

Probabilistic models for networks

Sophie Donnet

26-27 mars 2019, Nantes



Introduction

Descriptive statistics

Probabilistic model

Inference

Model selection

Network data



Networks can account for

- ▶ Ecological network : Food web , Co-existence networks, Host-parasite interactions, Plant-pollinator interactions,
- ▶ Social network
- ▶ Inventory datasets
- ▶ ...

Networks may be or not bipartite : Interactions between nodes belonging to the same or to different functional group(s).

Terminology

A network consists in :

- ▶ 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 / dyad.

A network may be

- ▶ directed / oriented (e.g. food web...),
- ▶ symmetric / undirected (e.g. coexistence network),
- ▶ with or without loops.

This distinction only makes sense for simple networks (not bipartite).

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.

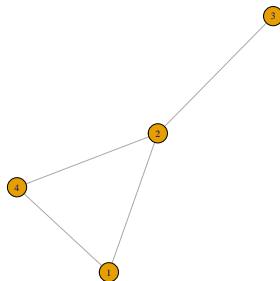
Goal



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

Network representation and adjacency matrix

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



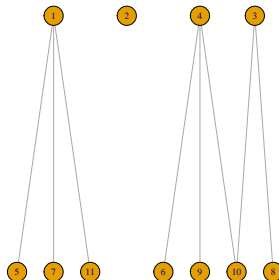
- ▶ n rows and n columns,
- ▶ symmetric or not.

Bipartite network and incidence matrix

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

- n rows and m columns, rectangular matrix.
- corresponding adjacency matrix $(n + m) \times (n + m)$:

$$\begin{pmatrix} 0 & Y \\ Y^T & 0 \end{pmatrix}$$



Introduction

Descriptive statistics

Probabilistic model

Inference

Model selection

Some common features studied on networks

- ▶ 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 & Santamaria \(2006\)](#)
- ▶ Betweenness centrality : for a node, numbers of shortest paths between any pair of nodes passing through this node. [Freeman \(1979\)](#)
- ▶ 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, Newman & Moore \(2004\)](#)

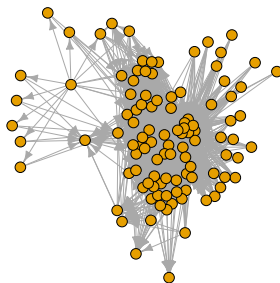
Some common features studied on networks

All this criterion shall be adapted to :

- ▶ directed network,
- ▶ bipartite network.

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

Example Chilean food web

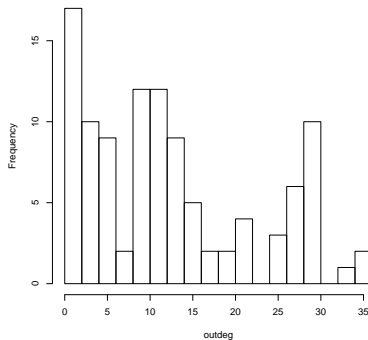


- $n = 106$ species / nodes,
- density of edges : 12.1%.

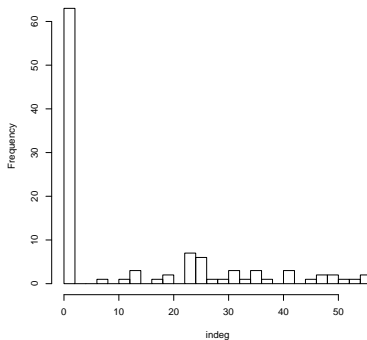
Kéfi, Miele, Wieters, Navarrete & Berlow (2016)

