency matrix





## Stochastic Block Model

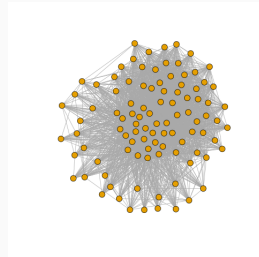
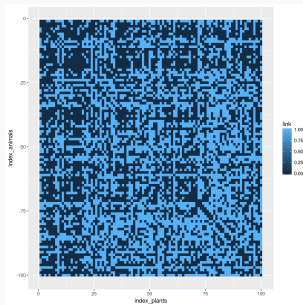
Let  $n$  nodes divided into

- $\mathcal{K} = \{\bullet, \bullet, \bullet\}$ ,  $\text{card}(\mathcal{K})$  known
- $\pi_{\bullet} = ?$ ,
- $\alpha_{\bullet\bullet} = ?$

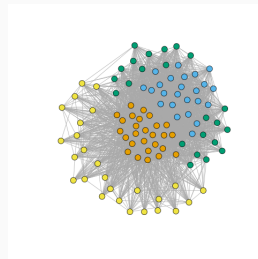
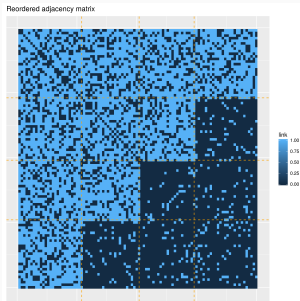
[Nowicki and Snijders, 2001], [Daudin et al., 2008]

R packages: blockmodels, sbm

From....



... to



## Statistician job

- Find the clusters
- Find the number of clusters
- Practical implementation
- Theoretical results

Introduction

Descriptive statistics

**Probabilistic models**

The stochastic Block Model

**Latent block models**

Some possible extensions

Frequentist and Bayesian inference

Conclusion

# Probabilistic model for binary bipartite networks

Let  $Y_{ij}$  be a bi-partite network. Individuals in row and cols are not the same.

## Latent variables : bi-clustering

- Nodes  $i = 1, \dots, n_1$  partitionned into  $K_1$  clusters, nodes  $j = 1, \dots, n_2$  partitionned into  $K_2$  clusters
- $$\begin{aligned} Z_i^1 &= k && \text{if node } i \text{ belongs to cluster (block) } k \\ Z_j^2 &= \ell && \text{if node } j \text{ belongs to cluster (block) } \ell \end{aligned}$$
- $Z_i^1, Z_j^2$  independent variables

$$\mathbb{P}(Z_i^1 = k) = \pi_k^1, \quad \mathbb{P}(Z_j^2 = \ell) = \pi_\ell^2$$

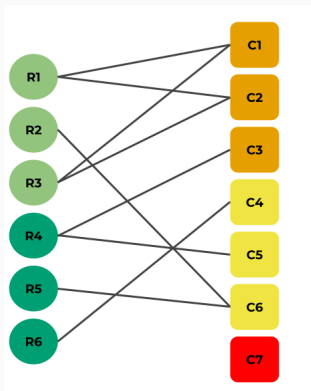
**Conditionally to**  $(Z_i^1)_{i=1,\dots,n_1}, (Z_j^2)_{j=1,\dots,n_2} \dots$

$(Y_{ij})$  independent and

$$Y_{ij}|Z_i^1, Z_j^2 \sim \text{Bern}(\alpha_{Z_i^1, Z_j^2}) \quad \Leftrightarrow \quad \mathbb{P}(Y_{ij} = 1 | Z_i^1 = k, Z_j^2 = \ell) = \alpha_{k\ell}$$

[Govaert and Nadif, 2008]

# Latent Block Model : illustration



$$\begin{aligned} Z_i^R &= \mathbf{1}_{\{i \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \pi^R), & \forall \bullet \in \mathcal{Q}_R, \\ Z_j^C &= \mathbf{1}_{\{j \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \pi^C), & \forall \bullet \in \mathcal{Q}_C, \\ Y_{ij} \mid \{i \in \bullet, j \in \bullet\} &\sim^{\text{ind}} \text{Bern}(\alpha_{\bullet\bullet}) \end{aligned}$$



Introduction

Descriptive statistics

**Probabilistic models**

The stochastic Block Model

Latent block models

**Some possible extensions**

Frequentist and Bayesian inference

Conclusion

# Valued-edge networks

## Values-edges networks

Information on edges can be something different from presence/absence. It can be:

