

Séminaire des stagiaires

Adaptation de colSBM aux réseaux bipartites

Louis Lacoste

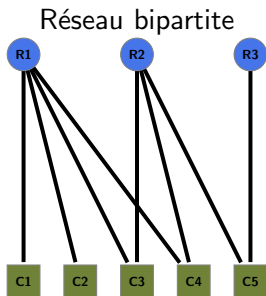
29 juin 2023

Contexte écologique

- De nombreux réseaux disponibles (« Web of Life : Ecological Networks Database », s. d.) et décrivant des interactions similaires
- Re-grouper les réseaux selon leur similarité (*clustering* de réseaux)
- Compléter d'éventuelles informations manquantes grâce à la collection
- Déterminer des structures d'interactions fines de manière agnostique
- Vérifier si le regroupement est lié à des co-variables

0. Pour combler les lacunes de
Chabert-Liddell et al., 2023

Réseaux bipartites¹



Matrice d'incidence

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

Permet de décrire des interactions impliquant deux agents dont les rôles sont de natures différentes.

Par exemple : hôtes-parasites, plantes-pollinisateurs, graines-disperseurs ...

1. Ou *bipartis*. Voir Larousse, s. d.

Latent Block Model (LBM²)

Proposé par Govaert et Nadif, 2005.

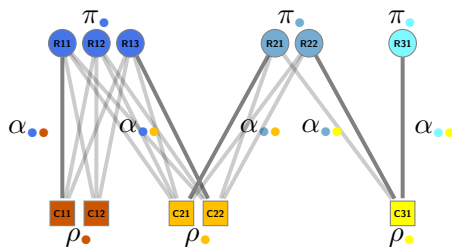


Figure – Exemple de LBM²

Pour

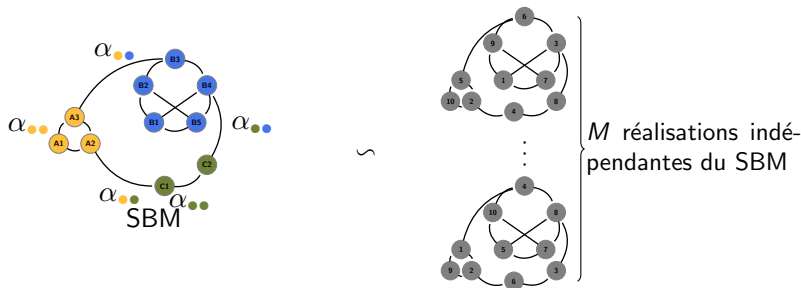
- $Q_1 = |\{\bullet, \bullet, \bullet\}|$ blocs fixés en ligne
- $Q_2 = |\{\bullet, \bullet, \bullet\}|$ blocs fixés en colonne

Paramètres

- $\pi_{\bullet} = \mathbb{P}(i \in \bullet)$ en ligne et
- $\rho_{\bullet} = \mathbb{P}(j \in \bullet)$ en colonne
- $\alpha_{\bullet, \bullet} = \mathbb{P}(i \leftrightarrow j | i \in \bullet, j \in \bullet)$

colSBM

Le modèle *colSBM* (Chabert-Liddell et al., 2023).

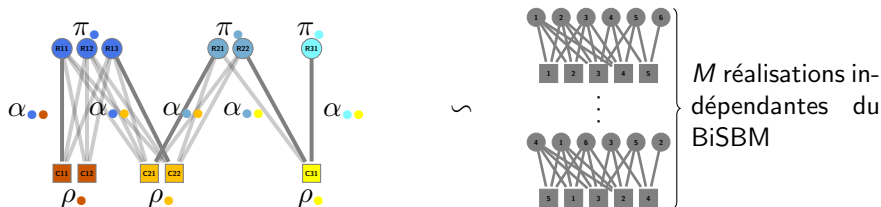


Pour $Q = |\{\bullet, \bullet, \bullet\}|$ blocs fixés :

Paramètres

- $\pi_{\bullet} = \mathbb{P}(i \in \bullet)$
- $\alpha_{\bullet, \bullet} = \mathbb{P}(i \leftrightarrow j | i \in \bullet, j \in \bullet)$

