

1 Ce que j'ai compris

Un modèle de graphes aléatoires qui à la fois :

- est un modèle de mélange
- décrit explicitement les liens entre les noeuds

Notations :

- $X_{ij} \in \{0, 1\}$ l'existence d'une arête entre les noeuds i et j .
- Q le nombre de classes, avec $\{\alpha_1, \dots, \alpha_Q\}$ les proportions de chaque classe.
- $Z_{iq} \in \{0, 1\}$ l'appartenance du noeud i à la classe q .
- $\pi_{ql} = \pi_{lq}$ la probabilité d'un lien entre un noeud de la classe q et un noeud de la classe l .

Hypothèses :

- Bonne définition des classes : $\sum_q Z_{iq} = 1$, $\sum_q \alpha_q = 1$ et $\alpha_q = P(Z_{iq} = 1)$.
- Pas de self loop : $X_{ii} = 0$
- Indépendances des arêtes conditionnellement aux classes : $X_{ij} | \{i \in q, j \in l\} \sim \mathcal{B}(\pi_{ql})$.

Propriétés importantes :

- Distribution des degrés binomiale (conditionnellement aux classes).

Extension : modèle indépendant, chaque classe a une propension à se lier avec les autres classes. On a alors $\pi_{ql} = \eta_q \eta_l$.

Log-vraisemblance du modèle (est-ce que le modèle observé est probable au vu des paramètres fixés?) :

$$\log \mathcal{L}(X, Z) = \sum_i \sum_q Z_{iq} \log \alpha_q + \frac{1}{2} \sum_{i \neq j} \sum_{q, l} Z_{iq} Z_{jl} \log b(X_{ij}, \pi_{ql})$$

avec $b(x, \pi) = \pi^x (1 - \pi)^{1-x}$.

Estimation : on veut estimer les paramètres du modèle en maximisant la log-vraisemblance. L'algorithme EM n'est pas utilisable ici. Le papier se concentre sur l'optimisation d'une borne inférieure de la log-vraisemblance, de façon similaire à l'algorithme EM.

- Étape E : on calcule les paramètres optimaux $\hat{\tau}_{iq}$ de la distribution conditionnelle approchée (point fixe).
- Étape M : on maximise la log-vraisemblance approchée par rapport aux paramètres du modèle.

On note que :

- On a une garantie de croissance de la log-vraisemblance approchée à chaque itération.
- Le nombre de classes est fixé à l'avance. Le papier propose un critère de sélection du nombre de classes.

2 Questions

Rentrer dans le détail des preuves ?

3 Objectifs

- Implémenter l'algorithme d'estimation.
- Tester l'implémentation sur le même dataset que le papier.
- Trouver des datasets pertinents sur lesquels tester l'algorithme.
- Proposer une méthode d'estimation alternative (idées, preuves, implémentation)
- Investiguer la pertinence du critère de sélection du nombre de classes. Implémenter d'autres critères, montrer leur pertinence en fonction du type de tâche.
- Étudier le sampling après estimation : est-ce que le modèle est capable de générer des graphes similaires à ceux du dataset ?

4 Ce qui sera simple

- Sampler un graphe aléatoire avec des paramètres fixés.
- Implémenter l'algorithme d'estimation.

5 Ce qui sera difficile

- Faire tourner l'algorithme efficacement.
- Trouver des méthodes d'estimation alternatives qui s'appliquent.

6 Idées, critiques

Comparaison avec d'autres modèles ?

7 Implémentation

E-step :

On a :

$$\hat{\tau}_{iq} \propto \alpha_q \prod_{j \neq i} \prod_l b(X_{ij}; \pi_{ql})^{\hat{\tau}_{jl}} \quad (1)$$