

DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **17 avril 2025**

Nom de famille et prénom de l'auteur. e : **Monsieur Walid MEGHERBI**

Titre de la thèse : Détection d'anomalies dans les flux de graphes

Résumé



La détection d'anomalies est devenue un domaine clé de la science des données, jouant un rôle essentiel dans l'identification des événements qui s'écartent significativement des comportements normaux. Ce champ d'étude est particulièrement crucial dans divers domaines d'application tels que la sécurité informatique, la détection de fraude, la surveillance de patients, ainsi que la maintenance prédictive, où la détection précoce des anomalies peut prévenir des conséquences potentiellement graves. Avec l'augmentation exponentielle des données structurées sous forme de graphes, la nécessité de développer des techniques avancées pour détecter les anomalies dans ces structures complexes est devenue impérative. Les graphes, par leur capacité à modéliser à la fois les entités et les relations complexes entre elles, offrent un cadre riche pour analyser et comprendre les anomalies. Cependant, cette richesse structurelle soulève des défis algorithmiques importants, notamment en raison de la diversité des types de graphes (orientés ou non, statiques ou dynamiques, etc.) et de la complexité des interactions au sein des données. Dans cette thèse, nous nous sommes intéressés à la détection d'anomalies dans les flux de graphes. Nous avons commencé par étudier les solutions existantes, que nous avons classifiées en différentes catégories en fonction de leurs approches

principales. Cette analyse de l'état de l'art nous a permis de proposer trois contributions. Dans la première contribution, nous proposons une nouvelle méthode de plongement de graphes basée sur des techniques de hachage qui a pour but la capture conjointe de la structure des graphes et de leur évolution temporelle. Cette approche permet de représenter efficacement des graphes complexes dans un espace mémoire limité. Les plongements ainsi générés sont capables de refléter la structure sous-jacente des graphes dynamiques, les rendant adaptés à des tâches spécifiques comme la détection d'anomalies de graphes. Combinée à des algorithmes de machine learning tels que l'isolation forest, cette technique permet de gérer de grands volumes de données tout en facilitant la détection d'anomalies en temps réel dans les graphes dynamiques. Dans notre deuxième approche, nous nous sommes intéressés à la détection conjointe d'anomalies de nœuds, d'arêtes et de sous-graphes. Notre approche tire partie des techniques de deep learning, plus précisément les Auto-encodeurs et les LSTMs afin de distinguer les aspects structurels et temporels des graphes, qui se complètent pour une détection plus fine. En combinant ces deux aspects, notre approche capture à la fois les motifs inhabituels dans la structure du graphe et des évolutions anormales sur la durée. Dans notre troisième contribution, nous nous sommes intéressés aux anomalies de nœuds. Nous proposons un nouveau plongement de nœuds combinant une composante obtenue par apprentissage profond et des informations de voisinage spécialement choisies pour améliorer la détection de sous-structures denses dans les graphes. En effet, les structures denses sont très importantes dans l'analyse de graphes et leur découverte constitue une étape clé dans la détection d'anomalies dues à des outliers. Nous avons implémenté, testé et comparé chacune de nos contributions. Le code correspondant est mis à disposition pour consultation et réutilisation publique.

Mots-clés : graphes, anomalies, intelligence artificielle