

Extraction de motifs dans des graphes orientés attribués en présence d'automorphisme

Claude Pasquier^{*,**}, Frédéric Flouvat^{*}
Jérémy Sanhes^{*}, Nazha Selmaoui-Folcher^{*}

^{*}Université de Nouvelle Calédonie
PPME, BP R4, F-98851 Nouméa, Nouvelle Calédonie
{frederic.flouvat,jeremy.sanhes,nazha.selmaoui}@univ-nc.nc,
<http://ppme.univ-nc.nc>

^{**}Institut de Biologie Valrose (IBV)
UNS - CNRS UMR7277 - INSERM U1091, F-06108 Nice cedex 2
claude.pasquier@unice.fr
<http://ibv.unice.fr>

Résumé. Les graphes orientés attribués sont des graphes orientés dans lesquels les nœuds sont associés à un ensemble d'attributs. De nombreuses données, issues du monde réel, peuvent être représentées par ce type de structure, mais encore peu d'algorithmes sont capables de les traiter directement. La fouille des graphes attribués est difficile, car elle nécessite de combiner l'exploration de la structure du graphe avec l'identification d'itemsets fréquents. De plus, du fait de l'explosion combinatoire des itemsets, les isomorphismes de sous-graphes, dont la présence impacte énormément les performances des algorithmes de fouille, sont beaucoup plus nombreux que dans les graphes étiquetés.

Dans cet article, nous présentons une nouvelle méthode de fouille de données qui permet d'extraire des motifs fréquents à partir d'un ou de plusieurs graphes orientés attribués. Nous montrons comment réduire l'explosion combinatoire provoquée par les isomorphismes de sous-graphes en traitant de manière particulière les motifs automorphes.

1 Introduction

Les graphes orientés sont des structures adaptées à la modélisation d'un grand nombre de données complexes présentes dans le monde réel. Les réseaux de régulation génétique, par exemple, sont des graphes orientés où les nœuds représentent des gènes et les arcs, des relations d'inhibition ou d'activation. Le Web est modélisé par un graphe orienté dans lequel les nœuds sont des pages et les arcs des liens hypertextes. La circulation des emails dans une organisation, les graphes sociaux dans lesquels un individu peut suivre les activités d'autres personnes, les réseaux de citations dans lesquels un article cite d'autres articles sont également représentés par des graphes orientés.

Du fait de ces nombreuses applications, ces structures sont beaucoup étudiées en théorie des graphes, et, plus récemment, dans le contexte de la fouille de données. La plupart des