

Que dire de la structure d'un réseau observé ?

Rencontres de la chaire MMB

Sarah Ouadah

LPSM, Sorbonne Université

29 janvier 2025

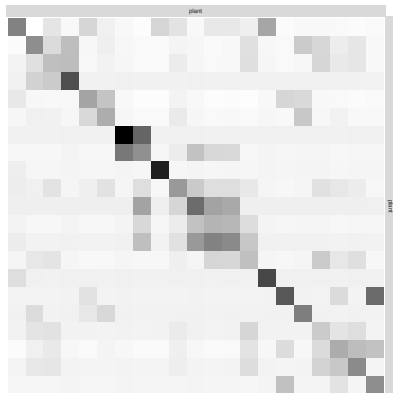
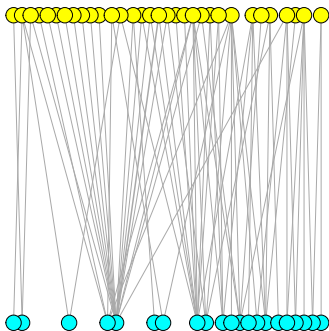


Étude de la topologie de graphes aléatoires pour l'analyse de réseaux sociaux et écologiques

- ★ **Objectifs** caractériser la structure d'un réseau observé, comparer plusieurs réseaux
- ★ **Mots-clés** modèles de graphes aléatoires, graphon, clustering de nœuds

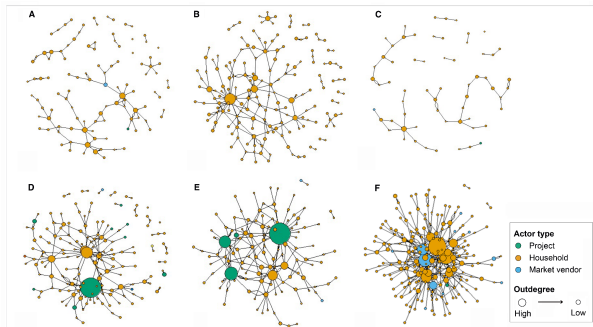
Réseaux écologiques

Plant-pollinator relationships at two altitudes in the Andes of Mendoza, Argentina
[MMD⁺02]



Réseaux sociaux

Seed circulation networks in a Bassari village in south-eastern Senegal [PFLAF⁺23]



A) Sorghum, B) Bambara groundnut, C) fonio, D) maize, E) rice, F) peanut

Théorie statistique des réseaux aux échelles micro, méso et macro

Questions À partir de l'observation d'un réseau

- ★ Comment se structure le réseau ?
- ★ Comment constituer des groupes de nœuds ?
- ★ Comment étudier les relations indirectes entre acteurs d'un réseau ?
- ★ Comment comparer plusieurs réseaux ?

Théorie statistique des réseaux aux échelles micro, méso et macro

Approches

- ★ *Embedding* de graphe/nœuds pour analyses multivariées et clustering
- ★ Tests d'adéquation à des **modèles de graphes aléatoires** basés sur des **métriques** appropriées
- ★ Tests de **comparaison de réseaux**

Statistique descriptive et modélisation pour caractériser la
structure d'un réseau observé

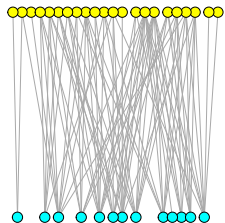
Métriques de réseaux

Métriques

- ★ degré et indicateurs de centralité (échelle locale)
- ★ motifs (échelle méso)
- ★ densité et modularité (échelle globale)

Motifs bipartites

Réseau plante-pollinisateur [MMD⁺02]



Motifs [SCB⁺19]¹

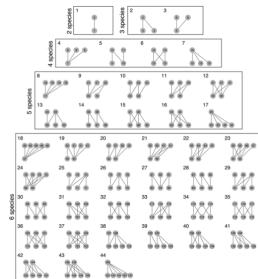


Figure 3. All possible two- to six-species bipartite motifs. Large numbers represent individual motifs. Small numbers within nodes represent unique position within motifs. In total there are 148 positions across 44 motifs.

