Stochastic block models for multilevel networks

Applications in ecology

Sophie Donnet for the Econet group, INRAE, MIA Paris-Saclay

April 2024

1. Introduction
2. Multiplex networks
3. Multipartite networks
4. Multilevel Networks

From simple and bipartite networks...

- SBM and LBM : probabilistic models for simple and bipartite networks.
- The SBM involves one group of nodes : we obtain a clustering based on the observation of their interactions.
- The LBM involves two groups of node : we obtain a bi- clustering based on the bipartite network.

... to more complex networks

Sometimes, we would like to study some more complex networks... For instance :

- study several types of relations at the same time
- study tripartite or more complex networks...
- study at the same time the relations between individuals and the connexions between the organizations they belong to.

1. Introduction

2. Multiplex networks

3. Multipartite networks

4. Multilevel Networks

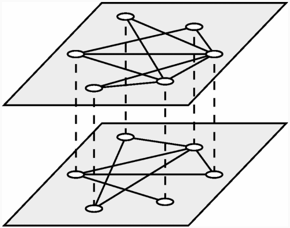
Multiplex networks

Definition

Given a set of vertices, we speak of a multiplex network if we study several relations simultaneously.

Example :

- Vertices : students
- Network 1: facebook
- Network 2: LinkedIn



In theory, each relationship can be oriented or not.

Example in agroecology : multiplex seed exchange network

- Vertices: homes
- Network 1: Sorgho exchange
- Network 2: Mil exchange

Example in ecology

Coexistence multiplex network

- Vertices : species
- Network 1: dry season coexistence, Network 2: coexistence in the wet season
- Network 1: competition, Network 2: coexistence

Formalization

Relationship between i and j described by two indicators :

$$Y_{ij} = (Y_{ij}^1, Y_{ij}^2) \quad \text{with} \quad Y_{ij}^1 \in \{0.1\} \quad \text{and} \quad Y_{ij}^2 \in \{0.1\}$$

- If 2 networks, Y_{ij} can take 4 possible values.
- If M networks, Y_{ij} can take 2^M possible values.

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

Objective

Understand / study the structure of the multiplex network

- Do all actors have the same behavior, or can we distinguish actors according to their behavior?
- Examples :
  - groups of highly connected individuals according to network 1 and no network 2.
  - groups of highly connected individuals according to the 2 networks
- For 2 non-oriented networks: R package sbm.
- For more networks: have to consider conditional independence between the layers

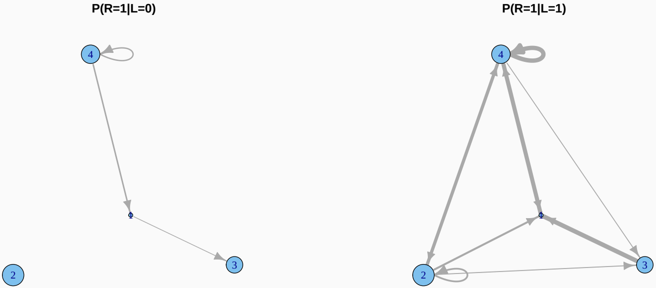
SBM for multiplex network

$$\begin{aligned} Z_i &\sim \mathcal{C}at(\pi) \\ P(Y_{ij} = w | Z_i = k, Z_j = \ell) &= \alpha_{k\ell}^{w} \end{aligned}$$

Researcher advisory networks

- Level 1: exchange of advice between researchers
- Level 2: Relationships through laboratories

Results



What you don't do with a multiplex network

- If we consider the multiplex version of the network, we don't want to "compare networks".
- Considering the multiplex version is really considering that the relationship is *multiple or complex*.

