



Université Claude Bernard



DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **15 décembre 2017**

Nom de famille et prénom de l'auteur : **NABTI Chemseddine**

Titre de la thèse : « *Isomorphisme de Sous-Graphes dans les graphes de données massifs* »

Résumé



L'interrogation de graphes de données est un problème fondamental qui connaît un grand intérêt, en particulier pour les données structurées massives où les graphes constituent une alternative prometteuse aux bases de données relationnelles pour la modélisation des grandes masses de données. Cependant, l'interrogation des graphes de données est différente et plus complexe que l'interrogation des données relationnelles à base de tables. La tâche principale impliquée dans l'interrogation de graphes de données est la recherche d'isomorphisme de sous-graphes qui est un problème NP-complet.

La recherche d'isomorphisme de sous-graphes est un problème très important impliqué dans divers domaines comme la reconnaissance de formes, l'analyse des réseaux sociaux, la biologie, etc. Il consiste à énumérer les sous-graphes d'un graphe de données qui correspondent à un graphe requête. Les solutions les plus connues de ce problème sont basées sur le retour arrière (backtracking). Elles explorent un grand espace de recherche, ce qui entraîne un coût de traitement élevé, notamment dans le cas de données massives.

Pour réduire le temps et la complexité en espace mémoire dans la recherche d'isomorphisme de sous-graphes, nous proposons d'utiliser des graphes compressés. Dans notre approche, la recherche d'isomorphisme de sous-graphes est réalisée sur une représentation compressée des graphes sans les décompresser. La compression des graphes s'effectue en regroupant les sommets en super-sommets. Ce concept est connu dans la théorie des graphes par la décomposition modulaire. Il sert à générer une représentation en arbre d'un graphe qui met en évidence des groupes de sommets qui ont les mêmes voisins. Avec cette compression, nous obtenons une réduction substantielle de l'espace de recherche et par conséquent, une économie significative dans le temps de traitement.

Nous proposons également une nouvelle représentation des sommets du graphe, qui simplifie le filtrage de l'espace de recherche. Ce nouveau mécanisme appelé "compact neighborhood Index (CNI)" encode l'information de voisinage autour d'un sommet en un seul entier. Cet encodage du voisinage réduit la complexité du temps de filtrage de cubique à quadratique. Ce qui est considérable pour les données massifs.

Nous proposons également un algorithme de filtrage itératif qui repose sur les caractéristiques des CNIs pour assurer un élagage global de l'espace de recherche.

Nous avons évalué nos approches sur plusieurs datasets et nous les avons comparés avec les algorithmes de l'état de l'art.