Bipartite networks

S. Donnet
INRA / AgroParisTech / univ. Paris-Saclay

Netbio, Octobre 2019

Introduction

- Focus of biologists shifted from the study of isolated biological component to the study of complex biological system.
- Graphs widely used to represent bio-entities (proteins, genes, small molecules...) as nodes and their interactions as edges.
- Special focus on bipartite networks

Reference

Pavlopoulos & al. *Bipartite graphs in systems biology and medicine : a survey of methods and applications* in GigaScience (2018)

A large variety of networks

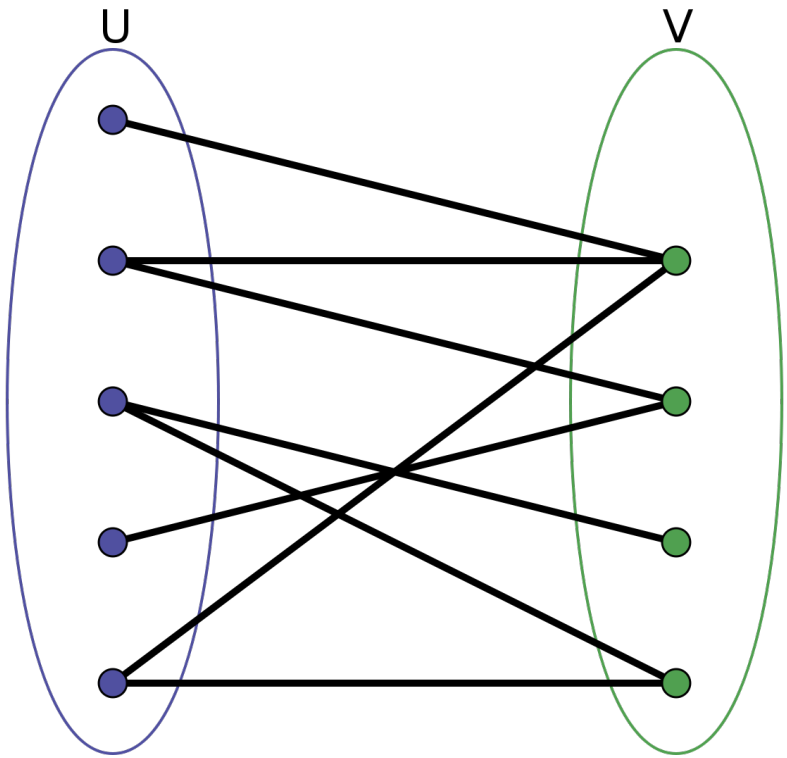
Depending on which biological entities and interactions are at stake : differents types of networks

- Proteins-Proteins interactions : simple undirected networks
- Gene regulation networks : simple directed networks
- Gene expression networks : weighted networks
- Multi-edged networks :
  - When nodes are connected in multiple ways
  - For instance : 2 proteins may interact physically and/or have a certain degree of sequence similarity

Bipartite networks

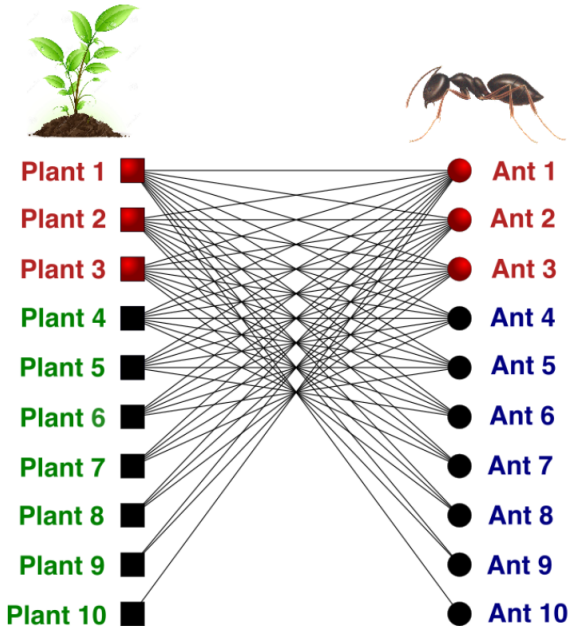
Definition

Bipartite networks (also refered as bigraphs) are graphs such that the nodes (vertices) are divided into two disjoint sets (U and V) and such that all the edges link one node from U and one node of V.



Examples of bipartite networks in ecology

Node = specie



By W. Dattilo

- Plant-pollinator network, (mutualistic relation)
- Plant-ant network (mutualistic relation)
- Host-Pathogen interactions (e.g. tree-fungus)

Examples of biomedical networks

- U = genes, drugs, environmental exposures
- V = diseases, symptoms, adverse drug effects
- Drug - protein target interactions, gene-drug interactions

More abstract networks

- Biomedical field now uses methods of network analysis to model factors that influence human diseases
- Traditionally analyzed with standard methods
- But : Networks offer a way to explore not only the molecular complexity of ONE disease but also the molecular relationships among diseases.
- Aim : design new therapeutic strategy

Examples of Biomolecular bipartite networks

- Representing interactions between biological molecules.
- In general, reconstructed networks from multi-omics data.
- Peptide / protein : edge : peptide involved in protein. Proteins may share peptides.
- Protein/ complexes : participation of proteins in identified complexes
- Gene expression regulation network : regulatory genes and target genes

Scope of the talk

- Object : bipartite network
- Directly observed or previously inferred
- Aim : understanding the structure of the network i.e.
  - we assume that all the nodes of the network do not play the same role and
  - we want to unravel these complicated structures (existence of specialists, communities, star...) $\rightarrow$ topology

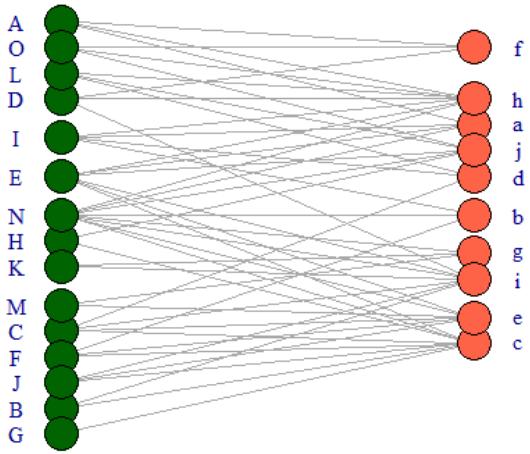
Matrix Representation

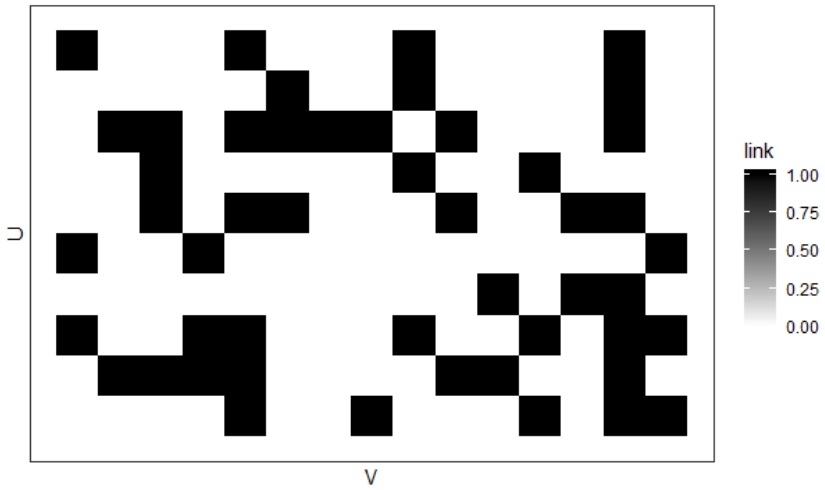
- Bipartite networks can be represented by an incidence matrix or bi-adjacency matrix
- For $i \in U, j \in V$,

$$Y_{ij} = \begin{cases} 1 & \text{if there is an interaction between } i \text{ and } j \\ 0 & \text{otherwise.} \end{cases}$$

- Rectangular matrix
- In most cases : $Y_{ij} \in \{0, 1\}$. However, sometimes $Y_{ij} \in \mathbb{R}$, weighted bipartite graph
- Directed bipartite graph : not classical. Proposition $Y_{ij} \in \{-1, 0, 1\}$

Matrix Representation





Projection I

Going back to simple graph

- From a bipartite network, one can define two (or more) simple graph where each one involves one set of nodes among the two sets.
- Because tools already developed for simple graph

- First projection $X^U = YY'$: network among U

$$\begin{aligned} X^U_{ii'} &= \sum_{j=1}^{|V|} Y_{ij} Y'_{ji'} = \sum_{j=1}^{|V|} Y_{ij} Y_{i'j} \\ &= \text{number of shared connections between } i \text{ and } i'. \end{aligned}$$

Host - parasite : number of parasites shared by any two species.

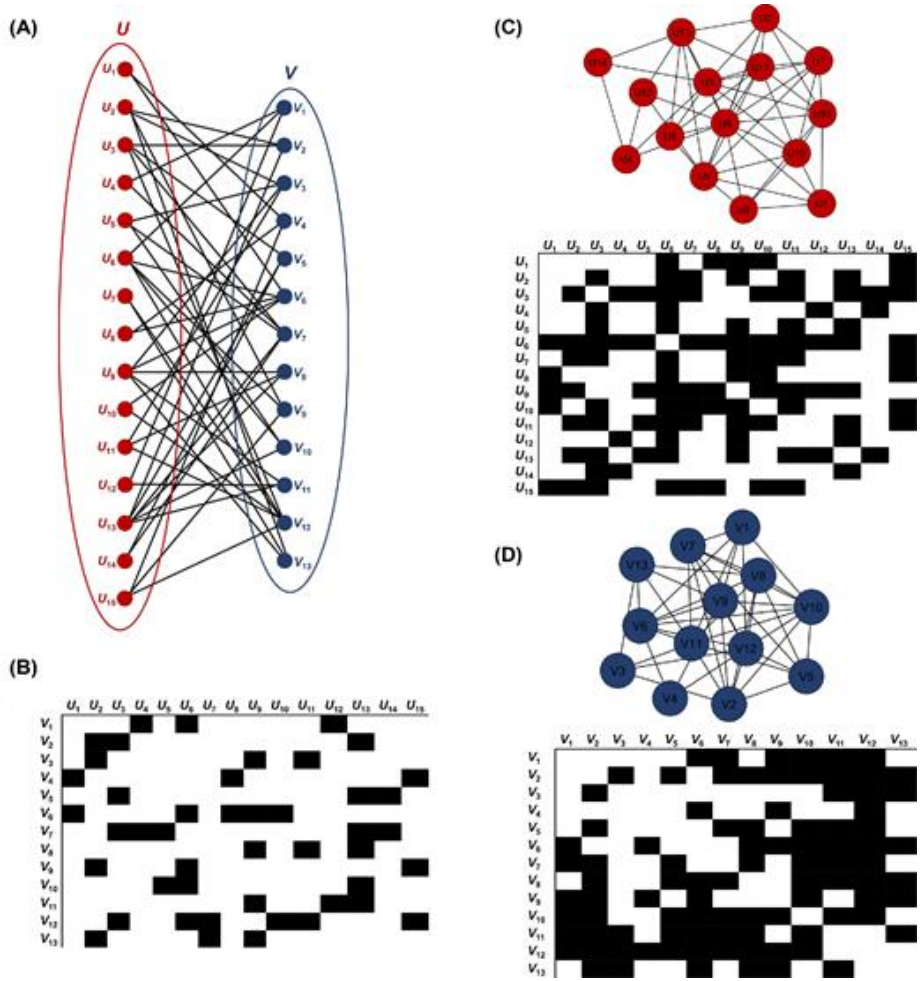
Projection II

- Second projection $X^V = Y'Y$: network among V

$$\begin{aligned} X^V_{jj'} &= \sum_{i=1}^{|U|} Y'_{ji} Y_{ij'} = \sum_{i=1}^{|U|} Y_{ij} Y_{ij'} \\ &= \text{number of shared connections between } j \text{ and } j'. \end{aligned}$$

Host - parasite : number of common species infested by any pair of parasites.

Projection III



[PKP+18]

Projection IV

Limitations

- Loosing a lot of information
- Meaning if Y is weighted ?
- Topology on X^U / X^V and Y difficult to relate ?

Metrics

Aim : give a short description of the network, give a hint about its structure, look for heterogeneity in the connections

- Many metrics supplied for simple networks
- Have been extended to bipartite networks

Libraries

R-packages

Name	Usage
Networksis	Tool to simulate bipartite networks
enaR	Provides algorithms for the analysis of ecological networks
Netpredictor	Prediction of missing links in any given bipartite network
biGRAPH	Extension of the igraph library for bipartite graphs
bipartite	Visualising Bipartite Networks and Calculating Some (Ecological) Indices

Degree

$$\begin{array}{rclcrcl} \deg(u) & = & \sum_{v \in V} (u \leftrightarrow v), & & \deg(v) & = & \sum_{u \in U} (u \leftrightarrow v) \\ \deg_i & = & \sum_{j=1}^{|V|} Y_{ij} & & \deg_j & = & \sum_{i=1}^{|U|} Y_{ij} \end{array}$$

- Nodes with high degree are hubs
- Nodes with null degree are isolated
- If edges are oriented : in- and out- degrees can be computed.

