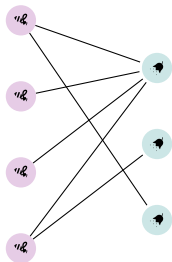


Joint analysis of bipartite networks collection

Présentation LSD

Louis Lacoste, under the supervision of Pierre Barbillon and Sophie
Donnet
Laboratoire MIA Paris-Saclay

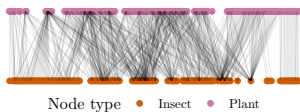
Why a network?



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

Associated
adjacency
matrix

Figure 1: Example of a network



Node type ● Insect ● Plant

- Modeling of various interactions, here ecosystems
- Structure necessary for: biodiversity monitoring, robustness, risk of collapse
- Increasingly available

Figure 2: Plant-pollinator network of Bristol
Baldock et al., 2019

Analysis methods for a network

Several methods :

- Metrics : degree, centrality, nesting . . .
- Network embedding with GNN
- *Clustering* of nodes with latent variable models

Analysis methods for a network

Several methods :

- Metrics : degree, centrality, nesting . . .
- Network embedding with GNN
- **Clustering of nodes with latent variable models**

Latent Block Model (LBM¹)

Govaert and Nadif, 2005.

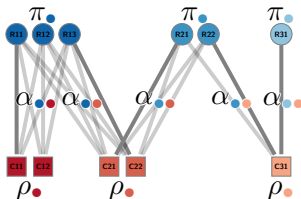


Figure 3: Example of LBM¹

Hierarchical model

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

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

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

where

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

Concise LBM formula

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

¹Which I will henceforth call BiSBM

Latent Block Model (LBM¹)

Govaert and Nadif, 2005.

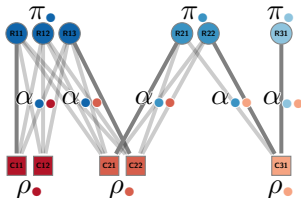


Figure 3: Example of LBM¹

With

- $Q_1 = |\{\bullet, \bullet, \bullet\}|$ fixed row blocks
- $Q_2 = |\{\bullet, \bullet, \bullet\}|$ fixed column blocks

Parameters

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

¹Which I will henceforth call BiSBM

Multiple networks

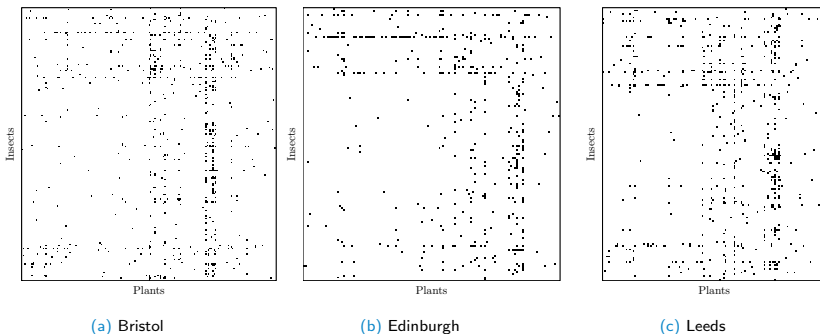


Figure 4: Adjacency matrices, Baldock et al., 2019

Bipartite collections

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

Bipartite collections

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

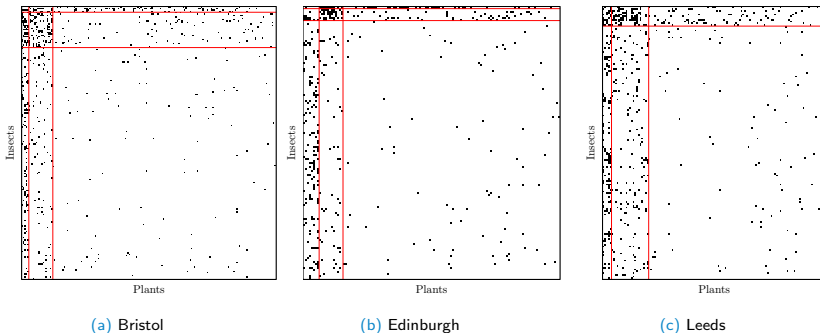


Figure 5: Reordered adjacency matrices, thanks to LBM

Different models

iid-colBiSBM

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

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

Different models

iid-colBiSBM

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

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

$\pi\rho$ -colBiSBM

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

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

Parameter estimation

Maximizing the log-likelihood?

log-likelihood and complete log-likelihood

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

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

Parameter estimation

Maximizing the log-likelihood?

log-likelihood and complete log-likelihood

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

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

Parameter estimation

Maximizing the log-likelihood?

log-likelihood and complete log-likelihood

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

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

So, classic algorithm \Rightarrow *Expectation-Maximization* (EM).