1. a count of the number of observed interactions,
2. a quantity interpreted as the interaction strength,

## Natural extensions of SBM and LBM

1. Poisson distribution:  $Y_{ij} \mid \{i \in \bullet, j \in \bullet\} \sim^{\text{ind}} \mathcal{P}(\lambda_{\bullet\bullet})$ ,
2. Gaussian distribution:  $Y_{ij} \mid \{i \in \bullet, j \in \bullet\} \sim^{\text{ind}} \mathcal{N}(\mu_{\bullet\bullet}, \sigma^2)$ ,  
[Mariadassou et al., 2010]
3. More generally,

$$Y_{ij} \mid \{i \in \bullet, j \in \bullet\} \sim^{\text{ind}} \mathcal{F}(\theta_{\bullet\bullet})$$

# Taking into account covariates

Sometimes covariates are available. They may be on:

- nodes,
- edges,
- both.

1. They can be used a posteriori to explain blocks inferred by SBM.
2. Extension of the SBM which takes into account covariates. Blocks are structure of interaction which is not explained by covariates !

If covariates are sampling conditions, case 2 be may more interesting.

# SBM with covariates

- As before :  $(Y_{ij})$  be an adjacency matrix
- Let  $x^{ij} \in \mathbb{R}^p$  denote covariates describing the pair  $(i, j)$

## Latent variables : as before

- The nodes  $i = 1, \dots, n$  are partitioned into  $K$  clusters
- $Z_i$  independent variables

$$\mathbb{P}(Z_i = k) = \pi_k$$

## Conditionally to $(Z_i)_{i=1, \dots, n} \dots$

$(Y_{ij})$  independent and

$$Y_{ij} | Z_i, Z_j \sim \text{Bern}(\text{logit}(\alpha_{Z_i, Z_j} + \theta \cdot x_{ij})) \quad \text{if binary data}$$

$$Y_{ij} | Z_i, Z_j \sim \mathcal{P}(\exp(\alpha_{Z_i, Z_j} + \theta \cdot x_{ij})) \quad \text{if counting data}$$

If  $K = 1$  : all the connection heterogeneity is explained by the covariates.

# Multiplex networks

Several kind of interactions between nodes . For instance :

- Love and friendship
- Working relations and friendship
- In ecology : mutualistic and competition

## Block model for multiplex networks

$$Y_{ij} \in \{0, 1\}^Q = (Y_{ij}^a, Y_{ij}^b), \forall w \in \{0, 1\}^2$$

$$\mathbb{P}(Y_{ij}^a, Y_{ij}^b = w | Z_i = k, Z_j = \ell) = \alpha_{k\ell}^w$$

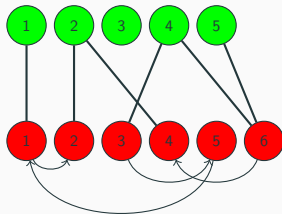
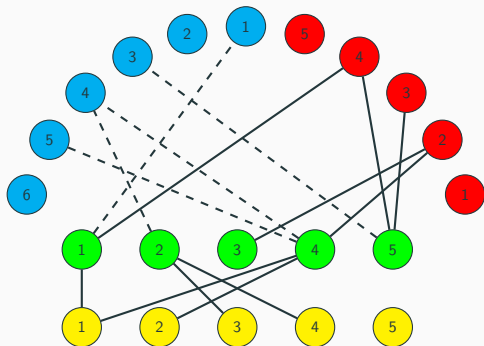
[Kéfi et al., 2016], [Barbillon et al., 2017]

In R package: `blockmodels` when two relations are at stake.

**Remark:** a particular case of multiplex network is dynamic network, [Matias and Miele, 2017].

# Multipartite networks

More than 2 types of nodes : Plant-pollinators, plant-ants interactions for instance



[Kéfi et al., 2016], [Bar-Hen et al., 0]

In R package: GREMLINS and sbm

Introduction

Descriptive statistics

Probabilistic models

Frequentist and Bayesian inference

- Statistical tasks and likelihood

- Frequentist inference

- Bayesian inference

Conclusion

Introduction

Descriptive statistics

Probabilistic models

Frequentist and Bayesian inference

- Statistical tasks and likelihood

- Frequentist inference

- Bayesian inference

Conclusion



- Selection of the number of clusters  $K$  for SBM or  $K_1, K_2$  for LBM
- Estimation of the parameters  $\pi, \theta$  for a given number of clusters
- Clustering  $\hat{\mathbf{Z}}$