Closeness centrality

Property on a node

Definition

Determine whether a node can communicate with other nodes of the network directly or through the short paths.

$$C(u) = \frac{1}{\sum_{w \in U \cup V} d(u, w)}$$

where $d(u, w)$ is the length of the shortest path between u and w (through the network).

Note that, for bipartite networks

- A node $u \in U$ can have a minimum distance of 1 with $v \in V$.
- A node $u \in U$ can have a minimum distance of 2 with $u' \in U$.
- All paths between nodes of the same set are of even length.

Betweenness centrality

Property on a node

Definition

Betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

The betweenness of a vertex v is computed as follows.

- For each pair of vertices (w, w'), compute the shortest paths between them. $\delta_{w,w'}$ is the number of shortest paths between (w, w')
- For each pair of vertices (w, w'), determine the fraction of shortest paths that pass through v : $\frac{\delta_{w,w'}(v)}{\delta_{w,w'}}$
- Sum this fraction over all pairs of vertices (w, w').

$$B(v) = \sum_{w \neq w' \neq v} \frac{\delta_{w,w'}(v)}{\delta_{w,w'}}$$

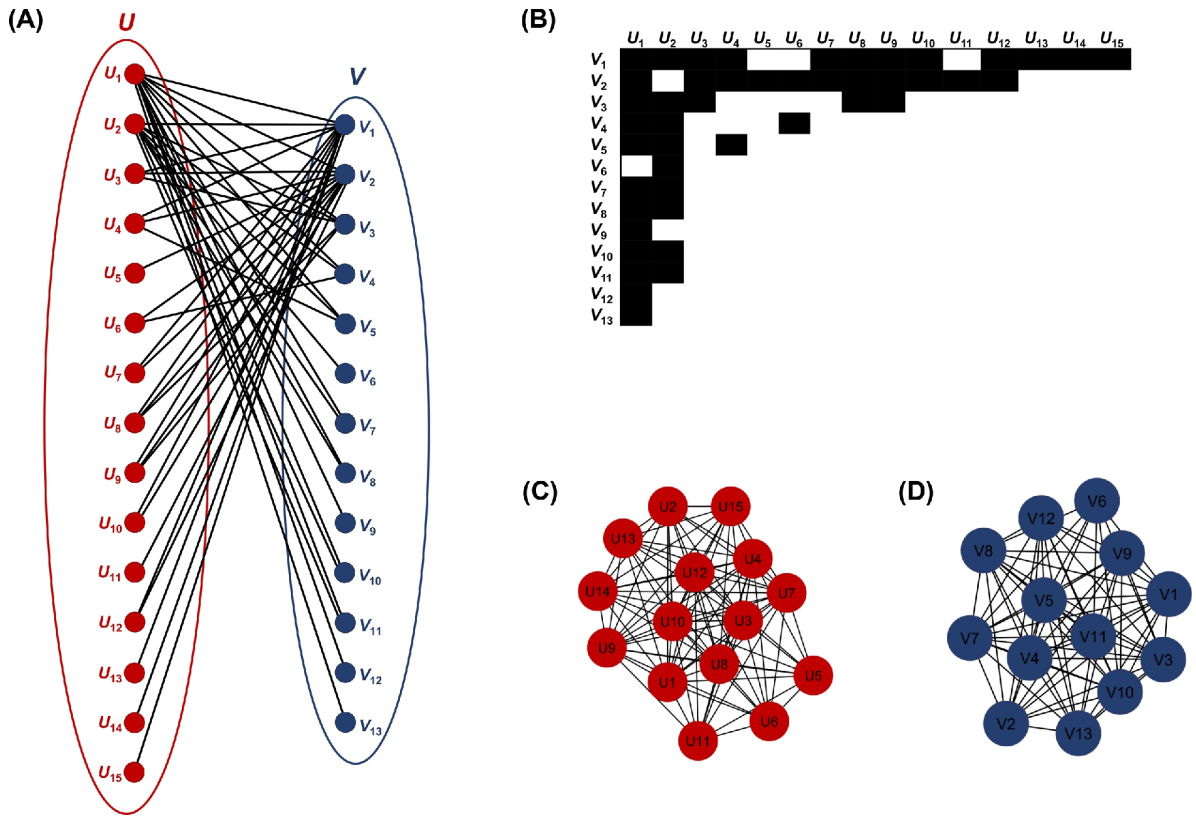
Nestedness

Property on the network

Definition

- Important property in ecology
- Defined as a pattern of interactions in which specialists (e.g. pollinators that visit few plant species) interact with plants that are visited by generalists.
- Mathematically, looking for a reordering of rows and columns such that Y is nested

Nestedness



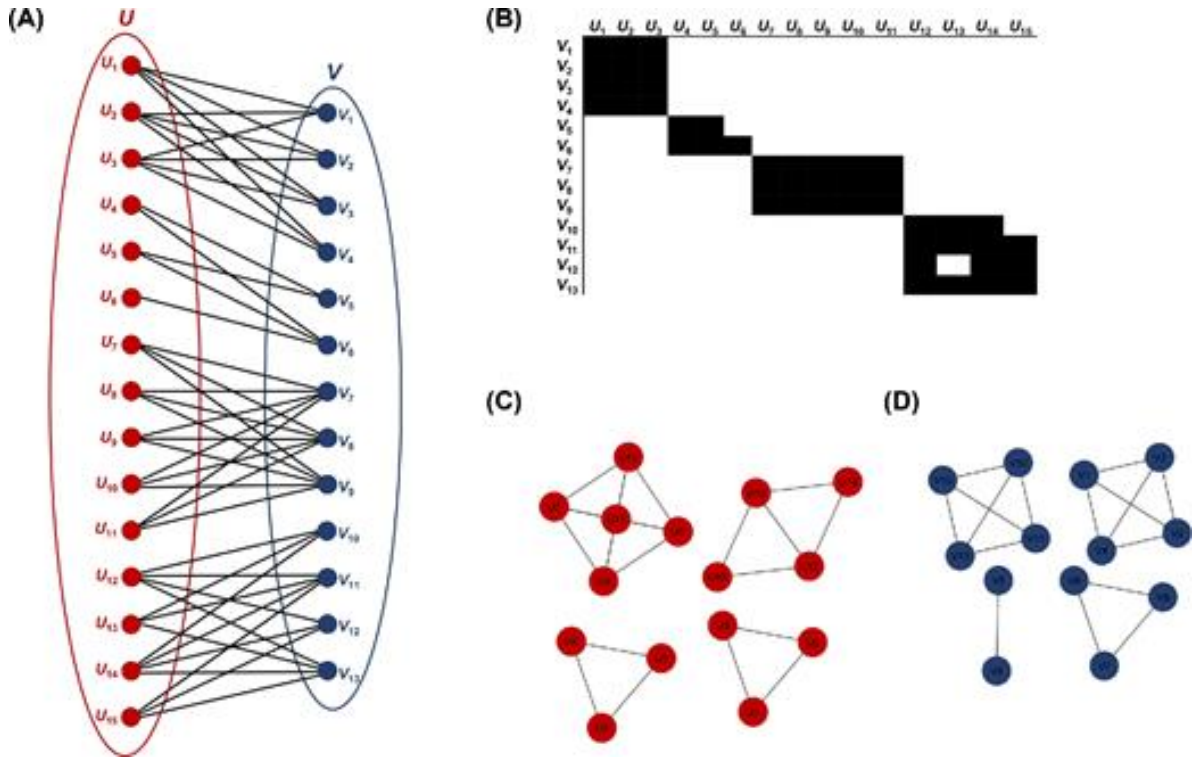
[PKP+18]

Modularity

Property on the network

Definition

Existence of clusters (blocks, module, communities) where nodes are much more connected than with other clusters



[PKP+18]

Bipartivity

- Consider a simple network with no prior partition of the nodes
- This network may be close to a bipartite network.
- Example : assume the nodes are men and women and edges represent sexual relationships. The resulting network is not exactly bipartite but not far from it.
- Measures of bipartivity exist : [PKP+18] are references there in (quite complex).

A first probabilistic model

- Context : our incidence matrix Y is the realization of a stochastic process.
- Aim : Propose a stochastic process is able to mimic heterogeneity in the connections.

Naive model

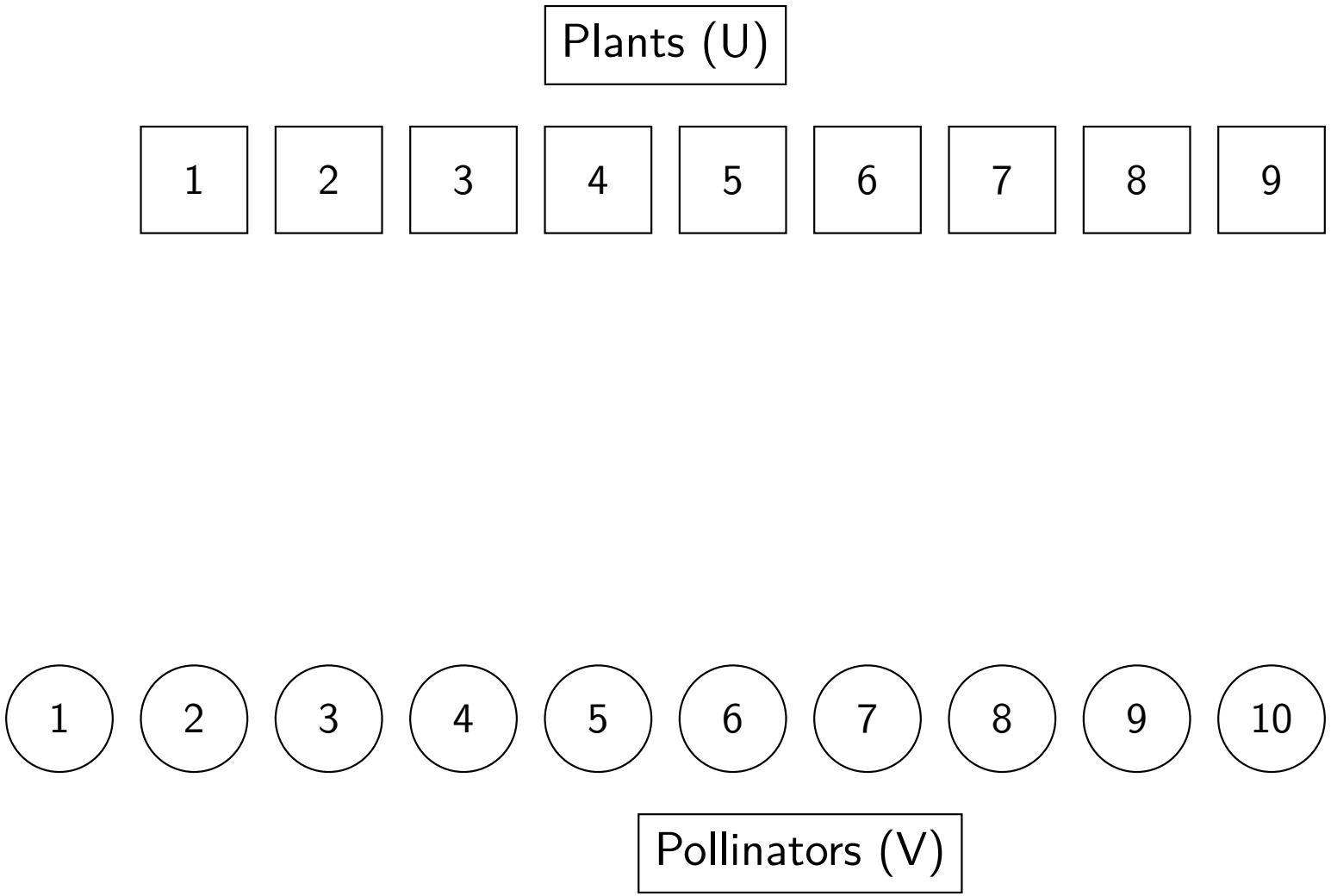
$$\forall (i,j) \in U \times V, \qquad Y_{ij} \sim \mathcal{B}\text{ern}(p)$$