¹Figure issue de [SCB⁺19]

Motifs bipartites

Motifs Sous-graphes représentant des configurations d'interactions entre un petit nombre de nœuds

- ★ Compter les motifs [PDK⁺08], [SCB⁺19], package bmotif
- ★ Distribution des comptages des motifs triangle [NW88, Sta01, PDK⁺08]
- ★ Les motifs dépendent de
 - ◇ taille du réseau
 - ◇ densité du réseau
 - ◇ distribution des degrés des nœuds du haut (insectes spécialistes vs généralistes)
 - ◇ distribution des degrés des nœuds du bas (plantes spécialistes vs généralistes)

Modèles de graphes aléatoires

Modèles

- ★ Erdős-Rényi [ER59]
- ★ attachement préférentiel [BA99]
- ★ exponentiel de graphe (ERGM) [WP96]
- ★ W -graphe [LS06, DJ08] dont les modèles à blocs stochastiques (SBM) [NS01], EDD [CL02], etc.

Modèle de W -graphe

$$\{U_i\}_{1 \leq i \leq m} \text{ iid}, U_1 \sim \mathcal{U}_{[0,1]}$$

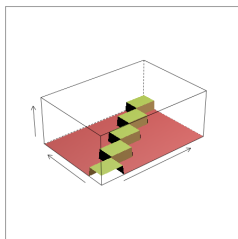
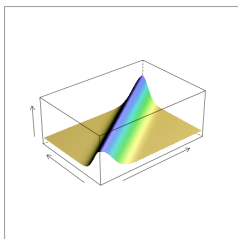
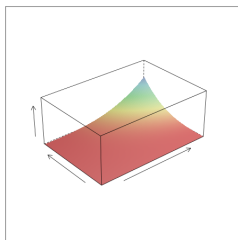
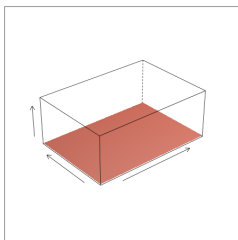
$$\{V_j\}_{1 \leq j \leq n} \text{ iid}, V_1 \sim \mathcal{U}_{[0,1]}$$

$$\{Y_{ij}\}_{i,j} \text{ ind} | \{U_i\}_i, \{V_j\}_j$$

$$Y_{ij} | U_i, V_j \sim \mathcal{B}(W(U_i, V_j))$$

$W : [0, 1]^2 \mapsto [0, 1]$ la fonction graphon

Graphons



Erdős-Rényi, *scale-free*, *small world*, comm. ²

²Figures de P. Latouche

Modèle bipartite de distribution des degrés attendu B-EDD [OLR22]

Modèle B-EDD [OLR22] (graphon produit)

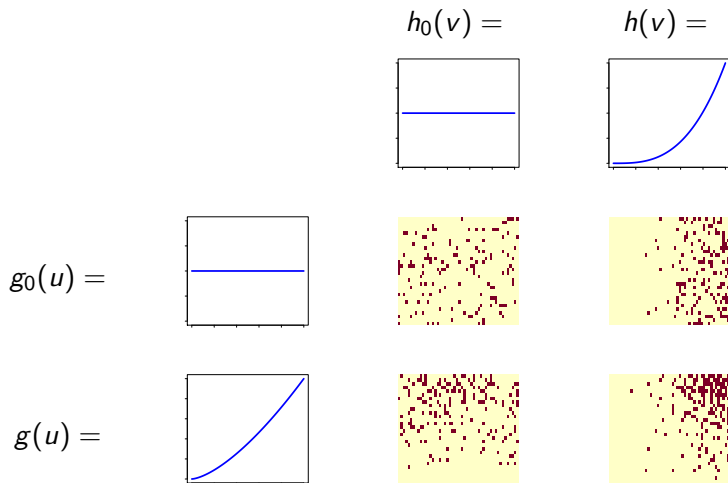
$$\{U_i\}_{1 \leq i \leq m} \text{ iid}, U_1 \sim \mathcal{U}_{[0,1]}$$

$$\{V_j\}_{1 \leq j \leq n} \text{ iid}, V_1 \sim \mathcal{U}_{[0,1]}$$

$$\{Y_{ij}\}_{i,j} \text{ ind} | \{U_i\}_i, \{V_j\}_j$$

$$Y_{ij} | U_i, V_j \sim \mathcal{B}(\rho g(U_i) h(V_j))$$

ρ densité du graphe, $g, h : [0, 1] \mapsto \mathbb{R}^+$ bornées, $\int g = \int h = 1$ traduisent l'hétérogénéité des degrés (nombre de connexions) des nœuds