1. Introduction

2. Multiplex networks

3. Multipartite networks

3.1 The type of data we are talking about

3.2 Modeling a collection of matrices

3.3 Inference

4. Multilevel Networks

1. Introduction

2. Multiplex networks

3. Multipartite networks

3.1 The type of data we are talking about

3.2 Modeling a collection of matrices

3.3 Inference

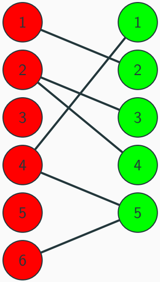
4. Multilevel Networks

Multipartite networks

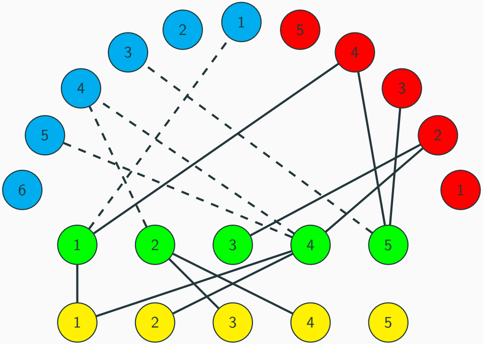
Definition

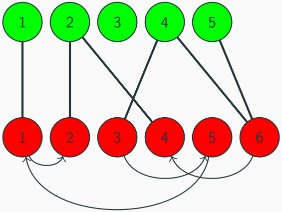
We talk about multipartite network if the vertices are divided into **several** subsets in advance.

From bipartite ...

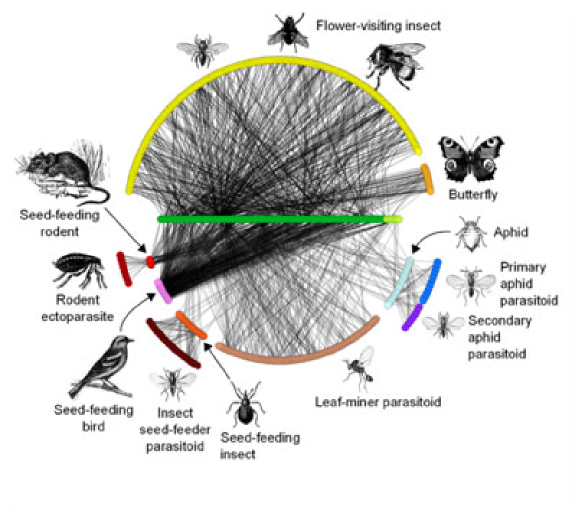


... to multipartite



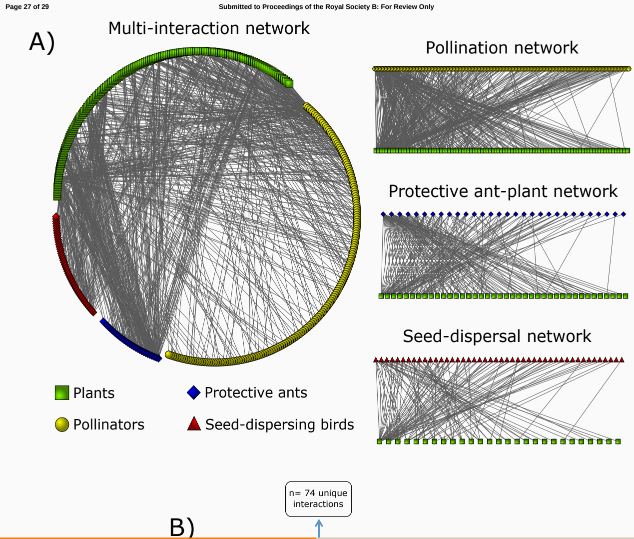


Ecology example: super multipartite network



Example in ecology: mutualist relations between animals and plants

[Dáttilo et al., 2016]



Example in ecology

- Vertices
  - Plants , Ants, Seed dispersing birds, Pollinators
- Interactions :

  [Plants / ants] – [Plants / Seed dispersing birds] – [Plants / Pollinators]
- 3 bipartite graphs
  - **Vertices/ nodes**: plants and animals (divided into functional groups $\Leftrightarrow$ multipartite graph)
  - **Edge** : if animal seen interacting with plant

$\Leftrightarrow$ 3 rectangular matrices (called *incidences*)

Data in matrices

$$
Y_{ij}^{1q} = \begin{cases} 1 & \text{if animal } j \text{ au functional group } q \text{ has been seen} \\ & \text{interacting with plant } i \\ 0 & \text{otherwise} \end{cases}
$$

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

Example 2: sociology / ecology

- Relationships between farmers (seed exchanges ...)
- Inventories of plants (species or varieties) cultivated by the farmers of the network
- 2 functional groups :
  - farmers : group 1
  - Plants: group 2
- Interactions :
  - farmers / farmers : oriented network
  - homes / Plants : bipartite network

[Thomas and Caillon, 2016]

Data in matrices for Example 2

Farmer 1			1			
Farmer 2			1			1
⋮		Y_{ij}^{11}			Y_{ij}^{12}	
Farmer n_1	1		1			1
	Farmer 1	…	Farmer n_1	Plant 1	… …	Plantn_2

Objectives

Aim

Identify subgroups of each functional group sharing the same interaction characteristics and simultaneously taking into account all the matrices.