- Homogeneity of the connections
- No hubs, no community, no nestedness

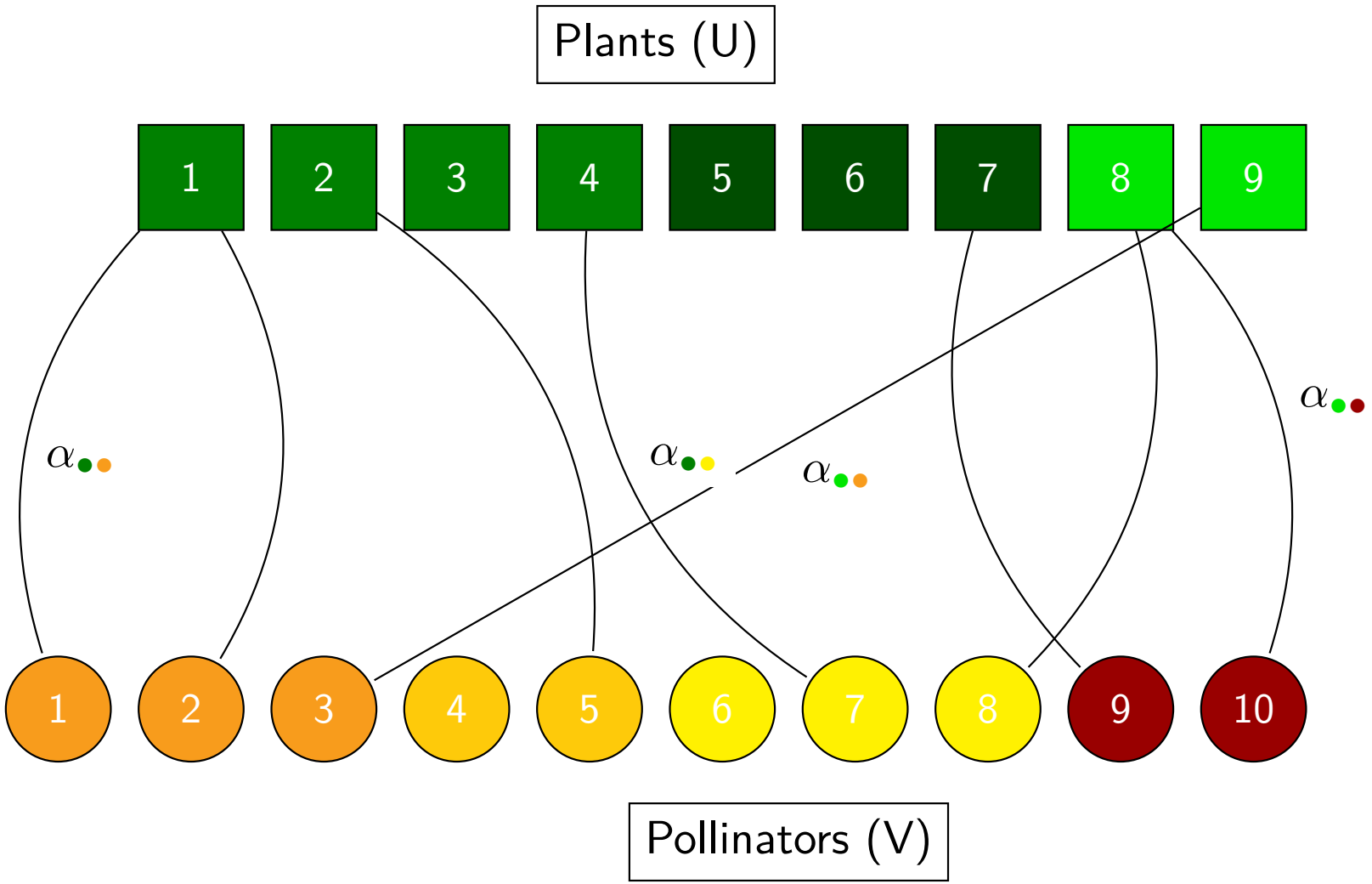
Latent Block Model

- Aim : introduce heterogeneity in the connections
- Tool : introduce blocks of nodes gathering entities that interact roughly similarly in the network

Latent Block Model : a generative model



Latent Block Model : a generative model



Latent Block Model with equations : latent variables I

- Each group of nodes (U and V) is divided into blocks / clusters
- K_U number of blocks in U and K_V number of blocks in V
- For any $i \in \{1, \ldots, |U|\}$, let Z_i^U be such that

$$Z_i^U = k \quad \text{if entity } i \text{ of group } U \text{ belongs to cluster } k$$

- For any $j \in \{1, \ldots, |V|\}$, let Z_j^V be such that

$$Z_j^V = \ell \quad \text{if entity } j \text{ of group } V \text{ belongs to cluster } \ell$$

Latent Block Model with equations : latent variables II

Random latent variables

$(Z_i^U)_{i=1\ldots|U|}$ and $(Z_j^V)_{j=1\ldots|V|}$ independent random variables, such that,

$$\begin{aligned} \mathbb{P}(Z_i^U = k) &= \pi_k^U, \\ \mathbb{P}(Z_j^V = \ell) &= \pi_\ell^V \end{aligned}$$

with $\sum_{k=1}^{K_U} \pi_k^U = 1$ and $\sum_{\ell=1}^{K_V} \pi_\ell^V = 1$

Latent Block Model with equations : connection probability

Conditionally to the latent variables...

$\boldsymbol{Z} = \{Z_i^U, i = 1 \dots |U|, Z_j^V, j = 1 \dots |V|\}$:

$$\mathbb{P}(Y_{ij} = 1 | Z_i^U = k, Z_j^V = \ell) = \alpha_{k\ell} \, .$$

Other emission distributions

- Previous model adapted to 0-1 network
- If Y_{ij} is a count

$$Y_{ij} | Z_i^U = k, Z_j^V = \ell \sim \mathcal{P}(\alpha_{k\ell})$$

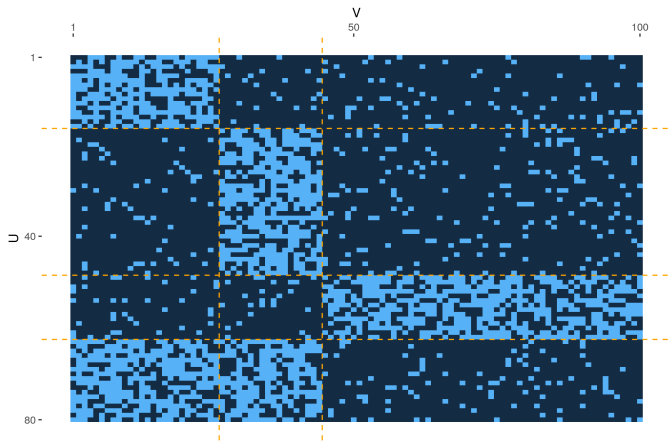
- If $Y_{ij} \in \mathbb{R}$

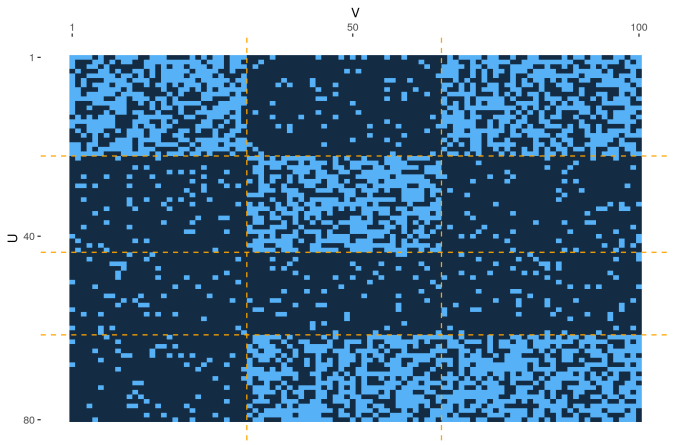
$$Y_{ij} | Z_i^U = k, Z_j^V = \ell \sim \mathcal{N}(\alpha_{k\ell}, \sigma_{k\ell})$$

[GN08]

A very flexible model I

LBM able to generate communities...

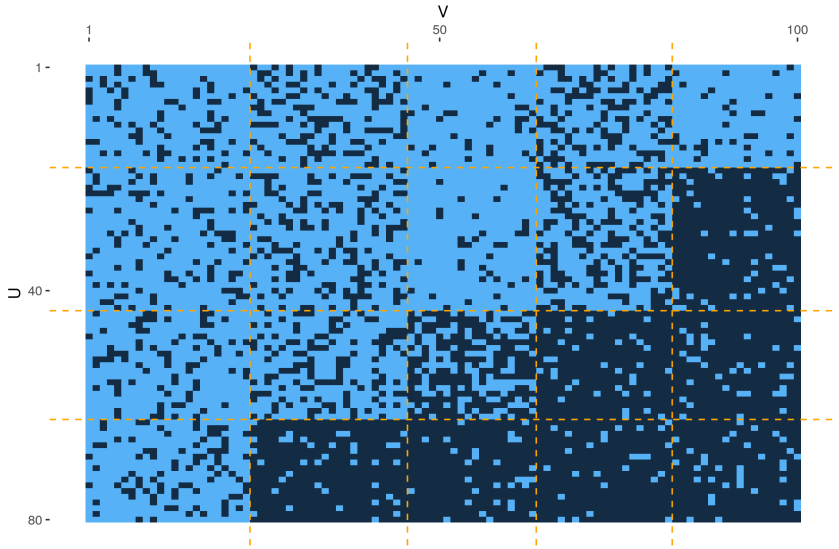




$$\alpha = \begin{pmatrix} 0.60 & 0.09 & 0.09 \\ 0.09 & 0.60 & 0.09 \\ 0.09 & 0.09 & 0.60 \\ 0.60 & 0.60 & 0.09 \end{pmatrix} \qquad \alpha = \begin{pmatrix} 0.60 & 0.09 & 0.60 \\ 0.09 & 0.60 & 0.09 \\ 0.09 & 0.09 & 0.09 \\ 0.09 & 0.60 & 0.60 \end{pmatrix}$$

A very flexible model II

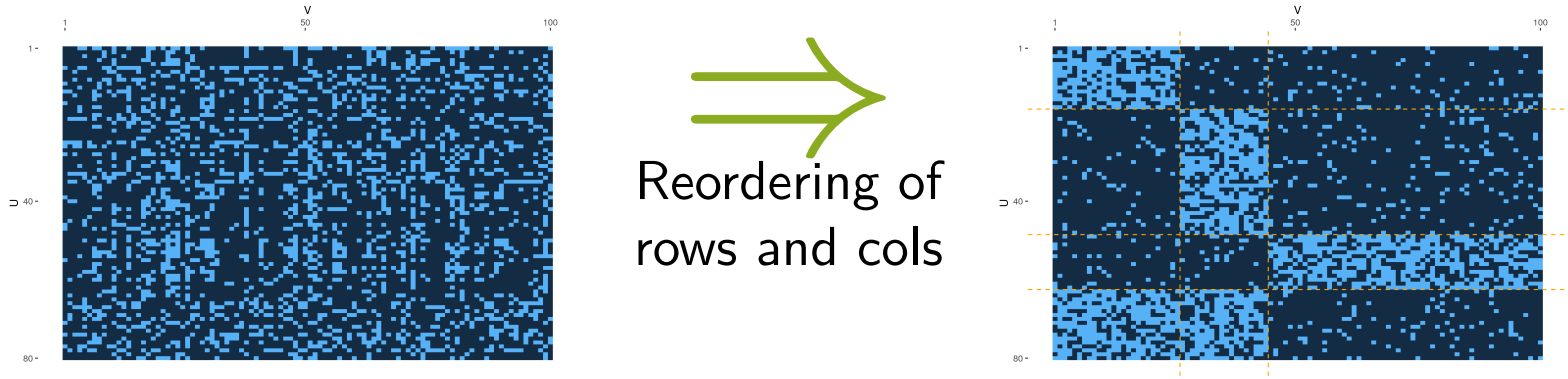
... or nested networks



$$\alpha = \begin{pmatrix} 0.80 & 0.70 & 0.90 & 0.60 & 0.90 \\ 0.80 & 0.70 & 0.90 & 0.60 & 0.09 \\ 0.80 & 0.70 & 0.40 & 0.09 & 0.09 \\ 0.80 & 0.09 & 0.09 & 0.09 & 0.09 \end{pmatrix}$$

Inference for LBM

Aim : From an incidence matrix, discovering the clusters



Remarks

- Looking for the blocks such that, under the assumption that my data come from the LBM model, the observed data Y is most probable (= most likely to occur)
- No specific prior structure
- Entities gathered because they have similar behavior in the network

Maximum likelihood inference

$$\begin{aligned}\widehat{\theta} &= \underset{\theta}{\operatorname{argmax}}\, p_\theta(Y) \\ &= \underset{\theta}{\operatorname{argmax}} \sum_{Z \in \{1,\dots,K\}^n} p_\theta(Y, Z)\end{aligned}$$

- Complete likelihood : $p_\theta(Y, Z)$ easy to compute
- Likelihood $p_\theta(Y)$: integration over all the possible clusterings $(Z_1, \dots, Z_n)$ (K^n)
- Latent variables : Expectation-Maximization
- Requires to evaluate $p(Z \mid Y)$
- No independence in this distribution
- Complicated distribution

Variational Inference

- Use a variational version of the Expectation-Maximization algorithm [DPR08, BKM17, MRV10]
- Penalized criterion to select the numbers of blocks K_U and K_V.
- R-package blockmodels [Leg15]

Variational inference

Principle [WJ08, BKM17].

