Probabilistic models for ecological networks

Sophie Donnet, Pierre Barbillon & Avner Bar-Hen

14/06/17

Outline

Network data

Networks can account for

- Food web,
- Co-existence networks,
- Host-parasite interactions,
- Plant-pollinator interactions,
- ...

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 reprensent individuals / species 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 and goal

Available data:

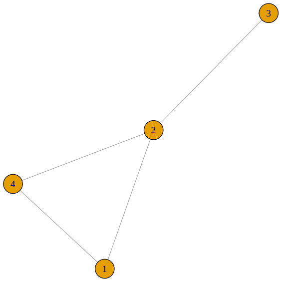
- the network provided as:
  - an adjacency matrix (for simple network) or an incidence matrix (for bipartite network),
  - a list of pair of nodes / dyads which are linked.
- some additional covariates on nodes, dyads which can account for sampling effort.

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

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



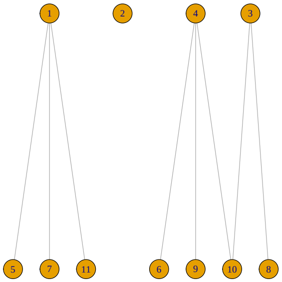
- n rows and n columns,
- symmetric or not.

Bipartite network and incidence matrix

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

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

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



Some common traits studied on networks

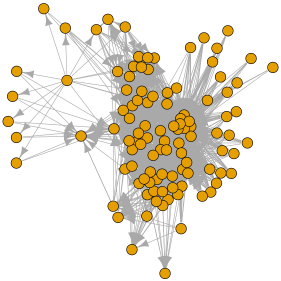
- Degree distribution, can be viewed as a measure of heterogeneity,
- Nestedness: a network is said to be nested when its nodes that have the smallest degree, are connected to nodes with the highest degree, Rodríguez-Gironés & Santamaria (2006)
- Betweenness centrality: for a node, numbers of shortest 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)

All this criterion shall be adapted to:

- directed network,
- bipartite network.

R packages: igraph, sna, vegan.

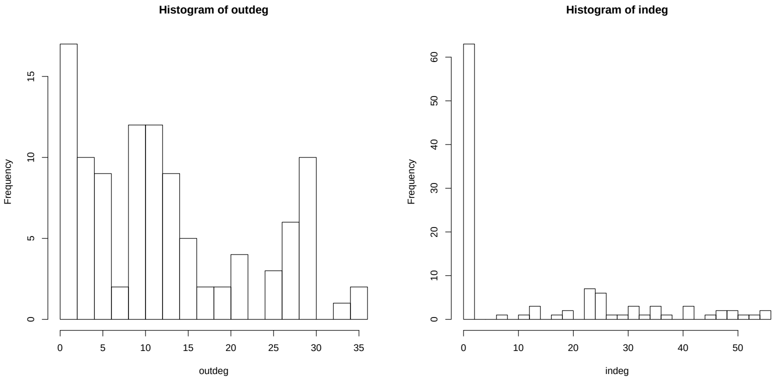
Example Chilean food web



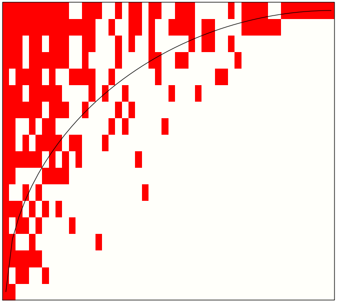
- $n = 106$ species / nodes,
- density of edges: 12.1%.

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

Degree distribution

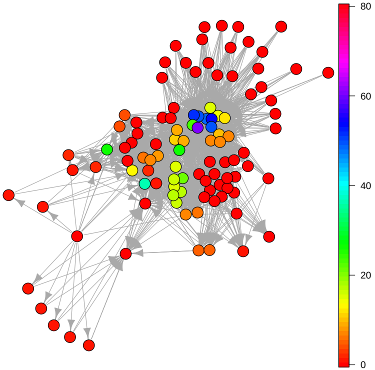


Nestedness



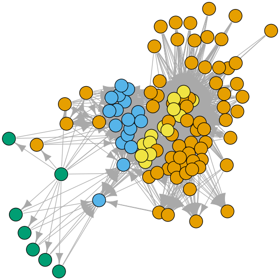
- more generally used on incidence matrices,
- significance of the nestedness index computed by random permutations of the matrix,
- this food web is found to be nested.

Betweenness




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

Modularity



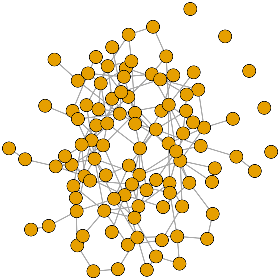
1	2	3	4
69	17	7	13
- very low modularity.