Existing solutions

- Calculate modularity
  - Detecting communities: making subgroups of individuals who connect more within the subgroup than outside it.
  - In general, people do it separately on each type of interaction and then compare the results between them.

Proposal

Use extensions of the Latent Block Models (LBM) and Stochastic Block Models (SBM) to propose a classification of individuals/agents based on the set of observations.

1. Introduction

2. Multiplex networks

3. Multipartite networks

3.1 The type of data we are talking about

3.2 Modeling a collection of matrices

3.3 Inference

4. Multilevel Networks

Data formatting

- Q functional groups.
- Each functional group q is of size n_q.
- **Data** : a collection of matrices (of adjacency or incidence) representing the relationships within and/or between functional groups:
  - $\mathcal{E}$ = list of pairs (q, q') for which a matrix of interaction between functional groups q and q' is observed.
  - $\mathbf{Y} = \{Y^{q'}, (q, q') \in \mathcal{E}\}$ where $Y^{q'}$ is a matrix of size $n_q \times n_{q'}$.
  - If $q = q'$ matrix of adjacency, symmetrical or not
  - If $q \neq q'$, incidence matrix, bipartite graph

Examples

- Example 1: $1 =$ plants, $2 =$ ants, $3 =$ birds, $4 =$ pollinators
- Example 2: $1 =$ farmers, $2 =$ plants

Latent variable probabilistic model

- In the spirit of LBM / SBM: mixing model to model edges
- Each functional group of nodes (or vertices) q is divided into K_q blocks.
- $\forall q = 1 \dots Q$, $Z_i^q = k$ if the entity i of the functional group q belongs to the block k.

Latent variables

$(Z_i^q)_{i=1\dots n_q}$ latent, independent random variables: $\forall k = 1 \dots K_q$, $\forall i = 1 \dots n_q$, $\forall q = 1 \dots Q$,

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

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

Latent variable probabilistic model

Conditionally...

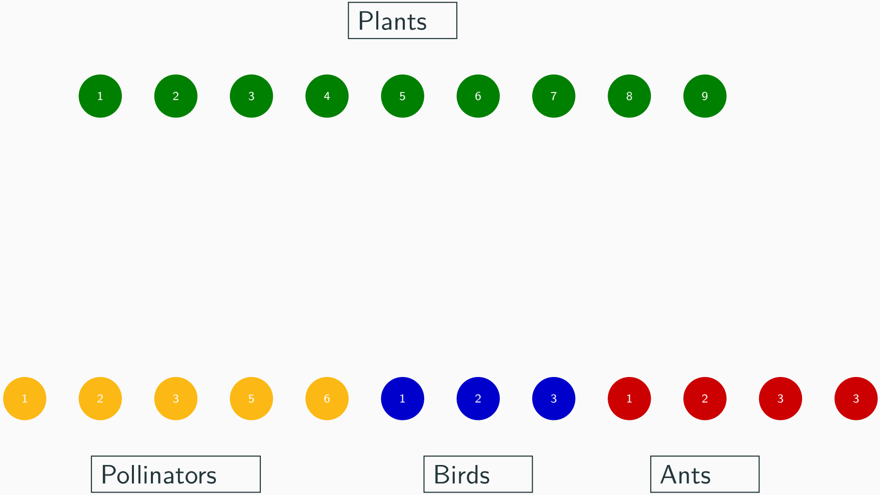
... to latent variables $\boldsymbol{Z} = \{Z_i^q, i = 1 \dots n_q, q = 1 \dots Q\}$:

$$Y_{ij}^{q'} | Z_i^q, Z_j^{q'} \sim_{i.i.d} \mathcal{F}(\alpha^{qq'}_{Z_i^q, Z_j^{q'}}) . \qquad (2)$$