By classic EM

At iteration (t):

- **E Step:** calculate

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

- **M Step:**

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

By classic EM

At iteration (t):

- **E Step:** calculate

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

- **M Step:**

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

Problem for classic EM

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

By *Variational EM*, as proposed by Daudin et al., 2008; Chabert-Liddell et al., 2024.

Variational approximation of $\mathbf{Z}, \mathbf{W} | \mathbf{Y}, \theta^{(t-1)}$

$\mathcal{R}_{Y^m, \tau}(\mathbf{Z}^m, \mathbf{W}^m) = \mathcal{R}_{Y^m, \tau}^1(\mathbf{Z}^m) \times \mathcal{R}_{Y^m, \tau}^2(\mathbf{W}^m) \Rightarrow$ independence rows, columns.

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

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

Developed formula of variational EM

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

Variational approximation

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

Variational Expectation Step

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

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

²Initialization of $\hat{\tau}$ with a *spectral clustering* on the networks.

Maximization Step

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

Connectivity parameters

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

Proportions for *iid*

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

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

Maximization Step

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

Connectivity parameters

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

Proportions for $\pi\rho$

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

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

Problem of choosing (Q_1, Q_2)

Need to select Q_1 and Q_2 . BIC-Like criterion²

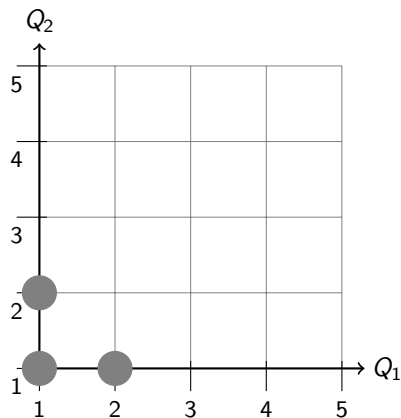
$$\begin{aligned}\text{BIC-L}(\mathbf{Y}, Q_1, Q_2) &= \max_{\theta} \mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \hat{\tau}}} [\ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \theta)] + \mathcal{H}(\mathcal{R}_{\mathbf{Y}, \hat{\tau}}) - \frac{1}{2} \text{pen}(\theta, Q_1, Q_2) \\ &= \max_{\theta} \mathcal{J}(\mathcal{R}_{\mathbf{Y},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

Exploration problems

- Exploration of \mathbb{N}^2 costly.
- Sensitivity to initializations.

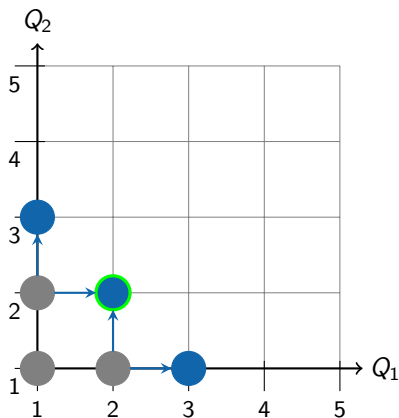
²ICL + Entropy + penalty

Choice of (Q_1, Q_2) - Greedy approach



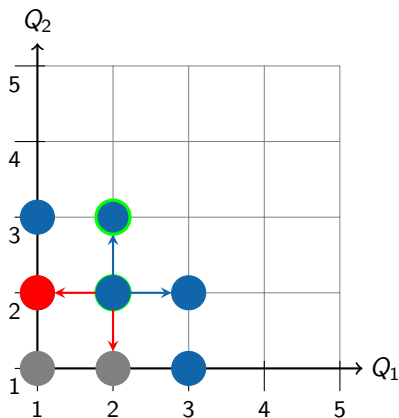
- Initial model :





Choice of (Q_1, Q_2) - Greedy approach



- Initial model :
●
- Model after *split* :
●
- Model maximizing the criterion :
○

Choice of (Q_1, Q_2) - Greedy approach



- Initial model :

- Model after *split* :

- Model maximizing the criterion :

- Model after *merge* :


