

Extraction de co-variations entre des propriétés de sommets et leur position topologique dans un graphe attribué

Adriana Prado*, Marc Plantevit**, Céline Robardet*, Jean-François Boulicaut*

*Université de Lyon, CNRS, INSA-Lyon, LIRIS UMR5205, F-69621, France

**Université de Lyon, CNRS, Université Lyon 1, LIRIS UMR5205, F-69622, France

Résumé. L'analyse de grands réseaux est très étudiée en fouille de données. Toutefois, les approches existantes proposent une analyse soit à un niveau macroscopique (étude des propriétés globales comme la distribution des degrés), soit à un niveau microscopique (extraction de sous-graphes fréquents ou denses). Nous proposons une nouvelle méthode qui effectue une analyse intermédiaire permettant de découvrir des motifs regroupant des propriétés microscopiques et macroscopiques du réseau. Ces motifs capturent des co-variations entre des propriétés numériques relatives aux sommets. Par exemple, un motif mésoscopique dans un réseau de co-auteurs peut être *plus le nombre de publications à EGC est important, plus la centralité des sommets correspondants dans le réseau l'est également*. Notre contribution est multiple. D'abord, ce travail est le premier à exploiter conjointement des propriétés locales et des propriétés topologiques. De plus, nous produisons de nouvelles avancées dans le domaine de l'extraction de co-variations en revisitant les motifs émergents dans ce contexte. Enfin, nous rapportons une analyse d'un réseau bibliographique réel issu de DBLP.

1 Introduction

De nombreux phénomènes réels peuvent être modélisés par des réseaux où les sommets représentent les entités, les arêtes représentant des relations entre elles. Des attributs sont souvent associés aux sommets, fournissant des informations supplémentaires. Ce type de données est devenu ubiquitaire. Par conséquent, permettre la découverte de connaissances dans de telles données, comme par exemple dans de grands réseaux sociaux ou biologiques, est devenu un défi pour la communauté fouille de données. Dans cet article, nous illustrons notre proposition sur un réseau de co-auteurs où les sommets décrivent les auteurs, les arêtes encodent la relation de co-publication. Les attributs rattachés aux sommets décrivent des caractéristiques des auteurs comme leur affiliation, leurs domaines d'expertise. De nombreux travaux se sont attaqués à la fouille de graphes. On peut classer les approches actuelles en deux types d'analyse : (i) celles qui considèrent le graphe à un niveau *macroscopique* se focalisant sur des caractéristiques statistiques des propriétés topologiques pour décrire de grands réseaux (voir par exemple Albert et Barabási (2000)) ; (ii) celles qui effectuent des analyses à un niveau *microscopique*, se focalisant sur l'extraction de motifs locaux comme des cliques ou des quasi-cliques (Liu et Wong (2008)), de sous-graphes fréquents dans une collection de graphes (Jiang et Pei (2009)) ;