



Université Claude Bernard



DIPLÔME NATIONAL DE DOCTORAT

(Arrêté du 25 mai 2016)

Date de la soutenance : **24 novembre 2022**

Nom de famille et prénom de l'auteur : **Monsieur OLUIGBO Ikenna Victor**

Titre de la thèse : « *Contributions aux approches basées sur les marches aléatoires pour l'apprentissage de représentations de graphes* »



Résumé

Un graphe ou réseau est une structure mathématique qui modélise les interactions entre différentes entités. Un graphe est constitué de sommets (ou nœuds) et d'arêtes (ou liens) reliant ces sommets. En raison de l'adaptabilité des graphes, ils peuvent être utilisés pour modéliser des données complexes, ainsi que pour servir de cadre pour comprendre les entités et leurs interactions dans divers domaines tels que les réseaux sociaux, les systèmes de recommandation, les réseaux biologiques, etc. L'apprentissage de représentations de graphes ou Embedding de graphes est une technique visant à apprendre une représentation sous forme de vecteurs de faible dimension pour les entités (nœuds), la connectivité (arêtes) entre les entités, ou les groupes d'entités (sous-graphes) dans un graphe, tout en essayant de préserver les caractéristiques du graphe comme la structure, les attributs des nœuds/arêtes, etc. Les modèles d'apprentissage de représentations de graphes existants sont confrontés à certaines limitations en ce qui concerne la capture des propriétés du graphe et l'embedding de nœuds. Premièrement, la plupart de ces modèles se concentrent principalement sur l'extraction et la préservation uniquement des modèles structurels du réseau sans tenir compte des riches attributs contextuels des nœuds. Deuxièmement, de nombreux modèles existants sont confrontés à des complexités de calcul élevées en temps et en mémoire. Troisièmement, préserver efficacement les structures de voisinage des nœuds est un défi, en particulier pour les modèles utilisant des techniques de marches aléatoires mal guidées. Quatrièmement, certains modèles s'appuient sur des paramètres réglables pour échantillonner les nœuds d'un réseau, ce qui nécessite une connaissance approfondie du domaine pour déterminer les bons paramètres. À la lumière de ces défis, notre objectif principal est de proposer de nouveaux algorithmes de représentations de graphes pour apprendre efficacement des embeddings de nœuds dans un réseau en utilisant leurs informations structurelles et leurs attributs contextuels. Nous nous appuyons principalement sur des méthodes basées sur les marches aléatoires à travers quatre contributions: Dans notre première contribution, nous avons passé en revue de manière approfondie divers modèles d'apprentissage de représentations de graphes et avons évalué les performances des embeddings appris sur différentes tâches d'apprentissage comme le clustering, la détection d'anomalies, et l'analyse de prédiction de liens. Dans notre deuxième contribution, nous proposons l'utilisation d'une nouvelle technique d'encodage de voisinage de nœuds pour améliorer la précision des embeddings dans les tâches

d'apprentissage. L'idée principale de cette approche est que nous essayons d'aborder les problèmes de précision et de dimensionnalité de l'embedding de nœuds en introduisant une représentation de voisinage compacte qui capture les propriétés sémantiques et structurelles du voisinage d'un nœud dans une seule dimension. Grâce à cet encodage, nous sommes en mesure d'améliorer la précision de l'embedding pour les tâches d'apprentissage. Dans notre troisième contribution, nous avons développé un nouveau modèle de représentation de nœuds qui utilise les attributs non linéaires des nœuds ainsi que leurs informations structurelles de voisinage pour capturer les similarités entre nœuds. Pour cela, nous avons amélioré l'approche basée sur les marches aléatoires en utilisant une fonction de décision semi-supervisée pour modéliser une marche de probabilité flexible, qui explore différents espaces de voisinage pour échantillonner des nœuds partageant des caractéristiques similaires. La force de cette approche est que nous sommes capables de préserver conjointement les structures de voisinage du nœud et les attributs riches du nœud de manière simple et efficace. Pour notre quatrième contribution, nous proposons un nouveau modèle d'apprentissage de représentations de graphes pour capturer l'identité structurelle des nœuds en utilisant une métrique de probabilité de Poisson avec des scores d'estimation de kl-divergence pour échantillonner la similarité d'identité structurelle du nœud dans une marche guidée.

Nous avons implémenté toutes les approches proposées et mené une expérimentation approfondie avec divers datasets de référence dans plusieurs domaines, pour montrer l'efficacité de nos techniques pour un apprentissage efficace de représentations de nœuds.