

Stage M2 (LS2N) :

Clustering de Réseaux d'ordre supérieur pour des données de transports maritimes (2024)

Contexte

Le contexte de ce stage est l'analyse des relations entre lieux à partir de trajectoires maritimes. Ces traces peuvent être définies comme des séquences discrètes d'événements (e.g. les différents ports visités par un navire). Dans ce cadre, on s'intéresse aux relations indirectes entre lieux e.g. sachant qu'un navire vient de Shanghai et est actuellement à Singapour, quelle est sa prochaine destination ? On va ainsi chercher à dépasser la représentation usuelle des déplacements sous la forme de graphes en utilisant des « réseaux d'ordre supérieur » (voir Fig. 1). Ces réseaux représentent les probabilités de transitions d'un lieu à un autre en tenant compte d'un historique de déplacement. Ainsi, ce nouveau type de graphe inclut des « noeuds-mémoire » correspondant à des sous-séquences de lieux. Par exemple, le nœud « Singapour/Tokyo » va encoder l'événement « Le navire, actuellement à Singapour, est arrivé de Tokyo ». Ces modèles reflètent mieux les données d'entrées que les modèles « sans mémoire » (chaîne de Markov d'ordre 1) où la probabilité de transition d'un port à l'autre ne dépendra que du port actuel.

La construction et la fouille de ces nouveaux modèles constituent d'important axes de recherche dans le domaine de la fouille de réseaux. Une application importante est la détection de communautés de lieux chevauchantes, qui est l'objectif de ce stage détaillé plus bas. Ce stage se situe dans la prolongation d'avancées récentes sur ce sujet [Saebi et al. 2020, Queiros et al, 2022]. Les résultats obtenus seront valorisés à travers des publications scientifiques et des outils logiciels.



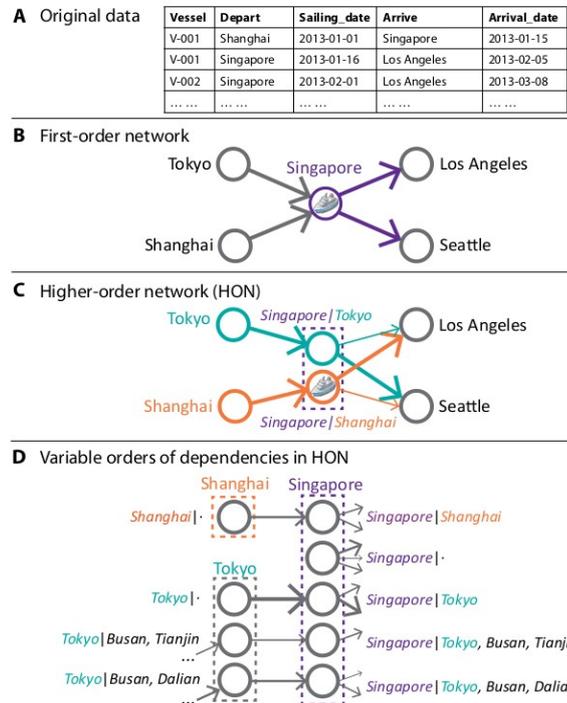


Fig 1. Exemple de construction de Réseau d'ordre supérieur. Image tirée de [Xu et al., 2016]

Objectif

Un premier objectif du stage est de proposer des outils permettant de produire un *clustering* chevauchant des lieux à partir d'un partitionnement des noeuds-mémoires d'un réseau d'ordre supérieur. En effet, un *cluster* de lieux doit représenter un groupe de lieux dont la majorité des flux est dirigée vers d'autres lieux du même groupe. Toutefois, certains lieux (typiquement des grands ports du réseau maritime international) vont pouvoir appartenir à différents groupes. Une problématique dans ce cadre est que la façon dont sont construit les réseaux d'ordre supérieur va beaucoup influencer les *clusters* découverts si on utilise directement des algorithmes classiques [Queiros et al, 2022]. Le stage consistera à développer un algorithme évitant cet écueil. On pourra par exemple proposer une alternative à l'algorithme de *clustering* Walktrap [Pons et Latapy, 2005] (voir Fig. 2). Cela impliquera toutefois d'utiliser une fonction de *scoring* qui est définie pour les *clusterings* chevauchant (par ex. [Esquivel et Rosvall, 2011]). L'algorithme développé pourra être intégré au paquet Python *honyx* (<https://pypi.org/project/honyx>).