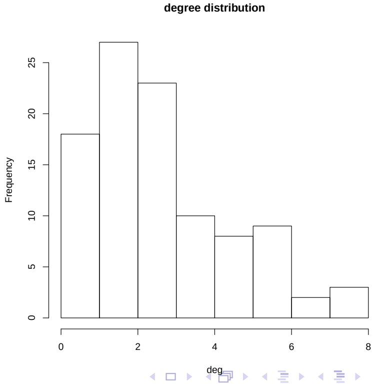
A first random graph model for network: Null model

Erdős-Rényi (1959) Model for n nodes

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

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



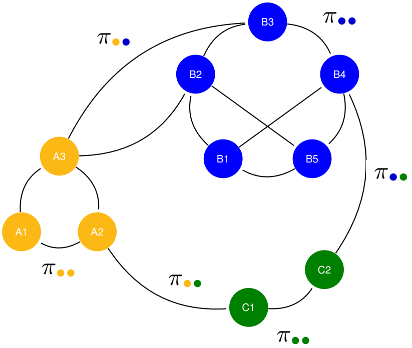


Limitations of an ER graph to describe real networks

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

Outline

Stochastic Block Model



Stochastic Block Model

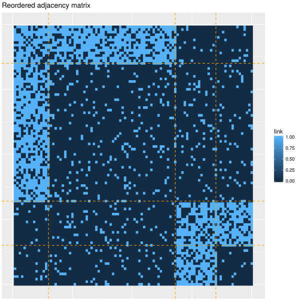
Let n nodes divided into

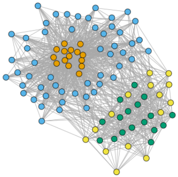
- $\mathcal{Q} = \{\bullet, \bullet, \bullet\}$ classes
- $\alpha_\bullet = \mathbb{P}(i \in \bullet), \bullet \in \mathcal{Q}, i = 1, \dots, n$
- $\pi_{\bullet\bullet} = \mathbb{P}(i \leftrightarrow j | i \in \bullet, j \in \bullet)$

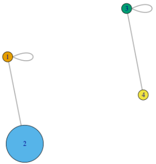
$$Z_i = \mathbf{1}_{\{i \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \alpha), \quad \forall \bullet \in \mathcal{Q},$$

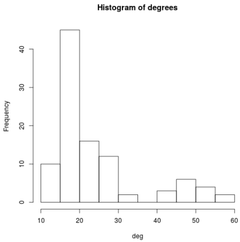
$$X_{ij} \mid \{i \in \bullet, j \in \bullet\} \sim^{\text{ind}} \mathcal{B}(\pi_{\bullet\bullet})$$

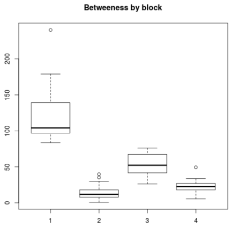
Some remarkable structure generated with SBM : networks with hubs



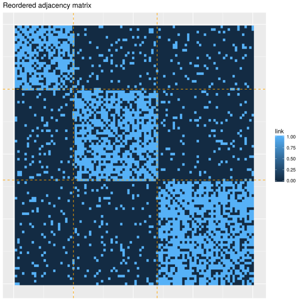


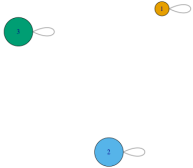


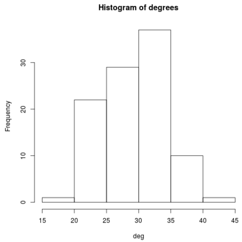




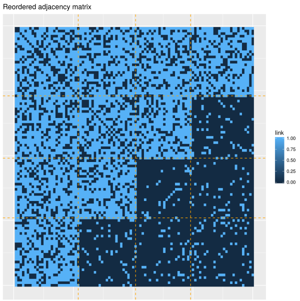
Some remarkable structure generated with SBM : community network

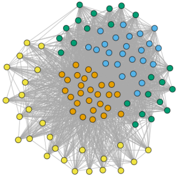


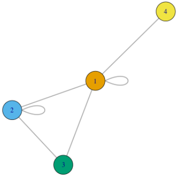


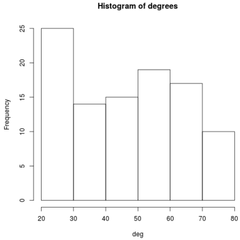


Some remarkable structure generated with SBM : nestedness

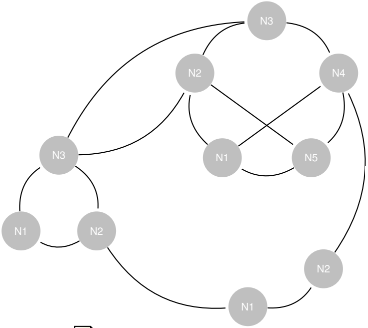








Statistical inference



Stochastic Block Model

Let *n* nodes divided into

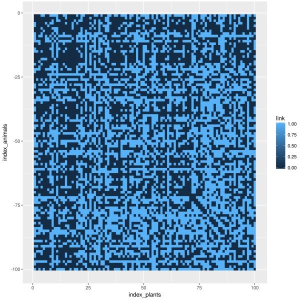
- $\mathcal{Q} = \{{\color{orange}\bullet}, {\color{blue}\bullet}, {\color{green}\bullet}\}$, card($\mathcal{Q}$) known
- $\alpha_{\bullet} = ?$,
- $\pi_{{\color{orange}\bullet}{\color{blue}\bullet}} = ?$

