

QUELQUES EXEMPLES D'ANALYSE DE RÉSEAU EN ÉCOLOGIE

Stephane ROBIN, INRA/AroParisTech/Paris-Saclay

Mots-clés : graphes aléatoires, modèles graphiques, statistiques, réseaux

L'analyse de réseau a pris une place prépondérante dans bon nombre de domaines d'application et notamment dans les sciences du vivant. De façon générale un réseau est un graphe dont les noeuds sont identifiées à des entités (gènes, espèces, molécules) et dans lequel la présence d'une arête indique l'existence d'une interaction entre deux entités. Le terme d'analyse de réseau recouvre un grand nombre de problèmes statistiques différents dont *l'inférence de réseau* qui vise, à partir d'observations faites sur les noeuds, à inférer l'existence des arêtes et *l'analyse topologique* qui cherche à comprendre l'organisation globale ou locale d'un réseau directement observé.

Dans cet exposé, on décrira rapidement ces deux problèmes. D'un point de vue statistique, l'inférence de réseau est généralement formulée en terme d'inférence de la structure d'un modèle graphique sous-jacent [4] et les modèles graphiques gaussiens y sont un cas d'école. Les méthodes statistiques développées pour l'analyse topologique reposent généralement sur la définition de modèles de graphes aléatoires hétérogènes [5].

On présentera ensuite deux développements dédiés à l'analyse de réseaux écologiques. Le premier porte sur l'inférence de réseau d'interactions entre espèces à partir de données d'abondance. L'approche proposée tire partie de méthodes développées dans le cadre gaussien pour les étendre à l'analyse de données de comptages dans le cadre du modèle Poisson log-normal [1]. L'inférence de ce modèle repose sur un approximation variationnelle qui bénéficie de la convexité partielle de la fonction objectif [2].

Le second développement porte sur la prise en compte de covariables pour expliquer la topologie d'un réseau. Plus précisément, il s'agit de déterminer si les covariables disponibles suffisent à expliquer la topologie observée. Ici encore, une première inférence repose sur une approximation (bayésienne) variationnelle. On montrera comment une approche par échantillonnage séquentiel permet d'utiliser le résultat de cette approximation comme point de départ pour échantillonner efficacement dans la loi *a posteriori* des paramètres [3].

Co-auteurs : Julien CHIQUET, Sophie DONNET et Mahendra MARIADASSOU.

Références

- [1] J. Aitchison and C.H Ho. The multivariate Poisson-log normal distribution. *Biometrika*, 76(4) :643–653, 1989.
- [2] J. Chiquet, M. Mariadassou, and S. Robin. Variational inference for sparse network reconstruction from count data. Technical Report 1806.03120, arXiv, 2018.
- [3] S. Donnet and S. Robin. Using deterministic approximations to accelerate SMC for posterior sampling. Technical Report 1612.06928, arXiv, 2017.
- [4] S. L. Lauritzen. *Graphical Models*. Oxford Statistical Science Series. Clarendon Press, 1996.
- [5] C. Matias and S. Robin. Modeling heterogeneity in random graphs through latent space models : a selective review. *ESAIM : Proc.*, 47 :55–74, 2014.

Stephane ROBIN, UMR MIA-Paris, AgroParisTech, INRA, Université Paris-Saclay, 75005 Paris, France