- Choose a divergence measure $D(q \,||\, p)$
- Choose a class of distributions $\mathcal{Q}$
- Maximize w.r.t. θ and $q \in \mathcal{Q}$ the lower bound

$$J(Y; \theta, q) = \log p_\theta(Y) - D(q(Z) \,||\, p_\theta(Z \mid Y)) \leq \log p_\theta(Y)$$

Popular choice for SBMs. [GN08, DPR08, Leg16, MM15]

- $D = KL$:

$$\begin{aligned} J(Y; \theta, q) &= \log p_\theta(Y) - KL(q(Z) \,||\, p_\theta(Z \mid Y)) \\ &= \mathbb{E}_q\left(\log p_\theta(Y, Z)\right) - \mathbb{E}_q\left(q(Z)\right) \\ &= \mathbb{E}_q\left(\log p_\theta(Y, Z)\right) + \mathcal{H}\left(q(Z)\right) \end{aligned}$$

- q factorizable : $\mathcal{Q} = \{q(Z) : q(Z) = \prod_i q_i(Z_i)\}$

$$\tau_{ik} = \mathbb{P}_q(Z_i = k)$$

$\rightarrow$ mean field approximation

Variational EM

Algorithm

At iteration (t), given $(\theta^{(t-1)}, q_{\tau^{(t-1)}})$,

- **Step 1** Maximization w.r.t. τ

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

- **Step 2** Maximization w.r.t. θ

$$\begin{aligned}\theta^{(t)} &= \arg\max_{\theta} J(Y;\theta,q_{\tau^{(t)}})\\ &= \arg\max_{\theta} \mathbb{E}_{q_{\tau^{(t)}}}\left[\log p_\theta(Y,Z)\right]\end{aligned}$$

Model selection

- Selection of the number of blocks (K_U, K_V)
- BIC : approximation of the marginal log-likelihood $m_c(Y, Z; \mathcal{M})$ where the parameters θ have been integrated out with a prior distribution

BIC for observed Z

Let $\mathcal{M} = \mathcal{M}_{K_0, K_1, \ldots, K_Q}$, then

$$\log m(Y, Z; \mathcal{M}) \approx_{n_Q \to \infty} \max_{\theta} \log p_\theta(Y, Z; \mathcal{M}) + pen_{\mathcal{M}}$$

with

$$pen_{\mathcal{M}} = -\frac{1}{2}\left\{(K_U - 1)\log n_U + (K_V - 1)\log n_V + K_U K_V \log(n_U n_V)\right\}$$

Penalized criterion for latent Z : ICL

- Imputation of the Z with the MAP [BCG00]

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

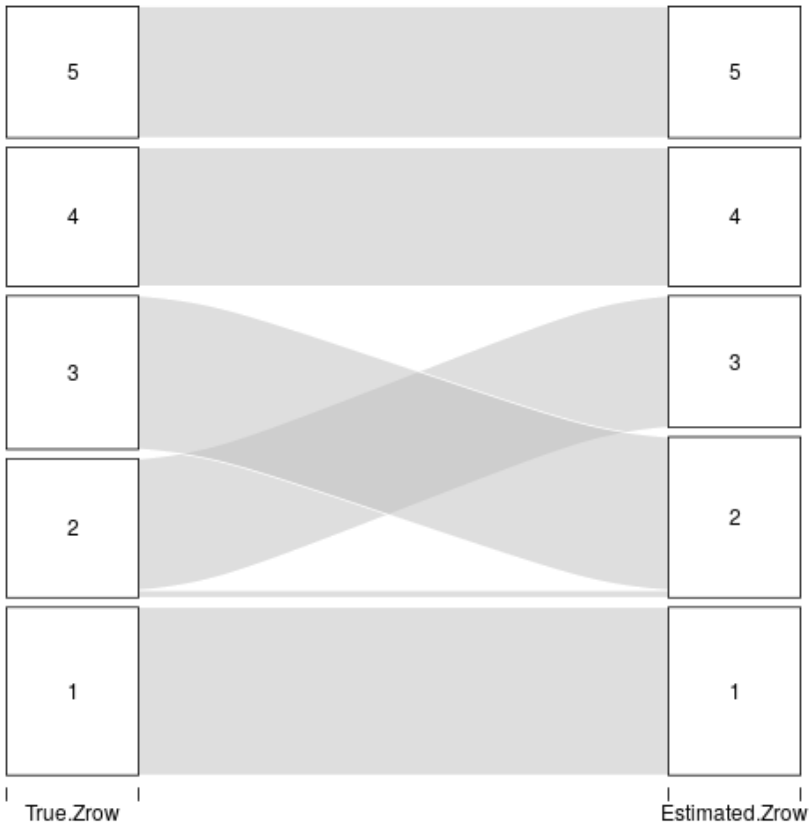
- Integration of the Z [DPR08, BDLBH16]
  - $ICL(\mathcal{M}) = E_{Z|Y;\widehat{\theta}\mathcal{M}}\left[\log \ell_c(Y,Z;\widehat{\theta},\mathcal{M})\right] + pen_{\mathcal{M}}$
  - $p(Z|Y;\widehat{\theta},\mathcal{M}) \Rightarrow \mathcal{R}_{Y,\widehat{\tau}}$
  - $\widehat{ICL}(\mathcal{M}_{K_0,K_1,\ldots,K_Q}) = E_{\mathcal{R}_{Y,\widehat{\tau}}}[\log \ell_c(Y,Z;\theta,\mathcal{M})] + pen_{\mathcal{M}}$

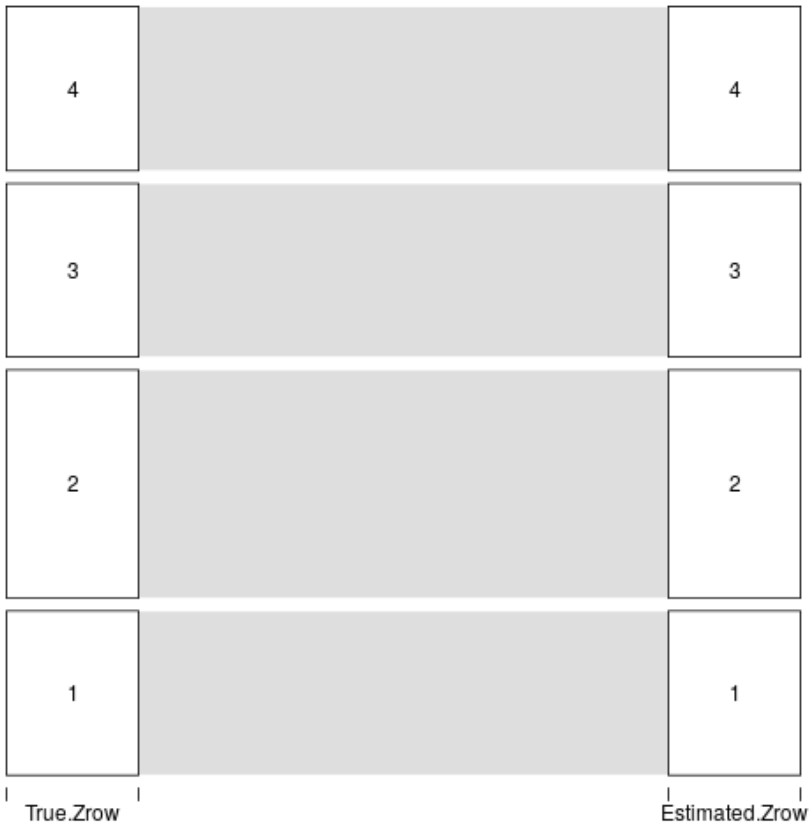
Results on my simulated networks

$\rightarrow$ Used blockmodels $\rightarrow$ Less than one minute to obtain the results.

$\rightarrow \hat{K}_U = K_U$, $\rightarrow$ Estimated blocks versus simulated blocks :

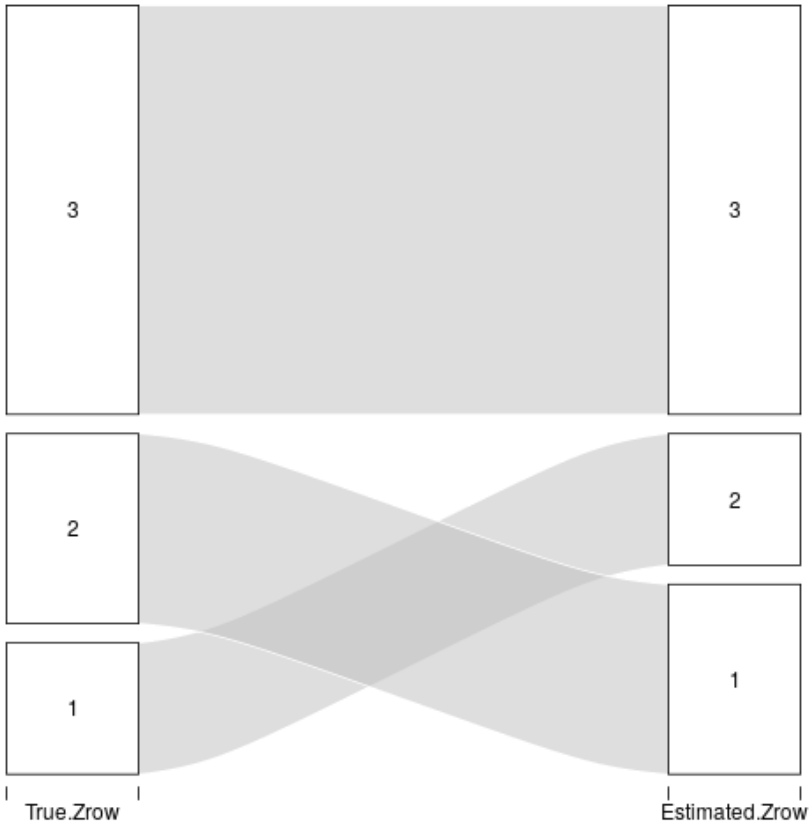
Nested network

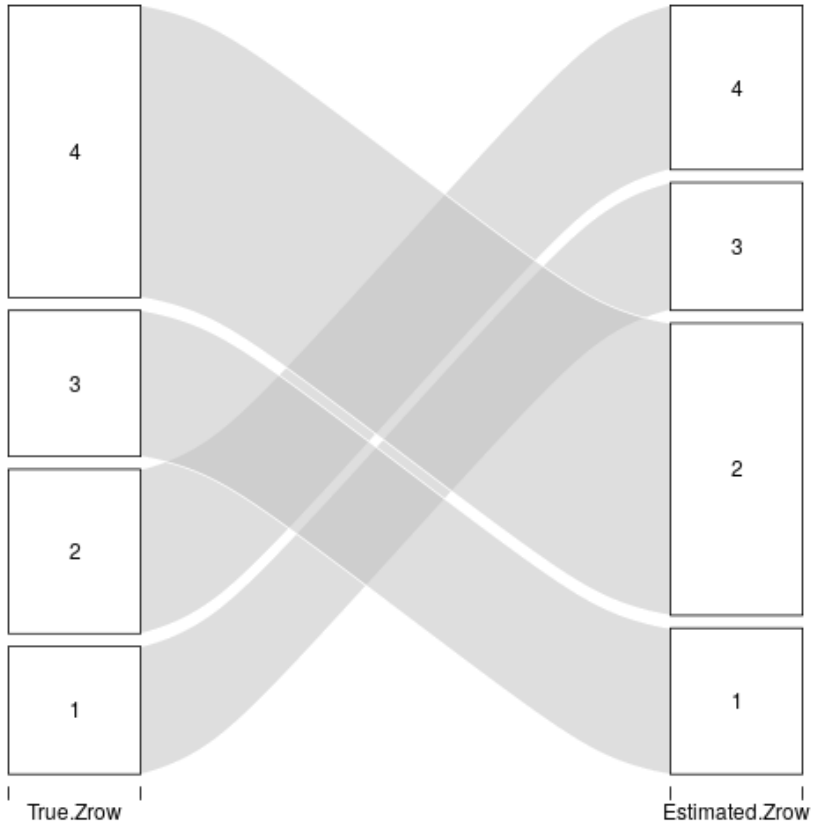




Results on my simulated networks

Community network



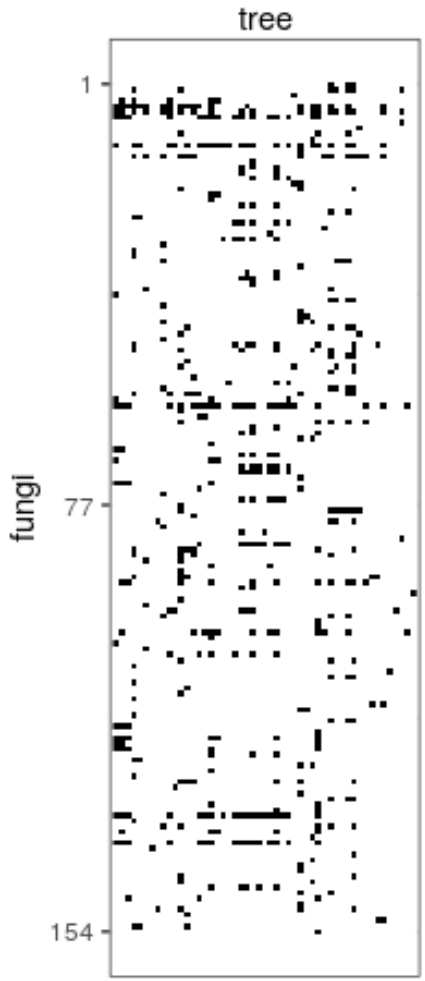


→ Numbering of blocks has no meaning. Up to label switching

Application to ecological data

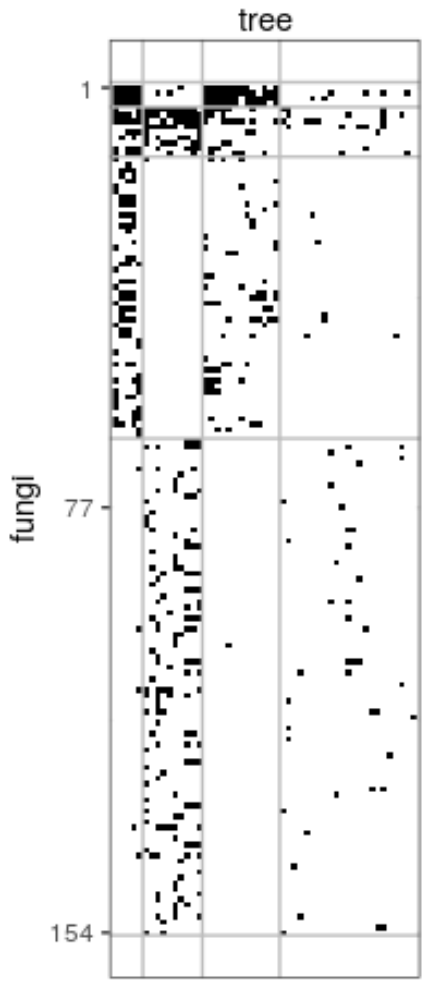
[VPDL08]

