

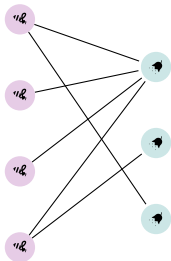
Analyse jointe de collections de réseaux bipartites

Soutenance de Master MathSV

Louis Lacoste, encadré par Pierre Barbillon et Sophie Donnet
Laboratoire MIA Paris-Saclay

29 août 2024

Pourquoi un réseau ?



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

Matrice
d'adjacence
associée

Figure 1 – Exemple
d'un réseau

- Modélisation d'interactions variées, ici d'écosystèmes
- Structure nécessaire pour : suivi biodiversité, robustesse, risque d'effondrement
- De plus en plus disponibles

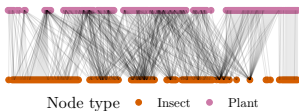


Figure 2 – Réseau plante-pollinisateur de Bristol
Baldock et al., 2019

Méthodes d'analyse pour un réseau

Plusieurs méthodes :

- Métriques : degré, centralité, emboîtement . . .
- Plongement des réseaux avec GNN
- *Clustering* des nœuds avec modèles à variables latentes

Méthodes d'analyse pour un réseau

Plusieurs méthodes :

- Métriques : degré, centralité, emboîtement . . .
- Plongement des réseaux avec GNN
- **Clustering des nœuds avec modèles à variables latentes**

Latent Block Model (LBM¹)

Govaert et Nadif, 2005.

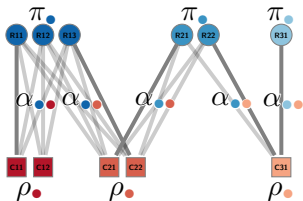


Figure 3 – Exemple de LBM¹

Modèle hiérarchique

$$\forall q \in \llbracket 1, Q_1 \rrbracket, \mathbb{P}(Z_i = q) = \pi_q$$

$$\forall r \in \llbracket 1, Q_2 \rrbracket, \mathbb{P}(W_j = r) = \rho_r$$

$$Y_{ij} | Z_i, W_j \sim \mathcal{F}(\alpha_{Z_i, W_j})$$

où $|\pi| = Q_1, |\rho| = Q_2, |\alpha| = Q_1 \times Q_2$

Formule concise LBM

$$Y \sim \mathcal{F}\text{-BiSBM}_{n_1, n_2}(Q_1, Q_2, \pi, \rho, \alpha)$$

1. Que j'appellerai par la suite BiSBM

Latent Block Model (LBM¹)

Govaert et Nadif, 2005.

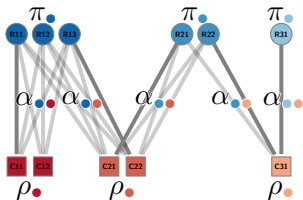


Figure 3 – Exemple de LBM¹

Avec

- $Q_1 = |\{\bullet, \bullet, \bullet\}|$ blocs fixés en ligne
- $Q_2 = |\{\bullet, \bullet, \bullet\}|$ blocs fixés en colonne

Paramètres

- $\pi_{\bullet} = \mathbb{P}(Z_i = \bullet)$
- $\rho_{\bullet} = \mathbb{P}(W_j = \bullet)$
- $\alpha_{\bullet\bullet} = \mathbb{P}(Y_{ij} = 1 | Z_i = \bullet, W_j = \bullet)$

1. Que j'appellerai par la suite BiSBM

Plusieurs réseaux

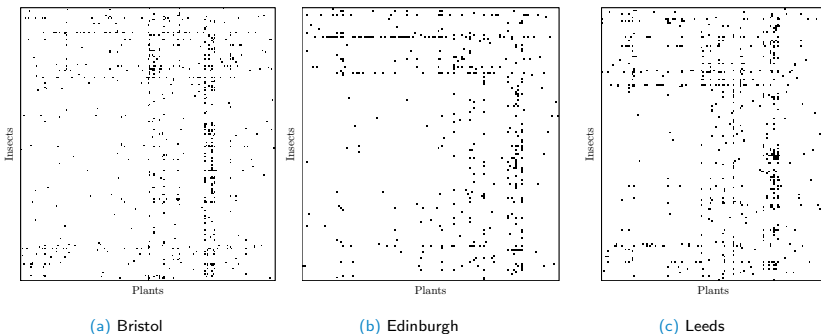


Figure 4 – Matrices d'adjacence, Baldock et al., 2019

Collections bipartites

$$\forall m \in \{1 \dots M\}, Y^m \stackrel{ind}{\sim} \mathcal{F}\text{-BiSBM}_{n_1^m, n_2^m}(Q_1^m, Q_2^m, \pi^m, \rho^m, \alpha^m)$$