## Complete likelihood ( $\mathbf{Y}$ ) et ( $\mathbf{Z}$ )

$$\begin{aligned}\ell_c(\mathbf{Y}, \mathbf{Z}; \theta) &= p(\mathbf{Y}|\mathbf{Z}; \alpha)p(\mathbf{Z}; \pi) \\ &= \prod_{i,j} f_{\alpha_{Z_i, Z_j}}(Y_{ij}) \times \prod_i \pi_{Z_i} \\ &= \prod_{i,j} \alpha_{Z_i, Z_j}^{Y_{ij}} (1 - \alpha_{Z_i, Z_j})^{1-Y_{ij}} \prod_i \pi_{Z_i}\end{aligned}$$

## Marginal likelihood ( $\mathbf{Y}$ )

$$\log \ell(\mathbf{Y}; \theta) = \log \sum_{\mathbf{Z} \in \mathcal{Z}} \ell_c(\mathbf{Y}, \mathbf{Z}; \theta). \quad (1)$$

Introduction

Descriptive statistics

Probabilistic models

**Frequentist and Bayesian inference**

Statistical tasks and likelihood

**Frequentist inference**

Bayesian inference

Conclusion

$$\log \ell(\mathbf{Y}; \theta) = \log \sum_{\mathbf{Z} \in \mathcal{Z}} \ell_c(\mathbf{Y}, \mathbf{Z}; \theta).$$

### Remark

$\mathcal{Z} = \{1, \dots, K\}^n \Rightarrow$  when  $K$  and  $n$  increase, impossible to compute.

**Standard tool to maximize the likelihood when latent variables involved** : EM algorithm.

# From EM to variational EM

## Standard EM

At iteration ( $t$ ) :

- **Step E:** compute

$$Q(\theta|\theta^{(t-1)}) = \mathbb{E}_{\mathbf{Z}|\mathbf{Y},\theta^{(t-1)}} [\log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta)]$$

- **Step M:**

$$\theta^{(t)} = \arg \max_{\theta} Q(\theta|\theta^{(t-1)})$$

# Limitations of standard EM

- Step  $E$  requires the computation of  $\mathbb{E}_{\mathbf{Z}|\mathbf{Y}, \theta^{(t-1)}} [\log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta)]$
- However, once conditioned by par  $\mathbf{Y}$ , the  $\mathbf{Z}$  are not independent anymore: complex distribution if  $K$  and  $n$  big.

# Variational EM : maximization of a lower bound

**Idea** : replace the complicated distribution  $p(\cdot|\mathbf{Y};\theta) = [\mathbf{Z}|\mathbf{Y},\theta]$  by a simpler one.

Let  $\mathcal{R}_{\mathbf{Y},\tau}$  be any distribution on  $\mathbf{Z}$

## Central identity

$$\begin{aligned}\mathcal{I}_{\theta}(\mathcal{R}_{\mathbf{Y},\tau}) &= \log \ell(\mathbf{Y};\theta) - \text{KL}[\mathcal{R}_{\mathbf{Y},\tau}, p(\cdot|\mathbf{Y};\theta)] \leq \log \ell(\mathbf{Y};\theta) \\ &= \mathbb{E}_{\mathcal{R}_{\mathbf{Y},\tau}} [\log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta)] - \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{Y},\tau}(\mathbf{Z}) \log \mathcal{R}_{\mathbf{Y},\tau}(\mathbf{Z}) \\ &= \mathbb{E}_{\mathcal{R}_{\mathbf{Y},\tau}} [\log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta)] + \mathcal{H}(\mathcal{R}_{\mathbf{Y},\tau}(\mathbf{Z}))\end{aligned}$$

**Note that:**

$$\mathcal{I}_{\theta}(\mathcal{R}_{\mathbf{Y},\tau}) = \log \ell(\mathbf{Y};\theta) \Leftrightarrow \mathcal{R}_{\mathbf{Y},\tau} = p(\cdot|\mathbf{Y};\theta)$$

- Maximization of  $\log \ell(\mathbf{Y}; \theta)$  w.r.t.  $\theta$  replaced by maximization of the lower bound  $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{Y}, \tau})$  w.r.t.  $\tau$  and  $\theta$ .
- **Benefit** : we choose  $\mathcal{R}_{\mathbf{Y}, \tau}$  such that the maximization calculus can be done explicitly
  - In our case: mean field approximation : neglect dependencies between the  $(Z_i)$

$$P_{\mathcal{R}_{\mathbf{Y}, \tau}}(Z_i = k) = \tau_{ik}$$



## Algorithm

At iteration  $(t)$ , given the current value  $(\theta^{(t-1)}, \mathcal{R}_{\mathbf{Y}, \tau^{(t-1)}})$ ,

