Joint analysis of bipartite networks collection

JdS 2025

Louis Lacoste, Pierre Barbillon and Sophie Donnet
Laboratoire MIA Paris-Saclay

May 27, 2025

Why a network?

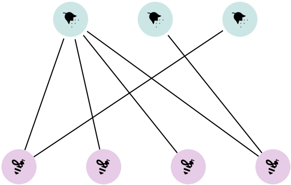


Figure 1: A toy network

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

Associated bi-adjacency matrix

- A bipartite graph $G = (U, V, E)$
- Can be encoded by a bi-adjacency matrix $Y \in \{0, 1\}^{n_1 \times n_2}$

Why a network?

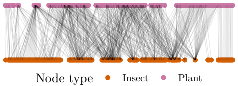


Figure 1: Plant-pollinator network from Bristol Baldock et al., 2019

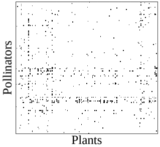
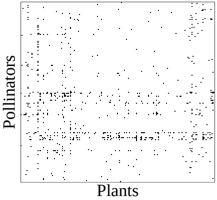


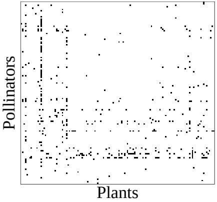
Figure 2: Adjacency matrix of the network

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

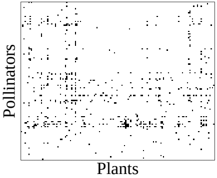
Multiple networks



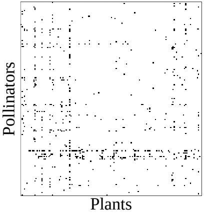
(a) Bristol



(b) Edinburgh



(c) Leeds



(d) Reading

Figure 3: Adjacency matrices, Baldock et al., 2019

Multiple networks

Figure 4: Map of the four cities

Analysis methods for a network

Several methods :

- Metrics at
  - node level: degree, centrality. . .
  - network level: density, nestedness. . .

  Kolaczyk, 2009
- Node embedding and/or clustering with latent variable models
  Snijders and Nowicki, 1997; Hoff et al., 2002
- Node or network embedding with Graph Convolutional Networks
  Kipf and Welling, 2016

Analysis methods for a network

Several methods :

- Metrics at
  - node level: degree, centrality...
  - network level: density, nestedness...

  Kolaczyk, 2009
- **Node embedding and/or clustering with latent variable models**
  Snijders and Nowicki, 1997; Hoff et al., 2002
- Node or network embedding with Graph Convolutional Networks
  Kipf and Welling, 2016

Bipartite Stochastic Block Model (BiSBM[1])

Govaert and Nadif, 2005

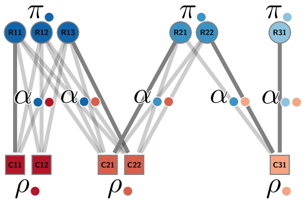


Figure 5: Example of LBM[1]

Hierarchical model

$$\forall q \in [\![1, Q_1]\!], \mathbb{P}(Z_i = q) = \pi_q$$

$$\forall r \in [\![1, Q_2]\!], \mathbb{P}(W_j = r) = \rho_r$$

$$Y_{ij} | Z_i = q, W_j = r \sim \mathcal{B}ern(\alpha_{q,r})$$

where

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

Concise BiSBM formula

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

[1] Commonly Known as *Latent Block Model* (LBM) in the literature.

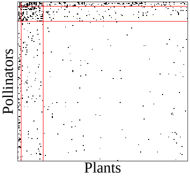
Model 0: sep-BiSBM

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

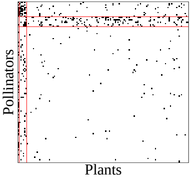
Model 0: sep-BiSBM

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

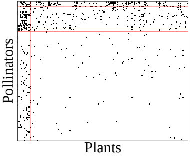
Model 0: sep-BiSBM



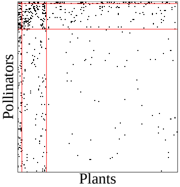
(a) Bristol, $Q_1 = 3$, $Q_2 = 3$



(b) Edinburgh, $Q_1 = 3$, $Q_2 = 3$



(c) Leeds, $Q_1 = 3$, $Q_2 = 2$



(d) Reading, $Q_1 = 3$, $Q_2 = 3$

Figure 6: Reordered adjacency matrices, using BiSBM for each network

Several joint models

iid-colBiSBM

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

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

Several joint models

iid-colBiSBM

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

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

$\pi\rho$-colBiSBM

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

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

Parameter estimation

How ?