Collections bipartites



Pour

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

Paramètres

- $\pi_{\bullet} = \mathbb{P}(i \in \bullet)$ en ligne et $\rho_{\bullet} = \mathbb{P}(j \in \bullet)$ en colonne
- $\alpha_{\bullet, \bullet} = \mathbb{P}(i \leftrightarrow j | i \in \bullet, j \in \bullet)$

Différents modèles I

iid-colBiSBM

$\pi = (\pi_1, \dots, \pi_{Q_1-1})$ et $\rho = (\rho_1, \dots, \rho_{Q_2-1})$, tous les réseaux partagent les mêmes paramètres³

π -*colBiSBM*

$\pi = ((\pi_1^m, \dots, \pi_{Q_1-1}^m))_{m=1, \dots, M}$ et $\rho = (\rho_1, \dots, \rho_{Q_2-1})$ avec
 $\forall q, m \in \llbracket 1, Q_1 - 1 \rrbracket \times \llbracket 1, M \rrbracket, \pi_q^m \in [0, 1]$

3. Dans tous les modèles la structure de connectivité est supposée identique au sein de la collection.

Différents modèles II

ρ -colBiSBM

$\boldsymbol{\pi} = (\pi_1, \dots, \pi_{Q_1-1})$ et $\boldsymbol{\rho} = ((\rho_1^m, \dots, \rho_{Q_2-1}^m))_{m=1, \dots, M}$ avec
 $\forall r, m \in \llbracket 1, Q_2 - 1 \rrbracket \times \llbracket 1, M \rrbracket, \rho_r^m \in [0, 1]$

$\pi\rho$ -colBiSBM

$\boldsymbol{\pi} = ((\pi_1^m, \dots, \pi_{Q_1-1}^m))_{m=1, \dots, M}$ et $\boldsymbol{\rho} = ((\rho_1^m, \dots, \rho_{Q_2-1}^m))_{m=1, \dots, M}$ avec
 $\forall q, m \in \llbracket 1, Q_1 - 1 \rrbracket \times \llbracket 1, M \rrbracket, \pi_q^m \in [0, 1]$ et
 $\forall r, m \in \llbracket 1, Q_2 - 1 \rrbracket \times \llbracket 1, M \rrbracket, \rho_r^m \in [0, 1]$

Borne inférieure de la vraisemblance

Maximisation de la borne inférieure de la log-vraisemblance des données observées.

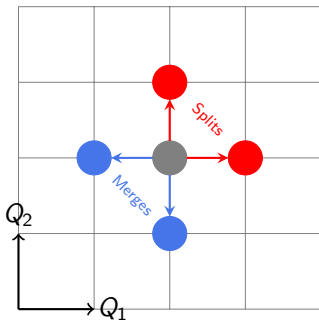
$$\begin{aligned}
 \ell(\mathbf{X}; \boldsymbol{\theta}) \geq & \sum_{m=1}^M \left(\sum_{i=1}^{n_1^m} \sum_{j=1}^{n_2^m} \sum_{q \in \mathcal{Q}_{1,m}} \sum_{r \in \mathcal{Q}_{2,m}} \tau_{i,q}^{1,m} \tau_{j,r}^{2,m} \log f(X_{ij}^m; \alpha_{qr}) \right. \\
 & + \sum_{i=1}^{n_1^m} \sum_{q \in \mathcal{Q}_{1,m}}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

Le premier terme correspond à la log-vraisemblance complétée et marginalisée sur la famille des distributions factorisables.

Parcours de la grille (Q_1, Q_2) - Approche gloutonne

Le VEM se fait à Q_1, Q_2 fixés, il faut donc déterminer les “meilleurs” coordonnées. Pour cela nous utilisons un BIC-L⁴ en adaptant les formules de Chabert-Liddell et al., 2023.

Détermination d'un premier mode par approche *gloutonne*



Exploration gloutonne

Pendant cette phase, après l'initialisation, pour chaque position Q_1, Q_2 nous calculons tous les modèles possible depuis le point courant. Le meilleur est alors celui avec le plus haut BIC-L et nous recommençons depuis ce point.