- **Step 1** Maximization w.r.t.  $\tau$

$$\begin{aligned}\tau^{(t)} &= \arg \max_{\tau \in \mathcal{T}} \mathcal{I}_{\theta^{(t-1)}}(\mathcal{R}_{\mathbf{Y}, \tau}) \\ &= \arg \min_{\tau \in \mathcal{T}} \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \tau}, p(\cdot | \mathbf{Y}; \theta^{(t-1)})]\end{aligned}$$

- **Step 2** Maximization w.r.t.  $\theta$

$$\begin{aligned}\theta^{(t)} &= \arg \max_{\theta} \mathcal{I}_{\theta}(\mathcal{R}_{\mathbf{Y}, \tau^{(t)}}) \\ &= \arg \max_{\theta} \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau^{(t)}}} [\log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta)]\end{aligned}$$

- Really fast
- Strongly depend on the initial values

# Penalized likelihood criterion

- Selection of the number of clusters  $K$  (or  $K_1, K_2$  in the LBM)
- Integrated Classification Likelihood (ICL) [Biernacki et al., 2000]

$$ICL(\mathcal{M}_{\mathbf{K}}) = \log \ell_c(\mathbf{Y}, \hat{\mathbf{Z}}; \hat{\theta}_{\mathbf{K}}) - \text{pen}(\mathcal{M}_{\mathbf{K}}) \quad (2)$$

where

$$\hat{Z}_i = \arg \max_{k \in \{1, \dots, K\}} \hat{\tau}_{ik}. \quad (3)$$

- Integrated Complete Likelihood (ICL)

$$ICL(\mathcal{M}_{\mathbf{K}}) = \mathbb{E}_{p(\cdot | \mathbf{Y}, \hat{\theta}_{\mathbf{K}})} [\log \ell_c(\mathbf{Y}, \hat{\mathbf{Z}}; \hat{\theta}_{\mathbf{K}}) - \text{pen}(\mathcal{M}_{\mathbf{K}})] \quad (4)$$

# Expression of the penalization

## For SBM

$$pen_{\mathcal{M}} = \begin{cases} -\frac{1}{2} \{ (K-1) \log(n) + K^2 \log(n^2 - n) \} & \text{for directed network} \\ -\frac{1}{2} \left\{ \underbrace{(K-1) \log(n)}_{\text{Clust.}} + \frac{K(K+1)}{2} \log\left(\frac{n^2-n}{2}\right) \right\} & \text{for undirected network} \end{cases}$$

## For LBM

$$pen_{\mathcal{M}} = -\frac{1}{2} \left\{ \underbrace{(K_1-1) \log(n_1) + (K_2-1) \log(n_2)}_{\text{Bi-Clust.}} + \underbrace{(K_1 K_2) \log(n_1 n_2)}_{\text{Connection}} \right\}$$

# Advantages of ICL

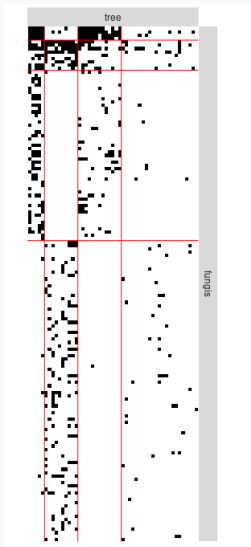
- its capacity to outline the clustering structure in networks
- Involves a trade-off between goodness of fit and model complexity
- ICL values : goodness of fit AND clustering sharpness.

