

Grands graphes aléatoires spatialisés sous-jacents à la propagation d'épidémies : estimation des paramètres et analyse de données

Encadrement : Patrick Hoscheit, Estelle Kuhn, Elisabeta Vergu (INRAE, UR1404 MaIAGE)

Contexte scientifique

Les réseaux (ou graphes), définis par un ensemble de nœuds (ou sommets), et par leurs connections (ou arêtes), permettent des représentations naturelles de systèmes d'entités et de leurs interactions et sont très abondants dans le monde actuel (réseaux sociaux, réseaux de contacts entre populations, réseaux de neurones, réseaux métaboliques, internet, etc).

Dans ce contexte, les modèles de graphes aléatoires jouent un rôle important dans la modélisation, l'analyse et la compréhension de phénomènes complexes sur des réseaux réels. Ils sont en particulier utilisés dans l'étude de la propagation d'épidémies pour modéliser l'un des facteurs majeurs influençant les dynamiques épidémiques, à savoir la structure de contacts entre individus ou populations. Cependant, la plupart de ces modèles ne considèrent pas la structure spatiale, qui est une propriété clé de nombreux réseaux du monde réel. La structure spatiale du graphe influence très souvent la propension d'un nœud à établir des liens avec d'autres nœuds en fonction de la distance physique.

Dans un travail récent (Dalmau & Salvi, 2019), un graphe aléatoire spatialisé a été proposé et ses propriétés mathématiques ont été étudiées. Dans ce graphe, les sommets sont placés aléatoirement dans l'espace selon un processus de Poisson et sont reliés entre eux par des arêtes avec une probabilité dépendant de leur distance et de leurs poids. Les poids sont ici des quantités aléatoires non observées rendant compte de l'importance des nœuds dans le réseau. Ce modèle de graphe sera appliqué à la modélisation des contacts entre les élevages de bovins en France via le commerce et à l'étude de la propagation d'épidémies sur une telle structure de contacts.

Ce stage s'inscrit dans le cadre du projet ANR Cadence, dont l'objectif est de développer de tels outils.

Objectifs

L'objectif de ce projet de stage est de :

- (i) proposer une méthode d'estimation des paramètres du modèle de graphe aléatoire spatialisé développé par une méthode basée sur la vraisemblance, inspirée de celles développées dans des modèles graphiques à variables latentes type SBM (Kuhn et al, 2020).
- (ii) ajuster le modèle de graphe aléatoire spatialisé aux données d'échanges marchands de bovins, préalablement analysées, afin d'évaluer qualitativement son adéquation ;
- (iii) implémenter et simuler un modèle épidémique simple de propagation sur ce réseau afin d'évaluer sa vulnérabilité ;
- (iv) évaluer par simulation (expériences de percolation) l'efficacité de mesures de maîtrise consistant à déconnecter du réseau certains nœuds ou certaines arêtes ciblés préférentiellement.

L'application repose sur les données extraites de la base BDNI (Dutta et al, 2014), répertoriant les informations concernant les mouvements de tous les bovins français (notamment les identifiants des exploitations source/vente et destination/achat), ainsi que les caractéristiques et les localisations des exploitations. Ces données exhaustives permettent une représentation détaillée et précisent des liens entre les exploitations sous forme de graphe.

Références

- J. Dalmau, Salvi M. Scale-free percolation in continuum space: quenched degree and clustering coefficient. <https://arxiv.org/abs/1902.05774>, 2019
- B.L. Dutta, P. Ezanno, E. Vergu. Characteristics of the spatio-temporal network of cattle movements in France over a 5-year period. *Preventive Veterinary Medicine*, 117:79-94, 2014.
- E. Kuhn, C. Matias, T. Rebaïka. [Properties of the Stochastic Approximation EM Algorithm with Mini-batch Sampling](#). *Statistics and Computing*, Vol. 30, pp 1725--1739, 2020.

Compétences recherchées

Master 2 ou stage de fin d'études ingénieur en mathématiques appliquées, statistiques, physique ou systèmes complexes.

Compétences attendues : connaissances en inférence statistique ; connaissances en théorie des graphes ; manipulation de gros jeux de données ; maîtrise de la programmation bas niveau (C/C++, Python), bonne connaissance du langage R ; appétence pour la modélisation appliquée à la biologie ; maîtrise de l'anglais

scientifique (lecture d'articles).

Déroulement du stage

Le stage se déroulera dans l'unité MaIAGE (Mathématiques et Informatique Appliquées du Génome à l'Environnement) du centre INRAE de Jouy-en-Josas (78350). La durée prévue est de 4 à 6 mois. Début possible en début d'année 2021. Gratification : 550€/mois environ (taux légal).

Candidature

Envoyer CV (inclure la liste des cours suivis en M1/M2 et notes disponibles) et lettre de motivation à patrick.hoscheit@inrae.fr, estelle.kuhn@inrae.fr, elisabeta.vergu@inrae.fr