

Topological Decomposition and Heuristics for High Speed Clustering of Complex Networks

Faraz Zaidi*, Guy Melançon**

*Karachi Institute of Economics and Technology (KIET)
Korangi Creek, Karachi, 75190, Pakistan
faraz@pafkiet.edu.pk

**CNRS UMR 5800 LaBRI & INRIA Bordeaux - Sud Ouest
351, cours de la Libération, 33405 Talence cedex, FRANCE
guy.melancon@labri.fr

Abstract

With the exponential growth in the size of data and networks, development of new and fast techniques to analyze and explore these networks is becoming a necessity. Moreover the emergence of scale free and small world properties in real world networks has stimulated lots of activity in the field of network analysis and data mining. Clustering remains a fundamental technique to explore and organize these networks. A challenging problem is to find a clustering algorithm that works well in terms of clustering quality and is efficient in terms of time complexity.

In this paper, we propose a fast clustering algorithm which combines some heuristics with a Topological Decomposition to obtain a clustering. The algorithm which we call Topological Decomposition and Heuristics for Clustering (TDHC) is highly efficient in terms of asymptotic time complexity as compared to other existing algorithms in the literature. We also introduce a number of Heuristics to complement the clustering algorithm which increases the speed of the clustering process maintaining the high quality of clustering. We show the effectiveness of the proposed clustering method on different real world data sets and compare its results with well known clustering algorithms.

Résumé

Avec l'accroissement exponentiel de la taille des données et des réseaux, il devient nécessaire de développer des techniques nouvelles et rapides d'analyse et d'exploration de ces réseaux. De plus, l'émergence de propriétés du petit monde et graphes sans échelle dans les réseaux du monde réel a grandement stimulé l'activité dans le domaine de l'analyse de réseau et d'exploitation des données. Le regroupement demeure une technique fondamentale pour explorer et organiser ces réseaux. La difficulté consiste à trouver un algorithme de regroupement qui fonctionne bien en termes de qualité de regroupement et qui soit efficace en termes de complexité de temps.

Topological Decomposition for High Speed Clustering

Dans cet article, nous proposons un algorithme de regroupement rapide qui combine certaines heuristiques avec une Décomposition Topologique pour obtenir un regroupement. L'algorithme que nous appelons Décomposition Topologique et Heuristiques pour Regroupement (TDHC) est très efficace en termes de complexité asymptotique de temps comparé aux autres algorithmes existant dans la littérature. Nous introduisons également un nombre d'Heuristiques pour compléter l'algorithme de regroupement qui accroît la vitesse du processus de regroupement en maintenant la haute qualité du regroupement. Nous montrons l'efficacité de la méthode de regroupement proposée sur différentes séries de données du monde réel et nous comparons ses résultats avec des algorithmes de regroupement bien connus.