Degree distribution

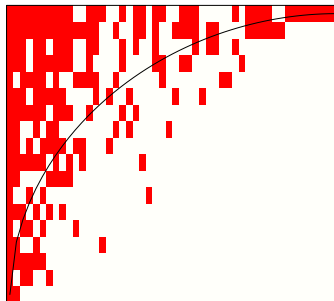
Histogram of outdeg



Histogram of indeg

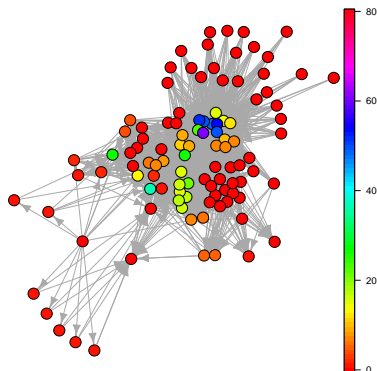


Nestedness



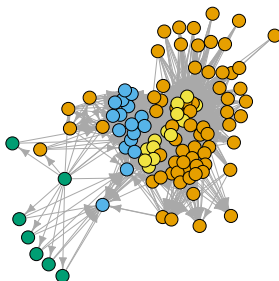
- ▶ more generally used on incidence matrices,
- ▶ significance of the nestedness index computed by random permutations of the matrix,
- ▶ this food web is found to be nested.

Betweenness



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.000	0.000	6.604	6.929	59.570

Modularity



	1	2	3	4
►	69	17	7	13

- ▶ very low modularity.

Introduction

Descriptive statistics

Probabilistic model

- SBM

- LBM

- Some possible extensions

Inference

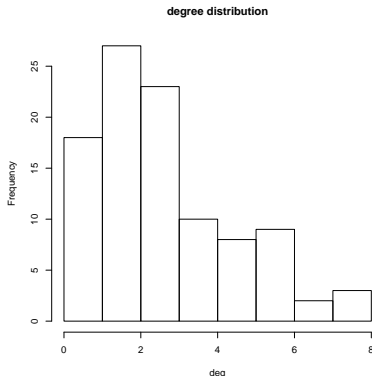
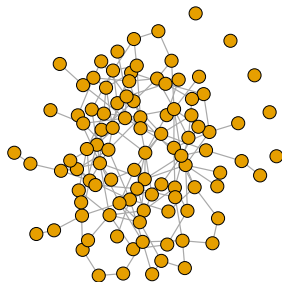
Model selection

A first random graph model for network : null model

Erdős-Rényi (1959) Model for n nodes

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

where b is the Bernoulli distribution and $p \in [0, 1]$ a probability for a link to exist.



Limitations of an ER graph to describe real networks

- ▶ Degree distribution too concentrated, no high degree nodes,
- ▶ All nodes are equivalent (no nestedness...),
- ▶ No modularity.

Stochastic Block Model

Nowicki, & 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 independent variables

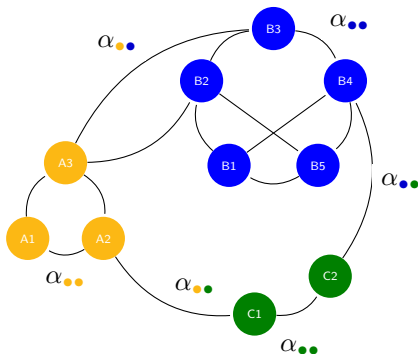
$$\mathbb{P}(Z_i = k) = \pi_k$$

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

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

$$Z_i = \mathbf{1}_{\{i \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \pi), \quad \forall \bullet \in \mathcal{K},$$

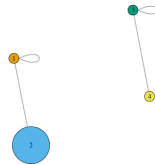
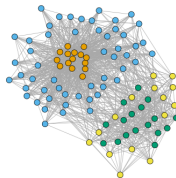
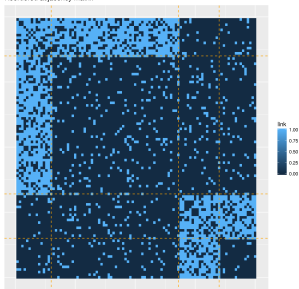
$$Y_{ij} \mid \{i \in \bullet, j \in \bullet\} \sim^{\text{ind}} \mathcal{B}(\alpha_{\bullet, \bullet})$$

SBM : A great generative model

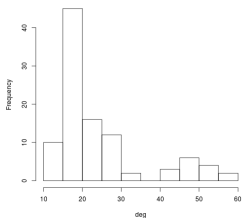
- ▶ Generative model : easy to simulate
- ▶ No a priori on the type of structure
- ▶ Combination of modularity, nestedness, etc...

