

RAPPORT DE STAGE DANS L'UMR MIA PARIS-SACLAY

LOUIS LACOSTE

13 avril 2023

Table des matières

1	Présentation de l'UMR	2
2	Adaption au cas bipartite : colBiSBM	3
2.1	<i>Etape VE de l'algorithme</i>	3
2.2	<i>Calcul des pénalités</i>	3
2.3	<i>Exploration de l'espace latent</i>	4
2.3.1	Initialisation et appariement des modèles	4
2.3.2	Exploration gloutonne pour trouver une estimation du mode	4
2.3.3	Fenêtre glissante pour mettre à jour les clusterings et les BIC-L	4

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* :

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

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

— $\rho\pi$:

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

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

avec Mask^m la matrice qui contient des 0 si la valeur est un NA et des 1 sinon.

2.2 Calcul des pénalités

iid-colBiSBM Dans le cas *iid-colBiSBM* les pénalités sont à adapter de la manière suivante :

— Pour les π et les ρ :

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

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

— Pour les α :

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

avec

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

Le BIC-L devient donc :

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

$\rho\pi$ -colBiSBM Dans le cas *$\rho\pi$ -colBiSBM* les pénalités deviennent les suivantes :

— Pour les pénalités dûes aux supports :

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

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

avec

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

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

— Pour les pénalités dûes aux ρ et π :

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

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

— Pour les pénalités dûes aux α :

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

Et alors le BIC-L est égal à :

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

2.3 Exploration de l'espace latent

2.3.1 Initialisation et appariement des modèles

2.3.2 Exploration gloutonne pour trouver une estimation du mode

2.3.3 Fenêtre glissante pour mettre à jour les clusterings et les BIC-L

Table des figures

Liste des tableaux