

# Modélisations de séquences spatialisées dans les réseaux d'ordre supérieur

Lysa Corcuff\*  
François Queyroi\*

\*LS2N, UMR CNRS 6004, Université de Nantes  
nom.prenom@univ-nantes.fr,  
<https://www.ls2n.fr/equipe/duke/>

**Résumé.** L'analyse des mobilités requiert souvent une représentation des flux en modèle markovien d'ordre 1. De nombreux travaux envisagent l'utilisation d'ordres supérieurs afin de construire des réseaux fournissant des meilleures représentations des séquences de déplacement observées. Nous proposons ici l'analyse et la comparaison des qualités prédictives et de la taille de tels modèles sur différents jeux de données géographiques. Nous allons également nous intéresser à la prise en compte de variables exogènes telles que la position ou des catégories de lieux visités qui offre des pistes de recherche intéressantes. Nos expériences indiquent notamment que le modèle HON (Xu et al. (2016)) permet d'obtenir des modèles parcimonieux qui conservent une bonne qualité prédictive même si certains résultats n'ont pu être reproduits. En particulier, aucune stratégie analysée ici ne permet d'obtenir des meilleures prédictions que le modèle d'ordre fixe (Rosvall et al. (2014)).

## 1 Introduction et Contexte

L'analyse des mobilités se base sur l'étude de séquences de déplacements entre lieux effectués par des personnes ou des véhicules. Dans ce cadre, la représentation par un réseau origine-destination est une approche courante qui peut être fouillée pour obtenir de l'information sur le système sous-jacent (Ducruet et Berli (2018)). Si les données en entrée correspondent à un ensemble de suites de lieux  $s_1 s_2 s_3 \dots$  alors ce réseau correspondra généralement à un graphe dont le poids d'une arête  $(s_1, s_2)$  est le nombre de transitions observées entre les lieux  $s_1$  et  $s_2$ . Les paires  $s_1 s_2$  et  $s_2 s_3$  sont ainsi considérées comme des mouvements indépendants. Cette représentation correspond aussi à un modèle prédictif. En effet, des marches aléatoires sur ce graphe peuvent être utilisées pour répondre à la question : quel sera le lieu visité après  $s_2$  ? En utilisant pour probabilités de transition les fréquences relatives, on aboutit à un modèle de Markov d'ordre 1 ou «sans-mémoire» dans lequel seule la dernière position du marcheur importe. Une question légitime est de savoir si ce modèle représente bien les séquences observées en réalité. En effet, de nombreux algorithmes de fouille de graphes se basent sur la notion de marche aléatoire (*e.g.* *PageRank*). Des travaux récents (Rosvall et al. (2014)) ont remis en cause la pertinence de ces modèles sans mémoire et proposent de construire des réseaux permettant de tenir compte des états antérieurs d'un marcheur et ainsi de capturer les dépendances