$$\ell(\mathbf{Y};\theta) = \sum_{m=1}^{M} \ell(Y^m;\theta)$$

$$= \sum_{m=1}^{M} \log \int_{\mathcal{Z}^m \times \mathcal{W}^m} \exp\{\ell_c(Y^m, Z^m, W^m;\theta)\} dZ^m dW^m$$

$$= \sum_{m=1}^{M} \log \int_{\mathcal{Z}^m \times \mathcal{W}^m} \exp\{\ell(Y^m|Z^m, W^m;\alpha)+$$

$$\ell(Z^m;\pi) + \ell(W^m;\rho)\} dZ^m dW^m$$

Parameter estimation

How ?

$$
\begin{aligned}
\ell(\mathbf{Y};\theta) &= \sum_{m=1}^{M} \ell(Y^m;\theta) \\
&= \sum_{m=1}^{M} \log \int_{\mathcal{Z}^m \times \mathcal{W}^m} \exp\{\ell_c(Y^m, Z^m, W^m;\theta)\} dZ^m dW^m \\
&= \sum_{m=1}^{M} \log \int_{\mathcal{Z}^m \times \mathcal{W}^m} \exp\{\ell(Y^m | Z^m, W^m;\alpha) + \\
&\quad \ell(Z^m;\pi) + \ell(W^m;\rho)\} dZ^m dW^m
\end{aligned}
$$

Parameter estimation

How ?

$$
\begin{aligned}
\ell(\mathbf{Y};\theta) &= \sum_{m=1}^{M} \ell(Y^m;\theta) \\
&= \sum_{m=1}^{M} \log \int_{\mathcal{Z}^m \times \mathcal{W}^m} \exp\{\ell_c(Y^m, Z^m, W^m;\theta)\} dZ^m dW^m \\
&= \sum_{m=1}^{M} \log \int_{\mathcal{Z}^m \times \mathcal{W}^m} \exp\{\ell(Y^m|Z^m, W^m;\alpha) + \\
&\quad \ell(Z^m;\pi) + \ell(W^m;\rho)\} dZ^m dW^m
\end{aligned}
$$

We would like to use Expectation-Maximization (EM) algorithm (Dempster et al., 1977) but the law of $\mathbf{Z}, \mathbf{W} | \mathbf{Y}, \theta^{(t-1)}$ is untractable due to dependence between rows and columns.

Parameter estimation

Solution

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

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

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

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

where $\theta = (\pi, \rho, \alpha)$ for *iid*-colBiSBM

Selection criterion for Q_1, Q_2

Biernacki et al., 2000 introduced the Integrated Classification Likelihood (ICL).

$$\begin{aligned}\mathrm{ICL}(\mathbf{Y}, Q_1, Q_2) &= \mathbb{E}[\ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \hat{\theta})] - \frac{1}{2}\mathrm{pen}(Q_1, Q_2)\\ &= \ell(\mathbf{Y}; \hat{\theta}) - \mathcal{H}(p(\mathbf{Z}, \mathbf{W}|\mathbf{Y}, \hat{\theta})) - \frac{1}{2}\mathrm{pen}(Q_1, Q_2)\end{aligned}$$

leads to low entropy clustering.

Selection criterion for Q_1, Q_2

Biernacki et al., 2000 introduced the Integrated Classification Likelihood (ICL).

$$\begin{aligned}\mathrm{ICL}(\mathbf{Y}, Q_1, Q_2) &= \mathbb{E}[\ell_c(\mathbf{Y}, \mathbf{Z}, \mathbf{W}; \hat{\theta})] - \frac{1}{2}\mathrm{pen}(Q_1, Q_2) \\ &= \ell(\mathbf{Y}; \hat{\theta}) - \mathcal{H}(p(\mathbf{Z}, \mathbf{W} | \mathbf{Y}, \hat{\theta})) - \frac{1}{2}\mathrm{pen}(Q_1, Q_2)\end{aligned}$$

leads to low entropy clustering.

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

because we want fuzzier clustering.

