

INFÉRENCE DU MODÈLE À BLOCS STOCHASTIQUES EN PRÉSENCE DE DONNÉES MANQUANTES.

Timothée Tabouy ¹, Pierre Barbillon ¹, Julien Chiquet ¹

¹ *UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005, Paris, France*
prenom.nom@agroparistech.fr

Résumé. Le modèle à blocs stochastiques ou *Stochastic Block Model* (SBM) (Nowicki and Snijders, 2001) est un modèle de graphe aléatoire généralisant le modèle d’Erdős-Reyni (Erdős and Renyi, 1959) à l’aide d’une structure latente sur les nœuds. L’utilisation de variables latentes dans le SBM permet de modéliser un large spectre de topologies de réseau, en particulier les graphes en affiliation, en étoile ou bipartite. L’inférence de ces modèles repose sur des modifications de l’algorithme EM (Expectation Maximization), comme par exemple l’approche EM variationnelle (Daudin et al., 2008) ou Bayésienne variationnelle (Latouche et al., 2012). Dans ces approches, le réseau est toujours considéré comme parfaitement observé, alors que de nombreux cas d’application (en particulier en sociologie) suggèrent que son observation est partielle et guidée par une stratégie d’échantillonnage dépendant du réseau lui-même.

Dans un précédent travail (Tabouy et al., 2017) nous avons constaté qu’un échantillonnage partiel du réseau peut induire un biais d’estimation dans le modèle SBM. Notre objectif était alors la modélisation de la stratégie d’échantillonnage utilisée et son intégration dans les procédure d’inférence. S’appuyant sur la théorie des données manquantes développée par Rubin (1976) nous l’avons adaptés au cadre du SBM.

Nous proposons la présentation du package **missSBM** qui implémente les méthodes décrites dans Tabouy et al. (2017) d’inférence du SBM binaire en présence de données manquantes pour 6 stratégies précises décrites dans notre travail (Tabouy et al., 2017), nous parlerons aussi de l’exploration du nombre de groupe a partir du critere ICL. Nous présenterons enfin la structure du package programmé a partir des classes R6 récemment introduites,

Mots-clés. classes R6 · parallélisation du code · Modèle à blocs stochastiques · données manquantes · EM variationnel

References

- A. Celisse, J.-J. Daudin, L. Pierre, et al. Consistency of maximum-likelihood and variational estimators in the stochastic block model. *Electron. J. Stat.*, 6:1847–1899, 2012.
- J.-J. Daudin, F. Picard, and S. Robin. A mixture model for random graphs. *Stat. comp.*, 18(2): 173–183, 2008.
- P. Erdős and A. Renyi. On random graphs. *Publicationes Mathematicae*, 6:290–297, 1959.
- P. Latouche, É. Birmelé, and C. Ambroise. Variational bayesian inference and complexity control for stochastic block models. *Stat. Modelling*, 12(1):93–115, 2012.
- K. Nowicki and T. A. B. Snijders. Estimation and prediction for stochastic blockstructures. *J. Am. Stat. Soc.*, 96(455):1077–1087, September 2001.

D. B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976.

T. Tabouy, P. Barbillon, and J. Chiquet. Variational inference of stochastic block model from sampled data. *submitted to JASA*, 2017.