Objectifs

- ★ comprendre l'organisation d'un réseau
- ★ comparer plusieurs réseaux

→ *Embedding* de graphe/nœuds pour analyses multivariées et clustering [en cours]

→ Tests d'adéquation à des **modèles de graphes aléatoires** basés sur des **métriques** appropriées [LRO18, ORL20, OLR22]

→ Tests de **comparaison de réseaux** [OLR22]

Motif-based tests for bipartite networks

[OLR22] *Electronic Journal of Statistics*

Fréquences de motif

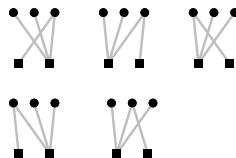
Positions

$$\alpha = 1, \dots, c_s$$

Nombre de positions

$$c_s := \binom{m}{p} \times \binom{n}{q} \times r_s$$

Automorphismes r_s



Fréquences de motif

$$F_s := \sum_{\alpha} Y_{s\alpha} / c_s$$

où $Y_{s\alpha} = 1$ si s en α , 0 sinon

Moments des fréquences de motif sous B-EDD

- ★ moments explicites : $\bar{\phi}_s$ fréquence attendue du motif s sous B-EDD

$$\begin{aligned}
 \bar{\phi}_9 &= \mathbb{P} \left(\begin{array}{c} \bullet \quad \bullet \quad \bullet \\ \diagdown \quad \diagup \quad \diagdown \\ \blacksquare \quad \blacksquare \end{array} \right) \\
 &= (\phi_1)^4 \times \frac{(\phi_1)^2 \phi_2}{(\phi_1)^4} \times \frac{\phi_1 \phi_4}{(\phi_1)^4} \\
 &= \frac{\phi_2 \phi_4}{\phi_1}
 \end{aligned}$$

avec $\phi_1 = \mathbb{P} \left(\begin{array}{c} \bullet \\ | \\ \blacksquare \end{array} \right)$, $\phi_2 = \mathbb{P} \left(\begin{array}{c} \bullet \\ \diagdown \quad \diagup \\ \blacksquare \quad \blacksquare \end{array} \right)$, $\phi_4 = \mathbb{P} \left(\begin{array}{c} \bullet \quad \bullet \quad \bullet \\ \diagdown \quad \diagup \quad \diagdown \\ \blacksquare \end{array} \right)$ les fréquences des motifs 1, 2 et 4

- ★ estimateurs des moments : $\bar{F}_9 = F_2 F_4 / F_1$

Tests d'adéquation à B-EDD ($W = \rho gh$) basés sur les motifs

Théorème

Sous des conditions de parcimonie du graphe $\mathbf{Y}$, pour tout motif s non étoilé : sous l'hypothèse $\{H_0 : \mathbf{Y} \text{ suit un modèle B-EDD}\}$

$$\frac{F_s - \hat{\mathbb{E}}(F_s)}{\sqrt{\hat{\mathbb{V}}(F_s)}} \xrightarrow{D} \mathcal{N}(0, 1), m \sim n \rightarrow \infty$$

où

F_s la fréquence empirique du motif s

$$\hat{\mathbb{E}}(F_s) = \bar{F}_s$$

$$\hat{\mathbb{V}}(F_s) = \kappa_{m,n,s} \bar{F}_s + \kappa'_{m,n,s} \bar{F}_s^2 + \sum_{S \in \mathcal{S}_2(s)} \kappa''_{m,n,s,S} \bar{F}_S - \bar{F}_s^2, \text{ avec } \mathcal{S}_2(s)$$

l'ensemble des super-motifs

- ★ estimateurs des moments consistants
- ★ propriétés des suites de différence de martingale

Tests de comparaison [OLR22]

Tests de comparaison de $\mathbf{Y}^A$ et $\mathbf{Y}^B$ e.g. $\{H_0 : g^A = g^B\}$