Practical problems of choosing Q_1, Q_2

Exploration problems

- Exploration of a 2D grid is costly.
- Sensitivity to initializations.

Practical problems of choosing Q_1, Q_2

Exploration problems

- Exploration of a 2D grid is costly. $\rightarrow$ **Greedy approach** and **sliding window**
- Sensitivity to initializations.

Practical problems of choosing Q_1, Q_2

Exploration problems

- Exploration of a 2D grid is costly. $\rightarrow$ **Greedy approach** and **sliding window**
- Sensitivity to initializations. $\rightarrow$ **Spectral clustering** and **split & merge** approach

Results Baldock et al., 2019

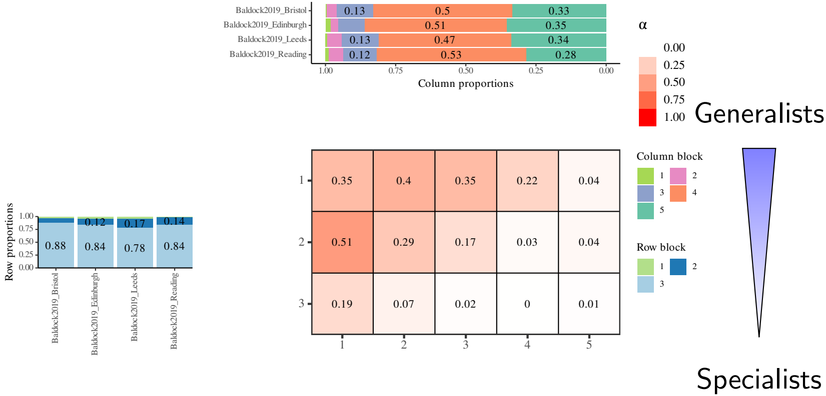
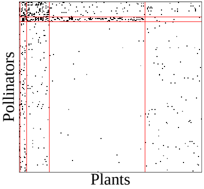
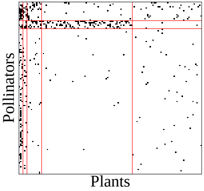


Figure 7: Shared structure (α matrix) and proportions (π and ρ) of the four networks