Collections bipartites

$$\forall m \in \{1 \dots M\}, Y^m \stackrel{ind}{\sim} \mathcal{F}\text{-BiSBM}_{n_1^m, n_2^m}(Q_1^m, Q_2^m, \pi^m, \rho^m, \alpha^m)$$

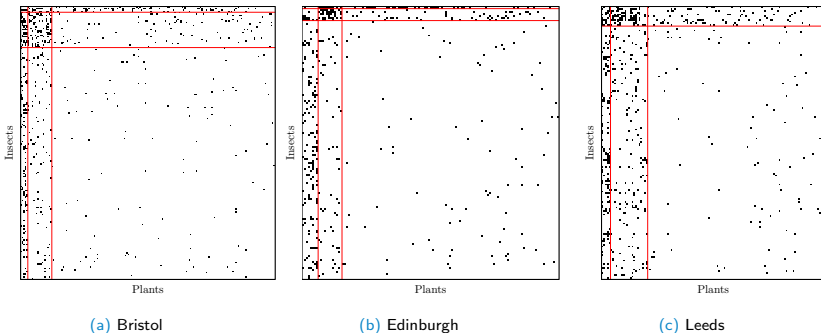


Figure 5 – Matrices d'adjacence réordonnées, grâce au LBM

Différents modèles

iid-colBiSBM

$$\forall m \in \{1 \dots M\}, Y^m \stackrel{iid}{\sim} \mathcal{F}\text{-BiSBM}_{n_1^m, n_2^m}(\mathbf{Q}_1, \mathbf{Q}_2, \pi, \rho, \alpha)$$

avec $\theta = (\pi, \rho, \alpha)$.

Différents modèles

iid-colBiSBM

$$\forall m \in \{1 \dots M\}, Y^m \stackrel{iid}{\sim} \mathcal{F}\text{-BiSBM}_{n_1^m, n_2^m}(Q_1, Q_2, \pi, \rho, \alpha)$$

avec $\theta = (\pi, \rho, \alpha)$.

$\pi\rho$ -colBiSBM

$$\forall m \in \{1 \dots M\}, Y^m \stackrel{ind}{\sim} \mathcal{F}\text{-BiSBM}_{n_1^m, n_2^m}(Q_1, Q_2, \pi^m, \rho^m, \alpha)$$

avec $\theta = ((\pi^m)_{m=1, \dots, M}, (\rho^m)_{m=1, \dots, M}, \alpha)$.

Estimation des paramètres

Maximisation de la log-vraisemblance ?

log-vraisemblance et log-vraisemblance complète

$$\ell(\mathbf{Y}; \theta) = \sum_{\mathbf{Z}, \mathbf{W} \in \mathcal{Z} \times \mathcal{W}} \ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta)$$

avec $\mathcal{Z} = \{1, \dots, Q_1\}^n$, $\mathcal{W} = \{1, \dots, Q_2\}^n$

Estimation des paramètres

Maximisation de la log-vraisemblance ?

log-vraisemblance et log-vraisemblance complète

$$\ell(\mathbf{Y}; \theta) = \sum_{\mathbf{Z}, \mathbf{W} \in \mathcal{Z} \times \mathcal{W}} \ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta)$$

avec $\mathcal{Z} = \{1, \dots, Q_1\}^n$, $\mathcal{W} = \{1, \dots, Q_2\}^n$

Estimation des paramètres

Maximisation de la log-vraisemblance ?

log-vraisemblance et log-vraisemblance complète

$$\ell(\mathbf{Y}; \theta) = \sum_{\mathbf{Z}, \mathbf{W} \in \mathcal{Z} \times \mathcal{W}} \ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta)$$

avec $\mathcal{Z} = \{1, \dots, Q_1\}^n$, $\mathcal{W} = \{1, \dots, Q_2\}^n$

Donc, algorithme classique \Rightarrow *Expectation-Maximization* (EM).