4. Bayesian Information Criterion - Like

Parcours de la grille (Q_1 , Q_2) - Fenêtre glissante

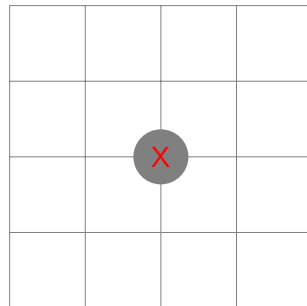
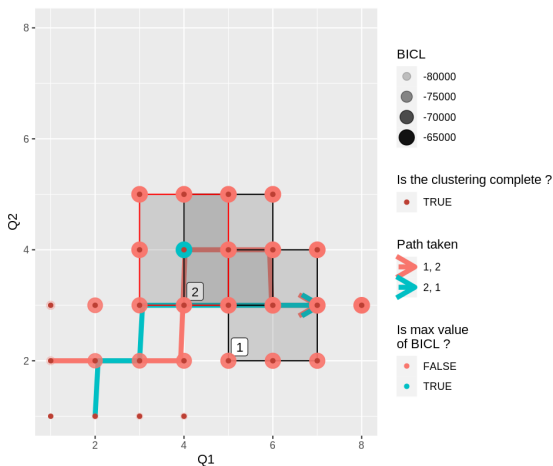


Figure – Exemple de parcours de fenêtre glissante

Parcours de la grille (Q_1, Q_2) - Fenêtre glissante

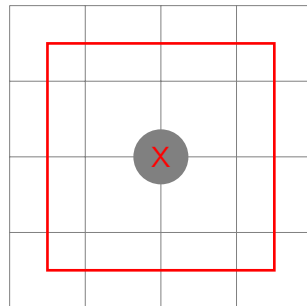
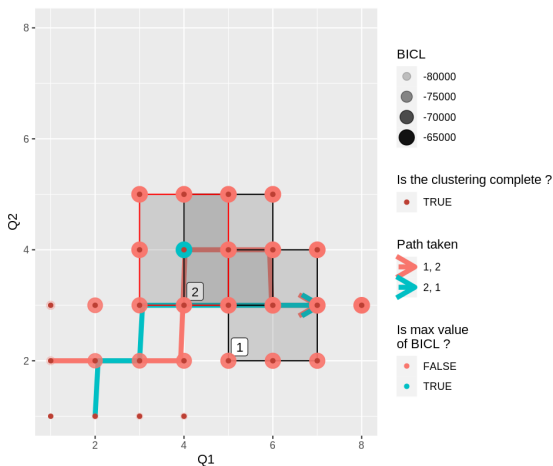


Figure – Exemple de parcours de fenêtre glissante

Parcours de la grille (Q_1, Q_2) - Fenêtre glissante

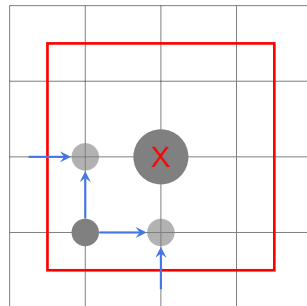
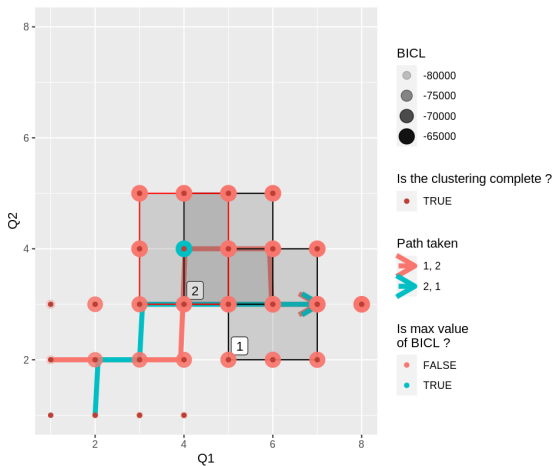


