

Données de réseaux : introduction

Sophie Donnet, François Massol, Nicolas Verzelen
MIA Paris, INRA, CNRS

Formation Réseaux MIRES / ReSodiv / COEX
18-19-06/2019



Tous les documents seront disponibles à cette adresse
https://github.com/Sophiedonnet/Formation_Reseau_MIRES

Contexte

Des données terrains au réseau

Les différents types de réseaux

Inférence statistique

Pourquoi un réseau ?

- ▶ Etudier/comprendre un **certain type de relations** entre des **acteurs donnés**
- ▶ **Exemples**
 - ▶ Dons de semences de riz entre agriculteurs de riz au Vietnam
 - ▶ Relations d'amitiés entre agriculteurs au Vanuatu
 - ▶ Relations de co-occurrence entre espèces animales dans un écosystème donné
 - ▶ Inventaires de culture : relations entre agricultures et des espèces.
 - ▶ Relation de pollinisation entre espèces animales et espèces végétales

Des données terrains au réseau

Définition d'un réseau

Définition

Un réseau est la donnée d'un ensemble de noeuds (ou sommets) et d'arêtes entre ces sommets.

- ▶ Un réseau est spécifique à *un type de relation fixé*.
- ▶ Les noeuds représentent des agents / individus / espèces animales
- ▶ Une arête existe entre deux sommets si ces deux agents ont une relation.



Du terrain à l'objet réseau

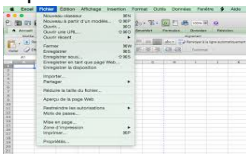
- Partant des interviews ou des relevés de terrain...



- ... on recupère des notes...



- ... qu'on transforme en fichier excel



Un exemple de données terrain

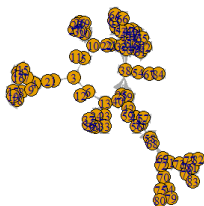
- ▶ Données de E. Pannier (IRD, UMR 208 Paloc)
- ▶ Circulation de variétés de riz au VietNam.
- ▶ Naturellement : liste des échanges *ayant eu lieu*

event tham duong.xlsx - Microsoft Excel

	A	B	C	D	E	F	G	H	I	J	K	L
	label	source	label	target	Type of transaction (code)	Type of transaction (code)	Nature of the material	Code nature of the material 1: selected seed 2: paddy	Variety	quantity	homogenized quantity (kg of seeds)	Nature of the return
1												
2	Vi Van Sau	1	Hoang Thanh Bich	2	loan		3	paddy	2	KTD	40 kg	40
3	Vi Van Sau	1	La Van Thao	3	loan		3	paddy	2	KTD	24kg	24
4	Vi Van Sau	1	Luc Van Thuong- Ba Men	4	sale		4	paddy	2	KTD	10kg	10
5	La Van Thao	3	La Van Tam	5	gift		1	seedling	3	KTD	1 ganh ma	
6	La Van Thao	3	La Van Ten	6	gift		1	seedling	3	BT	10 ganh	
7	Hoang Van Hu	7	Luc Van Thuong- Ba Men	4	sale		4	paddy	2	KTD	40	money
8	Hoang Van Hu	7	Luc Van Thuong- Ba Men	4	gift		1	selected seed	1	RVT		
9	Vi Van Sinh	8	Luc Van Thuong- Ba Men	4	sale		4	paddy	2	KTD	15kg	15 money
10	Ba Hoa-Phang	9	Luc Van Thuong- Ba Men	4	gift		1	seedling	3	KTD	2 ganh	
11	Ha Van Van	10	La Van Tam	5	gift		1	paddy	2	KTD	7kg	7
12	Ha Van Van	10	La Van Tam	5	exchange		2	paddy	2	BT	30kg	30 paddy
13	Ha Van Thinh	11	La Van Tam	5	loan		3	paddy	2	KTD	5kg	5 paddy
14	La Van Quyen	12	La Van Ten	6	gift		1	seedling	3	BT		
15	Nguyen Van Mung	13	La Van Ten	6	gift		1	seedling	3	BT		
16	Hoang Van Hu	7	La Van Nam	14	sale		4	selected seed	1	KTD	20kg	20 money
17	Hoang Van Hu	7	Hoang Van Thien	15	sale		4	selected seed	1	KTD	3,2kg	3,2 money
18	Hoang Van Hu	7	La Thi Phung	16	gift		1	selected seed	1	KTD	5kg	5
19	Vi Van Sinh	8	Nguyen Van Thuong	17	sale		4	selected seed	1	KTD	20kg	20 money
20	Vi Van Sinh	8	Mir X	18	sale		4	selected seed	1	KTD	10kg	10 money
21	Vi Van Sinh	8	Ha Van Tuan	19	sale		4	selected seed	1	KTD	5kg	5 money
22	Ha Van Van	10	La Van Khoong	20	exchange		2	selected seed	1	KTD	15kg	15

Représentation du réseau

- Liste des échanges : liste des arêtes
- Suffisant pour décrire et tracer le réseau



- Avantages : description fine des échanges (quantités, occasion, type de relation...)

