

Les capitalistes sociaux sur Twitter : détection via des mesures de similarité

Nicolas Dugué*, Anthony Perez*

* LIFO - Université d'Orléans

Résumé. Les réseaux sociaux tels que Twitter font partie du phénomène de *Déluge des données*, expression utilisée pour décrire l'apparition de données de plus en plus volumineuses et complexes. Pour représenter ces réseaux, des graphes orientés sont souvent utilisés. Dans cet article, nous nous focalisons sur deux aspects de l'analyse du réseau social de Twitter. En premier lieu, notre but est de trouver une méthode efficace et haut niveau pour stocker et manipuler le graphe du réseau social en utilisant des ressources informatiques raisonnables. Cet axe de recherche constitue un enjeu majeur puisqu'il est ainsi possible de traiter des graphes à échelle réelle sur des machines potentiellement accessibles par tous. Ensuite, nous étudions les *capitalistes sociaux*, un type particulier d'utilisateurs de Twitter observé par Ghosh et al. (2012). Nous proposons une méthode pour détecter et classifier efficacement ces utilisateurs.

1 Introduction

Contexte. Depuis plusieurs années, dans les secteurs de l'Internet, de l'analyse décisionnelle ou encore de la génétique sont collectées et analysées des données de plus en plus volumineuses et complexes. Ce phénomène connu sous le nom de *Déluge des données* (ou *Big Data*) soulève de nombreuses problématiques. En particulier, être capable de stocker, partager et analyser de telles quantités de données constitue un enjeu d'étude essentiel, comme le soulignent Schuett et Pierre (2012). La théorie des graphes est particulièrement appropriée pour étudier les réseaux sociaux, où les connexions entre utilisateurs peuvent facilement être représentées et analysées en utilisant des graphes, le plus souvent orientés. Dans cet article, nous considérons le *graphe des relations entre utilisateurs (anonymisés) de Twitter* créé en 2009 par Cha et al. (2010). Ce graphe orienté contient plus de 50 millions de sommets et près de 2 milliards d'arcs.

Les capitalistes sociaux. Nous nous intéressons particulièrement au comportement d'utilisateurs particuliers nommés *capitalistes sociaux*, observé par Ghosh et al. (2012). Ces utilisateurs, qui ne sont ni des spammeurs, ni des robots, partagent un objectif commun : acquérir un nombre maximum d'utilisateurs qui les suivent -*followers*. En effet, plus le nombre de *followers* d'un utilisateur est élevé, plus il peut être influent sur le réseau. Au-delà de cet intérêt évident, le nombre de *followers* a un impact direct sur le classement des *tweets* de l'utilisateur sur le moteur de recherche de Twitter. Ces utilisateurs ne sont pas sains pour un réseau social : en suivant des utilisateurs sans regarder le contenu de leurs tweets, les capitalistes sociaux donnent de l'influence à des utilisateurs tels que les spammeurs.