$\mathbf{Y}^A$ et $\mathbf{Y}^B$ générés sous B-EDD(ρ^A, g^A, h^A) et B-EDD(ρ^B, g^B, h^B) resp.
Contrastes entre F_9^A et l'estimation de

$$\bar{\phi}_9^{H_0, A} = \mathbb{P}_{H_0} \left(\begin{array}{c} \bullet \quad \bullet \quad \bullet \\ \diagdown \quad \diagup \quad \diagup \\ \blacksquare \quad \blacksquare \end{array} \right) = (\phi_1^A)^4 \times \frac{(\phi_1^B)^2 \phi_2^B}{(\phi_1^B)^4} \times \frac{\phi_1^A \phi_4^A}{(\phi_1^A)^4}$$

et entre F_9^B et l'estimation de

$$\bar{\phi}_9^{H_0, B} = (\phi_1^B)^4 \times \frac{(\phi_1^A)^2 \phi_2^A}{(\phi_1^A)^4} \times \frac{\phi_1^B \phi_4^B}{(\phi_1^B)^4}$$

Illustration

Réseaux écologiques de Web of Life database (www.web-of-life.es)

★ coefficients c_s , comptages N_s et fréquences F_s

plante-pollinisateur [Rob29] : $m = 546, n = 1044$							
	arête	étoiles du haut			étoiles du bas		
s	1	2	7	17	3	4	8
c_s	$4.76 \cdot 10^5$	$2.48 \cdot 10^8$	$8.62 \cdot 10^{10}$	$2.24 \cdot 10^{13}$	$1.08 \cdot 10^8$	$1.64 \cdot 10^{10}$	$1.86 \cdot 10^{12}$
N_s	$1.53 \cdot 10^4$	$2.61 \cdot 10^5$	$3.04 \cdot 10^6$	$2.72 \cdot 10^7$	$3.07 \cdot 10^5$	$6.82 \cdot 10^6$	$1.48 \cdot 10^8$
F_s	$3.20 \cdot 10^{-2}$	$1.05 \cdot 10^{-3}$	$3.52 \cdot 10^{-5}$	$1.21 \cdot 10^{-6}$	$2.84 \cdot 10^{-3}$	$4.16 \cdot 10^{-4}$	$7.99 \cdot 10^{-5}$

plante-disperseur [Sil02] : $m = 207, n = 110$							
	arête	étoiles du haut			étoiles du bas		
s	1	2	7	17	3	4	8
c_s	$2.28 \cdot 10^4$	$1.24 \cdot 10^6$	$4.47 \cdot 10^7$	$1.20 \cdot 10^9$	$2.35 \cdot 10^6$	$1.60 \cdot 10^8$	$8.17 \cdot 10^9$
N_s	$1.12 \cdot 10^3$	$6.50 \cdot 10^3$	$4.07 \cdot 10^4$	$2.32 \cdot 10^5$	$1.24 \cdot 10^4$	$1.31 \cdot 10^5$	$1.23 \cdot 10^6$
F_s	$4.92 \cdot 10^{-2}$	$5.23 \cdot 10^{-3}$	$9.11 \cdot 10^{-4}$	$1.94 \cdot 10^{-4}$	$5.28 \cdot 10^{-3}$	$8.16 \cdot 10^{-4}$	$1.50 \cdot 10^{-4}$

Illustration

Réseaux écologiques de Web of Life database (www.web-of-life.es)

★ statistique du test d'adéquation à B-EDD

plante-pollinisateur					
<i>s</i>	5	6	10	15	16
W_s	$-6.45 \cdot 10^{-2}$	$9.96 \cdot 10^{-1}$	$-6.63 \cdot 10^{-2}$	$7.52 \cdot 10^{-1}$	2.43

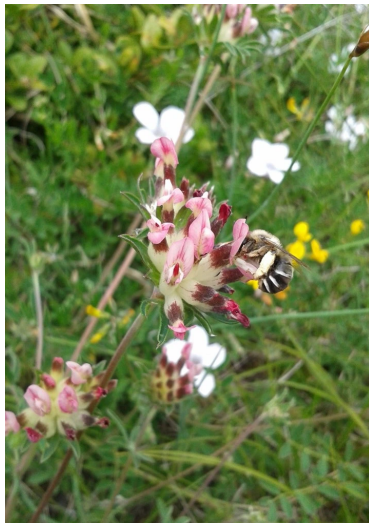
plante-disperseur					
<i>s</i>	5	6	10	15	16
W_s	$-2.14 \cdot 10^{-1}$	$-2.14 \cdot 10^{-1}$	$-2.93 \cdot 10^{-1}$	$-2.95 \cdot 10^{-1}$	$-3.56 \cdot 10^{-1}$

★ statistique du test de comparaison des réseaux plante-pollinisateur et plante-disperseur

<i>s</i>	5	6	10	15	16
W_s	-1.56	-1.56	-0.97	-1.28	-0.96

Travail en cours

Ecologie des plantes et pollinisateurs dans les prairies calcaires le long d'un gradient latitudinal en France ⁴ [dM19]



⁴ Images de N. de Manincor

Merci

Références I



A. L. Barabási and R. Albert.

Emergence of scaling in random networks.