# Application : interactions Tree-fungi



- **Fungi:** *Amphiporte leiphaemia*, *Apiognomonina errabunda*, *Apiognomonina veneta* , *Armillaria cepistipes* , *Armillaria gallica*,...
- **Trees:** *Abies alba*, *Abies grandis*, *Abies nordmanniana* , Large Maples (*Acer platanoides*, *Acer pseudoplatanus*) , Small Maples

## Application : reordered Tree-fungi



Introduction

Descriptive statistics

Probabilistic models

Frequentist and Bayesian inference

Statistical tasks and likelihood

Frequentist inference

Bayesian inference

Conclusion



$$\begin{aligned}(\pi_1, \dots, \pi_K) &\sim \text{Dir}(a_1, \dots, a_K) \\ \alpha_{kl} &\sim \text{Beta}(a_0, b_0)\end{aligned}$$

# Posterior sampling

- **Solution 1** : any MCMC sampler
- **Solution 2** : variational Bayes

[Latouche et al., 2012]

- **Idea:** approximate  $p(\mathbf{Z}, \theta | \mathbf{Y})$  by a simpler one  $q(\mathbf{Z}, \theta)$
- Approximating = Minimizing

$$D_{\text{KL}}(q(\mathbf{Z}, \theta), p(\mathbf{Z}, \theta | \mathbf{Y})) = \mathbf{E}_q \left[ \log \frac{q(\mathbf{Z}, \theta)}{p(\mathbf{Z}, \theta | \mathbf{Y})} \right]$$

# The same magik trick

$$D_{\text{KL}}(q(\mathbf{Z}, \theta), p(\mathbf{Z}, \theta | \mathbf{Y})) = \log \ell(\mathbf{Y}) - \underbrace{\left[ \mathbf{E}_q[\log \ell(\mathbf{Y} | \mathbf{Z}, \theta) p(\mathbf{Z} | \theta) \pi(\theta)] - \mathbf{E}_q[\log q(\mathbf{Z}, \theta)] \right]}_{\mathcal{F}(q)}$$

- $\log \ell(\mathbf{Y})$  independent of  $q$
- Minimizing the KL w.r. to  $q$  is equivalent to maximizing  $\mathcal{F}(q)$  with respect to  $q$

Choose a form for

$$q(\mathbf{Z}, \theta) = q_{\tau}(\mathbf{Z}) q_{Dir}(\pi) q_{Beta}(\alpha)$$

- Iteratively maximization in the various parameters
- Conjugacy helps to be able to explicitly maximize

# Bayesian inference for more complex models

- In presence of covariates, no more explicit maximization.
- **Our solution**
  - Build a pseudo posterior distribution using the Variational EM algorithm with estimation of the variance with Louis formula
  - Use this first posterior approximation to accelerate a sequential Monte Carlo sampler to get a sample from the true posterior
  - [Donnet and Robin, 2021]

Introduction

Descriptive statistics

Probabilistic models

Frequentist and Bayesian inference

Conclusion

## SBM/LBM

- generative models,
- flexible,
- comprehensive models which can be linked to a lot of classical descriptors.



## Other extensions

- Time evolving networks [Matias](#)
- Multipartite, Multiplexe networks
- Multilevel networks (individuals and organizations) [\[Chabert-Liddell et al., 2021b\]](#)
- Collections of networks ([on going work by CL](#))
- Link with robustness of ecological networks [\[Chabert-Liddell et al., 2021a\]](#)
- Missing data in the network [\[Tabouy et al., 2020\]](#)
- R-package [sbm](#)

# References i



Bar-Hen, A., Barbillon, P., and Donnet, S. (0).

**Block models for generalized multipartite networks: Applications in ecology and ethnobiology.**

*Statistical Modelling*, 0(0):1471082X20963254.



Barbillon, P., Donnet, S., Lazega, E., and Bar-Hen, A. (2017).

**Stochastic block models for multiplex networks: an application to a multilevel network of researchers.**

*Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(1):295–314.



Bickel, P., Choi, D., Chang, X., Zhang, H., et al. (2013).

**Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels.**

*The Annals of Statistics*, 41(4):1922–1943.



Biernacki, C., Celeux, G., and Govaert, G. (2000).

**Assessing a mixture model for clustering with the integrated completed likelihood.**

*IEEE transactions on pattern analysis and machine intelligence*, 22(7):719–725.



Celisse, A., Daudin, J.-J., and Pierre, L. (2012).

**Consistency of maximum-likelihood and variational estimators in the stochastic block model.**

*Electronic Journal of Statistics*, 6:1847–1899.



Chabert-Liddell, S.-C., Barbillon, P., and Donnet, S. (2021a).

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

*Environmetrics*.



Chabert-Liddell, S.-C., Barbillon, P., Donnet, S., and Lazega, E. (2021b).

**A stochastic block model approach for the analysis of multilevel networks: An application to the sociology of organizations.**

*Computational Statistics & Data Analysis*, 158:107179.



Clauset, A., Moore, C., and Newman, M. E. (2008).  
**Hierarchical structure and the prediction of missing links in networks.**  
*Nature*, 453(7191):98.



