

Extraction de Liens Fréquents dans les Réseaux Sociaux

Erick Stattner*, Martine Collard*

* Laboratoire LAMIA
Université des Antilles et de la Guyane, France
{estattne, mcollard}@univ-ag.fr

Résumé. Cet article présente *FLMin*, une nouvelle méthode d'extraction de motifs fréquents dans les réseaux sociaux. Contrairement aux méthodes traditionnelles qui s'intéressent uniquement aux régularités structurelles, l'originalité de notre approche réside dans sa capacité à exploiter la structure et les attributs des noeuds pour extraire des régularités, que nous appelons “*liens fréquents*”, dans les liens entre des noeuds partageant des caractéristiques communes.

1 Introduction

Les approches traditionnelles de la fouille de données reposent sur l'hypothèse implicite selon laquelle les données sont indépendantes et identiquement distribuées (*IID*). Si cette restriction s'avère être cohérente au regard du problème classique d'inférence statistique, elle ignore toutefois les relations de dépendance et de corrélation, inhérentes à de nombreux phénomènes du monde réel, qui émergent fréquemment d'interactions complexes entre des entités.

Ces dernières décennies ont ainsi vu naître la “science des réseaux” (Barabasi, 2002), une discipline qui vise à l'étude des relations entre des entités. Récemment, le domaine du “*link mining*” (Getoor et Diehl, 2005) a tenté d'appliquer les concepts du data mining aux réseaux, en s'intéressant entre autres à l'extraction de motifs. Dans ce domaine, les travaux se sont limités à l'exploitation de la structure du réseau, à travers la recherche de sous-graphes fréquents.

Cet article présente une approche nouvelle du problème de la recherche de motifs fréquents dans les réseaux sociaux. L'originalité de notre solution est qu'elle combine à la fois la structure et les attributs des noeuds, pour extraire des motifs réguliers, appelés “*liens fréquents*”, au sein des liens qui connectent des noeuds possédant des caractéristiques communes.

Ce papier est organisé en cinq sections. La Section 2 passe en revue les travaux sur l'extraction de motifs dans les réseaux et définit le problème de la recherche des liens fréquents. La Section 3 détaille et discute *FLMin*, la méthode proposée. Dans la Section 4, nous évaluons les performances de notre solution. Enfin, la Section 5 conclut et présente nos travaux futurs.

2 Contexte

2.1 État de l'art

Dans le domaine de la modélisation réseau, la définition la plus courante et la plus largement admise d'un motif est celle du “sous-graphe” (Inokuchi et al., 2000; Kuramochi et