

Extraction de communautés ego-centrées par apprentissage supervisé d'espaces prétopologiques

Gaëtan Caillaut*, Guillaume Cleuziou*, Nicolas Dugué**

*Université d'Orléans, INSA Centre Val de Loire, LIFO EA 4022, Orléans, France
prénom.nom@univ-orleans.fr

**Le Mans Université, LIUM, EA 4023, Le Mans, France
prénom.nom@univ-lemans.fr

Résumé. Nous proposons une méthode d'extraction de communautés ego-centrées reposant sur l'apprentissage d'un modèle de propagation prétopologique. Là où les méthodes classiques ne considèrent souvent qu'un aspect de la structuration du réseau pour en extraire ses communautés, la prétopologie permet une analyse multi-critères du réseau. Notre démarche consiste à apprendre de façon supervisée un espace prétopologique défini par une combinaison logique de descripteurs du réseau. Une communauté locale à chaque nœud peut alors être extraite par une opération définie sur l'espace prétopologique appris. La qualité de chaque communauté locale est ensuite évaluée selon une communauté de référence. Nous avons comparé notre approche aux approches existantes sur des réseaux synthétiques et du réel et montrons ainsi sa pertinence.

1 Introduction

La théorie des réseaux complexes a mis en avant l'existence de propriétés communes aux réseaux modélisant des systèmes réels. En particulier, la plupart de ces réseaux possèdent une structure communautaire, i.e. une partition de l'ensemble des nœuds telle que les nœuds de chaque partie sont plus connectés entre eux qu'avec l'extérieur (Newman, 2006). Un cas typique est celui des réseaux sociaux, où les utilisateurs se regroupent autour de thèmes. La structure de communautés est particulièrement importante pour l'étude du réseau puisqu'elle permet de se placer à un niveau intermédiaire (mésoscopique) entre le niveau local (voisinage uniquement) et le niveau global (la totalité du réseau).

Une façon courante d'extraire les communautés d'un réseau consiste à trouver une partition de ses nœuds maximisant la modularité (Newman, 2006), c'est à dire qui maximise la densité des liens au sein des communautés en minimisant le nombre de liens entre communautés. Il est dans ce cas nécessaire de connaître la totalité du réseau afin d'en déterminer ses communautés. Dans le cas des très grands réseaux tels que l'internet ou les réseaux sociaux en ligne, cette condition est parfois impossible à remplir, soit parce qu'on ne connaît pas le réseau complet, soit parce qu'il est difficile de le stocker en mémoire. De plus, définir l'ensemble des communautés d'un réseau comme une partition stricte de ses nœuds est souvent éloigné de la réalité, puisque cela empêche un nœud d'appartenir à plusieurs communautés (Palla et al., 2005).