- 154 fungi species, 51 tree species
- Binary fungus-tree interactions



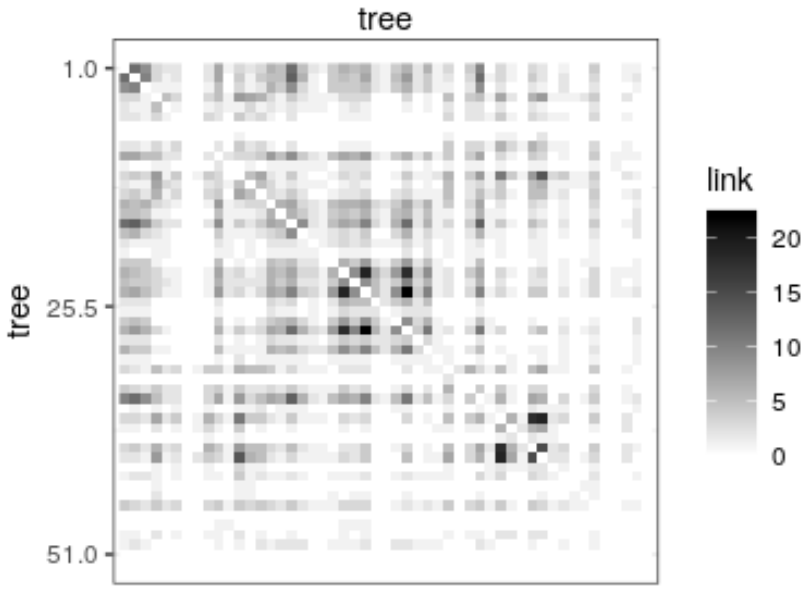
Application to ecological data : LBM inference

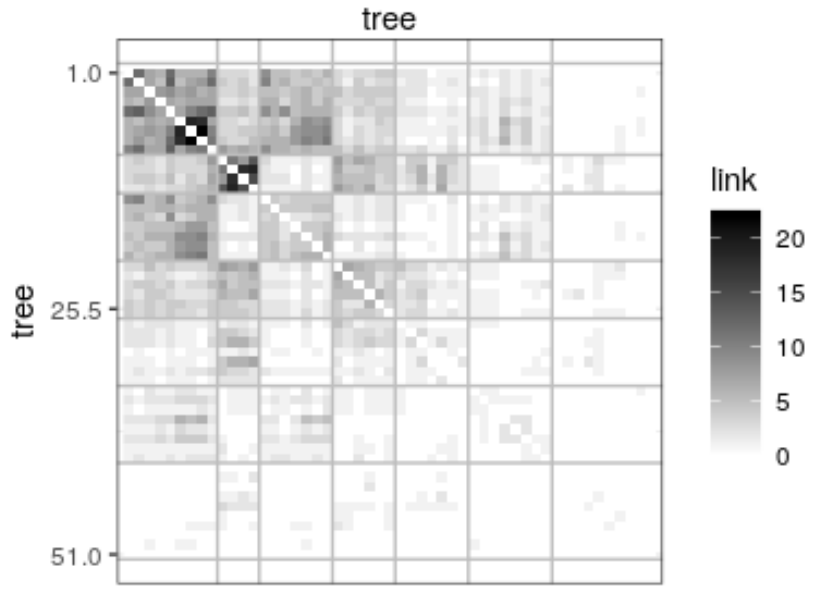
4 blocks of trees, 4 blocks of fungi



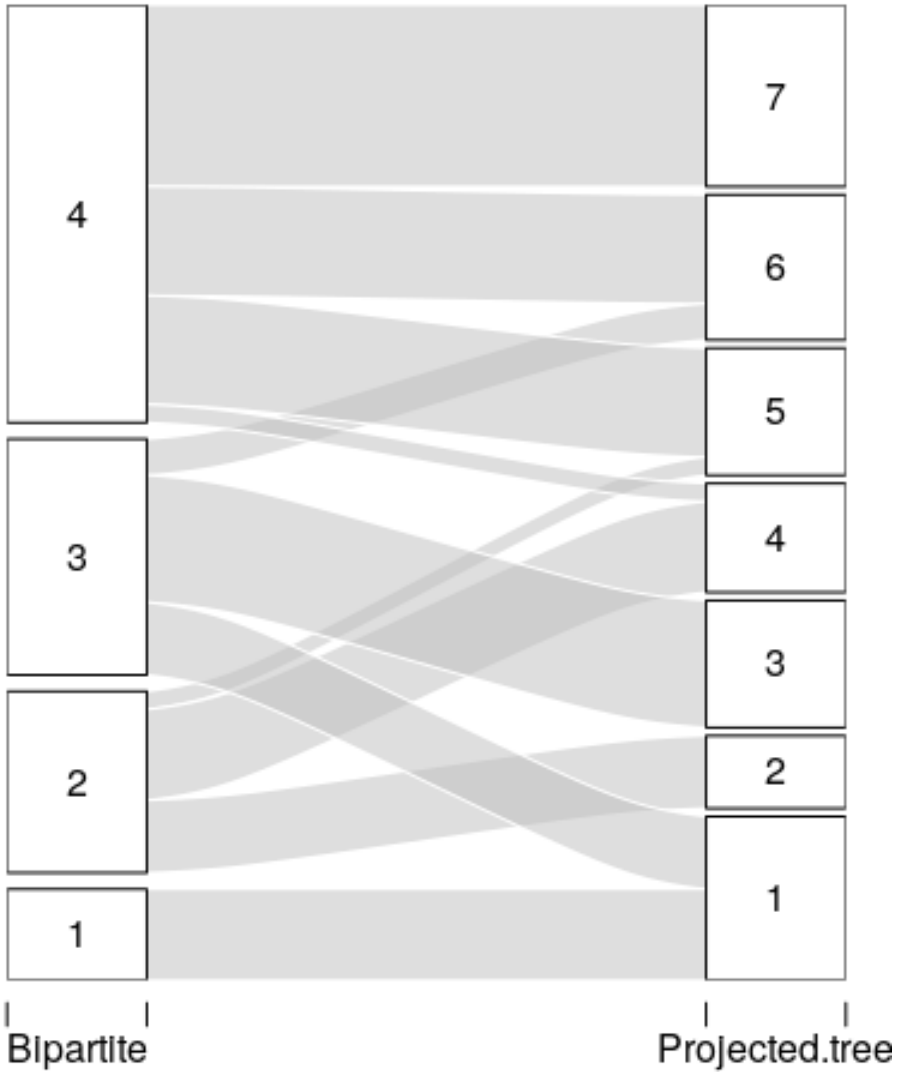
Blocks from the projected matrix

7 clusters of trees





Comparison of the cluterings



About the model

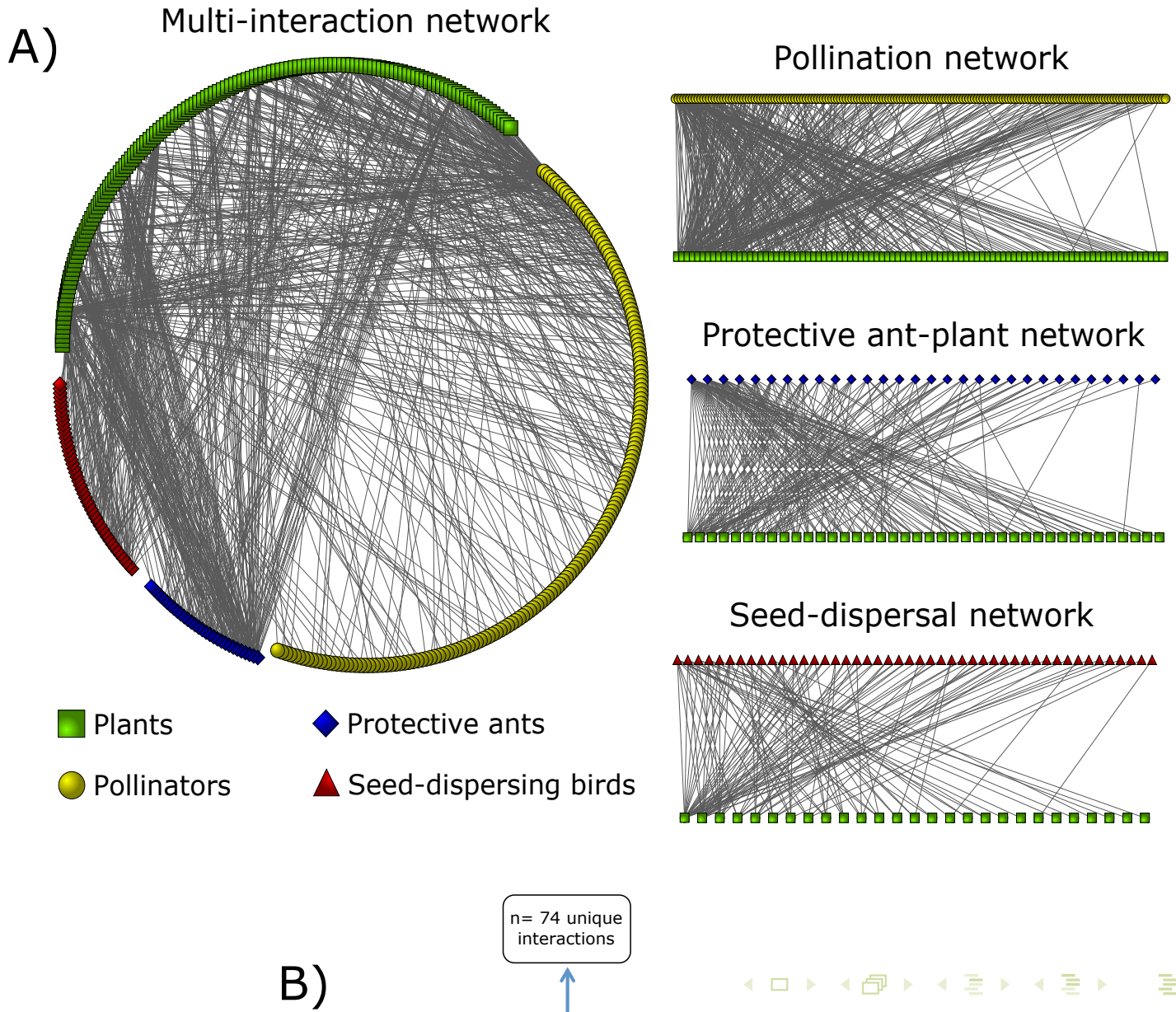
- Clusters in rows and columns : Biclustering
- If weighted networks : extension of LBM to Poisson
- If overdispersed count data : see talk by Julie Aubert this afternoon
- About the blocks
  - Blocks may be due to inherent properties of entities at stake
  - Such properties not taken into account in the model
  - Blocks can be analized as posteriori with respect to some covariates
- Taking into account covariates

More complicated metabolic networks

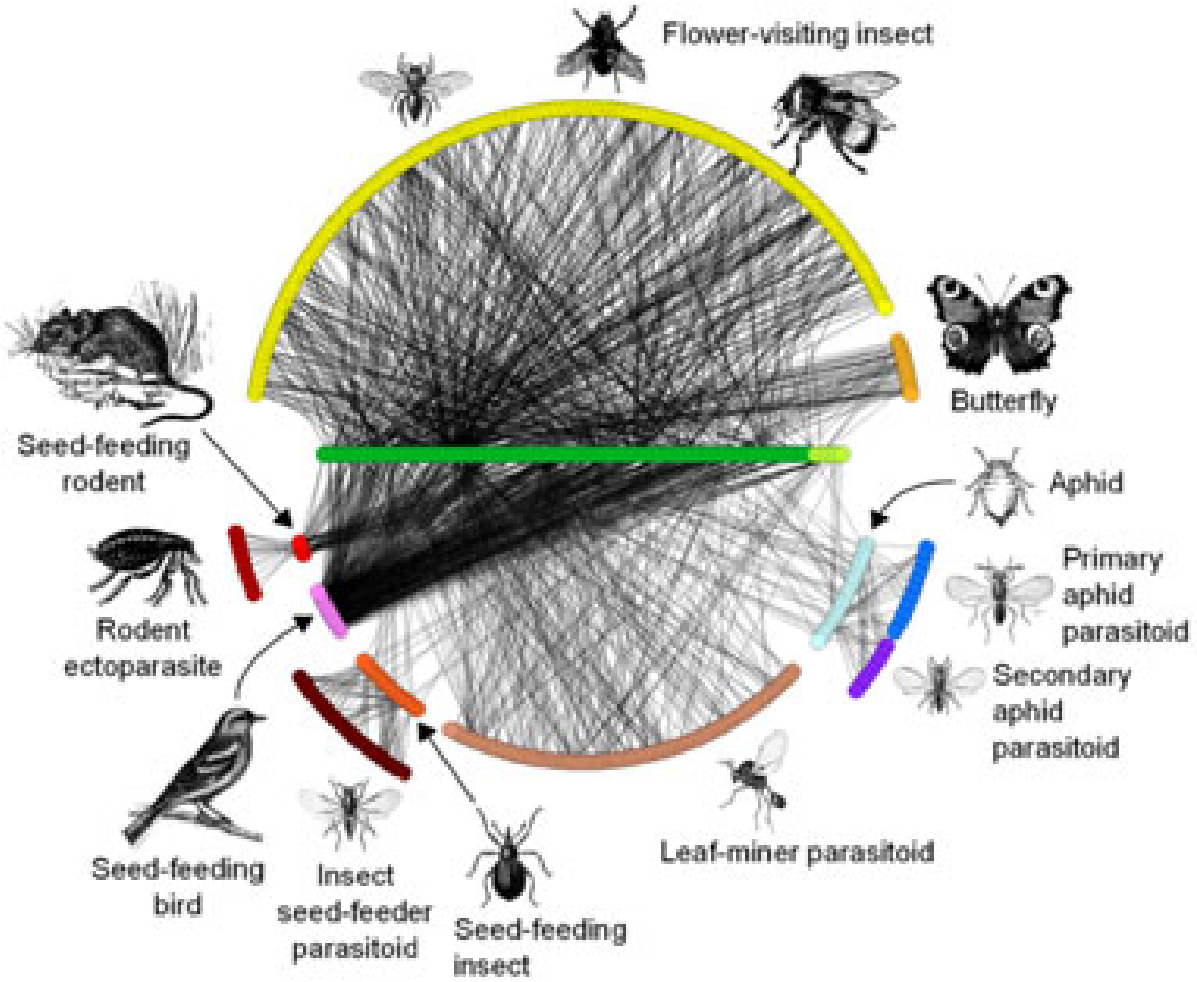
- Two types of vertices : metabolites and metabolic reactions
- Edges joining each metabolite to the reaction in which it participates.
- Edges are directed : some metabolites (the substrates) go into the reaction and some (the products) come out of it.
- Enzymes incorporated : adding a third class of vertex to represent them, with undirected edges connecting them to the reactions they catalyze.
- Resulting graph : mixed (directed and undirected) tripartite network.