Choice of (Q_1, Q_2) - Sliding window

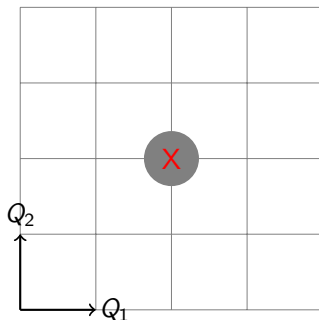


Figure 6: Sliding window

Choice of (Q_1, Q_2) - Sliding window

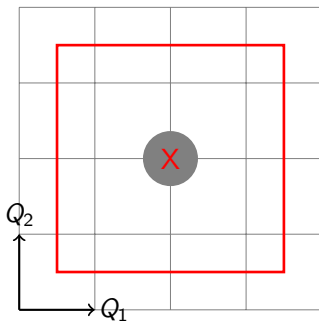


Figure 6: Sliding window

Choice of (Q_1, Q_2) - Sliding window

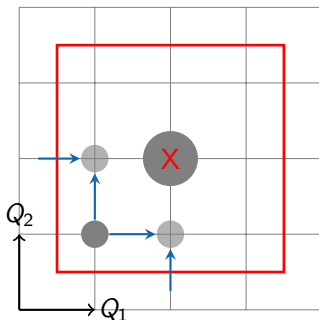


Figure 6: Sliding window

Initialization of the model if necessary

Choice of (Q_1, Q_2) - Sliding window

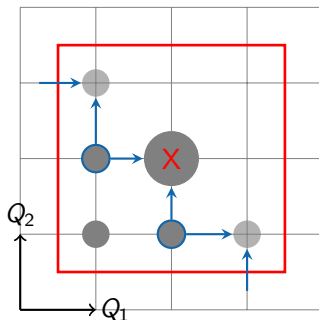


Figure 6: Sliding window

Choice of (Q_1, Q_2) - Sliding window

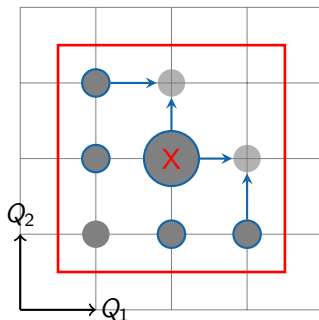


Figure 6: Sliding window

Choice of (Q_1, Q_2) - Sliding window

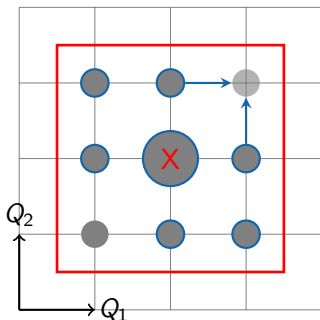


Figure 6: Sliding window

Choice of (Q_1, Q_2) - Sliding window

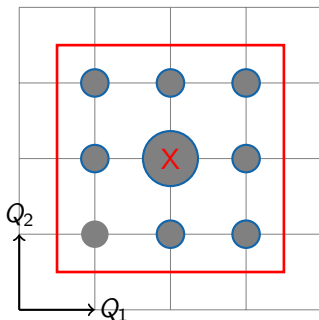


Figure 6: Sliding window

Choice of (Q_1, Q_2) - Sliding window

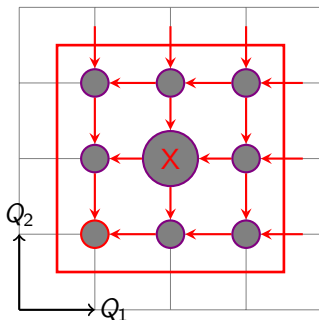


Figure 6: Sliding window

Choice of (Q_1, Q_2) - Sliding window

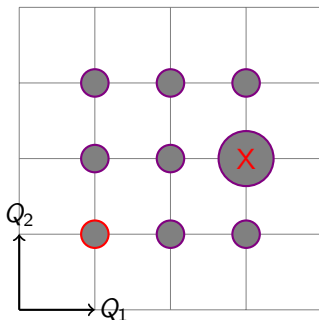


Figure 6: Sliding window

Localization of the new mode

Choice of (Q_1, Q_2) - Sliding window

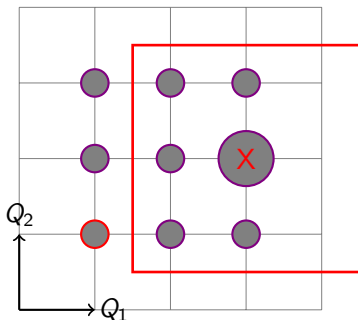
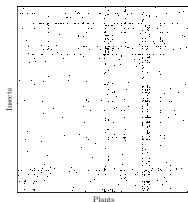