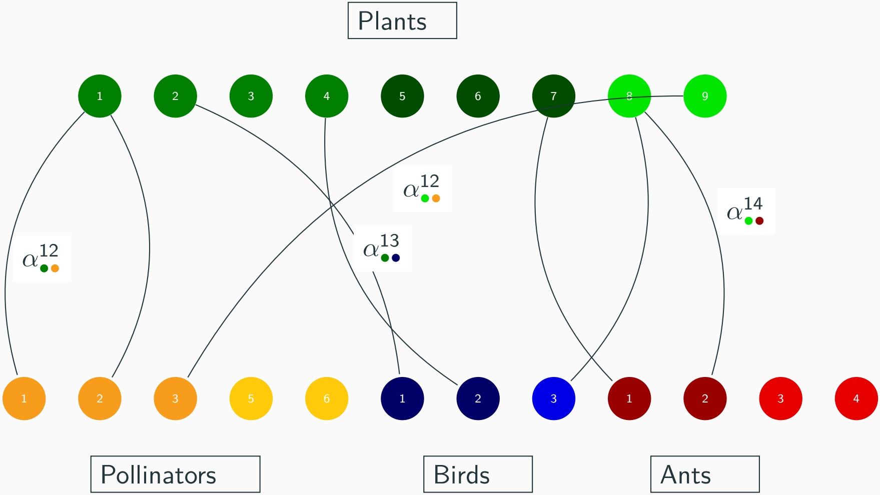
- Law of the interaction phenomenon depends on the i and j membership groups
- In the examples, $\mathcal{F} = \mathcal{B}ern$ but other possible laws (Fish...)
- Special cases
  - If only one functional group and $\mathcal{E} = \{(1,1)\}$: SBM
  - If two functional groups and $\mathcal{E} = \{(1,2)\}$: LBM

[Bar-Hen et al., pear]

Synthetic scheme for plants/insects networks



Synthetic scheme for plants/insects networks



Résumé

Latent variables

Each functional group q divided into K_q clusters

- $\forall q = 1 \dots Q, \forall i = 1 \dots n_q,\ Z_i \in \{1, \dots, K_q\}$ Latent variables
- $\pi_k^q = \mathbb{P}(Z_i^q = k),\ \forall i, \forall k, \forall q$
- $\sum_{k=1}^{K_q} \pi_k^q = 1$
- i.i.d. variables

Connection distribution

Conditionally to the latent variables : $\forall (q, q') \in \mathcal{E}$

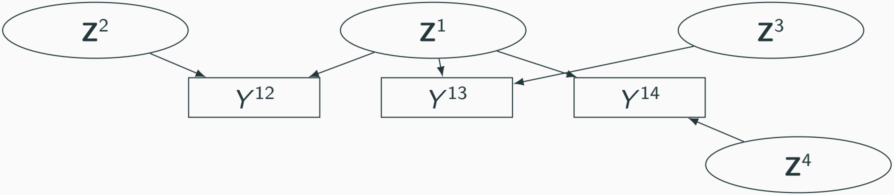
$$Y_{ij}^{qq'} | Z_i^q, Z_j^{q'} \sim_{ind} \mathcal{F}(\alpha^{qq'}_{Z_i^q, Z_j^{q'}}) \,.$$

Dependencies between matrices

- If $K_q = 1$ for all q then all the entries of all the matrices are independent random variables: homogeneous connection.
- Otherwise, integration of the random variables $\Rightarrow$ dependence between the elements of the matrices
- Dependence between matrices
- **Consequences** on $\mathbf{Z}^q | \mathbf{Y}$
  - The obtained clustering depends on all interaction matrices.
  - Few simplifications possible