Multipartite network in ecology [DLRJ+16]

Page 27 of 29 Submitted to Proceedings of the Royal Society B: For Review Only



Super Multipartite network in ecology



Multipartite Networks

Joint work with P. Barbillon and A. Bar-Hen

- Q functional group : each functional group q of size n_q
- Multipartite network : a collection of networks
- Each network involves one or two functional groups : indexed by pairs (q, q') (q and q' in $[\![1, Q]\!]$).
- $\mathcal{E}$ denotes the list of pairs of observed networks
- Each network encoded in a matrix $X^{qq'}$

$$Y_{ii'}^{qq'} = \begin{cases} 1 & \text{if entity } i \text{ of group } q \text{ is in interaction} \\ & \text{with entity } i' \text{ of group } q'. \\ 0 & \text{otherwise} \end{cases}$$

- $Y = \left\{ \left(Y^{qq'} \right), (q, q') \in \mathcal{E} \right\}$.

Example in ecology

$$Y_{ii'}^{1q'} = \begin{cases} 1 & \text{if animal specie } i' \text{ of functional group } q' \text{ has been observed} \\ & \text{in interaction with plant } i \\ 0 & \text{otherwise} \end{cases}$$

$q' = 2, 3, 4.$

Plant 1	X_{ij}^{11}	X_{ij}^{12}	X_{ij}^{13}
Plant 2			
$\vdots$			
Plant n_1			
	Ant 1 … Ant n_2	Seed dispersing bird 1 … Seed dispersing bird n_3	Pollinator 1 … Pollinator n_4

$Y_{\cdot\cdot}^{q} \in \{0, 1\}$ to avoid sampling issues

Example in ethnobiology

- Ethnobiology : scientific study of the relations between environment and people
- [TC16] : understand how seed exchanges relations between farmers structure and guaranty biodiversity in the cultivated crop species.
- Functional groups : farmers and crop species
- Relations :
  - Between farmers : seed exchange (oriented relation) = Simple graph
  - Betwen farmers and crop species : bipartite network. Edge = farmer grows crop specie.

Block model : Mixture model on the $Y_{ii'}^{qq'}$

Latent variables

- Each functional group q divided into K_q blocks or clusters
- $\forall q \in [\![1, Q]\!], \forall i \in [\![1, n_q]\!]$, $Z_i^q = k$ if individual i of functional group q belongs to cluster k.
- Independant random variables :

$$\mathbb{P}(Z_i^q = k) = \pi_k^q, \tag{1}$$

with $\sum_{k=1}^{K_q} \pi_k^q = 1$ for any $q = 1, \dots Q$.

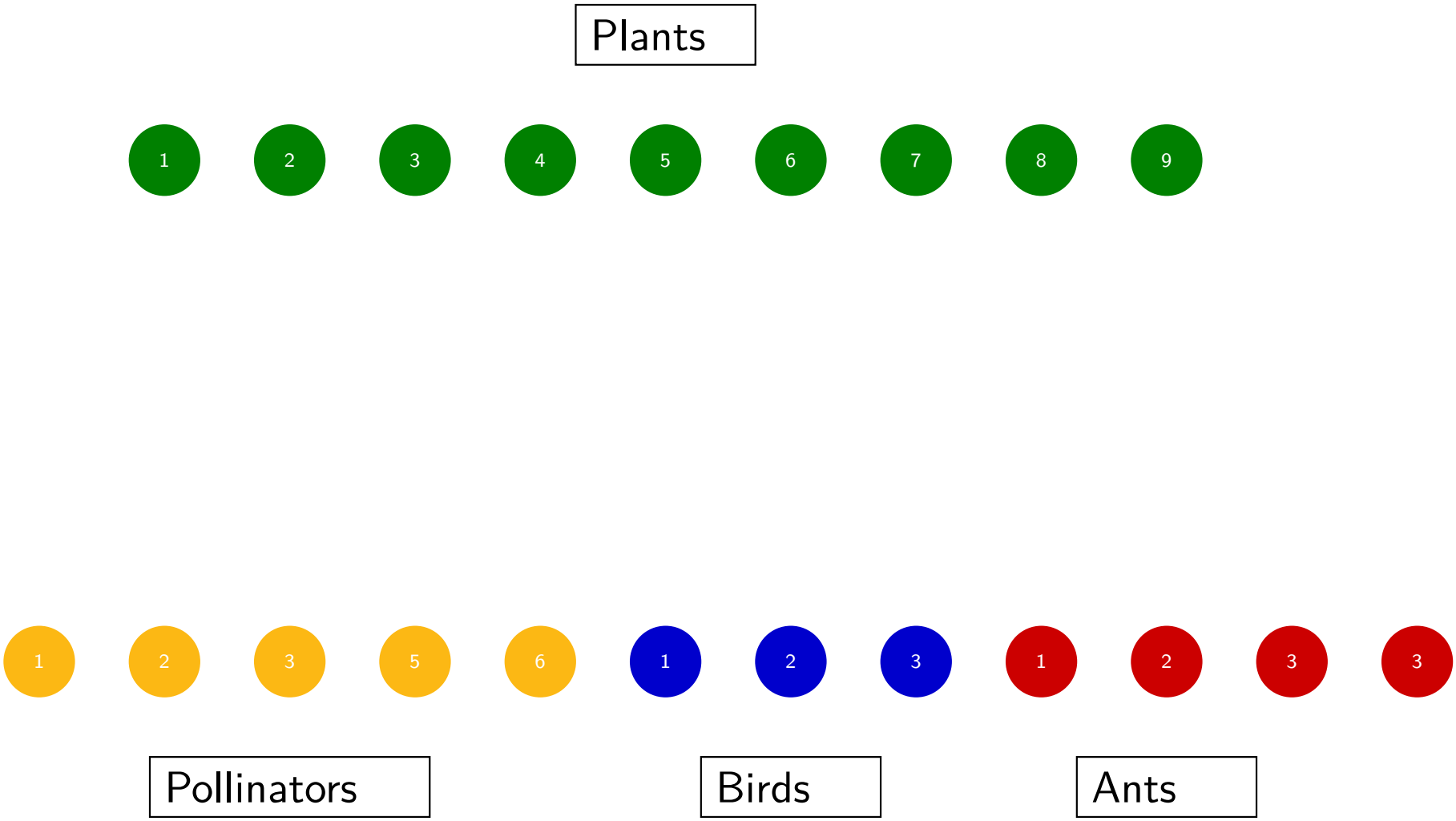
- $Z = (Z_i^q)_{i \in [\![1,n_q]\!], q \in [\![1,Q]\!]}$.

Conditionally to the latent variables

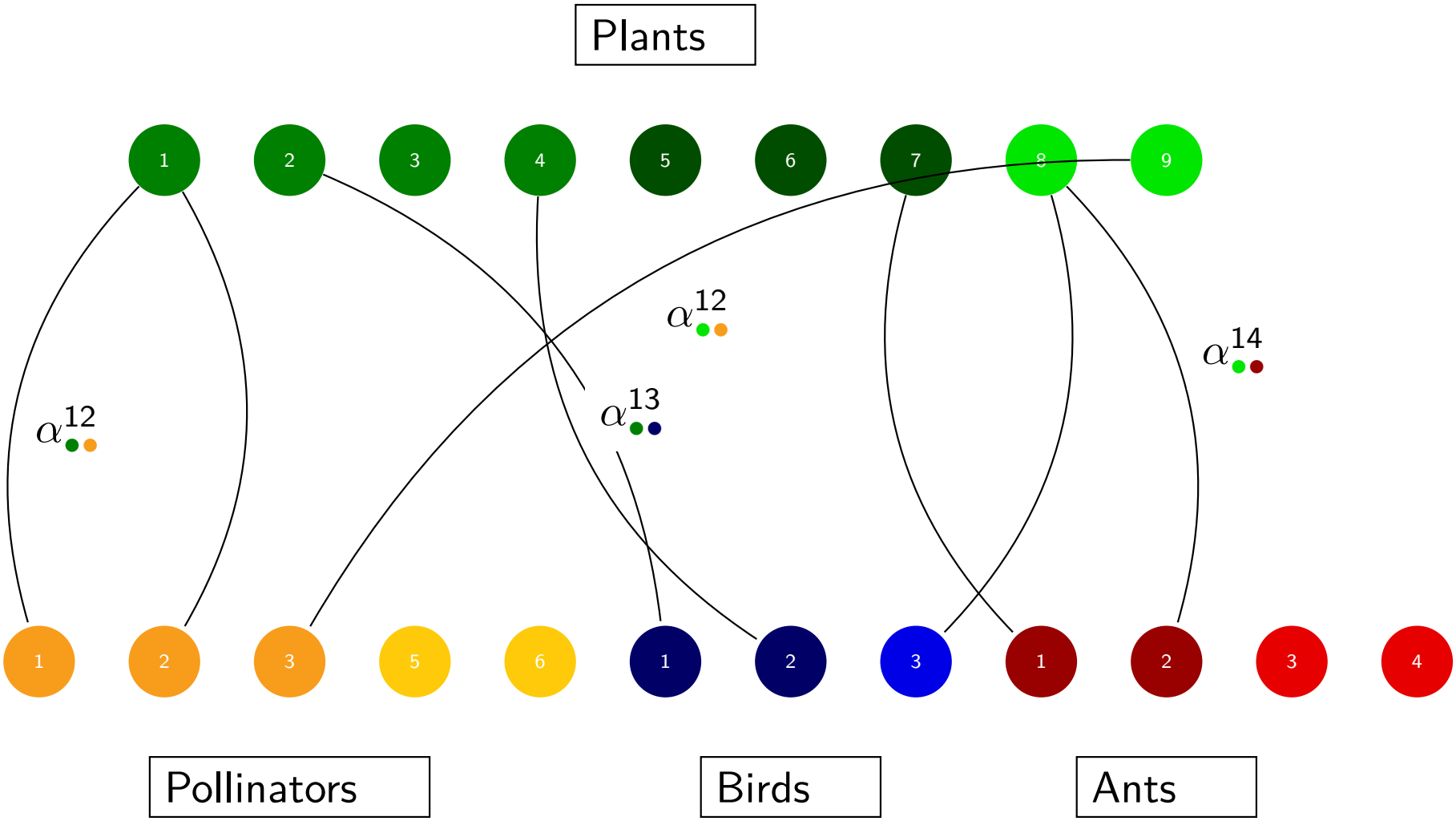
$\forall (i, i', q, q')$, entries of the matrices independant and

$$\mathbb{P}(Y_{ii'}^{qq'} = 1 | Z_i^q = k, Z_{i'}^{q'} = k') = \alpha_{kk'}^{qq'} \tag{2}$$

Generative model illustration



Generative model illustration



Statistical Inference

- Parameters π, α for given numbers of clusters $K_1, \ldots, K_Q$.
- Clustering of the agents
- Numbers of blocks $K_1, \ldots, K_Q$.