Figure – Exemple de parcours de fenêtre glissante

Parcours de la grille (Q_1, Q_2) - Fenêtre glissante

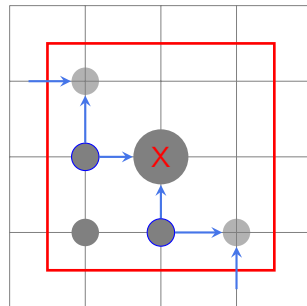
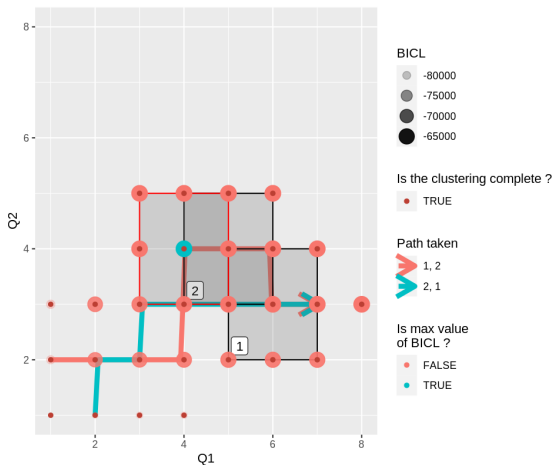


Figure – Exemple de parcours de fenêtre glissante

Parcours de la grille (Q_1, Q_2) - Fenêtre glissante

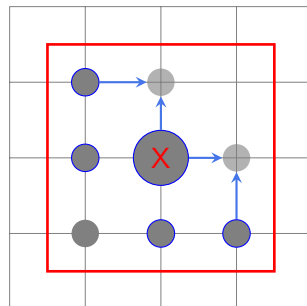
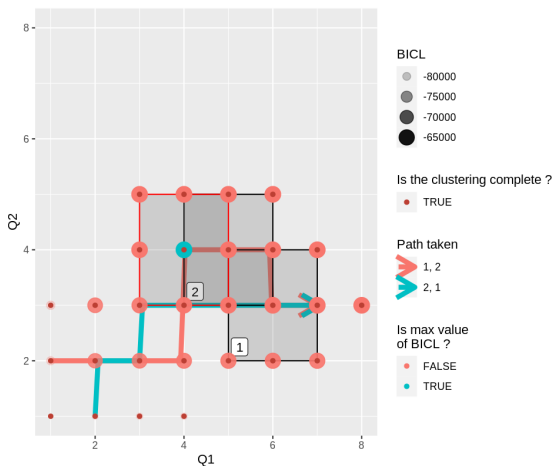


Figure – Exemple de parcours de fenêtre glissante

Parcours de la grille (Q_1, Q_2) - Fenêtre glissante

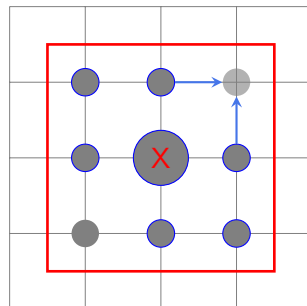
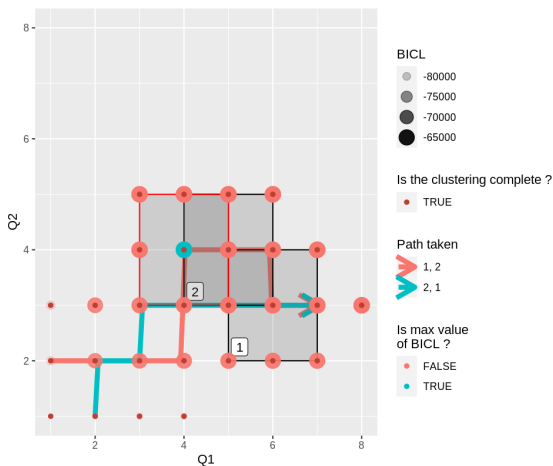


Figure – Exemple de parcours de fenêtre glissante

Parcours de la grille (Q_1, Q_2) - Fenêtre glissante

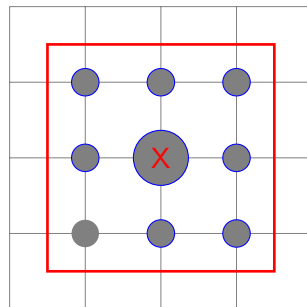
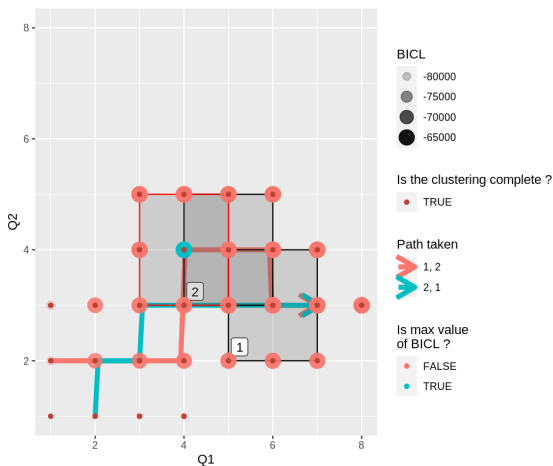
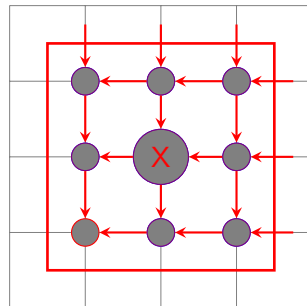