Par EM classique

A l'itération (t) :

- **Étape E** : calculer

$$Q(\theta|\theta^{(t-1)}) = \mathbb{E}_{\mathbf{Z}, \mathbf{W} | \mathbf{Y}, \theta^{(t-1)}} [\ell_c(\mathbf{Y}, \mathbf{W}, \mathbf{Z}; \theta)]$$

- **Étape M** :

$$\theta^{(t)} = \arg \max_{\theta} Q(\theta|\theta^{(t-1)})$$

Par EM classique

A l'itération (t) :

- **Étape E** : calculer

$$Q(\theta|\theta^{(t-1)}) = \mathbb{E}_{\mathbf{Z}, \mathbf{W}|\mathbf{Y}, \theta^{(t-1)}} [\ell_c(\mathbf{Y}, \mathbf{W}, \mathbf{Z}; \theta)]$$

- **Étape M** :

$$\theta^{(t)} = \arg \max_{\theta} Q(\theta|\theta^{(t-1)})$$

Problème pour l'EM classique

Loi de $\mathbf{Z}, \mathbf{W}|\mathbf{Y}, \theta^{(t-1)}$ inaccessible

Par *Variational EM*, comme proposé par Daudin et al., 2008 ;
Chabert-Liddell et al., 2024.

Approximation variationnelle de $\mathbf{Z}, \mathbf{W} | \mathbf{Y}, \theta^{(t-1)}$

$\mathcal{R}_{Y^m, \tau}(\mathbf{Z}^m, \mathbf{W}^m) = \mathcal{R}_{Y^m, \tau}^1(\mathbf{Z}^m) \times \mathcal{R}_{Y^m, \tau}^2(\mathbf{W}^m) \Rightarrow$ indépendance lignes,
colonnes.

$$\ell(\mathbf{Y}; \theta) \geq \sum_{m=1}^M \left(Q^m(\theta | \theta^{(t)}) + \mathcal{H}(\mathcal{R}_{Y^m, \theta^{(t)}}(\mathbf{Z}^m, \mathbf{W}^m)) \right) =: \mathcal{J}(\tau; \theta)$$

où $Q^m(\theta | \theta^{(t)}) = \mathbb{E}_{\mathbf{Z}^m, \mathbf{W}^m \sim \mathcal{R}_{Y^m, \tau}(\cdot)} [\ell_c(Y^m, \mathbf{Z}^m, \mathbf{W}^m | \theta)]$

Formule développée de l'EM variationnel

$$\begin{aligned}
 \ell(\mathbf{Y}; \theta) \geq & \sum_{m=1}^M \left(\sum_{i=1}^{n_1^m} \sum_{j=1}^{n_2^m} \sum_{q \in \mathcal{Q}_{1,m}} \sum_{r \in \mathcal{Q}_{2,m}} \tau_{i,q}^{1,m} \tau_{j,r}^{2,m} \log f(Y_{ij}^m; \alpha_{qr}) \right. \\
 & + \sum_{i=1}^{n_1^m} \sum_{q \in \mathcal{Q}_{1,m}} \tau_{i,q}^{1,m} \log \pi_q^m + \sum_{j=1}^{n_2^m} \sum_{r \in \mathcal{Q}_{2,m}} \tau_{j,r}^{2,m} \log \rho_r^m \\
 & \left. - \sum_{i=1}^{n_1} \tau_{i,q}^{1,m} \log \tau_{i,q}^{1,m} - \sum_{j=1}^{n_2} \tau_{j,r}^{2,m} \log \tau_{j,r}^{2,m} \right) =: \mathcal{J}(\tau; \theta),
 \end{aligned}$$

Approximation variationnelle

$$\tau_{iq}^{1,m} = \mathcal{R}_{Y^m, \tau}^1(Z_{iq}^m = 1) \text{ et } \tau_{jr}^{2,m} = \mathcal{R}_{Y^m, \tau}^2(W_{jr}^m = 1)$$

Étape *Variational Expectation*

$$\hat{\tau}^{(t+1)} = \arg \max_{\tau} \mathcal{J}(\tau, \hat{\theta}^{(t)}) \Leftrightarrow \arg \min_{\tau \in \mathcal{T}} \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \tau}, \mathbb{P}(\cdot | \mathbf{Y})]$$

$$\begin{cases} \hat{\tau}_{iq}^{1,m} \propto \hat{\pi}_q^{m(t)} \prod_{j=1}^{n_2^m} \prod_{r \in \mathcal{Q}_2^m} f(Y_{ij}^m; \hat{\alpha}_{qr}^{(t)}) \hat{\tau}_{jr}^{2,m(t+1)} & \forall i = 1, \dots, n_1^m, q \in \mathcal{Q}_1^m \\ \hat{\tau}_{jr}^{2,m} \propto \hat{\rho}_r^{m(t)} \prod_{i=1}^{n_1^m} \prod_{q \in \mathcal{Q}_1^m} f(Y_{ij}^m; \hat{\alpha}_{qr}^{(t)}) \hat{\tau}_{iq}^{1,m(t+1)} & \forall j = 1, \dots, n_2^m, r \in \mathcal{Q}_2^m \end{cases}$$

