

Contributions à une efficiente classification non-supervisée de réseaux et de graphes

(Accessit du Prix de Thèse EGC 2025)

Chakib Fettal*

* Université Paris Cité

Les graphes sont une structure de données importante utilisée dans de nombreux domaines car ils constituent un outil puissant pour la modélisation et l'analyse de systèmes complexes. Ils sont utilisés pour représenter les relations entre les entités, comme les individus dans un réseau social ou les nœuds dans un réseau informatique. Les graphes ont été utilisés dans diverses applications dans différents domaines, tels que l'analyse des réseaux sociaux, la bioinformatique, l'épidémiologie et bien d'autres encore. Dans l'analyse des réseaux sociaux, par exemple, les graphes peuvent être utilisés pour étudier les modèles d'interactions entre les individus dans un réseau social et identifier les groupes d'individus ayant des intérêts ou des comportements similaires. Cela peut être utile pour le marketing ou les recommandations ciblées.

Le partitionnement de graphes, également connu sous le nom de détection de communautés, est une technique importante dans l'analyse des données de graphes. Il permet d'identifier des groupes de nœuds similaires dans le graphe. Cela peut révéler des motifs et des structures sous-jacents dans le graphe qui ne sont pas immédiatement apparents. Par exemple, dans un réseau social, le clustering peut révéler des groupes d'individus ayant des intérêts ou des comportements similaires, et en bioinformatique, il peut révéler des modules fonctionnels dans les réseaux d'interaction protéine-protéine.

Cette thèse tente de résoudre les problèmes de scalabilité des modèles de clustering de graphes de l'état de l'art et présente des approches pour le clustering et l'apprentissage de représentation de différents types de graphes, y compris les graphes classiques, les graphes bipartis, les graphes attribués, les graphes attribués bipartis et les graphes attribués multivues. À cette fin, nous exploitons des techniques simples telles que : les projections linéaires, le lissage laplacien, le transport optimal, etc. Les approches proposées partagent toutes trois caractéristiques clés : simplicité, efficacité et peu d'hyperparamètres. Grâce à leur nature simple mais efficace, les méthodes proposées sont compétitives par rapport à l'état de l'art tout en étant généralement plus efficaces en termes de calcul. Nous démontrons l'efficacité et l'efficience de nos modèles par rapport à l'état de l'art par le biais d'une expérimentation approfondie et de tests de Significativité statistique.