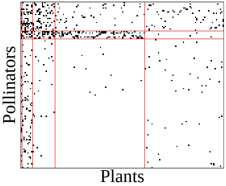
Results Baldock et al., 2019



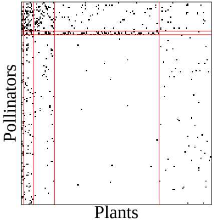
(a) Bristol, $Q_1 = 3$, $Q_2 = 5$



(b) Edinburgh, $Q_1 = 3$, $Q_2 = 5$



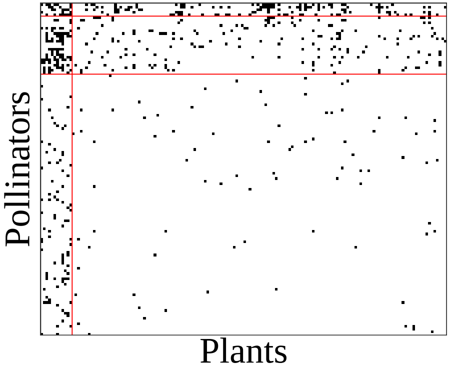
(c) Leeds, $Q_1 = 3$, $Q_2 = 5$



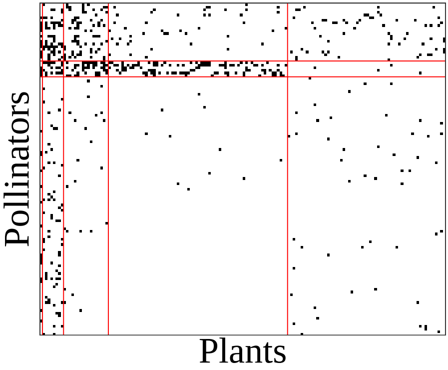
(d) Reading, $Q_1 = 3$, $Q_2 = 5$

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

Focus on Leeds



(a) sep-BiSBM, $Q_1 = 3, Q_2 = 2$



(b) *iid*-colBiSBM, $Q_1 = 3, Q_2 = 5$

Bombus

(a) *Bombus Hortorum* or garden bumblebee

(b) *Bombus Lapidarius* or red-tailed bumblebee

Bombus

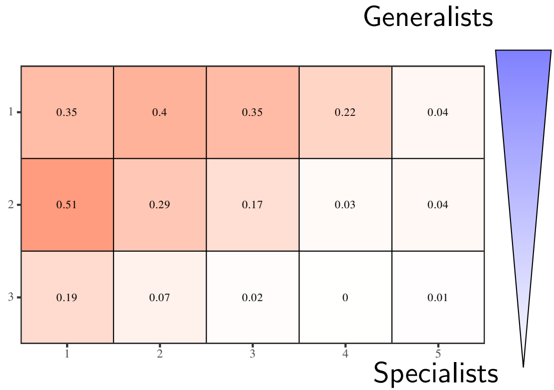


Figure 10: Shared structure (α matrix) of the four networks

Bombus

(a) *Bombus Hortorum* or garden bumblebee

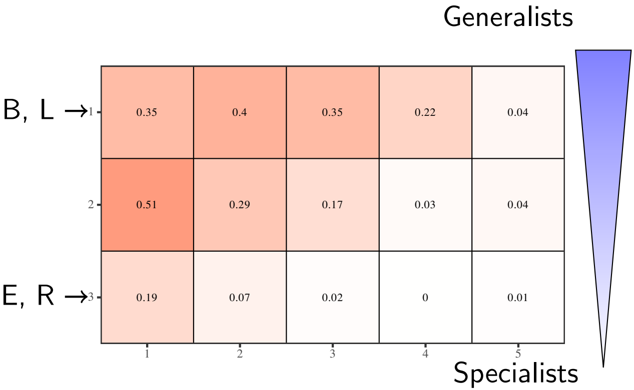


Figure 10: Shared structure (α matrix) of the four networks

Bombus

(b) *Bombus Lapidarius* or red-tailed bumblebee

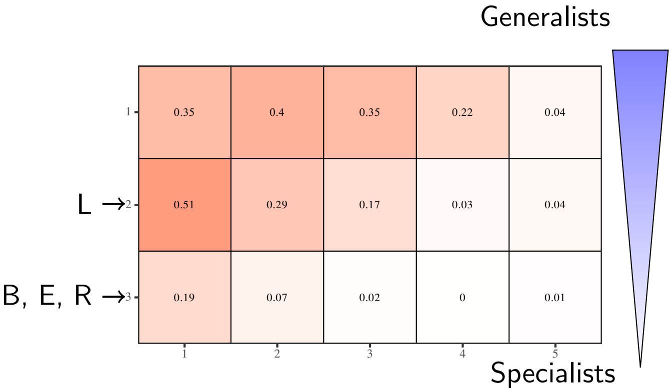


Figure 10: Shared structure (α matrix) of the four networks

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.

Package and applications

- Article in redaction
- R package `colSBM` on Github[a]
- Apply clustering to data from Pichon et al., 2024; Doré et al., 2021 to tell if interaction drives the structure of the network.

[a] https://github.com/GrossSBM/colSBM

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

Kolaczyk, E. D. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer New York. https://doi.org/10.1007/978-0-387-88146-1 Read_Status: New Read_Status_Date: 2025-05-26T11:42:27.939Z.

Snijders, T. A., & Nowicki, K. (1997). Estimation and Prediction for Stochastic Blockmodels for Graphs with Latent Block Structure. *Journal of Classification*, *14*(1), 75–100. https://doi.org/10.1007/s003579900004

References II

Hoff, P. D., Raftery, A. E., & Handcock, M. S. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, *97*(460), 1090–1098. https://doi.org/10.1198/016214502388618906

Kipf, T. N., & Welling, M. (2016, November 21). *Variational Graph Auto-Encoders*. arXiv: 1611.07308 [stat]. https://doi.org/10.48550/arXiv.1611.07308
Read_Status: New
Read_Status_Date: 2025-05-09T11:54:37.094Z.

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

References III

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, *39*(1), 1–38. Retrieved May 27, 2025, from https://www.jstor.org/stable/2984875
Read_Status: New
Read_Status_Date: 2025-05-27T16:20:41.925Z.

