

Défi EGC 2016

Vues Conceptuelles des Collaborations aux Conférences EGC depuis 2004: Une modélisation descriptive

Erick Stattner

Université des Antilles
erick.stattner@univ-ag.fr,
<http://www.erickstattner.com>

Résumé. Dans ce travail, nous analysons les données concernant les articles publiés à la conférence EGC. Notre objectif est d'identifier et de comprendre les tendances en matière de collaborations. Pour ce faire, nous adoptons une modélisation descriptive, à travers une approche réseau qui consiste à générer tout d'abord le réseau de collaborations des auteurs à partir des données. Nous enrichissons ensuite les noeuds de ce réseau d'une dizaine d'attributs individuels extraits à partir des données. Enfin, nous recherchons des vues conceptuelles, une approche récente de clustering de liens, qui permet de synthétiser des réseaux en mettant en évidence les ensembles d'attributs retrouvés fréquemment liés dans le réseau. Les résultats obtenus montrent les tendances existantes dans les comportements de collaborations. Dans ce papier, nous présentons ces tendances et montrons comment elles évoluent selon différents seuils d'extraction.

1 Introduction

L'étude des réseaux est devenue un des axes de recherche très actif du 21^e siècle, qu'on retrouve dans la littérature comme étant la *Science des réseaux* (Barabasi et Crandall, 2003). En effet, grâce à l'amélioration des capacités de stockage et de calculs et l'hétérogénéité des données qui sont actuellement extraites de systèmes en ligne, de plus en plus de travaux se sont intéressés à des approches qui combinent plusieurs sources d'informations et qui redéfinissent les schémas traditionnels de connaissance. C'est par exemple le cas des approches de clustering de réseaux. Alors que les méthodes traditionnelles de clustering dédiées aux réseaux ont exploité uniquement la structure du réseau pour extraire des communautés (Fortunato, 2010), les approches récentes se sont, elles, intéressées à la fois à la structure du réseau et aux attributs des noeuds pour identifier de nouveaux types de groupes (Zhou et al., 2009).

Ainsi dans ce travail, nous adoptons une approche réseau, et plus particulièrement une approche de clustering, dans le but d'analyser les données qui concernent les articles publiés à la conférence internationale francophone *Extraction et Gestion des Connaissances (EGC)* de 2004 à 2015. Notre objectif est de mettre en évidence les tendances en matière de collaboration en recherchant et en identifiant des régularités cachées dans la co-écriture des articles.

Pour ce faire, nous adoptons une modélisation descriptive qui s'appuie sur une méthodologie à trois niveaux. (i) Nous générons, après pré-traitement des données, le réseau de collaborations des auteurs à partir des articles publiés. (ii) Les noeuds du réseau ainsi obtenu sont ensuite enrichis d'une dizaine d'attributs individuels qui viennent caractériser chaque auteur et qui sont générés à partir de la fréquence