

# Recherche de communautés dans les grands réseaux sociaux

Emmanuel Viennet

Université de Paris-Nord, L2TI - Institut Galilée  
99, avenue Jean-Baptiste Clément  
93430 Villetaneuse, France  
emmanuel.viennet@univ-paris13.fr

**Résumé.** Cet article décrit quelques méthodes récentes pour la recherche de communautés dans les grands réseaux sociaux (dizaines voire centaines de millions de nœuds). Après avoir rappelé quelques notions de base sur ce sujet, nous décrivons quelques approches récentes pour l'extraction de micro-communautés et de communautés globales, et montrons quelques résultats prouvant que ces méthodes sont parfaitement utilisables pour la fouille d'ensembles de données parmi les plus grands rencontrés aujourd'hui dans les applications industrielles.

## 1 Introduction

Un “réseau social” est un graphe dont les nœuds sont des individus ou organisations, connectés par des liens représentant une relation “sociale” : appartenance à la même famille, échange de messages, goûts communs... (voir l'article de P. Kuntz et F. Picarougne dans ce même numéro). L'étude des réseaux sociaux est très active depuis quelques années (voir par exemple Barabasi (2002)), et les techniques automatiques permettent d'étudier les propriétés statistiques de réseaux de très grandes tailles, comme celui formé par les sites web de l'Internet ou l'ensemble des appels téléphoniques passés sur un opérateur de Télécommunications. L'étude des réseaux sociaux intéresse depuis quelques décennies les chercheurs en sciences sociales (voir par exemple Peter J. Carrington (2005)) et a réuni depuis la fin des années 90 une importante communauté de chercheurs d'horizons divers, attirés tant par la découverte d'intéressantes propriétés théoriques de ces structures que par la richesse des applications potentielles.

La connaissance des liens entre individus (étude de la formation de communautés, de la propagation des rumeurs ou modes au sein de celles-ci, etc.) est d'une grande importance pour la fouille de données : les applications sont nombreuses en marketing, en bioinformatique, en analyse de données textuelles.

Les premiers travaux sur les réseaux sociaux ont cherché à caractériser ceux-ci (types de structures) et à décrire des classes de nœuds : individus “influents” (*hubs*) ou “suiveurs”, etc., puis à développer des outils collaboratifs exploitant ces structures. L'analyse des réseaux sociaux rejoint aussi les préoccupations de la communauté des chercheurs en *Link Analysis* (voir par exemple Workshop on Link Analysis (2006)), qui s'intéressent à des problèmes dans lesquels les données sont hétérogènes et arrivent de sources variées, incluant des représentations