



Université Claude Bernard



# DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **01 décembre 2021**

Nom de famille et prénom de l'auteur : **Madame BOUHENNI Sarra**

Titre de la thèse : « *Algorithmes parallèles et distribués pour le matching dans les grands graphes* »

## Résumé



L'appariement des sous-graphes (ASG) est un problème classique, souvent modélisé à l'aide de l'isomorphisme de sous-graphes. Il est utilisé dans différents domaines d'application tels que la reconnaissance de motifs et la détection de communautés dans les réseaux sociaux. Néanmoins, en plus du fait qu'il soit NP-Complet, l'isomorphisme de sous-graphes s'avère très strict pour l'ASG dans le contexte actuel des grands graphes. Par conséquent, de nouveaux modèles d'ASG relaxé ont apparu comme la Graph Simulation, permettant d'avoir des résultats intéressants dans un temps polynomial. De plus, les graphes massifs qui sont issus des réseaux sociaux requièrent un stockage et un traitement distribués sur plusieurs machines, d'où la nécessité de revisiter les algorithmes d'ASG relaxé en adoptant de nouveaux paradigmes, dédiés au traitement des grands graphes, notamment le Think-Like-A-Vertex et ses variantes.

Dans cette thèse, nous étudions l'intérêt des systèmes et paradigmes distribués de traitement des grands graphes dans l'évaluation des requêtes d'ASG. L'objectif est d'identifier les modèles de programmation qui sont les mieux adaptés pour ce problème. Par ailleurs, nous visons à proposer de nouveaux algorithmes d'ASG qui sont parallèles, distribués et offrant une scalabilité linéaire.

Nos contributions se résument en quatre points : (1) nous proposons une nouvelle classification des approches distribuées d'ASG, en nous basant sur plusieurs critères tels que le modèle d'ASG et le paradigme de programmation, (2) nous introduisons le nouveau modèle d'ASG relaxé BDSim qui permet de mieux capturer les similarités entre les graphes, tout en ayant une complexité cubique. En plus, nous proposons des algorithmes distribués centré sommet pour l'évaluation de BDSim sur des grands graphes, (3) nous développons le premier algorithme scalable et complètement distribué pour évaluer Strong Simulation, un modèle d'ASG relaxé offrant un compromis entre la flexibilité et la faisabilité, (4) enfin, nous proposons la première approche parallèle et centrée arêtes pour évaluer Graph Simulation et Dual Simulation dans les graphes massifs et distribués. Nous validons les différents algorithmes proposés théoriquement et expérimentalement sur des graphes massifs synthétiques et réels.

A travers ce travail de recherche, nous avons confirmé que différents modèles de programmation peuvent être utilisés pour la conception d'algorithmes d'ASG et cela dépend du modèle d'ASG adopté. Effectivement, l'isomorphisme de sous-graphes et Strong Simulation sont des modèles basés sur la localité et le voisinage à plusieurs sauts, ce qui nécessite un paradigme centré sommet ou encore centré sous-graphe. En revanche, les

algorithmes les plus efficaces pour évaluer Graph Simulation et Dual Simulation effectuent des traitements centrés arêtes et garantissent une scalabilité linéaire.