

Estimation de la densité d'arcs dans les graphes de grande taille: une alternative à la détection de clusters

Marc Boullé *

* Orange Labs, 2 avenue Pierre Marzin, 22300 Lannion
marc.boulle@orange-ftgroup.com,
<http://perso.rd.francetelecom.fr/boulle/>

Résumé. La recherche de structures dans les graphes est un sujet étudié depuis longtemps, qui a bénéficié d'un regain d'intérêt avec la mise à disposition de graphes de grande taille sur le web, tels les réseaux sociaux. De nombreuses méthodes de recherche de clusters "naturels" dans les graphes ont été proposées, fondées notamment sur la modularité de Newman. On introduit dans cet article une nouvelle façon de résumer la structure des graphes de grande taille, en utilisant des estimateurs de densité des arcs exploitant des modèles en grille, basés sur un co-partitionnement des noeuds source et cible des arcs. Les structures identifiées par cette méthode vont au delà de la "classique" détection de clusters dans les graphes, et permettent d'estimer asymptotiquement la densité des arcs. Les expérimentations confirment le potentiel de l'approche, qui permet d'identifier des structures fortement informatives dans les graphes, sans faire l'hypothèse d'une décomposition en clusters denses.

1 Introduction

Le partitionnement de graphe est un sujet étudié depuis longtemps dans le domaine de la recherche opérationnelle. Une des plus ancienne approche est celle de la coupe minimale, dans laquelle les noeuds d'un graphe sont partitionnés en un nombre prédéterminé de clusters, habituellement de tailles approximativement égales, de telle façon que le nombre d'arcs inter-clusters soit minimisé. Ce problème combinatoire est intéressant dans de nombreuses applications, telles que le partitionnement de graphe de télécommunications, la conception de circuits VLSI (very large-scale integration) ou la distribution des traitements en calculs parallèles de façon à minimiser les échanges entre processeurs. Ce problème étant NP-dur (Garey et Johnson, 1979), de nombreuses heuristiques ont été proposées dans la littérature. Par exemple, l'algorithme de (Kernighan et Lin, 1970) est souvent utilisé pour améliorer localement des bi-partitionnements. De nombreuses méta-heuristiques ont également été utilisées pour le partitionnement de graphe, comme le recuit simulé (Johnson et al., 1989), les algorithmes génétiques ou la recherche tabou (Battiti et Bertossi, 1999). L'approche multi-niveau (Hendrickson et Leland, 1995) est particulièrement adaptée aux très grands graphes avec des contraintes fortes sur le temps de calcul. Ces familles d'heuristique représentent un large éventail d'options pour le compromis entre temps de calcul et qualité de la solution.