Encodage en matrice

- ▶ Définir un tableau avec en ligne et en colonne les individus.
- ▶ Dans chaque case (i, j) :

$$y_{ij} = \begin{cases} 1 & \text{si relation entre } i \text{ et } j \\ 0 & \text{sinon} \end{cases}$$

The screenshot shows a spreadsheet application with a large table of data. The table has columns labeled A through V, each corresponding to an individual ID (e.g., IND_001, IND_002, ..., IND_175). The rows also represent these individuals. The cells contain binary values (0 or 1), representing the presence or absence of a relationship between the individuals. The spreadsheet is titled 'Individus' and shows a large grid of data.

(données du Vanuatu)

- ▶ Passage de liste d'arête à la matrice d'adjacence/incidence par une fonction R : `igraph`
- ▶ Ne permet pas de stocker les descripteurs de la relation.

Contexte

Des données terrains au réseau

Les différents types de réseaux

Inférence statistique

Les réseaux simples

- ▶ Relation au sein d'un groupe d'acteurs : amitiés, échange de semences, co-occurrences d'espèces, réseaux trophiques
 - ▶ Matrice d'adjacence : matrice carrée

Les réseaux dirigés ou non

Relation peut être *orientée (dirigée)* ou non

	Réseaux orientés	Réseaux non-orientés
Exemples	Don, réseau trophique	Amitiés, co-occurrence d'espèces
Arête	flèche	trait
Matrice	carrée, non-symétrique $y_{ij} \neq y_{ji}$	carrée, symétrique $y_{ij} = y_{ji}$

Les réseaux valués

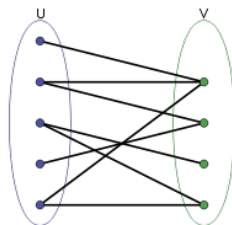
La relation peut être décrite par un nombre, orientée ou non.

- ▶ Exemple : nombre de fois où i a donné des semences à j
- ▶ Arête : épaisseur du trait proportionnelle à la valeur
- ▶ Matrice : symétrique ou non, pas seulement de 0/1
- ▶ $y_{ij} \in \mathbb{N}$, $y_{ij} \in \mathbb{R}$

Réseaux bipartites

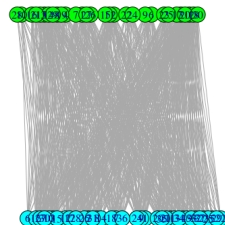
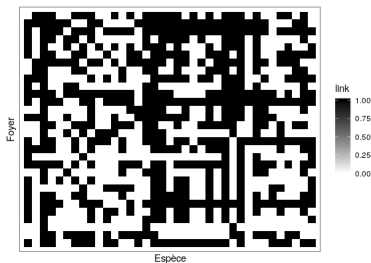
Définition

Un réseau est dit bipartite si les sommets sont divisés en deux sous-ensembles et une arête a une extrémité dans un des sous-ensemble et l'autre dans l'autre sous-ensemble.



Exemple de réseau bipartite : inventaire de cultures

- ▶ Entre de sommets U : foyers
- ▶ Entre de sommets V : espèces végétales
- ▶ Relation : culture
- ▶ Arête entre i et j si foyer i cultive espèces j .



Contexte

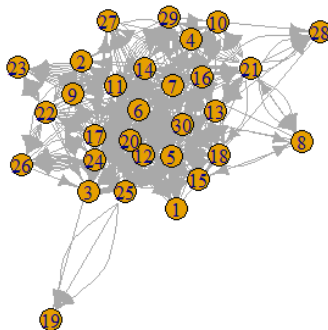
Des données terrains au réseau

Les différents types de réseaux

Inférence statistique

Statistiques : pour quoi faire ?

Comprendre ça :



Statistiques : pour quoi faire ?

Comprendre / étudier la structure du réseau

- ▶ Vue générale :
 - ▶ très connecté / peu connecté (densité)...
 - ▶ Existence de star ?
 - ▶ Communauté (= sous-groupes d'acteurs plus connectés entre eux qu'avec le reste des acteurs).
- ▶ Du point de vue des individus : existence de généralistes et de spécialistes, centralité, etc..
- ▶ Avoir une vue résumée du réseau

Comment faire ?

Si le réseau est petit (10-15 sommets)...

Possible de le faire à la main

Si plus de sommets...

- ▶ Statistiques descriptives
- ▶ Approche egocentrée
- ▶ Modèles probabilistes : modèles à blocs stochastiques Outils automatiques pour faire des groupes de sommets qui ont le même rôle dans le réseau