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

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

References IV

Biernacki, C., Celeux, G., & Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(7), 719–725. https://doi.org/10.1109/34.865189

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

Developed formula of variational EM

$$\ell(\mathbf{Y};\theta) \geq \sum_{m=1}^{M} \Bigg(\sum_{i=1}^{n_1^m} \sum_{j=1}^{n_2^m} \sum_{q \in \mathcal{Q}_{1,m}} \sum_{r \in \mathcal{Q}_{2,m}} \tau_{i,q}^{1,m} \tau_{j,r}^{2,m} \log f(Y_{ij}^m; \alpha_{qr})$$

$$+ \sum_{i=1}^{n_1^m} \sum_{q \in \mathcal{Q}_{1,m}} \tau_{i,q}^{1,m} \log \pi_q{}^m + \sum_{j=1}^{n_2^m} \sum_{r \in \mathcal{Q}_{2,m}} \tau_{j,r}^{2,m} \log \rho_r{}^m$$

$$- \sum_{i=1}^{n_1} \tau_{i,q}^{1,m} \log \tau_{i,q}^{1,m} - \sum_{j=1}^{n_2} \tau_{j,r}^{2,m} \log \tau_{j,r}^{2,m} \Bigg) =: \mathcal{J}(\tau;\theta),$$

Variational approximation

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

Variational Expectation Step

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

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

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

Maximization Step

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

Connectivity parameters

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

Proportions for *iid*

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

Maximization Step

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

Connectivity parameters

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

Proportions for $\pi \rho$

$$\widehat{\pi}^{m}{}_{q} = \frac{\sum_{i=1}^{n_1^m} \tau_{iq}^{1,m}}{n_1^m} \qquad \widehat{\rho}^{m}{}_{r} = \frac{\sum_{j=1}^{n_2^m} \tau_{jr}^{2,m}}{n_2^m}$$

Why does VE minimizes KL ?

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

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

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

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

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

On the BIC-L

$$\mathrm{ICL}(\hat{\theta}) = \mathbb{E}_{\mathbf{Z},\mathbf{W}|\mathbf{Y}}[\ell_c(\mathbf{Y},\mathbf{Z},\mathbf{W};\hat{\theta})] - \frac{1}{2}\mathrm{pen}(\dots)$$

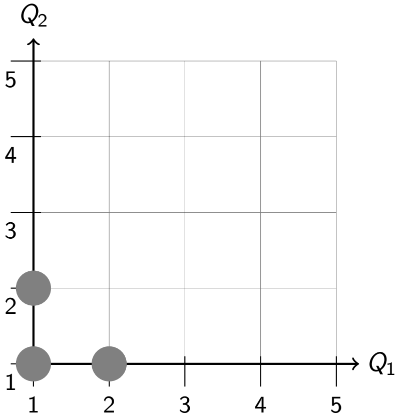
$$\mathbb{E}_{\mathbf{Z},\mathbf{W}|\mathbf{Y}}[\ell_c(\mathbf{Y},\mathbf{Z},\mathbf{W};\hat{\theta})] = \log p(\mathbf{Y};\hat{\theta}) - \mathcal{H}(p(\mathbf{Z},\mathbf{W}|\mathbf{Y}))$$

$$\text{And thus, } \mathrm{ICL}(\hat{\theta}) = \log p(\mathbf{Y};\hat{\theta}) - \mathcal{H}(p(\mathbf{Z},\mathbf{W}|\mathbf{Y})) - \frac{1}{2}\mathrm{pen}(\dots)$$

Recalling that $\mathbf{Z},\mathbf{W}|\mathbf{Y}$ is inaccessible, we use the *variational approximation* $\mathcal{R}_{\mathbf{Y},\hat{\tau}}$ and not penalizing the entropy of the distribution we derive the BIC-Like criterion:

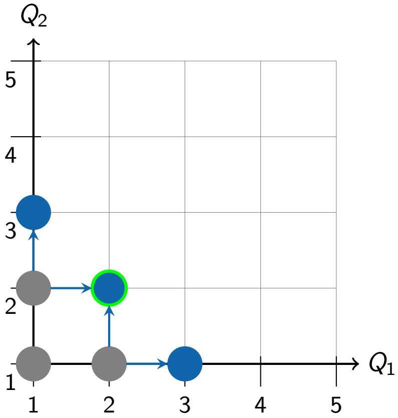
$$\text{BIC-L}(\hat{\theta},\hat{\tau}) = \mathbb{E}_{\mathcal{R}_{\mathbf{Y},\hat{\tau}}}[\ell_c(\mathbf{Y},\mathbf{Z},\mathbf{W};\hat{\theta}^{\mathrm{var}})] + \mathcal{H}(\mathcal{R}_{\mathbf{Y},\hat{\tau}}) - \frac{1}{2}\mathrm{pen}(\dots)$$

