

Proposition thèse

« Clustering structurel de grands graphes distribués et évolutifs »

Responsable : *Engelbert Mephu Nguifo*, LIMOS CNRS-UMR 6158

Contexte :

Les graphes sont omniprésents dans de nombreuses applications, allant du domaine des transports à celui des réseaux sociaux ou de la fouille de données. Le volume important des graphes de plus en plus collectés nécessite un traitement distribué et parallèle avec des milliers d'ordinateurs. L'exemple des grands graphes le plus populaire est le World Wide Web. Des multiples applications des grands graphes sont présentes de nos jours. En fait, les réseaux sociaux, les protéines, les réseaux routiers et autres domaines peuvent être modélisés sous forme de graphes afin d'être résolus en utilisant un ensemble des algorithmes dédiés. Les graphes offrent une abstraction flexible pour la description des relations entre les objets. La structure des graphes est également riche puisqu'elle peut montrer certaines caractéristiques difficiles à découvrir avec d'autres telles que les sous-graphes denses dans les grands graphes. Au cours des dernières années, la découverte des sous-graphes denses dans les grands graphes est devenue l'une des problématiques majeures, encore appelée clustering de graphes.

Le clustering de graphes vise principalement à détecter tous les groupes dans un graphe donné. Un tel traitement peut toucher de nombreuses applications. Par exemple, il peut être utile pour détecter des structures cachées dans un graphe. Plus précisément, dans les réseaux sociaux (par exemple Facebook), les groupes dans un graphe peuvent être considérés comme étant des communautés. Dans un réseau de collaboration (par exemple DBLP), un cluster peut être un groupe de chercheurs ayant des intérêts de recherche similaires. Alors que pour la biologie computationnelle, le calcul de grappes fonctionnelles de gènes en appliquant ce traitement peut aider les biologistes à mener une meilleure étude du micro réseaux de gènes.

On peut distinguer trois types d'algorithmes de clustering de graphes qui sont : le clustering basé sur la modularité, le clustering spectral et le clustering structurel. Nous nous intéressons au clustering structurel où plusieurs algorithmes ont été proposés, et sont malgré tout limités pour le traitement de grands graphes évolutifs.

Notre objectif est de proposer un algorithme de clustering structurel de graphes pour supporter les grands graphes évolutifs. Le travail consistera :

- Explorer la génération de grands graphes évolutifs : en effet ceci est un challenge actuellement dans la communauté scientifique, afin de pouvoir fournir des exemples de grands graphes évolutifs pour l'évaluation de modèles.
- Etudier les propriétés structurelles des graphes pour concevoir un algorithme de clustering de grands graphes évolutifs et distribués.

Références :

1. Wissem Inoubli, Sabeur Aridhi, Engelbert Mephu Nguifo, et al., 2018, An Experimental Survey on Big Data Frameworks, *Future Generation Computer Systems*, 86:546-564, Elsevier.
2. Karim Alami, Radu Ciucanu, Engelbert Mephu Nguifo. 2017. EGG: A Framework for Generating Evolving RDF Graphs. *ISWC (Posters, Demos & Industry Tracks) 2017*
3. Karim Alami, 2017, Generation and Querying Evolving Graphs, Master thesis technical report, University Clermont Auvergne, August.
4. Yollande Kamga Molle, 2018, Génération de graphes évolutifs, Rapport technique Master Informatique, Université Clermont Auvergne, Août.
5. Shiokawa, H., Fujiwara, Y., & Onizuka, M. (2015). SCAN++: efficient algorithm for finding clusters, hubs and outliers on large-scale graphs. *Proc. VLDB Endowment*, 8(11), 1178-1189
6. Wen, D., Qin, L., Zhang, Y., Chang, L., & Lin, X. (2017). Efficient structural graph clustering: an index-based approach. *Proceedings of the VLDB Endowment*, 11(3), 243-255.