Un deuxième objectif sera la constitution de jeux de données de trajectoires maritimes par l'utilisation d'API de sites recensant les positions des navires.

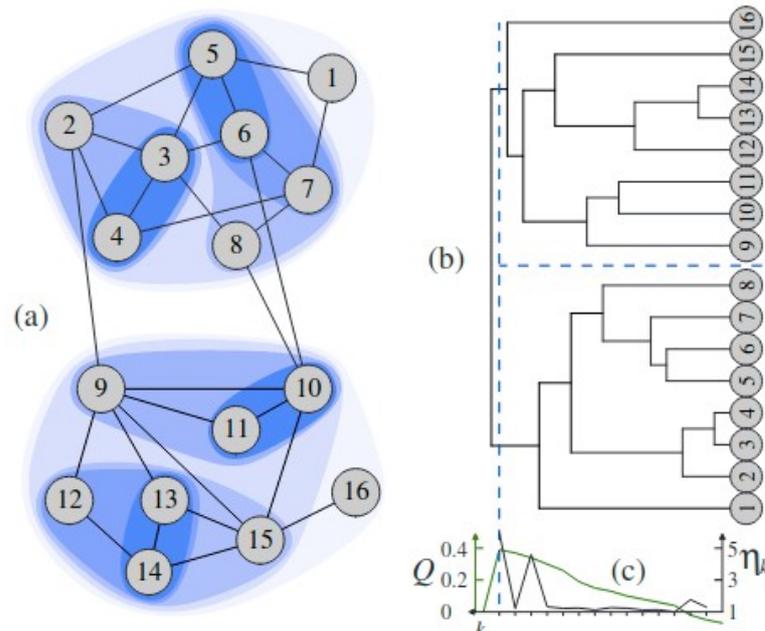


Fig 2. Exemple de clustering d'un graphe non-orientée avec l'algorithme Walktrap. Image tirée de [Pons et Latapy, 2005]

Bibliographie

Xu, Jian, Thanuka L. Wickramaratne, and Nitesh V. Chawla. "Representing higher-order dependencies in networks." *Science advances* 2.5 (2016): e1600028.

Saebi, Mandana, et al. "Efficient modeling of higher-order dependencies in networks: from algorithm to application for anomaly detection." *EPJ Data Science* 9.1 (2020): 15.

Queiros, Julie, Célestin Coquidé, and François Queyroi. "Toward Random Walk Based Clustering of Variable-Order Networks." *Network Science* (2022).

Pons, Pascal, and Matthieu Latapy. "Computing communities in large networks using random walks." *International symposium on computer and information sciences*. Springer, Berlin, Heidelberg, 2005.



Esquivel, Alcides Viamontes, and Martin Rosvall. "Compression of flow can reveal overlapping-module organization in networks." *Physical Review X* 1.2 (2011): 021025.

Mot-clés

fouilles de séquences, réseaux d'ordre supérieur, *graph clustering*, extraction de données

Profil

Nous cherchons un.e étudiant.e en M2 mathématique/informatique (ou équivalent) intéressé.e et ayant des compétences dans l'analyse de données et la fouille de graphes.

- intérêt pour la Recherche et le travail à la fois en équipe et en autonomie
- bonne maîtrise de Python
- capacités rédactionnelles
- bon niveau d'Anglais

Déroulement

- **Durée** : 5/6 mois
- **Gratification** : taux légal en vigueur
- **Localisation** : LS2N (site de Polytech Nantes)

Contact

email : francois.queyroi@univ-nantes.fr