Networks with hubs generated by SBM

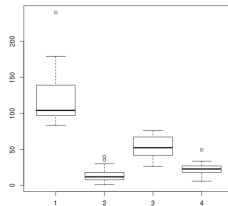
Reordered adjacency matrix



Histogram of degrees

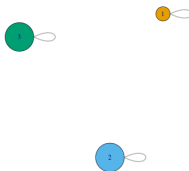
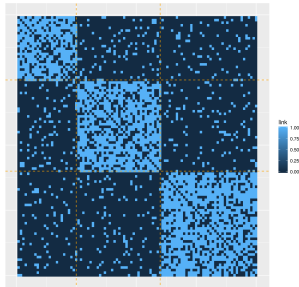


Betweenness by block

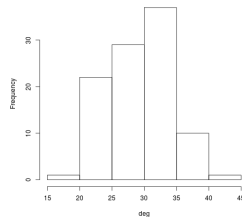


Community network generated by SBM

Reordered adjacency matrix

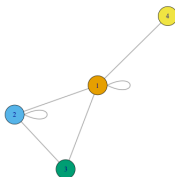
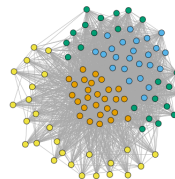
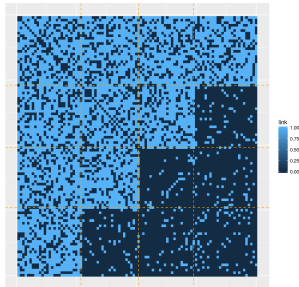


Histogram of degrees

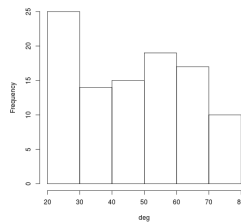


Nestedness generated by SBM

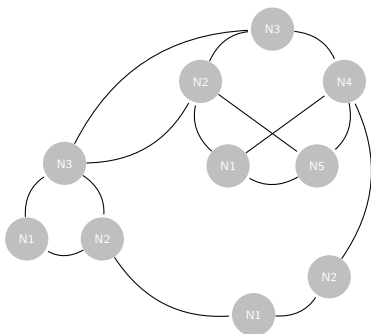
Reordered adjacency matrix



Histogram of degrees



Statistical inference



Stochastic Block Model

Let n nodes divided into

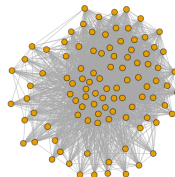
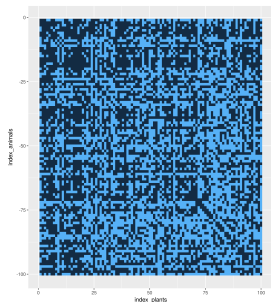
- ▶ $\mathcal{Q} = \{\bullet, \bullet, \bullet\}$, $\text{card}(\mathcal{Q})$ known
- ▶ $\alpha_{\bullet} = ?$,
- ▶ $\pi_{\bullet, \bullet} = ?$

Nowicki, & Snijders (2001), Daudin et al. (2008)

R package : blockmodels.