2. Initialisation des $\hat{\tau}$ avec un *spectral clustering* sur les réseaux.

Étape *Maximization*

$$\hat{\theta}^{(t+1)} = \arg \max_{\theta} \mathcal{J}(\hat{\tau}^{(t+1)}, \theta)$$

Paramètres de connectivité

$$\hat{\alpha}_{qr} = \frac{\sum_{m=1}^M \sum_{i=1}^{n_1^m} \sum_{j=1}^{n_2^m} \tau_{iq}^{1,m} \tau_{jr}^{2,m} Y_{ij}^m}{\sum_{m=1}^M \sum_{i=1}^{n_1^m} \sum_{j=1}^{n_2^m} \tau_{iq}^{1,m} \tau_{jr}^{2,m}}$$

Proportions pour *iid*

$$\hat{\pi}_q = \frac{\sum_{m=1}^M \sum_{i=1}^{n_1^m} \tau_{iq}^{1,m}}{\sum_{m=1}^M n_1^m}$$

$$\hat{\rho}_r = \frac{\sum_{m=1}^M \sum_{j=1}^{n_2^m} \tau_{jr}^{2,m}}{\sum_{m=1}^M n_2^m}$$

Étape *Maximization*

$$\hat{\theta}^{(t+1)} = \arg \max_{\theta} \mathcal{J}(\hat{\tau}^{(t+1)}, \theta)$$

Paramètres de connectivité

$$\hat{\alpha}_{qr} = \frac{\sum_{m=1}^M \sum_{i=1}^{n_1^m} \sum_{j=1}^{n_2^m} \tau_{iq}^{1,m} \tau_{jr}^{2,m} \mathbf{Y}_{ij}^m}{\sum_{m=1}^M \sum_{i=1}^{n_1^m} \sum_{j=1}^{n_2^m} \tau_{iq}^{1,m} \tau_{jr}^{2,m}}$$

Proportions pour $\pi\rho$

$$\hat{\pi}_q^m = \frac{\sum_{i=1}^{n_1^m} \tau_{iq}^{1,m}}{n_1^m} \qquad \hat{\rho}_r^m = \frac{\sum_{j=1}^{n_2^m} \tau_{jr}^{2,m}}{n_2^m}$$

Problème choix de (Q_1, Q_2)

Besoin sélectionner Q_1 et Q_2 . Critère BIC-Like²

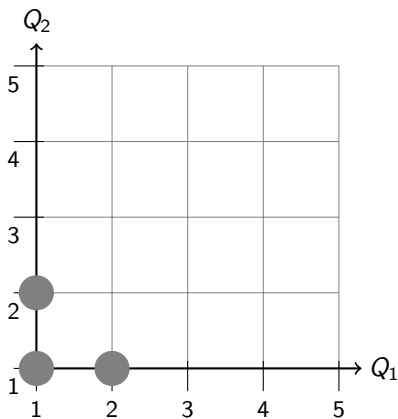
$$\begin{aligned} \text{BIC-L}(\mathbf{Y}, Q_1, Q_2) &= \max_{\theta} \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \hat{\tau}}} [\ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta)] + \mathcal{H}(\mathcal{R}_{\mathbf{Y}, \hat{\tau}}) - \frac{1}{2} \text{pen}(\theta, Q_1, Q_2) \\ &= \max_{\theta} \mathcal{J}(\mathcal{R}_{\mathbf{Y}, \hat{\tau}}, \theta) - \frac{1}{2} \text{pen}(\theta, Q_1, Q_2) \end{aligned}$$


Problèmes de l'exploration

- Exploration de N^2 coûteux.
- Sensibilité initialisations.