Likelihood function

Complete likelihood of (Y, Z)

$$
\begin{aligned}
p(Y, Z; \theta) &= p(Y|Z; \boldsymbol{\alpha}) p(Z; \boldsymbol{\pi}) \\
&= \prod_{q,q' \in \mathcal{E}} \prod_{i=1}^{n_q} \prod_{j=1}^{n_{q'}} (\alpha^{qq'}_{Z_i^q, Z_j^{q'}})^{X_{ij}^{qq'}} (1 - \alpha^{q}_{Z_i^q, Z_j^{q'}})^{1 - X_{ij}^{qq'}} \qquad (3) \\
&\quad \times \prod_{q=1}^{Q} \prod_{i=1}^{n_q} \pi^q_{Z_i^q} \, . \qquad (4)
\end{aligned}
$$

Observed likelihood (Y)

$$
\log p(Y; \theta) = \log \sum_{Z \in \boldsymbol{\mathcal{Z}}} p(Y, Z; \theta) \, . \qquad (5)
$$

Maximisation using variational EM

Model selection : penalized likelihood criteria

- Selection of the numbers of blocks $K_1, \ldots, K_Q$
- ICL : Integrated Completed Likelihood
  - BIC computed on the complete log-likelihood
  - Integration of the latent variables
- $ICL(\mathcal{M}) = \mathbb{E}_{Z|Y;\widehat{\theta}\mathcal{M}}\left[\log p(Y, Z; \widehat{\theta}, \mathcal{M})\right] + pen_{\mathcal{M}}$

$$pen_{\mathcal{M}} = -\frac{1}{2}\left\{\sum_{q=1}^{Q}(K_q - 1)\log(n_q) + \left(\sum_{(q,q')\in\mathcal{E}} K_{qq'}\right)\log\left(\sum_{(q,q')\in\mathcal{E}} n_{qq'}\right)\right\}$$

[DPR08, BDLBH16]

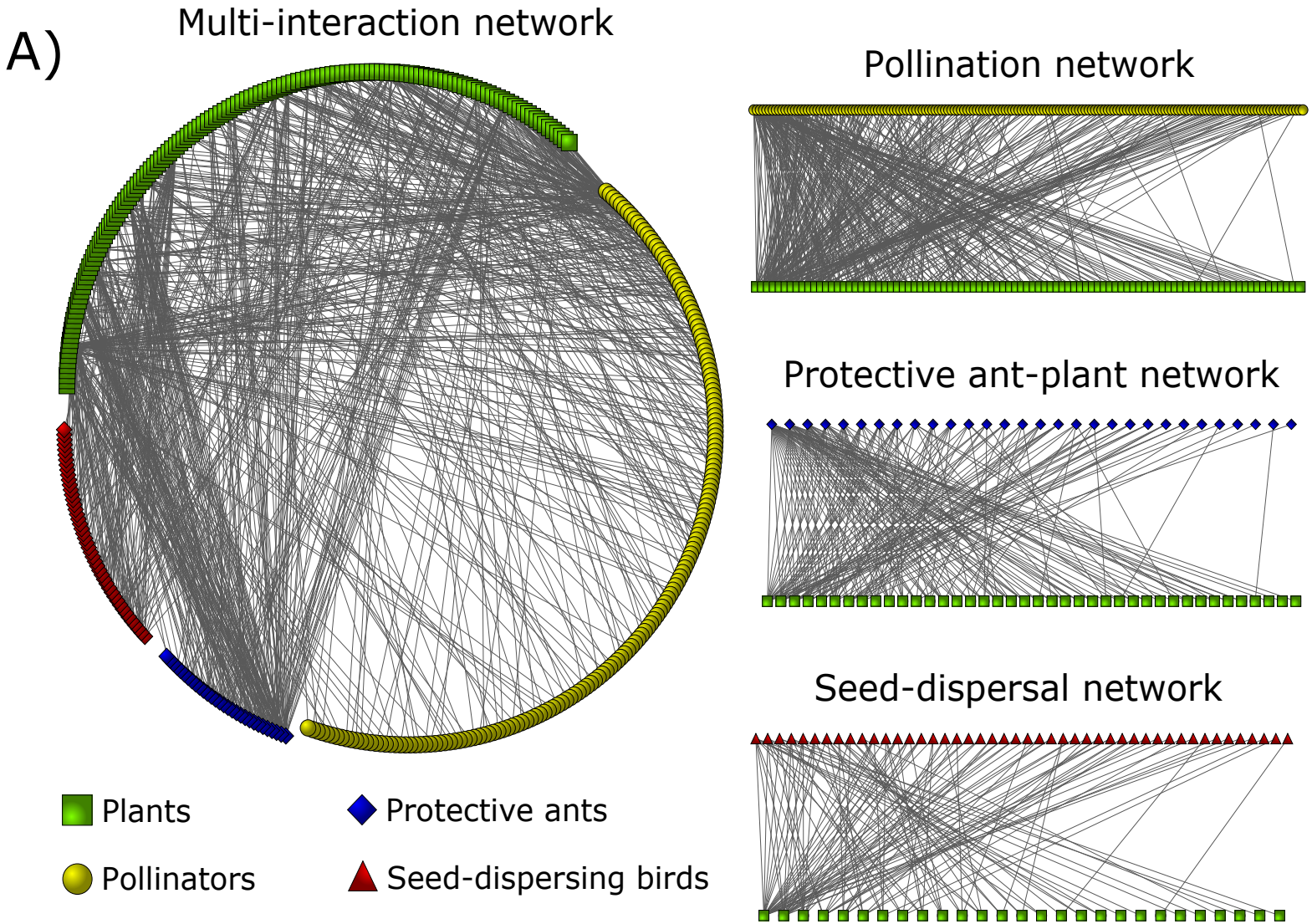
- **In practice** $\widetilde{ICL}(\mathcal{M}) = \mathbb{E}_{\mathcal{R}_{\hat{\tau}},Z}\left[\log p(Y, Z; \widehat{\theta}, \mathcal{M})\right] + pen_{\mathcal{M}}$
- Stepwize algorithm to select the better model

The dataset

- Weasley Dattilo, Inecol, Jalapa, Mexique [DLRJ+16]
- - $n_0 = 141$ plants species
  - $n_1 = 30$ ants species
  - $n_2 = 46$ bird species
  - $n_3 = 173$ pollinators species

Data

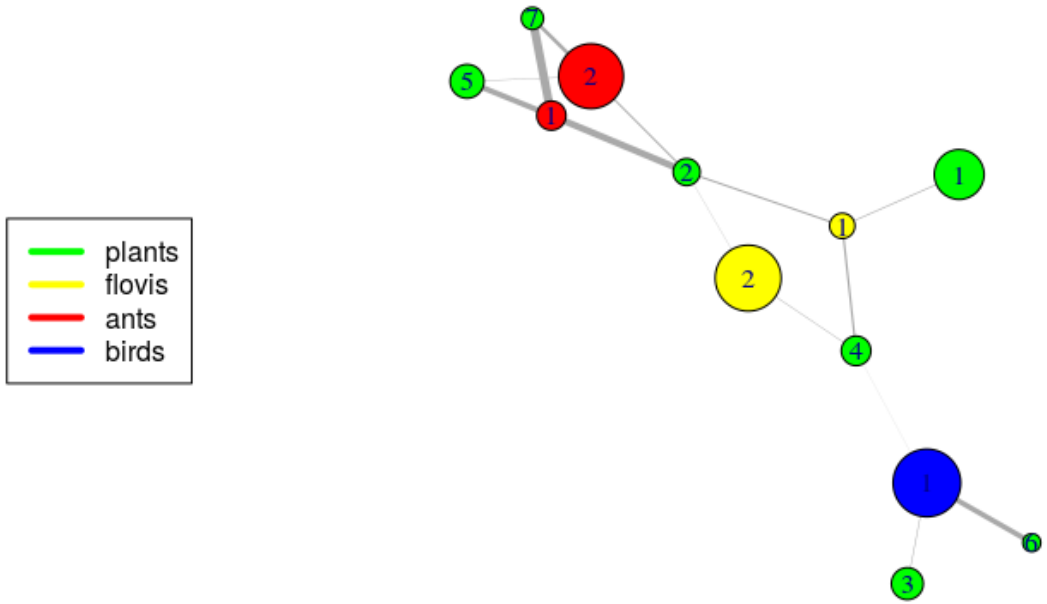
Page 27 of 29 Submitted to Proceedings of the Royal Society B: For Review Only



Results : a mesoscopic view of the network

With our model and model selection (a few minutes)

- 7 blocks of plants
- 2 blocks of pollinators
- 1 block of birds
- 2 blocks of ants

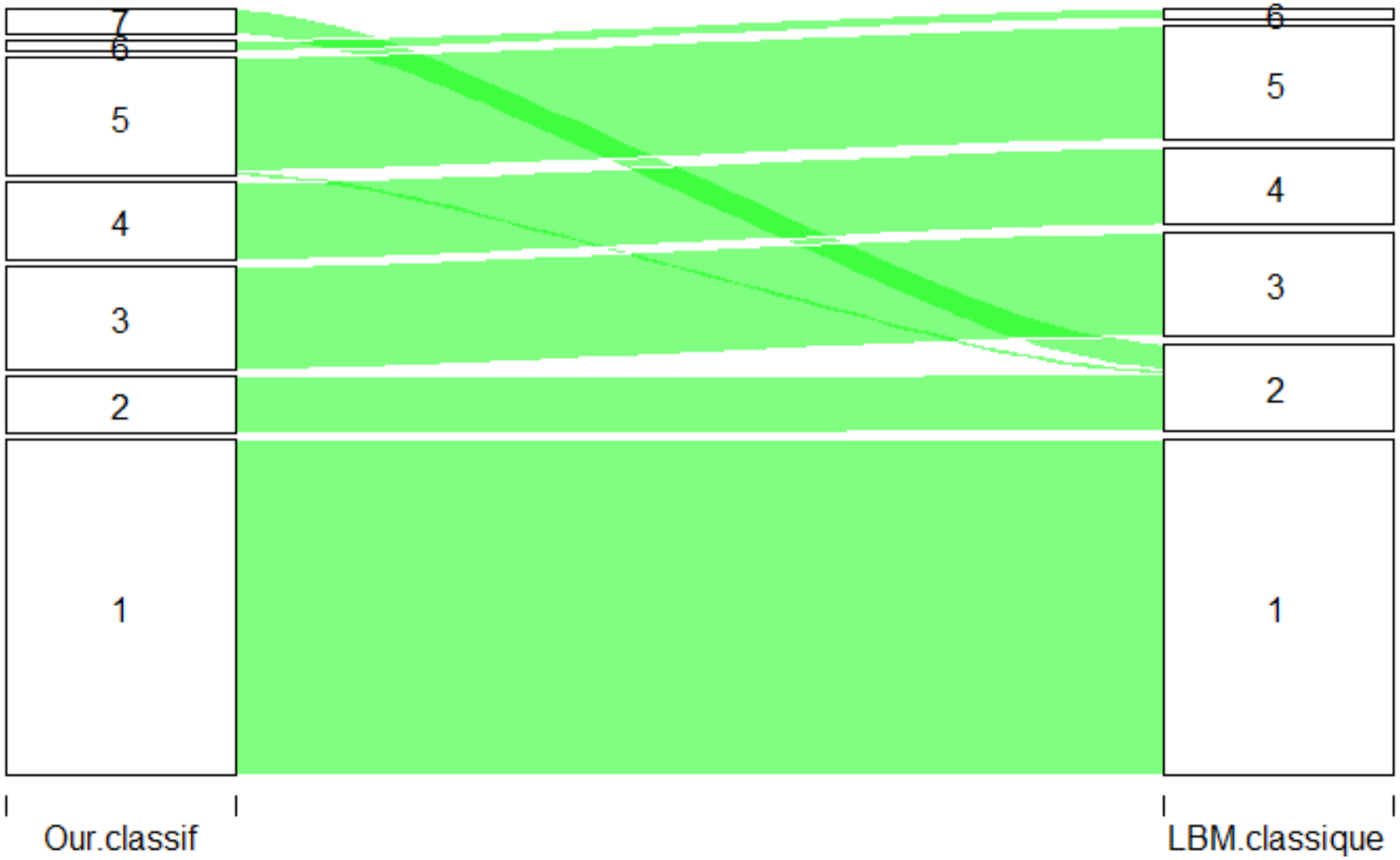


Comparison with existing methods

- Studying each matrix separately by LBM.
  - 1 clustering of ants
  - 1 clustering of pollinators
  - 1 clustering of birds
  - 3 clusterings of plants
- Creating an artificial clustering of plants taking in account the three matrices by intersection
- Comparing the clusterings by Adjusted Rand Index ($= 1$ if clusterings are equal, up to label switching)

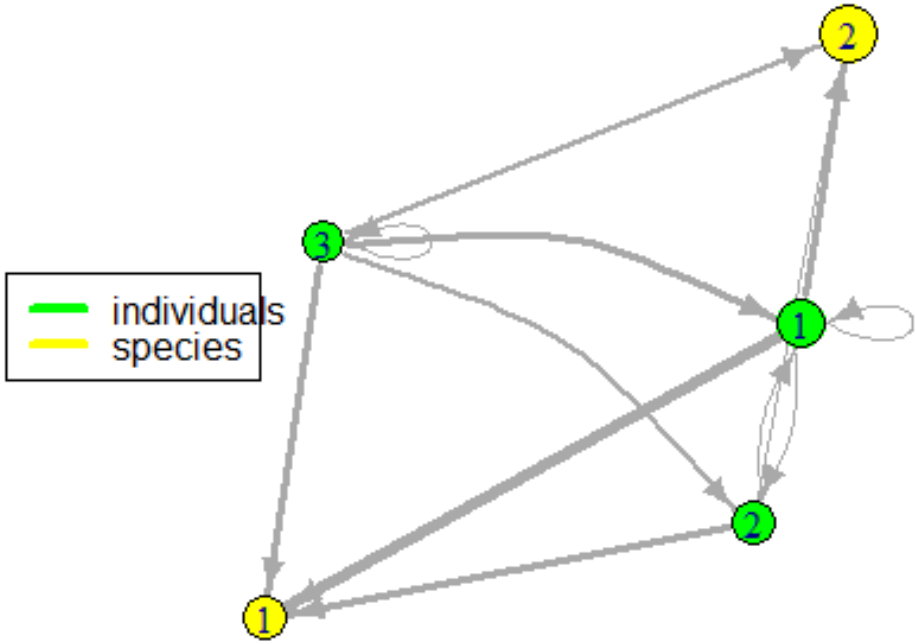
	Full/Poll.	Full/Ants	Full/Birds	Full/Inter
Plants	(7/3) 0.118	(7/3) 0.415	(7/3) 0.163	(7/12) 0.617
Poll.	(2/3) 0.997			
Ants		(2/2) 1.000		
Birds			(1/1) 1.000	