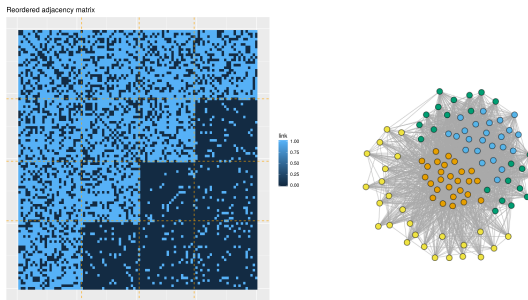
Statistical inference

From....



Statistical inference

... to



Statistician job

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

Details given later.

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
- ▶

$$\begin{aligned} Z_i^1 &= k && \text{if node } i \text{ belongs to cluster (block) } k \\ Z_j^2 &= k && \text{if node } j \text{ belongs to cluster (block) } k \end{aligned}$$
- ▶ Z_i^1, Z_j^2 independant variables

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

Probabilistic model for binary bipartite networks

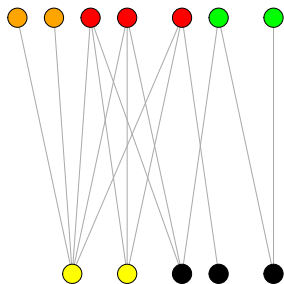
Conditionally to $(Z_i^1)_{i=1,\dots,n_1}, (Z_j^2)_{j=1,\dots,n_2}\dots$

(Y_{ij}) independent and

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

Govaert & Nadif (2008)

Latent Block Model : illustration



Latent Block Model

- ▶ n_1 row nodes $\mathcal{K}_1 = \{\bullet, \bullet, \bullet\}$ classes
- ▶ $\pi_{\bullet}^1 = \mathbb{P}(i \in \bullet), \bullet \in \mathcal{K}_1, i = 1, \dots, n$
- ▶ n_2 column nodes $\mathcal{K}_2 = \{\bullet, \bullet\}$ classes
- ▶ $\pi_{\bullet}^2 = \mathbb{P}(j \in \bullet), \bullet \in \mathcal{K}_2, j = 1, \dots, m$
- ▶ $\alpha_{\bullet, \bullet} = \mathbb{P}(i \leftrightarrow j | i \in \bullet, j \in \bullet)$

$$Z_i = \mathbf{1}_{\{i \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \alpha), \quad \forall \bullet \in \mathcal{Q}_1,$$

$$W_j = \mathbf{1}_{\{j \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \beta), \quad \forall \bullet \in \mathcal{Q}_2,$$

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

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,

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

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

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

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

Kefi et al. (2016), Barbillon et al. (2016)

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

Remark : a particular case of multiplex network is dynamic network, Matias & Miele (2015).

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 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 partitionned 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.

Introduction

Descriptive statistics

Probabilistic model

Inference

Model selection

Statistical Inference

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

Likelihood for SBM

Complète 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})

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

Marginal likelihood : remark

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

Remark

$\mathcal{Z} = \otimes_{q=0 \dots Q} \{1, \dots, K_q\}^{n_q} \Rightarrow$ when K and n increase, impossible to compute.

Standard tool to maximise the likelihood when latent variable :
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 : complexe distribution if K and n big.

Variationnel EM : maximisation 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

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

Note that :

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

Proof

By Bayes

$$\begin{aligned}\log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta) &= \log p(\mathbf{Z}|\mathbf{Y}; \theta) + \log \ell(\mathbf{Y}; \theta) \\ \log \ell(\mathbf{Y}; \theta) &= \log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta) - \log p(\mathbf{Z}|\mathbf{Y}; \theta)\end{aligned}$$

By integration against $\mathcal{R}_{\mathbf{Y}, \tau}$:

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

As a consequence :

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

Variational EM

- ▶ Maximization of $\log \ell(\mathbf{Y}; \theta)$ w.r.t. θ replaced by maximization of the lower bound $\mathcal{I}_\theta(\mathcal{R}_{\mathbf{Y}, \tau})$ w.r.t. τ and θ .
- ▶ **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)