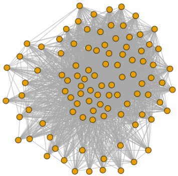
Nowicki, Snijders, JASA, 2001
Estimation and prediction for stochastic blockstructures.

Daudin, Picard, Robin, Statistics and Computing, 2008
A mixture model for random graphs.

Statistical inference

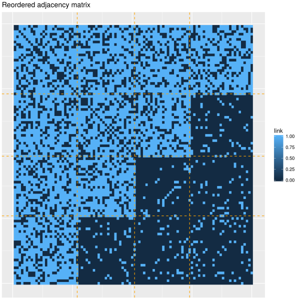
From....

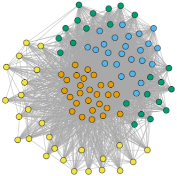




Statistical inference

... to

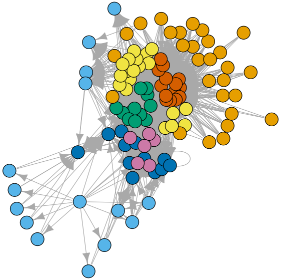




Statistician job

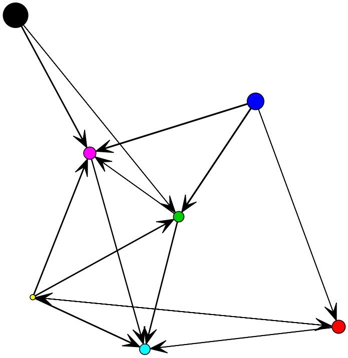
- Find the clusters
- Find the number of clusters
- Theoretical and practical constraints

Application to the Chilean food web



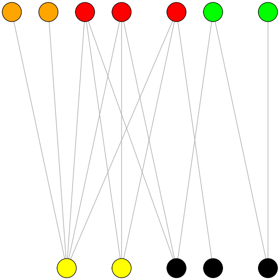
- 7 groups/blocks/clusters found,
1	2	3	4	5	6	7
28	15	12	19	12	14	6

Application to the Chilean food web



Outline

Latent Block Model



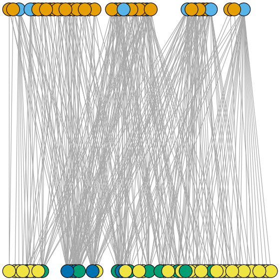
Latent Block Model

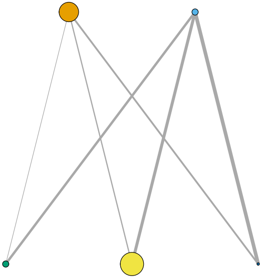
- n row nodes $\mathcal{Q}_1 = \{\bullet, \bullet, \bullet\}$ classes
- $\alpha_\bullet = \mathbb{P}(i \in \bullet), \bullet \in \mathcal{Q}_1, i = 1, \dots, n$
- m column nodes $\mathcal{Q}_2 = \{\bullet, \bullet\}$ classes
- $\beta_\bullet = \mathbb{P}(j \in \bullet), \bullet \in \mathcal{Q}_2, j = 1, \dots, m$
- $\pi_{\bullet\bullet} = \mathbb{P}(i \leftrightarrow j | i \in \bullet, j \in \bullet)$

$$Z_i = \mathbf{1}_{\{i \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \alpha), \quad \forall \bullet \in \mathcal{Q}_1,$$
$$W_j = \mathbf{1}_{\{j \in \bullet\}} \sim^{\text{iid}} \mathcal{M}(1, \beta), \quad \forall \bullet \in \mathcal{Q}_2,$$
$$X_{ij} \mid \{i \in \bullet, j \in \bullet\} \sim^{\text{ind}} \mathcal{B}(\pi_{\bullet\bullet})$$

Govaert & Nadif (2008)

LBM for ant-plant data





- 2 blocks found over the 41 ant species,
- 3 blocks found over the 51 plant species.

Blüthgen, Stork & Fiedler (2004)

Outline

- Valued edges: abundance count, weighted interactions...
- multiple interactions between nodes,
- multipartite networks: plants, pollinator, seed dispersers, ants...
- Taking into account sampling conditions (through covarites...).