Comparison with a classical LBM

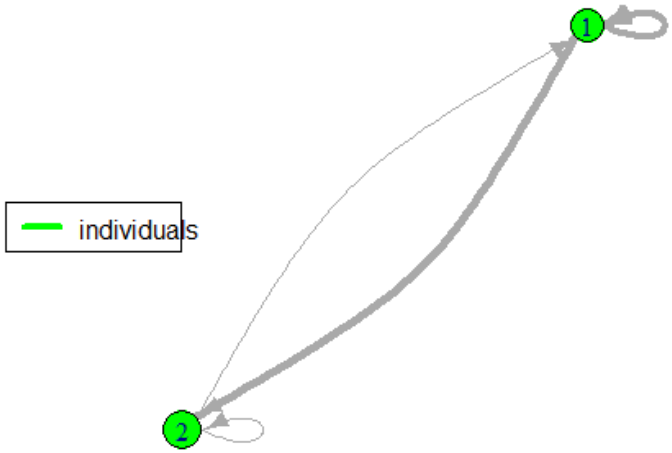


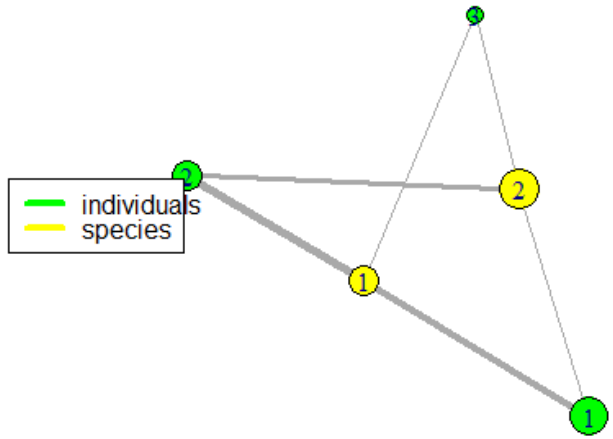
Ethnobiology data : results

2 blocks or clutivated species and 3 blocks of individuals

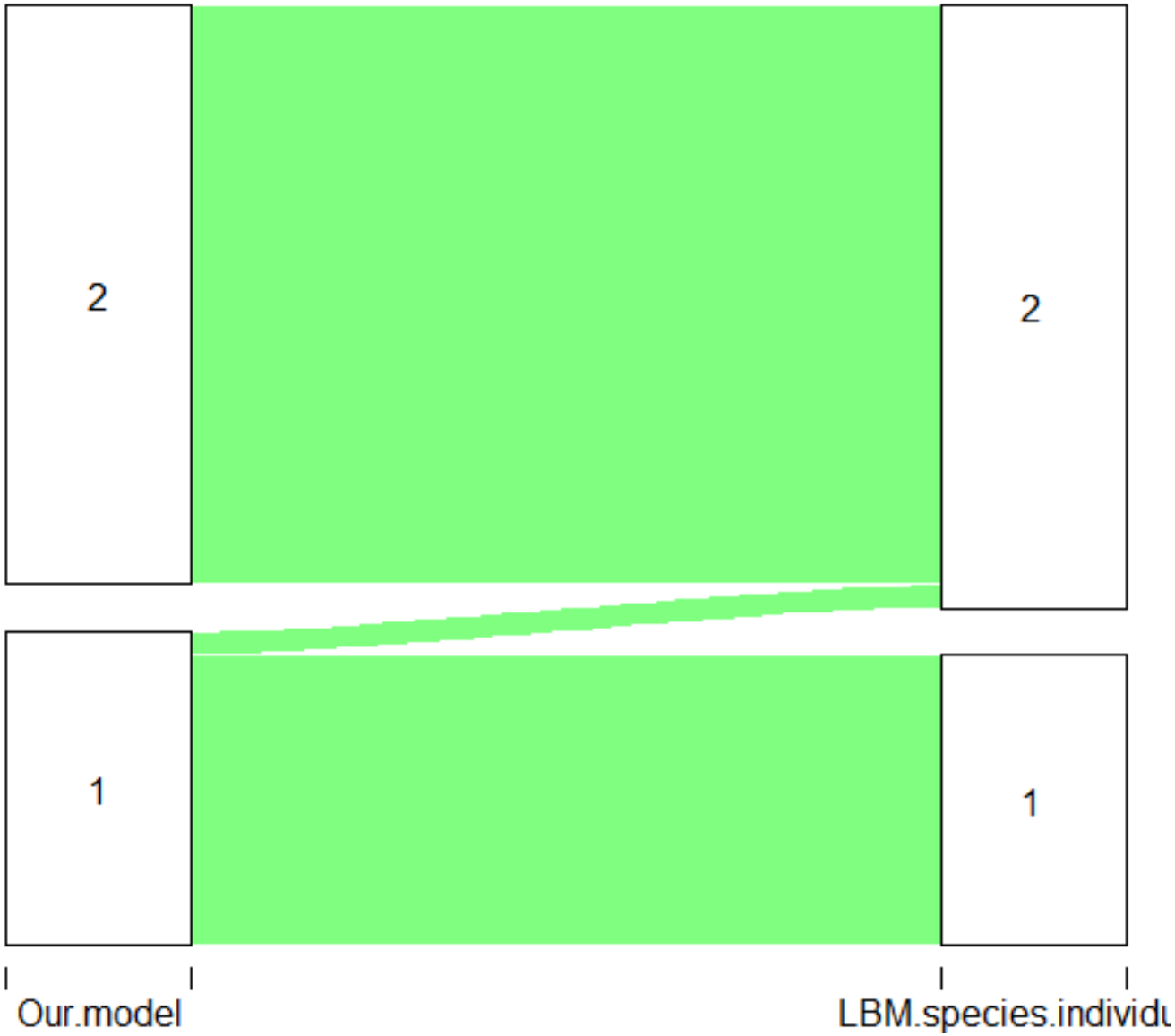


Comparison with the blocks obtained with a individuals-species or a individual-individual network

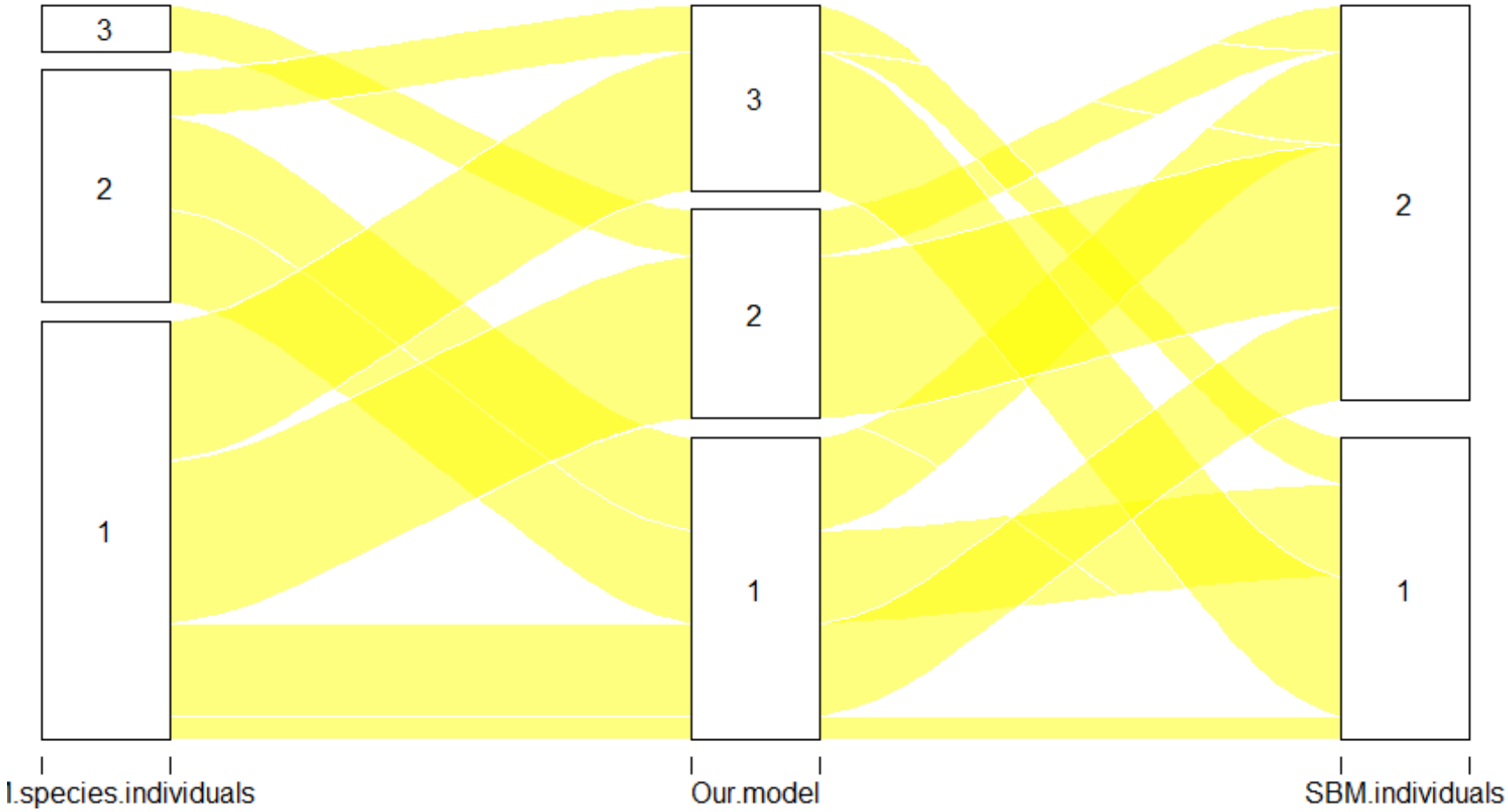




Comparison of the clusterings for plants



Comparison of individual classifications



Conclusions and perspectives

- Package R : GREMLIN (github) able to handle any type of multipartite networks (binary or counting interactions)
- Bipartite networks : a lot of examples. Require a specific treatment but with known tools

References I

C. Biernacki, G. Celeux, and G. Govaert.

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

Pattern Analysis and Machine Intelligence, IEEE Transactions on, 22(7) :719–725, Jul 2000.

Pierre Barbillon, Sophie Donnet, Emmanuel Lazega, and Avner Bar-Hen.

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, 2016.

D. M. Blei, A. Kucukelbir, and J. D. McAuliffe.

Variational inference : A review for statisticians.

Journal of the American Statistical Association, 112(518) :859–877, 2017.

References II

Wesley Dáttilo, Nubia Lara-Rodríguez, Pedro Jordano, Paulo R. Guimarães, John N. Thompson, Robert J. Marquis, Lucas P. Medeiros, Raul Ortiz-Pulido, Maria A. Marcos-García, and Victor Rico-Gray.

Unravelling darwin's entangled bank : architecture and robustness of mutualistic networks with multiple interaction types.

Proceedings of the Royal Society of London B : Biological Sciences, 283(1843), 2016.

J.-J. Daudin, F. Picard, and S. Robin.

A mixture model for random graphs.

Stat. Comput., 18(2) :173–83, 2008.

Gérard Govaert and Mohamed Nadif.

Block clustering with bernoulli mixture models : Comparison of different approaches.

Comput. Stat. Data Anal., 52(6) :3233–3245, February 2008.

Jean-Benoist Leger.

blockmodels : Latent and Stochastic Block Model Estimation by a 'V-EM' Algorithm, 2015.

R package version 1.1.1.

References III

Jean-Benoist Leger.

Blockmodels : A R-package for estimating in latent block model and stochastic block model, with various probability functions, with or without covariates.

Technical report, arXiv :1602.07587, 2016.

M. Mariadassou and C. Matias.

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

Bernoulli, 21(1) :537–573, 2015.

M. Mariadassou, S. Robin, and C. Vacher.

Uncovering latent structure in valued graphs : a variational approach.

The Annals of Applied Statistics, pages 715–742, 2010.

Georgios A Pavlopoulos, Panagiota I Kontou, Athanasia Pavlopoulou, Costas Bouyioukos, Evripides Markou, and Pantelis G Bagos.

Bipartite graphs in systems biology and medicine : a survey of methods and applications.

GigaScience, 7(4), 02 2018.

giy014.

References IV

Mathieu Thomas and Sophie Caillon.

Effects of farmer social status and plant biocultural value on seed circulation networks in Vanuatu.

Ecology and Society, 21(2), 2016.

Corinne Vacher, Dominique Piou, and Marie-Laure Desprez-Loustau.

Architecture of an antagonistic tree/fungus network : The asymmetric influence of past evolutionary history.

PLOS ONE, 3(3) :1–10, 03 2008.

M. J. Wainwright and M. I. Jordan.

Graphical models, exponential families, and variational inference.

Found. Trends Mach. Learn., 1(1–2) :1–305, 2008.