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

Variational EM

Algorithm

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

- **Step 1** Maximization w.r.t. τ

$$\begin{aligned}
 \tau^{(t)} &= \arg \max_{\tau \in \mathcal{T}} \mathcal{I}_{\theta^{(t-1)}}(\mathcal{R}_{\mathbf{Y}, \tau}) \\
 &= \arg \max_{\tau \in \mathcal{T}} \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}} \left[\log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta^{(t-1)}) \right] + \mathcal{H}(\mathcal{R}_{\mathbf{Y}, \tau}(\mathbf{Z})) \\
 &= \arg \max_{\tau \in \mathcal{T}} \log \ell(\mathbf{Y}; \theta^{(t-1)}) - \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \tau}, p(\cdot | \mathbf{Y}; \theta^{(t-1)})] \\
 &= \arg \min_{\tau \in \mathcal{T}} \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \tau}, p(\cdot | \mathbf{Y}; \theta^{(t-1)})]
 \end{aligned}$$

Variational EM

Algorithm

- **Step 2** Maximization w.r.t. θ

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

In practice

- ▶ Really fast
- ▶ Strongly depend on the initial values

Introduction

Descriptive statistics

Probabilistic model

Inference

Model selection

Penalized likelihood criterion

- ▶ Selection of the number of clusters K (or K_1, K_2 in the LBM)
- ▶ Bayesian information criterion (BIC) : approximation of the marginal likelihood $m_c(\mathbf{Y}, \mathbf{Z}; \mathcal{M})$ where the parameters θ have been integrated out with a "well chosen" prior distribution

BIC if \mathbf{Z} are observed

Let $\mathcal{M} = \mathcal{M}_K$

$$\log m_c(\mathbf{Y}, \mathbf{Z}; \mathcal{M}) \approx_{n \rightarrow \infty} \max_{\theta} \log \ell_c(\mathbf{Y}, \mathbf{Z}; \theta, \mathcal{M}) + \text{pen}_{\mathcal{M}}$$

Expression of the penalization if \mathbf{Z} are observed

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

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

Penalized criterion when \mathbf{Z} non-observed : ICL

- ▶ Imputation of the \mathbf{Z} par the MAP [?]

$$\hat{\mathbf{Z}} = \arg \max_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{Y}; \hat{\theta}, \mathcal{M}) \approx \arg \max_{\mathbf{Z}} \mathcal{R}_{\hat{\tau}}(\mathbf{Z}|\mathbf{Y}; \hat{\theta}, \mathcal{M})$$

- ▶ Integration of the \mathbf{Z} [?, ?]

- ▶ $ICL(\mathcal{M}) = E_{\mathbf{Z}|\mathbf{Y}; \hat{\theta}, \mathcal{M}} \left[\log p_c(\mathbf{Y}, \mathbf{Z}; \hat{\theta}, \mathcal{M}) \right] + pen_{\mathcal{M}}$
- ▶ $p(\mathbf{Z}|\mathbf{Y}; \hat{\theta}, \mathcal{M}) \Rightarrow \mathcal{R}_{\mathbf{Y}, \hat{\tau}}$
- ▶ $\widehat{ICL}(\mathcal{M}_{\mathbf{K}}) = E_{\mathcal{R}_{\mathbf{Y}, \hat{\tau}}} [\log p_c(\mathbf{Y}, \mathbf{Z}; \theta, \mathcal{M})] + pen_{\mathcal{M}}$

Comments on the ICL versus BIC

Conjecture

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

with the same penalty

- Under this conjecture

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

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



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

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

Other extensions

- ▶ Time evolving networks [Matias](#)
- ▶ Multipartite networks ([R-package GREMLIN](#), [Bar-Hen](#), [Barbillon](#), [Donnet](#))
- ▶ Multilevel networks (individuals and organisations) ([Chabbert-Liddell](#))
- ▶ Missing data in the network,

Probabilistic model for networks in a nutshell

SBM/LBM

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