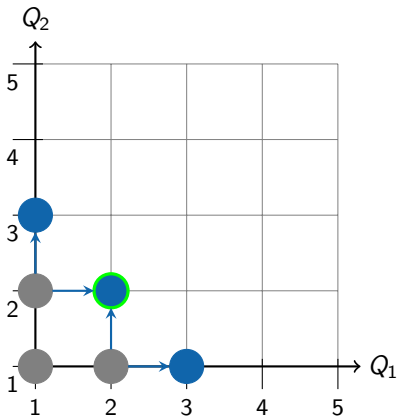
2. ICL + Entropie + pénalité

Choix de (Q_1, Q_2) - Approche gloutonne



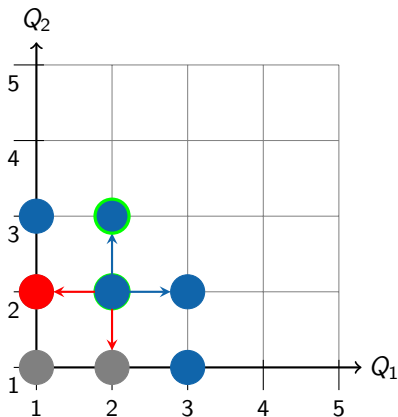
- Modèle initialisé : 

Choix de (Q_1, Q_2) - Approche gloutonne



- Modèle initialisé :
●
- Modèle après *split* :
●
- Modèle maximisant le critère :
○

Choix de (Q_1, Q_2) - Approche gloutonne



- Modèle initialisé :
●
- Modèle après *split* :
●
- Modèle maximisant le critère :
○
- Modèle après *merge* :
●

