

THÈSE DE DOCTORAT DE

NANTES UNIVERSITÉ

ÉCOLE DOCTORALE N° 641
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : « Informatique »

Par

Julie QUEIROS

**« Analyse de réseaux d'ordre supérieur construits à partir de
séquences historio-géographiques »**

Thèse présentée et soutenue à ..., le ...

Unité de recherche : Laboratoire des Sciences du Numérique de Nantes (LS2N) UMR 6004,
Nantes Université

Rapporteurs avant soutenance :

Cécile BOUTHOREL Professeur de l'École des Mines, IMT Atlantique Brest
Jean-Loup GUILLAUME Professeur des Universités, Université de la Rochelle

Composition du Jury :

Invité(s) :

Titre : Analyse de réseaux d'ordre supérieur construits à partir de séquences historio-géographiques

Mot clés : Analyse de Réseaux, Séquences discrètes, Ordre supérieur, Centralité, Clustering

Résumé : L'analyse de réseaux permet d'extraire de l'information sur un système à partir des relations existantes entre ses entités. Ces relations peuvent correspondre à des flux tirés de séquences d'états (*i.e.* des trajectoires d'agents entre ces entités). La représentation de ces données sous la forme de graphes suppose généralement que ces trajectoires respectent la propriété de Markov ; seul l'état courant est suffisant pour déterminer l'état futur d'un agent. De nombreux travaux ont remis en cause cette hypothèse et proposé d'autres modèles permettant de passer outre : les réseaux d'ordre supérieur ou HON. Dans les HON, un état peut être représenté par différents nœuds-mémoires encodant les états précédemment visités avant

d'aboutir à l'état courant. Une première problématique traitée dans cette thèse est la construction des HON, qui est un problème de modélisation statistique : on cherche un bon compromis entre taille et qualité du modèle. Nous proposons un modèle de réseaux d'ordre variable plus parcimonieux que les modèles existants. Une deuxième problématique est l'analyse des HON. Un avantage souvent mis en avant est que les algorithmes classiques de graphes peuvent être utilisés avec peu de modifications. Nous montrons, dans le cas de mesures de centralité ou de clustering de graphes, que ce n'est pas le cas. Nous affirmons qu'il faut au contraire adapter les algorithmes aux modèles.

Title: Analysis of Higher-Order Networks built from historio-geographic sequences

Keywords: Network Analysis, Discrete Sequences, Higher-Order, Centrality, Clustering

Abstract: Network analysis extracts information about a system from the relationships between its actors. These relationships can correspond to flows drawn from sequences of states (*i.e.* agent trajectories between these entities). Representing these data as graphs generally assumes that these trajectories respect the Markov property; only the current state is sufficient to determine an agent's future state. Numerous studies have challenged this assumption and proposed other models that allow us to overcome it: higher-order networks or HON. In it, a state can be represented by different memory nodes encoding

the states previously visited before arriving at the current state. A first problem addressed in this thesis is the construction of HON, which is a statistical modeling problem: finding a good compromise between model size and quality. We propose here a model of variable-order networks that is more parsimonious than existing models. A second problem is the analysis of HON. An advantage often put forward is that classical graph algorithms can be used with few modifications. We show, in the case of centrality measures or graph clustering, that this is not the case and argue that algorithms should be adapted to models.