



Université Claude Bernard



DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **6 juin 2019**

Nom de famille et prénom de l'auteur : **BERNARD Jocelyn**

Titre de la thèse : « *Gérer et analyser les grands graphes des entités nommées* ».



Résumé

Dans ce papier de thèse nous étudierons des problématiques de graphes. Nous allons étudier des problématiques théoriques en recherche de pattern et des problématiques appliquées en diffusion d'informations. Nous proposons deux études théoriques sur la recherche et l'énumération de sous-graphes denses que sont les cliques et quasi-cliques. Ensuite nous proposons une étude appliquée sur la propagation d'information dans un graphe d'entités nommées.

Dans un premier temps nous allons étudier la recherche de cliques dans des graphes compressés. Les problèmes MCE (pour l'anglais *Maximal Clique Enumeration*) et MCP (pour l'anglais *Maximum Clique Problem*) sont des problèmes que l'on rencontre dans l'analyse des graphes de données. Ce sont des problèmes difficiles (NP-difficile pour MCE et NP-Complet pour MCP) pour lesquels des solutions adaptées doivent être conçues pour les grands graphes. Nous proposons de répondre à ces problèmes en travaillant sur une version compressée du graphe initial. Nous montrons les bons résultats obtenus par notre méthode pour l'énumération de cliques maximales sur des graphes compressés.

Dans un second temps nous étudierons l'énumération de quasi-cliques maximales. Nous proposons un algorithme distribué qui énumère l'ensemble des quasi-cliques maximales du graphe. Nous démontrons que cet algorithme liste l'ensemble des quasi-cliques maximales du graphe. Nous proposons également une heuristique qui liste un ensemble de quasi-cliques plus rapidement. Nous montrons l'intérêt de l'énumération de ces quasi-cliques par une évaluation des relations en regardant la co-occurrence des noeuds dans l'ensemble des quasi-cliques énumérées.

Dans un troisième temps nous travaillerons sur la diffusion d'événements dans un graphe d'entités nommées. De nombreux modèles existent pour simuler des problèmes de diffusion de rumeurs ou de maladies dans des réseaux sociaux ainsi que des problèmes de propagation de faillites dans les milieux bancaires. Nous proposons de répondre au problème de diffusion d'événement marquant dans des réseaux hétérogènes représentant un environnement économique du monde. Nous proposons un problème de diffusion, nommé problème de classification de l'infection, qui consiste à déterminer quelles entités sont concernées par un événement. Pour résoudre ce problème nous proposons deux modèles inspirés du modèle de seuil linéaire auxquels nous ajoutons différentes fonctionnalités. Finalement, nous testons et validons nos modèles sur un ensemble d'événements.

Mots-clefs: Théorie des graphes, problème d'énumération, cliques, quasi-cliques, graphes compressés, diffusion d'événement, propagation d'information, problème de classification, modèles de seuils.