
La méthode de Louvain générique : un algorithme adaptatif pour la détection de communautés sur de très grands graphes

Romain Campigotto^{*†1}, Patricia Conde Céspedes², and Jean-Loup Guillaume¹

¹Laboratoire d'Informatique de Paris 6 (LIP6) – Université Pierre et Marie Curie (UPMC) - Paris VI,
CNRS : UMR7606 – 4 Place JUSSIEU 75252 PARIS CEDEX 05, France

²Laboratoire de Statistique Théorique et Appliquée (LSTA) – Université Pierre et Marie Curie (UPMC)
- Paris VI : EA3124 – Université Pierre et Marie Curie (Paris 6) Tour 15-25 2-ième étage Boite courrier
158 4, place Jussieu 75252 Paris Cedex 05, France

Résumé

Les réseaux complexes, appelés aussi "graphes de terrain", apparaissent dans divers contextes tels que l'informatique (e.g. réseaux de pages Web), la sociologie (e.g. réseaux collaboratifs) ou la biologie (e.g. réseaux de régulation de gènes). Une propriété importante et commune à tous ces graphes est qu'ils sont généralement constitués de sous-graphes fortement interconnectés appelés "communautés".

La détection de communautés (qui peut être vue comme un problème de partitionnement de graphes) est une problématique très largement étudiée ces dernières années et, dans ce sens, beaucoup d'algorithmes ont été proposés. Bon nombre d'entre eux cherchent à maximiser la fonction de qualité de Newman-Girvan, appelée aussi "modularité" (il s'agit d'un problème NP-difficile). C'est le cas notamment de l'algorithme de Louvain, qui est actuellement le meilleur algorithme en terme de complexité pour calculer des communautés sur de très grands graphes (il est capable de traiter en moins de 3h des graphes ayant plus d'un milliard de sommets et d'arêtes).

La modularité présente toutefois plusieurs inconvénients. En la maximisant, on peut trouver des partitions de bonne qualité dans des graphes qui n'ont pas de communautés (des graphes réguliers, des graphes aléatoires, des arbres...). Elle privilégie également des communautés de grande taille.

Nous avons ainsi adapté la méthode de Louvain à d'autres critères de qualité (comme le critère de Zahn-Condorcet par exemple). nous obtenons globalement les mêmes performances en terme de temps de calcul que l'algorithme de Louvain classique. De plus, nous avons exhibé une condition suffisante pour savoir si un critère pouvait être intégré à la méthode de Louvain.

Mots-Clés: algorithmique, grands graphes de terrain, détection de communautés, partitionnement

^{*}Intervenant

[†]Auteur correspondant: romain.campigotto@lip6.fr