Choice of (Q_1, Q_2) - Greedy approach



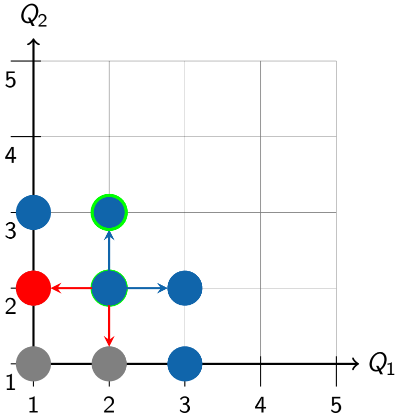
- Initial model :

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 : (grey dot)
- Model after *split* : (blue dot)
- Model maximizing the criterion : (green circle)
- Model after *merge* : (red dot)

Choice of (Q_1, Q_2) - Sliding window

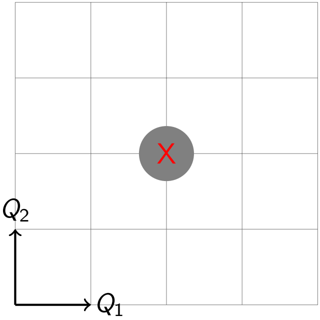


Figure 11: Sliding window

Choice of (Q_1, Q_2) - Sliding window

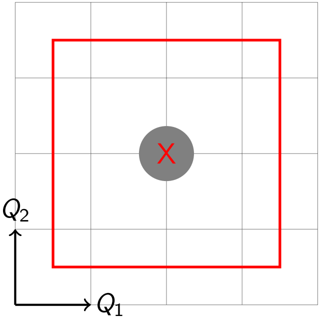


Figure 11: Sliding window

Choice of (Q_1, Q_2) - Sliding window

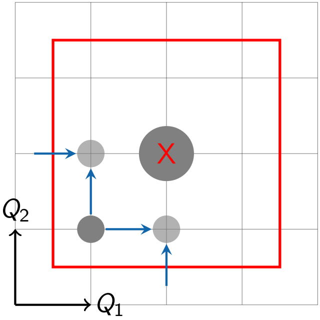


Figure 11: Sliding window

Initialization of the model if necessary

Choice of (Q_1, Q_2) - Sliding window

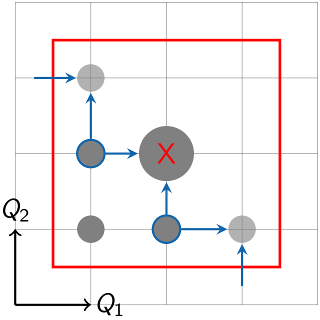


Figure 11: Sliding window

Choice of (Q_1, Q_2) - Sliding window

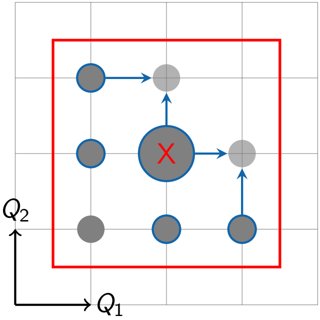


Figure 11: Sliding window

Choice of (Q_1, Q_2) - Sliding window

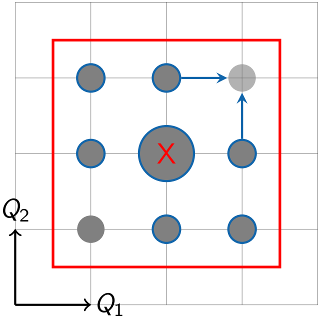


Figure 11: Sliding window

Choice of (Q_1, Q_2) - Sliding window

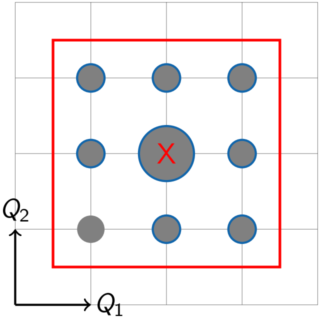


Figure 11: Sliding window

Choice of (Q_1, Q_2) - Sliding window