Figure – Exemple de parcours de fenêtre glissante

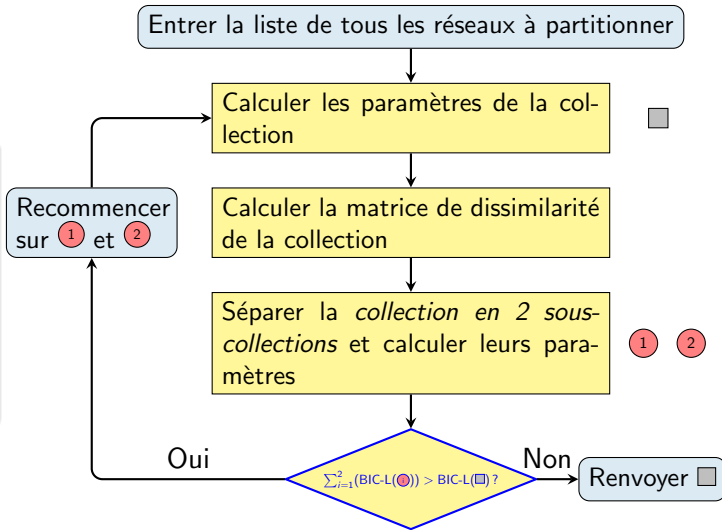


10 / 12

Clustering de réseaux

Objectif

Déterminer une partition qui maximise la somme du BICL de ses sous-collections.



. Même approche que Chabert-Liddell et al., 2023

Application, données plantes pollinisateurs

Voici des résultats du modèles *iid-colBiSBM* sur des données plantes-pollinisateurs (Doré et al., 2021 et Thébault et Fontaine, 2020)

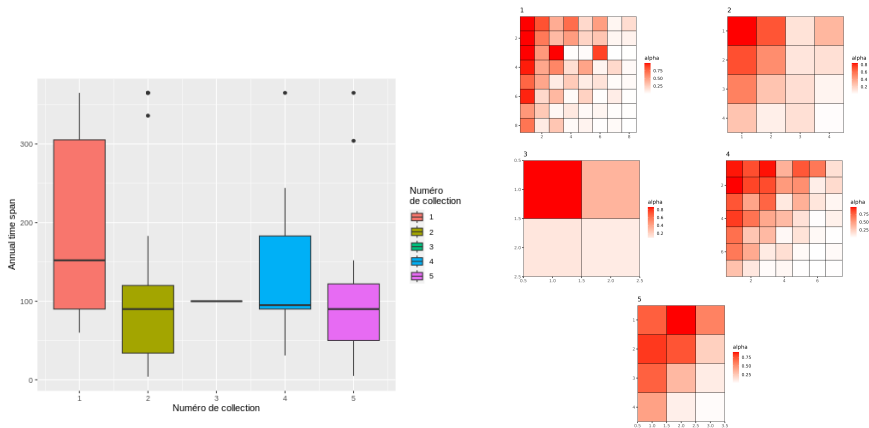








Figure – Connectivités de la partition

Bibliographie I

-  Web of Life : Ecological Networks Database. (s. d.).
-  Chabert-Liddell, S.-C., Barbillon, P., & Donnet, S. (2023). Learning Common Structures in a Collection of Networks. An Application to Food Webs. <https://doi.org/10.48550/arXiv.2206.00560>
-  Larousse, É. (s. d.). Définitions : biparti, bipartite - Dictionnaire de français Larousse.
-  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>
-  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>
-  Thébault, E., & Fontaine, C. (2020). A Database of Plant-Pollinator Networks. <https://doi.org/10.5281/zenodo.4300427>