

Algorithmes et structures de données pour les cœurs de communautés

Jean-Loup Guillaume

stages@complexnetworks.fr

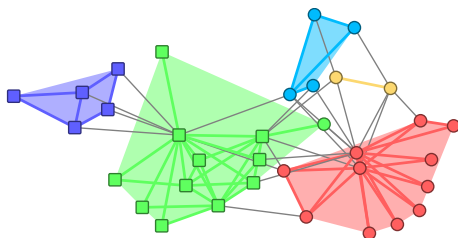
<http://complexnetworks.fr>

LIP6 – CNRS et UPMC – Paris

Dans de nombreux contextes applicatifs, on rencontre des graphes de grande taille. Citons notamment les réseaux sociaux en ligne ou réels, les graphes issus du web (pages web et liens hypertextes entre elles) ou des échanges pair-à-pair. L'analyse de ces grands graphes connaît un essor spectaculaire depuis quelques années.

Il est apparu notamment qu'ils sont composés de groupes de sommets fortement liés entre eux, appelés *communautés*, avec peu de liens vers les sommets des autres groupes. De nombreux algorithmes ont été proposés pour calculer automatiquement de telles communautés mais ils sont souvent non-déterministes et deux exécutions d'un même algorithme fournissent donc généralement des résultats différents.

La question se pose alors de savoir dans quelle mesure ces résultats sont vraiment différents et d'identifier d'éventuels points communs. Plusieurs travaux ont déjà montré qu'il y a effectivement de nombreuses similarités, en calculant un grand nombre de fois des communautés sur un graphe donné, puis en évaluant pour chaque paire de sommets (u, v) la proportion des exécutions $p(u, v)$ pour lesquelles ces deux sommets u et v ont été regroupés. Il a été montré qu'on observe souvent des valeurs de $p(u, v)$ élevées et on appelle cœur de communauté un groupe de sommets qui sont souvent mis ensemble lors des calculs de communautés.



Pour étudier en détails les similarités on souhaiterait pouvoir contruire une matrice contenant toutes ces proportions $p(u, v)$ pour toutes les paires de sommets (u, v) . Cela n'est pas faisable pour des graphes de grande taille et la méthode se trouve limitée à des graphes de quelques milliers de sommets.

L'objectif du stage est de permettre le traitement de graphes de plus grande taille et notamment d'identifier les algorithmes et structures de données qui, à partir d'un ensemble de découpages en communautés, permettent de répondre à des questions simples du type : "quels sont les paires de sommets (u, v) qui sont regroupés dans au moins $x\%$ des exécutions".

Suivant les compétences spécifiques du stagiaire et ses centres d'intérêt, l'accent pourra être orienté dans diverses directions.