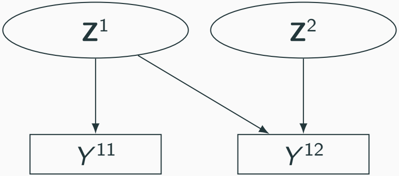
DAG for example 1

- $1 =$ plants
- $2 =$ ants
- $3 =$ farmers
- $2 =$ cultivated plants



DAG pour l'exemple 2

- $1 =$ farmers
- $2 =$ species



1. Introduction

2. Multiplex networks

3. Multipartite networks

3.1 The type of data we are talking about

3.2 Modeling a collection of matrices

3.3 Inference

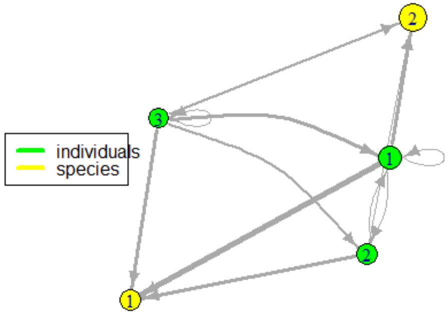
4. Multilevel Networks

Estimation and model selection

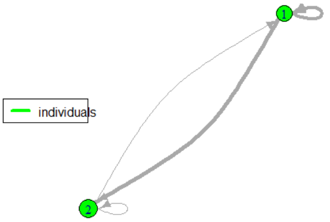
- Likelihood maximized by an adapted version of the VEM algorithm
- Numbers of blocks $(K_1, \dots; K_Q)$ chosen with an adapted ICL criterion (penalized likelihood)
- Method implemented in R package sbm

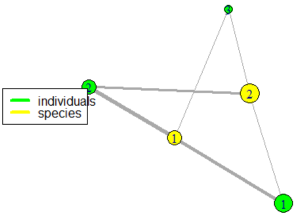
Results on data MIRES

2 groups of crop species, 3 groups of farmers

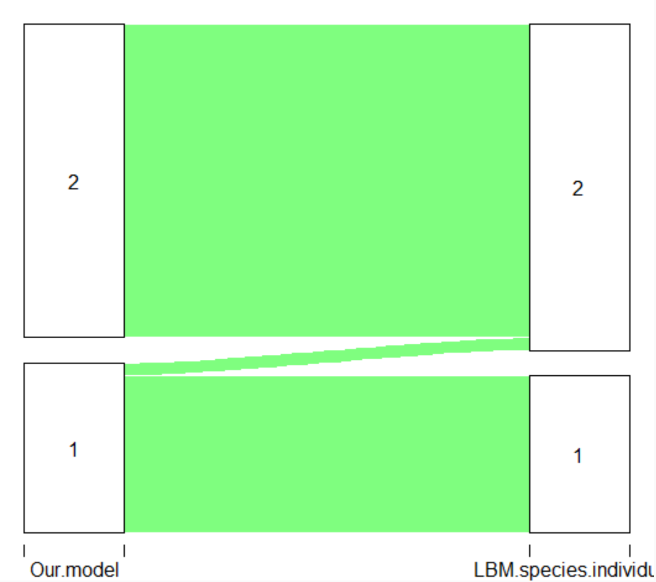


Comparison with a LBM or SBM

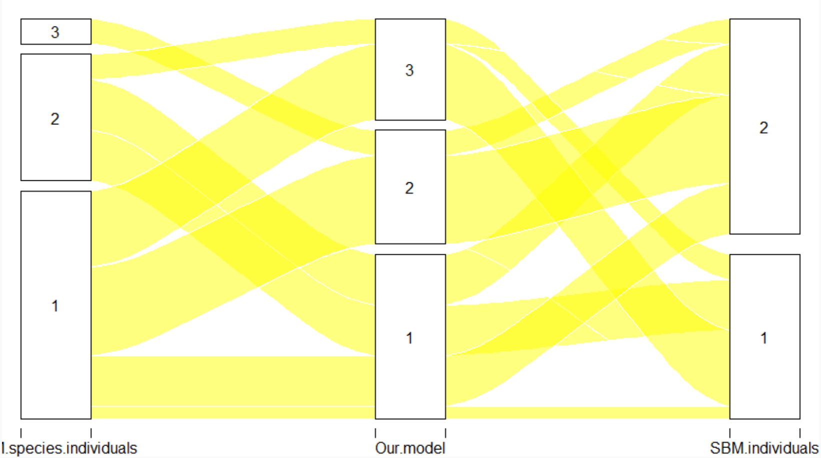




Comparison of crop species classifications



Comparison of individual classifications



1. Introduction

2. Multiplex networks

3. Multipartite networks

4. Multilevel Networks

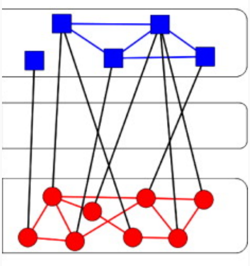
Multilevel networks

Definition

A multilevel network is said to exist if the vertices are divided into **several** subsets in advance and there is a **hierarchical** relationship between the vertices.