Science, 286:509–512, 1999.



F. Chung and L. Lu.

Connected components in random graphs with given expected degree sequences.

Annals of combinatorics, 6(2):125–145, 2002.



P. Diaconis and S. Janson.

Graph limits and exchangeable random graphs.

Rend. Mat. Appl., 7(28):33–61, 2008.



N. de Manincor.

Ecologie des plantes et des pollinisateurs dans les prairies calcaires le long d'un gradient latitudinal en france: diversité des espèces et structure des réseaux d'interaction plantes-pollinisateurs.

Thèse de doctorat, 2019.



P. Erdős and A. Rényi.

On random graphs.

I Publicationes Mathematicae (Debrecen), 6:290–297, 1959.



P. Latouche, S. Robin, and S. Ouadah.

Goodness of fit of logistic regression models for random graphs.

Journal of Computational and Graphical Statistics, 27(1):98–109, 2018.



L. Lovász and B. Szegedy.

Limits of dense graph sequences.

Journal of Combinatorial Theory, Series B, 96(6):933–957, 2006.

Références II



D. Medan, N.H Montaldo, M. Devoto, A. Maniese, V. Vasellati, G. G Roitman, and N. H Bartoloni.
Plant-pollinator relationships at two altitudes in the andes of mendoza, argentina.
Arctic, Antarctic, and Alpine Research, 34(3):233–241, 2002.



K. Nowicki and T.A.B. Snijders.
Estimation and prediction for stochastic block-structures.
J. Amer. Statist. Ass., 96:1077–87, 2001.



K. Nowicki and J. C. Wierman.
Subgraph counts in random graphs using incomplete u-statistics methods.
Discrete Math., 72(1):299–310, 1988.



S. Ouadah, P. Latouche, and S. Robin.
Motif-based tests for bipartite networks.
Electronic Journal of Statistics, 16(1):293 – 330, 2022.



S. Ouadah, S. Robin, and P. Latouche.
Degree-based goodness-of-fit tests for heterogeneous random graph models: Independent and exchangeable cases.
Scandinavian Journal of Statistics, 47(1):156–181, 2020.



F. Picard, J.-J. Daudin, M. Koskas, S. Schbath, and S. Robin.
Assessing the exceptionality of network motifs.,
J. Comput. Biol., 15(1):1–20, 2008.



A. Porcuna-Ferrer, V. Labeyrie, S. Alvarez-Fernandez, L. Calvet-Mir, Ndèye F. Faye, S. Ouadah, and V. Reyes-García.
Crop biocultural traits shape seed networks: Implications for social-ecological resilience in south eastern senegal.
Agricultural Systems, 211:103750, 2023.

Références III



C. Robertson.

Flowers and insects: lists of visitors to four hundred and fifty-three flowers. carlinville, il, usa, c. robertson. *National Center for Ecological Analysis and Synthesis Interaction Web Database*, 456, 1929.



B. I. Simmons, A. R. Cirtwill, N. J. Baker, H. S. Wauchope, L. V. Dicks, D. B. Stouffer, and W. J. Sutherland.

Motifs in bipartite ecological networks: uncovering indirect interactions. *Oikos*, 128(2):154–170, 2019.



WR Silva.

Patterns of fruit-frugivore interactions in two atlantic forest bird communities of south-eastern brazil: implications for conservation.

Seed dispersal and frugivory: ecology, evolution and conservation, pages 423–435, 2002.



D. Stark.

Compound poisson approximations of subgraph counts in random graphs.

Random Structures & Algorithms, 18(1):39–60, 2001.



S. Wasserman and P. Pattison.

Logit models and logistic regressions for social networks: I. an introduction to markov graphs andp. *Psychometrika*, 61(3):401–425, 1996.