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



Donnet, S. and Robin, S. (2021).  
**Accelerating bayesian estimation for network poisson models using frequentist variational estimates.**  
*Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 70(4):858–885.



Freeman, L. C. (1978).  
**Centrality in social networks conceptual clarification.**  
*Social Networks*, 1(3):215 – 239.



Govaert, G. and Nadif, M. (2008).  
**Block clustering with bernoulli mixture models: Comparison of different approaches.**  
*Computational. Statistics and Data Analysis*, 52(6):3233–3245.



Kéfi, S., Miele, V., Wieters, E. A., Navarrete, S. A., and Berlow, E. L. (2016).  
**How structured is the entangled bank? the surprisingly simple organization of multiplex ecological networks leads to increased persistence and resilience.**  
*PLOS Biology*, 14(8):1–21.



Latouche, P., Birméle, E., and Ambroise, C. (2012).  
**Variational bayesian inference and complexity control for stochastic block models.**  
*Statistical Modelling*, 12(1):93–115.



Mariadassou, M. and Matias, C. (2015).

**Convergence of the groups posterior distribution in latent or stochastic block models.**

*Bernoulli*, 21(1):537–573.



Mariadassou, M., Robin, S., and Vacher, C. (2010).

**Uncovering latent structure in valued graphs: a variational approach.**

*The Annals of Applied Statistics*, 4(2):715–742.



Matias, C. and Miele, V. (2017).

**Statistical clustering of temporal networks through a dynamic stochastic block model.**

*Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4):1119–1141.



Nowicki, K. and Snijders, T. A. B. (2001).

**Estimation and prediction for stochastic blockstructures.**

*Journal of the American Statistical Association*, 96(455):1077–1087.



Rodríguez-Gironés, M. and Santamaría, L. (2006).

**Rodríguez-gironés ma, santamaría l.. a new algorithm to calculate the nestedness temperature of presence-absence matrices. j biogeogr 33: 924-935.**

*Journal of Biogeography*, 33:924 – 935.



Tabouy, T., Barbillon, P., and Chiquet, J. (2020).

**Variational inference for stochastic block models from sampled data.**

*Journal of the American Statistical Association*, 115(529):455–466.

# Annexe 1: Comments on the ICL versus BIC

## Conjecture

$$BIC(\mathcal{M}) = \log \ell(\mathbf{Y}; \hat{\theta}, \mathcal{M}) - \text{pen}(\mathcal{M})$$

with the same penalty

- Under this conjecture

$$\begin{aligned} ICL(\mathcal{M}) &= BIC(\mathcal{M}) + \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{Y}; \hat{\theta}_{\mathbf{K}}) \log p(\mathbf{Z}|\mathbf{Y}; \hat{\theta}_{\mathbf{K}}) \\ &= BIC(\mathcal{M}) - \mathcal{H}(p(\cdot|\mathbf{Y}; \theta)) \end{aligned}$$

- As a consequence, because of the entropy, ICL will encourage clustering with well-separated groups
- 

$$\widehat{ICL}(\mathcal{M}) = BIC(\mathcal{M}) + \sum_{\mathbf{Z}} \mathcal{R}_{\mathbf{Y}}(\mathbf{Z}, \hat{\tau}) \log \mathcal{R}_{\mathbf{Y}, \hat{\tau}}(\mathbf{Z}) - \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \hat{\tau}}, p(\cdot|\mathbf{Y}; \hat{\theta})].$$

## Annexe 2: Algorithm in practice

- Going through the models and initiate VEM at the same time
- Bounds on  $K$  :  $\{K_{\min}, \dots, K_{\max}\}$

### Stepwise procedure

Starting from  $K$

- **Split** : if  $K < K_{\max}$ 
  - Maximize the likelihood (lower bound) of  $\mathcal{M}_{K+1}$
  - $K$  initializations of the VEM are proposed : split each cluster into 2 clusters
- **Merge** : If  $K > K_{\min}$ 
  - Maximize the likelihood (lower bound) of model  $\mathcal{M}_{K-1}$
  - $\frac{K(K-1)}{2}$  initializations of the VEM are proposed : merging all the possible pairs of clusters

## Annexe 3: Theoretical properties for SBM

- Identifiability and a first consistency result by [Celisse et al., 2012]
- Consistency of the posterior distribution of the latent variables [Mariadassou and Matias, 2015]
- Consistency and properties of the variational estimators [Bickel et al., 2013]