Choix de (Q_1, Q_2) - Fenêtre glissante

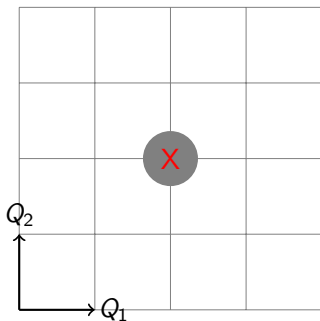


Figure 6 – Fenêtre glissante

Choix de (Q_1, Q_2) - Fenêtre glissante

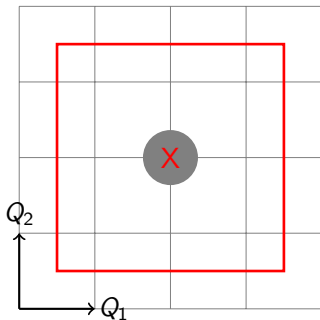


Figure 6 – Fenêtre glissante

Choix de (Q_1, Q_2) - Fenêtre glissante

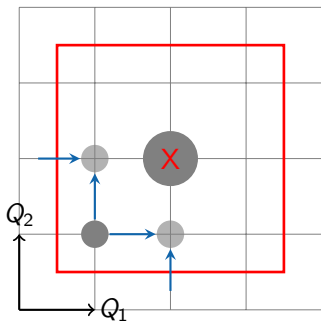


Figure 6 – Fenêtre glissante

Initialisation du modèle si nécessaire

Choix de (Q_1, Q_2) - Fenêtre glissante

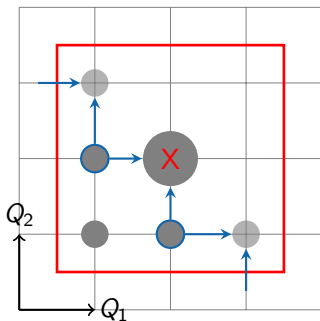


Figure 6 – Fenêtre glissante

Choix de (Q_1, Q_2) - Fenêtre glissante

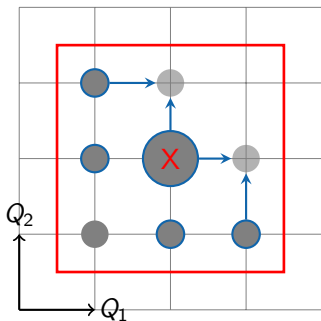


Figure 6 – Fenêtre glissante

Choix de (Q_1, Q_2) - Fenêtre glissante

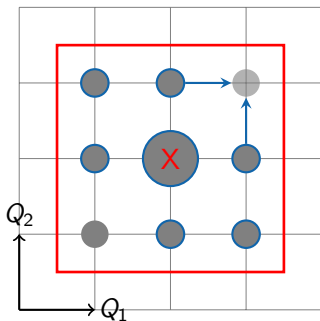


Figure 6 – Fenêtre glissante

Choix de (Q_1, Q_2) - Fenêtre glissante

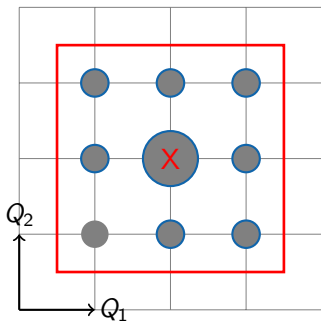


Figure 6 – Fenêtre glissante

Choix de (Q_1, Q_2) - Fenêtre glissante

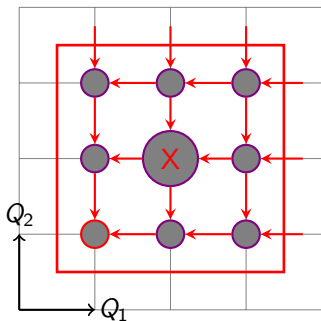


Figure 6 – Fenêtre glissante

Choix de (Q_1, Q_2) - Fenêtre glissante

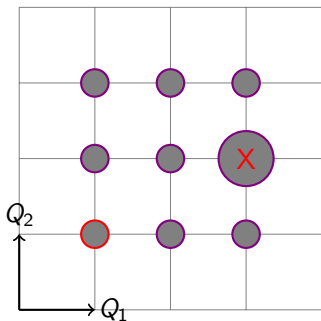


Figure 6 – Fenêtre glissante

Localisation du nouveau mode

Choix de (Q_1, Q_2) - Fenêtre glissante

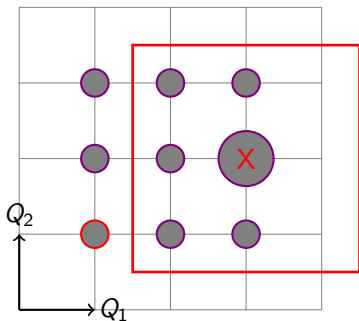
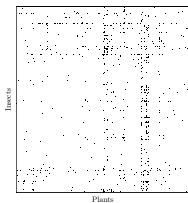


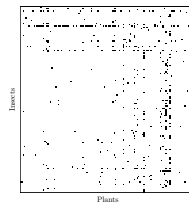
Figure 6 – Fenêtre glissante

Déplacement sur le nouveau mode puis itération

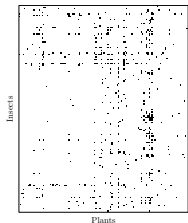
Résultats Baldock et al., 2019



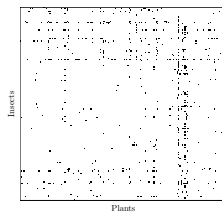
(a) Bristol



(b) Edinburgh



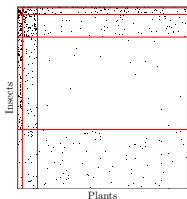
(c) Leeds



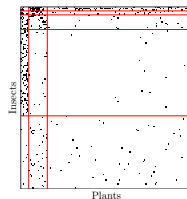
(d) Reading

Figure 7 – Matrices d'adjacence, Baldock et al., 2019

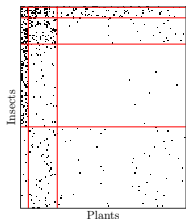
Résultats Baldock et al., 2019



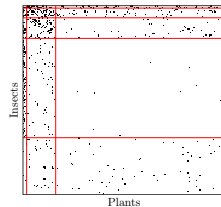
(a) Bristol



(b) Edinburgh



(c) Leeds



(d) Reading

Figure 7 – Matrices d'adjacence réordonnée par *iid*-colBiSBM, Baldock et al., 2019

Clustering de réseaux

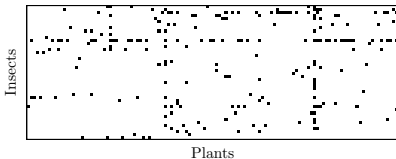
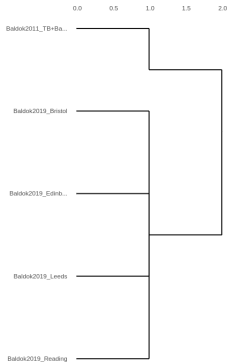
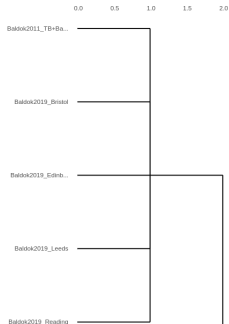


