

Validation et optimisation d'une décomposition hiérarchique de graphes

François Queyroi*

*Université de Bordeaux I, CNRS, LaBRI, INRIA Bordeaux – Sud-Ouest, France
francois.queyroi@labri.fr,
<http://www.labri.fr>

Résumé. De nombreux algorithmes de fragmentation de graphes fonctionnent par agrégations ou divisions successives de sous-graphes menant à une décomposition hiérarchique du réseau étudié. Une question importante dans ce domaine est de savoir si cette hiérarchie reflète la structure du réseau ou si elle n'est qu'un artifice lié au déroulement de la procédure. Nous proposons un moyen de valider et, au besoin, d'optimiser la décomposition multi-échelle produite par ce type de méthode. On applique notre approche sur l'algorithme proposé par Blondel et al. (2008) basé sur la maximisation de la modularité. Dans ce cadre, une généralisation de cette mesure de qualité au cas multi-niveaux est introduite. Nous testons notre méthode sur des graphes aléatoires ainsi que sur des exemples réels issus de divers domaines.

1 Introduction

La compréhension de l'organisation des communautés dans les réseaux est une problématique importante dans le domaine de l'analyse de ces structures. Différents travaux (voir, entre autres, Simon (1962) et Pumain (2006)) suggèrent que ces systèmes que l'on peut qualifier de complexes se composent de groupes d'individus rassemblés à leur tour en groupes d'individus plus grands menant à une hiérarchie de communautés.

Beaucoup d'algorithmes de fragmentation de graphes tentent de reproduire ce phénomène de décompositions de groupes ou d'agrégations de groupes. Les algorithmes agglomératifs basés sur des mesures de similarité (voir Fortunato (2010)) en sont un bon exemple puisqu'ils cherchent à regrouper à chaque étape les deux ensembles d'individus les plus proches. Toutefois, la hiérarchie obtenue est rarement pertinente car chaque niveau ne correspond qu'à la division d'un seul groupe en deux sous-ensembles. Des travaux tels que ceux de Pons et Latapy (2010) permettent dans ce cadre la détection des divisions pertinentes sans toutefois évaluer globalement la hiérarchie produite.

D'autres approches produisent une hiérarchie de communautés directement "exploitable" du point de vue de l'analyse. On peut notamment citer les algorithmes de Blondel et al. (2008), Lancichinetti et al. (2011) et Rosvall et Bergstrom (2011). Dans le premier, la hiérarchie n'est pas réellement l'objectif final de la méthode mais une construction nécessaire à l'obtention du résultat (en l'occurrence un découpage simple des individus). On peut néanmoins se demander si cette hiérarchie a du sens et reflète bien la structure du réseau étudié.