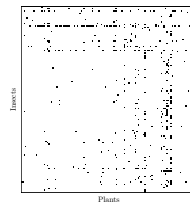
Figure 6: Sliding window

Move to the new mode then iterate

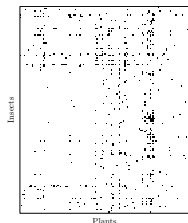
Results Baldock et al., 2019



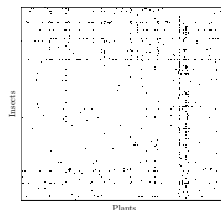
(a) Bristol



(b) Edinburgh



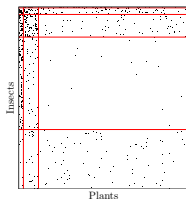
(c) Leeds



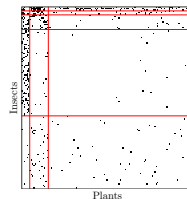
(d) Reading

Figure 7: Adjacency matrices, Baldock et al., 2019

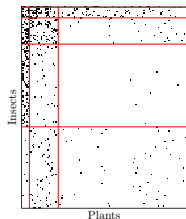
Results Baldock et al., 2019



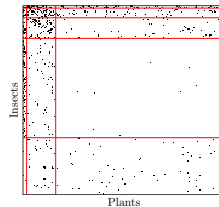
(a) Bristol



(b) Edinburgh



(c) Leeds



(d) Reading

Figure 7: Reordered adjacency matrices by *iid*-colBiSBM, Baldock et al., 2019

Network clustering

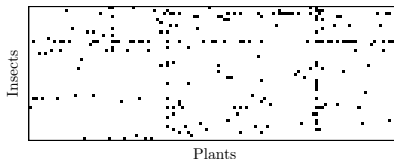
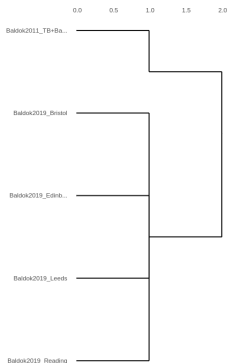
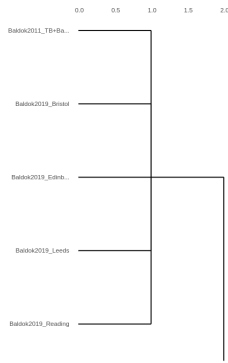


Figure 8: Adjacency matrix, Baldock et al., 2011

Application to Baldock et al., 2019, 2011 I



(a) Model iid



(b) Model $\pi\rho$

Figure 9: Partitioning of networks of Baldock et al., 2019, 2011

Application to Baldock et al., 2019, 2011 II

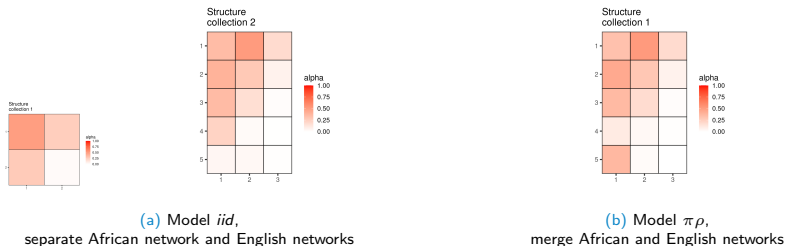
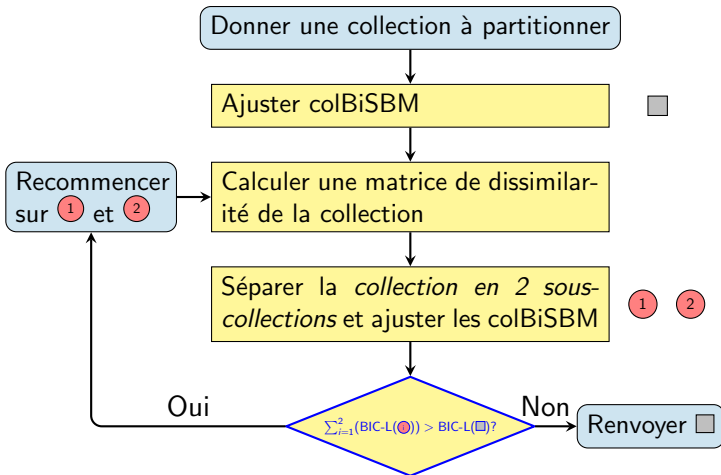