Figure 8 – Matrice d'adjacence, Baldock et al., 2011

Application à Baldock et al., 2019, 2011 I



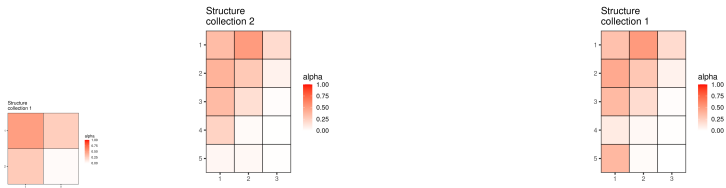
(a) Modèle *iid*



(b) Modèle $\pi\rho$

Figure 9 – Partitionnement des réseaux de Baldock et al., 2019, 2011

Application à Baldock et al., 2019, 2011 II

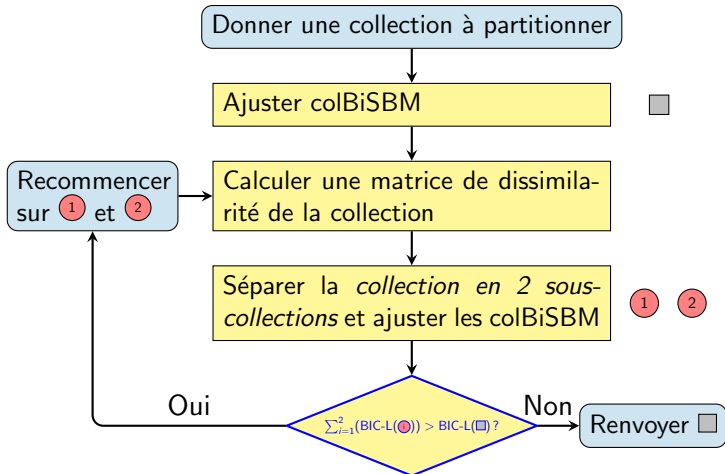


(a) Modèle *iid*,
séparet réseau africain et réseaux anglais

(b) Modèle $\pi\rho$,
fusionnent réseaux africain et anglais

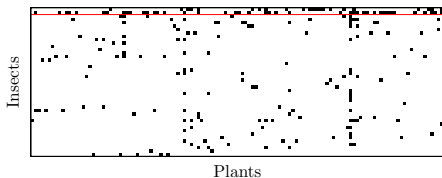
Figure 10 – Structures détectées pour les réseaux de Baldock et al., 2019, 2011

Algorithme du clustering

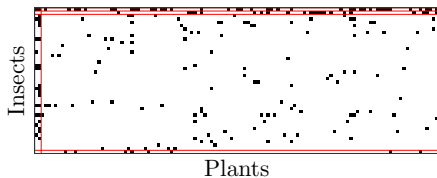


$$D_{\mathcal{M}}(m, m') = \sum_{q=1}^{Q_1} \sum_{r=1}^{Q_2} \max(\tilde{\pi}_q^m, \tilde{\pi}_q^{m'}) \left(\tilde{\alpha}_{qr}^m - \tilde{\alpha}_{qr}^{m'} \right)^2 \max(\tilde{\rho}_r^m, \tilde{\rho}_r^{m'})$$

Résultats



(a) Réordonnée par LBM



(b) Réordonnée par $\pi\rho$ -colBiSBM

Figure 11 – Matrice d'adjacence réordonnée par $\pi\rho$ -colBiSBM, Baldock et al., 2011

Conclusion et perspectives

Capacités

- 4 modèles dont 3 qui ont une flexibilité sur au moins une des dimensions (adaptabilité aux données).
- Détecter structures classiques et moins classique de façon agnostique.
- Partitionner un ensemble de réseaux selon leurs structures.

Perspectives

- Investiguer stabilité face à l'aléatoire et aux *optima* locaux.
- Preuve d'identifiabilité du modèle $\pi\rho$.

Package et applications

- Intégration au package `colSBM`, amélioration interface utilisateur et ajout retours écologiques
- Publication CRAN
- Intégrer possibilité d'un critère supplémentaire pour le clustering
- Appliquer clustering données de [Pichon et al., 2024](#) ; [Doré et al., 2021](#)

Merci pour votre attention !