Example :

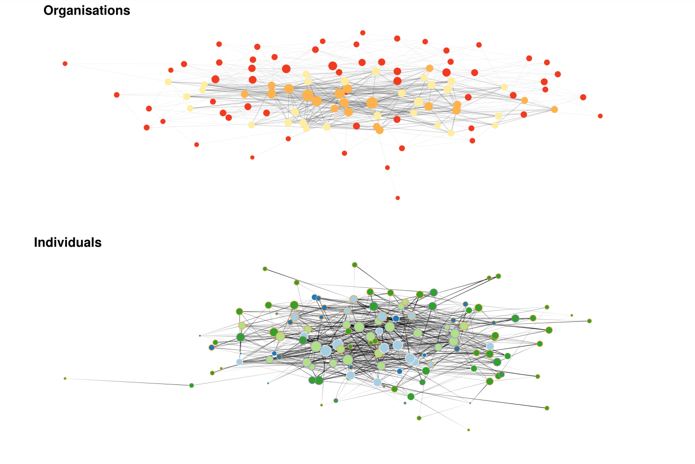
- Network of encounters between dogs
- Network of encounters of their owners



Each relationship can be oriented or not.

Example in sociology

- Informal inter-individual network (counseling, oriented)
- Formal inter-organizational network (contract, not oriented)
- Relationship of affiliation of each individual to a single organization



Objectives

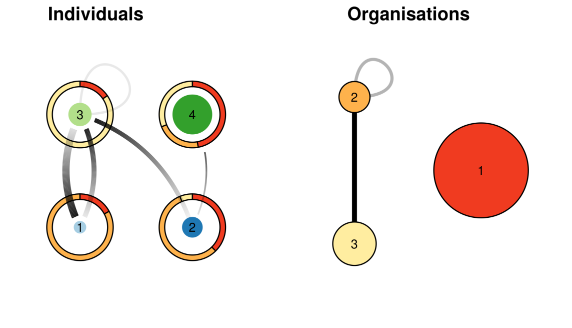
- Obtain a joint classification of individuals and organizations
- Stating on a dependency of the connection structures between the two levels.

[Chabert-Liddell et al., 2019]

Results in Sociology

Package R : `MLVSBM` which can manage two-level multilevel networks with binary data.

- 4 groups of individuals
- 3 groups of organizations
- Dependent inter-individual and inter-organizational relationships



Conclusions and perspectives

Not evoked multiple networks

- Dynamic networks: networks that evolve over time.
  - Either: network photo in discrete times
  - Either: observation of connections in continuous time
- Spatial variation of a network of interest: observation at different locations of the "same" network.

The evoked networks: multilayer networks

- List of networks *multi* that we are able to model and infer
- Which ones are not included in this catalog?
- Does it make sense to take into account several networks at the same time?
- Do we prefer comparing networks? How do we compare networks defined on different sets of nodes.

Références i

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

Block models for multipartite networks.applications in ecology and ethnobiology.

Statistical Modelling.

Barbillon, P., Donnet, S., Lazega, E., and Bar-Hen, A. (2016).

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.

Chabert-Liddell, S.-C., Barbillon, P., Donnet, S., and Lazega, E. (2019).

A stochastic block model for multilevel networks: Application to the sociology of organizations.

Dáttilo, W., Lara-Rodríguez, N., Jordano, P., Guimarães, P. R., Thompson, J. N., Marquis, R. J., Medeiros, L. P., Ortiz-Pulido, R., Marcos-García, M. A., and Rico-Gray, V. (2016).

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).

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.

Thomas, M. and Caillon, S. (2016).

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

Ecology and Society, 21(2).