

Stage (LS2N) : Conception d'outils pour la visualisation de trajectoires spatiales agrégées (2021)

Contexte

Le projet régional S&N (Sequences & Networks) a pour objectif l'étude et la conception d'outils pour l'analyse de trajectoires (déplacement routiers, maritimes *etc.*). Ces traces peuvent être définies comme des séquences discrètes d'événements (*e.g.* les différents ports visités par un navire). Dans ce cadre, le projet S&N 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 (voir Fig. 1) en utilisant des « réseaux d'ordre supérieur » (voir Fig. 2).

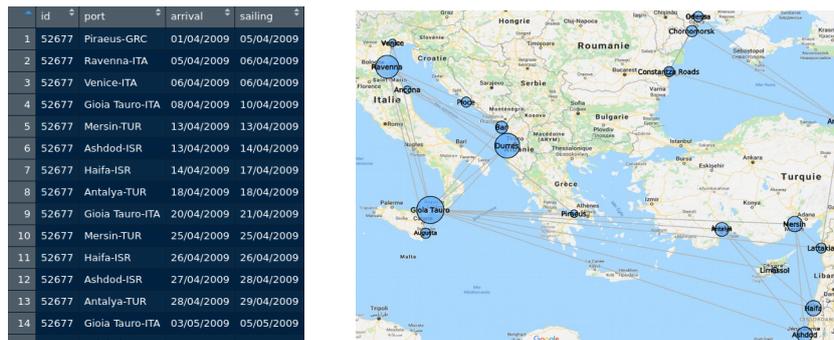


Fig 1. Exemple de données séquentielles correspondant aux ports visités par le navire MSC Koper en 2009 et la représentation de cette trace sous forme de graphe géolocalisé.

Objectif

L'objectif du stage est de proposer des outils permettant de visualiser efficacement les séquences spatiales. Des diagrammes comme les « alluvial graphs » (voir Fig. 2c et 2f) sont parfois utilisés dans ce contexte. Toutefois la taille des jeux de données rend ce type de visualisation difficile. On va donc s'intéresser à l'agrégation automatique des



événements / lieux permettant d'obtenir des modèles qui représentent bien les relations indirectes entre lieux. Ainsi, par exemple, si les navires venant des ports chinois ont des destinations similaires en quittant le port de Singapour, on pourra les considérer comme un objet unique dans le diagramme. Ce travail se basera sur des travaux existants sur la modélisation de trajectoires.

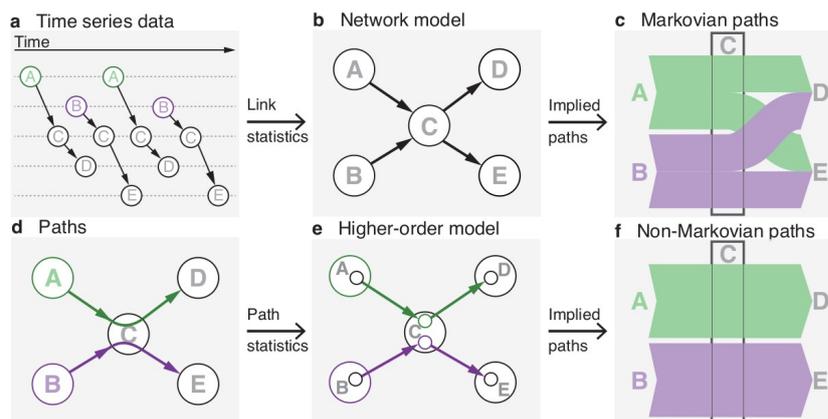


Fig 2. Différence entre la représentation graphique classique et celle par réseaux d'ordres supérieur sur un exemple simple de données séquentielles (image tirée de Lambiotte, R., Rosvall, M., & Scholtes, I. (2018). Understanding complex systems: From networks to optimal higher-order models. arXiv preprint arXiv:1806.05977). **a)** Les données séquentielles en entrée **b)** La représentation agrégée classique **c)** Prédiction des flux en utilisant une marche aléatoire sur la représentation classique (on observe une sur-représentation importante des flux $A \rightarrow C \rightarrow E$ et $B \rightarrow C \rightarrow D$) **d)** Représentation « avec mémoire » des données séquentielles **e)** Réseau d'ordre supérieur encodant les données séquentielles **f)** Prédiction des flux en utilisant une marche aléatoire sur le réseau d'ordre supérieur.

Mot-clés

visualisation de données, fouilles de séquences, *clustering*, trajectoires spatialisées

Profil

Nous cherchons un.e étudiant.e en L3/M1/M2 informatique (ou équivalent) intéressé.e et ayant des compétences dans l'analyse de données.

- bonne maîtrise de Python
- capacités rédactionnelles





LABORATOIRE
DES SCIENCES
DU NUMÉRIQUE
DE NANTES

- bon niveau d'Anglais

Déroulement

- **Durée** : 2/6 mois
- **Gratification** : taux légal en vigueur
- **Localisation** : LS2N (site de Polytech Nantes)

Contact

email : francois.queyroi@univ-nantes.fr



Adresse administrative : LS2N • UFR de Sciences et Techniques
2, rue de la Houssinière • BP 92208 • 44322 Nantes Cedex 03 • France

Unité Mixte UMR 6004 : Université de Nantes, CNRS, École Centrale de Nantes, IMT Atlantique