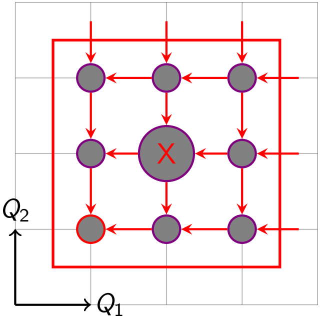


Figure 11: Sliding window

Choice of (Q_1, Q_2) - Sliding window

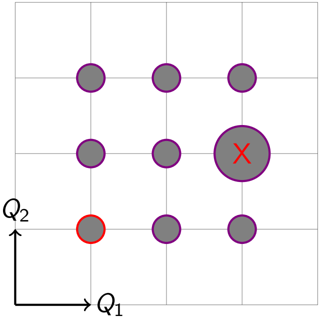


Figure 11: Sliding window

Localization of the new mode

Choice of (Q_1, Q_2) - Sliding window

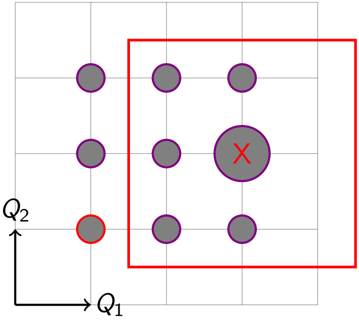
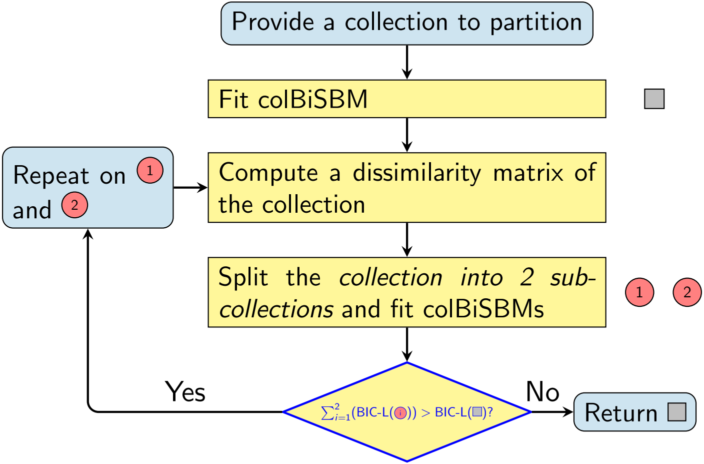


Figure 11: Sliding window

Move to the new mode then iterate

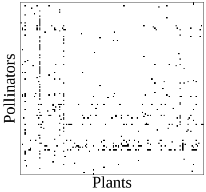
Clustering algorithm



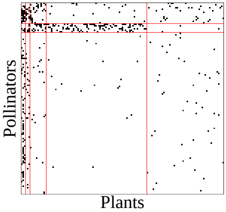
$$D_{\mathcal{M}}(m, m') = \sum_{q=1}^{Q_1} \sum_{r=1}^{Q_2} \max(\widetilde{\pi}_q^m, \widetilde{\pi}_q^{m'}) \left(\widetilde{\alpha}_{qr}^m - \widetilde{\alpha}_{qr}^{m'} \right)^2 \max(\widetilde{\rho}_r^m, \widetilde{\rho}_r^{m'})$$

(a) Donnée (b) Reordered

Figure 12: Bristol

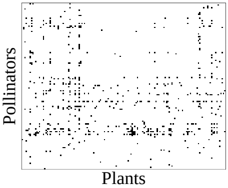


(a) Donnée

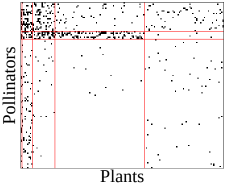


(b) Reordered

Figure 13: Edinburgh

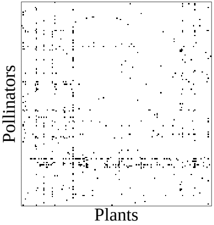


(a) Donnée

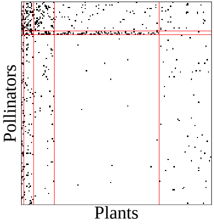


(b) Réordonnée

Figure 14: Leeds



(a) Donnée



(b) Réordonnée

Figure 15: Reading

Appendices references I