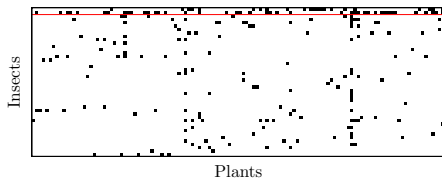
Figure 10: Structures detected for networks of Baldock et al., 2019, 2011

Clustering algorithm

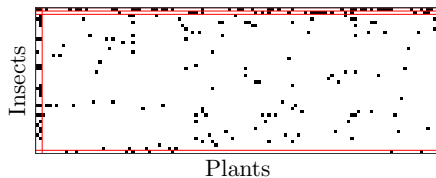


$$D_{\mathcal{M}}(m, m') = \sum_{q=1}^{Q_1} \sum_{r=1}^{Q_2} \max(\tilde{\pi}_q^m, \tilde{\pi}_q^{m'}) \left(\tilde{\alpha}_{qr}^m - \tilde{\alpha}_{qr}^{m'} \right)^2 \max(\tilde{\rho}_r^m, \tilde{\rho}_r^{m'})$$

Results



(a) Reordered by LBM



(b) Reordered by $\pi\rho$ -colBiSBM

Figure 11: Reordered adjacency matrix by $\pi\rho$ -colBiSBM, Baldock et al., 2011

Conclusion and perspectives

Capabilities

- 4 models including 3 with flexibility on at least one of the dimensions (adaptability to data).
- Detect classic and less classic structures in an agnostic way.
- Partition a set of networks according to their structures.

Perspectives

- Investigate stability against randomness and local *optima*.
- Proof of identifiability of the $\pi\rho$ model.

Package and applications

- Integration into the colSBM package, improvement of user interface and addition of ecologists' feedback
- CRAN publication
- Integrate the possibility of an additional criterion for clustering
- Apply clustering to data from [Pichon et al., 2024](#); [Doré et al., 2021](#)

Thank you for your attention !

References 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. *Nature Ecology & Evolution*, 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. *Statistics and Computing*, 18(2), 173–183.
<https://doi.org/10.1007/s11222-007-9046-7>

References 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>

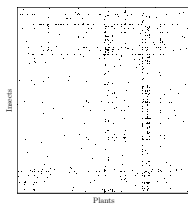
Why does VE minimizes KL ?

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

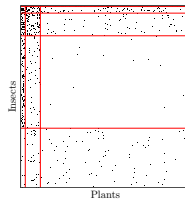
$$\begin{aligned}
 \text{But } \mathbf{KL}[\mathcal{R}_{\mathbf{Y}, \tau}, \log \mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)] &= -\mathbb{E}_{\mathcal{R}_{\mathbf{Y}, \tau}}\left[\log \frac{\mathbb{P}(\mathbf{Z}, \mathbf{W} | \mathbf{Y}; \theta)}{\mathcal{R}_{\mathbf{Y}, \tau}}\right] \\
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

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

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

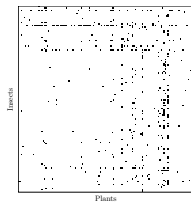


(a) Donnée

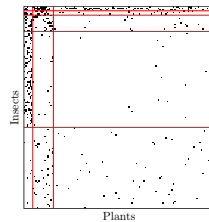


(b) Reordered

Figure 12: Bristol

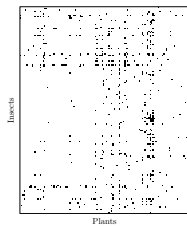


(a) Donnée

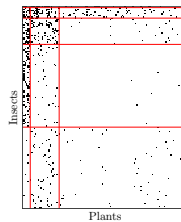


(b) Reordered

Figure 13: Edinburgh

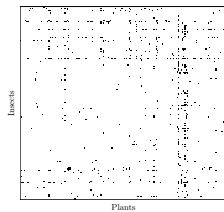


(a) Donnée

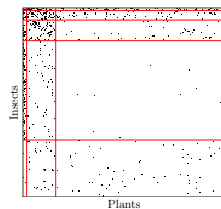


(b) Réordonnée

Figure 14: Leeds



(a) Donnée



(b) Réordonnée

Figure 15: Reading

Appendices references I