

# 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.

## Summary

The analysis of large networks is a popular data mining task. However, in most of the graph mining approaches, the analysis is focused either on a macroscopic level (e.g., studying global properties as degree distributions) or on a microscopic level (e.g., looking for frequent subgraphs or dense subgraphs). We present an algorithm that performs an intermediate analysis of graphs by discovering patterns among microscopic and macroscopic properties over large networks. Such patterns capture co-variations between numerical vertex properties. For instance, such a pattern in a co-authorship graph could be *the higher the number of publications in EGC, the higher the centrality of the vertex within the graph*. Our contributions is threefold. First, this work is the first attempt to exploit local and topological properties over graphs. Next, we provide new insights into co-variation pattern mining. Finally, we report a case study on a large real-life co-authorship network.