

Nouvelle approche de fouille de graphes AC-réduits fréquents

Brahim Douar^{*,**} Michel Liquière^{*}, Chiraz Latiri^{**}, Yahya Slimani^{**}

^{*}Equipe COCONUT, LIRMM, 161 rue Ada 34095 - Montpellier, France
{douar,liquiere}@lirmm.fr

^{**}URPAH, Département informatique, Faculté des Sciences de Tunis, Tunisie
chiraz.latiri@gnet.tn, yahya.slimani@fst.rnu.tn

Résumé. La fouille de graphes est devenue une piste de recherche intéressante et un défi réel en matière de fouille de données. Parmi les différentes familles de motifs de graphes, les graphes fréquents permettent une caractérisation intéressante des groupes de graphes, ainsi qu'une discrimination des différents graphes lors de la classification ou de la segmentation. A cause de la NP-complétude du test d'isomorphisme de sous-graphes et de l'immensité de l'espace de recherche, les algorithmes de fouille de graphes sont exponentiels en temps d'exécution et/ou occupation mémoire. Dans cet article, nous étudions un nouvel opérateur de projection polynomial nommé AC-projection basé sur une propriété clé du domaine de la programmation par contraintes, à savoir l'arc consistance. Cet opérateur est censé remplacer l'utilisation de l'isomorphisme de sous-graphes en établissant un biais sur la projection. Cette étude est suivie d'une évaluation expérimentale du pouvoir discriminant des patterns AC-réduits découverts.

1 Introduction

Avec la croissance importante du besoin d'analyser une grande masse de données structurées tels que les composés chimiques, les structures de protéines ou même les documents XML, pour n'en citer que quelques-uns, la fouille de graphes est devenue une problématique de recherche intéressante et un défi réel en matière de fouille de données. En effet, la découverte de sous-graphes fréquents est un réel challenge vu leur nombre exponentiel. A ceci s'ajoute la NP-complétude du problème d'isomorphisme d'un sous-graphe général (Garey et Johnson, 1979). Dans cet article, nous introduisons la notion de biais de projection afin de proposer un opérateur similaire à l'isomorphisme de sous-graphes mais ayant une complexité polynomiale et des contraintes relaxées. Nous utilisons ensuite cet opérateur dans un processus de fouille de graphes.

2 Fondements et travaux connexes

Nous présentons dans ce qui suit les fondements mathématiques du domaine de la fouille de graphes. Cette présentation est accompagnée d'une brève citation des travaux connexes et