

Une approche multidimensionnelle basée sur les comportements individuels pour la prédiction de la diffusion de l'information sur Twitter

Adrien Guille*, Hakim Hacid**, Cécile Favre*

*ERIC, Université Lumière Lyon 2, Lyon, France
{prénom.nom}@univ-lyon2.fr

**Bell Labs France, Alcatel-Lucent, Nozay, France
hakim.hacid@alcatel-lucent.com

Résumé. Aujourd'hui, les réseaux sociaux en ligne sont devenus des outils très puissants de propagation de l'information. Ils favorisent la diffusion rapide à grande échelle de contenu et les conséquences d'une information inexacte voire fausse peuvent alors prendre une ampleur considérable. Par conséquent il devient indispensable de proposer des moyens d'analyser le phénomène de diffusion de l'information dans ces réseaux. De nombreuses études récentes ont traité de la modélisation du processus de diffusion de l'information, essentiellement d'un point de vue topologique et dans une perspective théorique, mais les facteurs impliqués sont encore méconnus. Nous proposons ici une solution pratique dont l'objectif est de prédire la dynamique temporelle de la diffusion au sein de Twitter, basée sur des techniques d'apprentissage automatique. Notre approche repose sur l'inférence de probabilités de diffusion tirées d'une analyse multidimensionnelle des comportements individuels. Les expérimentations menées montrent l'intérêt de la modélisation proposée.

1 Introduction

Les réseaux sociaux en ligne sont des plateformes où l'information est publiée et mise à jour rapidement par des commentaires, des réponses, des transferts, etc. Par conséquent, l'information se déplace, d'un nœud à un autre du réseau, d'une communauté à une autre, etc. C'est le phénomène bien connu de diffusion, ou propagation de l'information, qui a suscité et suscite encore un grand intérêt au sein de la communauté de recherche (Bakshy et al., 2011; Kwak et al., 2010; Saito et al., 2011; Yang et Leskovec, 2010).

Par le passé, la propagation a surtout été étudiée dans le domaine de l'épidémiologie, afin de comprendre le déroulement du processus de diffusion d'un virus selon certaines conditions. Ainsi, la majorité des travaux actuels portant sur la propagation de l'information s'appuient largement sur ces travaux. Mais avec l'émergence des réseaux sociaux en ligne, les mécanismes de diffusion de l'information se sont complexifiés, en raison des spécificités suivantes : (i) la grande ampleur des réseaux, (ii) l'importante diversité des profils des utilisateurs et (iii) la particularité des lois régissant ces structures. Par conséquent, les modèles traditionnels apparaissent comme inefficaces voire dénués de sens. En réponse, nous proposons ici une solution