DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **26 novembre 2019**

Nom de famille et prénom de l'auteur : **HABI Abdelmalek**

Titre de la thèse : « *Recherche et agrégation dans les graphes massifs* ».

Résumé

Ces dernières années ont connu un regain d'intérêt pour l'utilisation des graphes comme moyen fiable de représentation et de modélisation des données, et ce, dans divers domaines de l'informatique. En particulier, pour les grandes masses de données, les graphes apparaissent comme une alternative prometteuse aux bases de données relationnelles. Plus particulièrement, le recherche de sous-graphes s'avère être une tâche cruciale pour explorer ces grands jeux de données.

Dans cette thèse, nous étudions deux problématiques principales. Dans un premier temps, nous abordons le problème de la détection de motifs dans les grands graphes. Ce problème vise à rechercher les k-meilleures correspondances (top-k) d'un graphe motif dans un graphe de données. Pour cette problématique, nous introduisons un nouveau modèle de détection de motifs de graphe nommé la *Simulation Relaxée de Graphe* (*RGS*), qui permet d'identifier des correspondances de graphes avec un certain écart (structurel) et ainsi éviter le problème de réponse vide. Ensuite, nous formalisons et étudions le problème de la recherche des k-meilleures réponses suivant deux critères, la pertinence (la meilleure similarité entre le motif et les réponses) et la diversité (la dissimilarité entre les réponses). Nous considérons également le problème des k-meilleures correspondances diversifiées et nous proposons une fonction de diversification pour équilibrer la pertinence et la diversité. En outre, nous développons des algorithmes efficaces basés sur des stratégies d'optimisation en respectant le modèle proposé. Notre approche est efficiente en terme de temps d'exécution et flexible en terme d'applicabilité. L'analyse de la complexité des algorithmes et les expérimentations menées sur des jeux de données réelles montrent l'efficacité des approches proposées.

Dans un second temps, nous abordons le problème de recherche agrégative dans des documents XML. Pour un arbre requête, l'objectif est de trouver des motifs correspondants dans un ou plusieurs documents XML et de les agréger dans un seul agrégat. Dans un premier temps nous présentons la motivation derrière ce paradigme de recherche agrégative et nous expliquons les gains potentiels par rapport aux méthodes classiques de requêtage. Ensuite nous proposons une nouvelle approche qui a pour but de construire, dans la mesure du possible, une réponse cohérente et plus complète en agrégeant plusieurs résultats provenant de plusieurs sources de données. Les expérimentations réalisées sur plusieurs ensembles de données réelles montrent l'efficacité de cette approche en termes de pertinence et de qualité de résultat.

Mots clés:

Appariement de graphes, recherche de motifs de graphe, simulation de graphes, simulation