# RAPPORT DE STAGE DANS L'UMR MIA PARIS-SACLAY

LOUIS LACOSTE

22 mai 2023

# Table des matières

# Chapitre 1

# Présentation de l'UMR

# Chapitre 2

# Adaption au cas bipartite : colBiSBM

## 2.1 Etape VE de l'algorithme

Formule du point fixe pour la distribution de Bernoulli

— *iid* :

$$\boldsymbol{\tau}^{m,1} = {}^t\pi + \exp[(\mathrm{Mask}^m \odot A^m)\boldsymbol{\tau}^{m,2\ t}(\mathrm{logit}(\alpha)) + \mathrm{Mask}^m \boldsymbol{\tau}^{m,2\ t}\log(\mathbf{1}-\alpha)]$$

$$\log(\boldsymbol{\tau}^{m,2}) = {}^t\log(\rho) + {}^t(\mathrm{Mask}^m \odot A^m)\boldsymbol{\tau}^{m,1}\mathrm{logit}(\alpha) + {}^t\mathrm{Mask}^m \boldsymbol{\tau}^{m,1}\log(\mathbf{1}-\alpha)$$

— $\rho\pi$ :

$$\log(\boldsymbol{\tau}^{m,1}) = {}^t\log(\pi^m) + (\mathrm{Mask}^m \odot A^m)\boldsymbol{\tau}^{m,2\ t}(\mathrm{logit}(\alpha)) + \mathrm{Mask}^m \boldsymbol{\tau}^{m,2\ t}\log(\mathbf{1}-\alpha)$$

$$\log(\boldsymbol{\tau}^{m,2}) = {}^t\log(\rho^m) + {}^t(\mathrm{Mask}^m \odot A^m)\boldsymbol{\tau}^{m,1}\mathrm{logit}(\alpha) + {}^t\mathrm{Mask}^m \boldsymbol{\tau}^{m,1}\log(\mathbf{1}-\alpha)$$

avec $\mathrm{Mask}^m$ la matrice qui contient des 0 si la valeur est un NA et des 1 sinon.

## 2.2 M step of the algorithm

## 2.3 Computation of the variational bound

## 2.4 Penalties

*iid-colBiSBM* For the *iid-colBiSBM* the penalties were modified in the following way :

— For the $\pi$s and $\rho$s :

$$\mathrm{pen}_\pi(Q_1) = (Q_1 - 1)\log(\sum_{m=1}^{M} n_r^{(m)})$$

$$\mathrm{pen}_\rho(Q_2) = (Q_2 - 1)\log(\sum_{m=1}^{M} n_c^{(m)})$$

— For the $\alpha$s :

$$\text{pen}_\alpha(Q_1, Q_2) = Q_1 \times Q_2 \log(N_M)$$

avec

$$N_M = \sum_{m=1}^{M} n_r^{(m)} \times n_c^{(m)}$$

And thus the $BIC - L$ formula is now :

$$BIC - L(\boldsymbol{X}, Q_1, Q_2) = \max_\theta \mathcal{J}(\hat{\mathcal{R}}, \boldsymbol{\theta}) - \frac{1}{2}[\text{pen}_\pi(Q_1) + \text{pen}_\rho(Q_2) + \text{pen}_\alpha(Q_1, Q_2)]$$

*$\rho\pi$-colBiSBM* For the *$\rho\pi$-colBiSBM* the penalties are the following :

— The support penalties are :

$$\text{pen}_{S_1}(Q_1) = -2 \log p_{Q_1}(S_1)$$

$$\text{pen}_{S_2}(Q_2) = -2 \log p_{Q_2}(S_2)$$

with

$$\log p_{Q_1}(S_1) = -M \log(Q_1) - \sum_{m=1}^{M} \log \binom{Q_1}{Q_1^{(m)}}$$

$$\log p_{Q_2}(S_2) = -M \log(Q_2) - \sum_{m=1}^{M} \log \binom{Q_2}{Q_2^{(m)}}$$

— Penalties for the $\rho$s and $\pi$s :

$$\text{pen}_\pi(Q_1, S_1) = \sum_{m=1}^{M} (Q_1^{(m)} - 1) \log n_r^{(m)}$$

$$\text{pen}_\rho(Q_2, S_2) = \sum_{m=1}^{M} (Q_2^{(m)} - 1) \log n_c^{(m)}$$

— Penalties for the $\alpha$s :

$$\text{pen}_\alpha(Q_1, Q_2, S_1, S_2) = (\sum_{q=1}^{Q_1} \sum_{r=1}^{Q_2} \mathbb{1}_{(S_1)'S_2 > 0}) \log(N_M)$$

And the corresponding $BIC - L$ formula :

$$\begin{aligned} BIC - L(\boldsymbol{X}, Q_1, Q_2) = \max_{S_1, S_2} [ & \max_{\theta_{S_1,S_2} \in \Theta_{S_1,S_2}} \mathcal{J}(\hat{\mathcal{R}}, \theta_{S_1,S_2}) \\ & - \frac{1}{2}(\text{pen}_\pi(Q_1, S_1) + \text{pen}_\rho(Q_2, S_2) \\ & \quad + \text{pen}_\alpha(Q_1, Q_2, S_1, S_2) \\ & \quad + \text{pen}_{S_1}(Q_1) + \text{pen}_{S_2}(Q_2))] \end{aligned}$$

## 2.5 Latent space exploration and model selection

In order to explorer the bi-dimensional latent space $(Q_1, Q_2)$ we use the following strategies.

### 2.5.1 Model selection

In the following steps the model selection consists of using the $BIC - L$ criterion to select the model. We choose among the proposed models the one that maximizes the $BIC - L$

### 2.5.2 Initialization and pairing of the models

First to combine the information from the $M$ networks we fit a collection model for each network at the two points $Q = (1, 2)$ and $Q = (2, 1)$. Using the previously described VEM algorithm we obtain for each network its parameters $(\rho, \pi, \alpha)$.

We then compute the marginal laws for each dimension, for each network. Then we order the network blocks by the probabilities obtained in decreasing order.

— For the memberships on the columns : $col\ order_m = order\left(\pi_m \times \alpha_m\right)$

— For the memberships on the rows : $row\ order_m = order\left(\rho_m \times\ {}^t(\alpha_m)\right)$

Using this order we relabel the memberships for the $M$ fitted collection of a single network. Then we use the $M$ memberships to fit a collection containing the $M$ networks.

### 2.5.3 Greedy exploration to find an estimation of the mode

Using the previously fitted models for $Q = (1, 2)$ and $Q = (2, 1)$ we choose to perform a greedy exploration to find a first mode.

Meaning that for a given $Q = (Q_1, Q_2)$ we will compute all the possible memberships for the points $Q = (Q_1 + 1, Q_2)$ and $Q = (Q_1, Q_2 + 1)$, fit the corresponding models and choose the one that maximizes the $BIC - L$ as the next point from which to repeat the procedure. We repeat the procedure until the $BIC - L$ stops increasing 3 times in a row.

When this first estimation of the $BIC - L$ mode has been find we apply the moving window on it.

### 2.5.4 Fenêtre glissante pour mettre à jour les clusterings et les $BIC - L$

## 2.6 Clustering des réseaux

### 2.6.1 Adaptation de la distance entre les paramètres du modèle

La distance pondère désormais avec les $\pi$ et les $\rho$.

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

# Table des figures

# Liste des tableaux