Bibliographie I

- Baldock, K. C. R., Goddard, M. A., Hicks, D. M., Kunin, W. E., Mitschunas, N., Morse, H., Osgathorpe, L. M., Potts, S. G., Robertson, K. M., Scott, A. V., Staniczenko, P. P. A., Stone, G. N., Vaughan, I. P., & Memmott, J. (2019). A systems approach reveals urban pollinator hotspots and conservation opportunities. *Nat Ecol Evol*, 3(3), 363-373.
<https://doi.org/10.1038/s41559-018-0769-y>
- Govaert, G., & Nadif, M. (2005). An EM Algorithm for the Block Mixture Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4), 643-647.
<https://doi.org/10.1109/TPAMI.2005.69>
- Daudin, J.-J., Picard, F., & Robin, S. (2008). A mixture model for random graphs. *Stat Comput*, 18(2), 173-183.
<https://doi.org/10.1007/s11222-007-9046-7>

Bibliographie II

- Chabert-Liddell, S.-C., Barbillon, P., & Donnet, S. (2024). Learning Common Structures in a Collection of Networks. An Application to Food Webs. *The Annals of Applied Statistics*, 18(2), 1213-1235. <https://doi.org/10.1214/23-AOAS1831>
- Baldock, K. C. R., Memmott, J., Ruiz-Guajardo, J. C., Roze, D., & Stone, G. N. (2011). Daily temporal structure in African savanna flower visitation networks and consequences for network sampling. *Ecology*, 92(3), 687-698. <https://doi.org/10.1890/10-1110.1>
- Pichon, B., Le Goff, R., Morlon, H., & Perez-Lamarque, B. (2024). Telling mutualistic and antagonistic ecological networks apart by learning their multiscale structure. *Methods in Ecology and Evolution*, 15(6), 1113-1128. <https://doi.org/10.1111/2041-210X.14328>
- Doré, M., Fontaine, C., & Thébault, E. (2021). Relative effects of anthropogenic pressures, climate, and sampling design on the structure of pollination networks at the global scale. *Global Change Biology*, 27(6), 1266-1280. <https://doi.org/10.1111/gcb.15474>

Pourquoi VE minimise KL ?

$$\ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta) = \log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta) + \ell(\mathbf{Y}; \theta)$$

$$\Leftrightarrow \ell(\mathbf{Y}; \theta) = \ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta) - \log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)$$

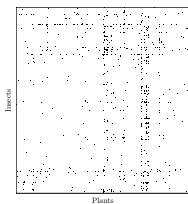
$$\Leftrightarrow \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}[\ell(\mathbf{Y}; \theta)] = \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}[\ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta)] - \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}[\log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)]$$

$$\Leftrightarrow \ell(\mathbf{Y}; \theta) = \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}[\ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta)] - \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}[\log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)]$$

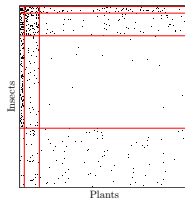
$$\begin{aligned} \text{Or } \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \tau}, \log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)] &= -\mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}\left[\log \frac{\mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)}{\mathcal{R}_{\mathbf{Y}, \tau}}\right] \\ &= -\mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}[\log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)] + \underbrace{\mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}[\log \mathcal{R}_{\mathbf{Y}, \tau}]}_{-\mathcal{H}(\mathcal{R}_{\mathbf{Y}, \tau})} \end{aligned}$$

$$\Leftrightarrow \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \tau}, \log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)] + \mathcal{H}(\mathcal{R}_{\mathbf{Y}, \tau}) = -\mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}[\log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)]$$

$$\text{D'où } \ell(\mathbf{Y}; \theta) - \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \tau}, \log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)] = \mathcal{J}(\tau; \theta) \quad \square$$

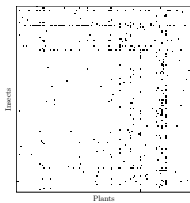


(a) Donnée

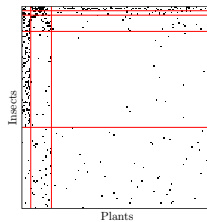


(b) Réordonnée

Figure 12 – Bristol

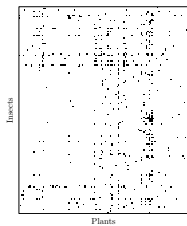


(a) Donnée

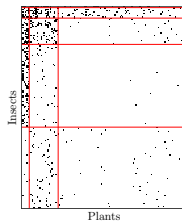


(b) Réordonnée

Figure 13 – Edinburgh

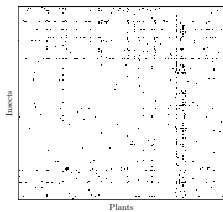


(a) Donnée

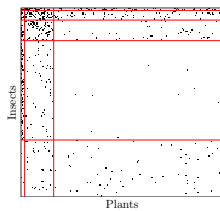


(b) Réordonnée

Figure 14 – Leeds



(a) Donnée



(b) Réordonnée

Figure 15 – Reading

Bibliographie des annexes I