

Évaluation comparative d’algorithmes de centralité pour la détection d’influenceurs

Kévin Deturck*
Damien Nouvel**
Frédérique Segond***

*Viseo Technologies 4, avenue Doyen Louis Weil 38000 Grenoble
INaLCO-ERTIM 2, rue de Lille 75007 Paris
kevin.deturck@viseo.com

**INaLCO-ERTIM 2, rue de Lille 75007 Paris
damien.nouvel@inalco.fr

***Viseo Technologies 4, avenue Doyen Louis Weil 38000 Grenoble
frederique.segond@viseo.com

L’influence, d’un point de vue social, peut être définie comme le pouvoir d’un individu qui mobilise des individus cibles pour des actions concrètes ou une opinion donnée. Détecter automatiquement les influenceurs dans les réseaux sociaux fournit des points d’entrée efficaces afin d’avoir un impact dans le cadre d’applications comme la diffusion ciblée d’une information de santé publique, la promotion d’un produit ou encore la réputation en ligne. Nos travaux s’intègrent au projet européen SOMA¹ qui a pour but d’enrichir les connaissances client de systèmes CRM² par de l’information issue de médias sociaux ; l’influence en fait partie afin d’évaluer l’impact potentiel d’un client par exemple sur un produit donné.

Il est possible d’analyser un réseau social selon sa structure ou le contenu diffusé, ces deux aspects étant les marqueurs de relations interpersonnelles essentielles à l’influence. Les approches qui analysent le contenu s’intéressent au texte en tenant compte de marqueurs discursifs, comme l’argumentation, qui tendent à influencer (Rosenthal et al., 2012). La structure du réseau est utilisée par les mesures de centralité, qui identifient les nœuds dominants d’un réseau d’après leurs liens. Elles sont beaucoup utilisées parce qu’elles ne requièrent principalement qu’un graphe d’interactions et sont suffisamment variées pour modéliser différentes modalités d’influence (Benyahia et Largeron, 2015). Cette variété requiert de pouvoir appréhender les meilleurs cas d’usage des différentes mesures en les évaluant et en les comparant notamment sur des données aux applications diverses (Ghazzali et Ouellet, 2017).

Nous voulons déterminer les mesures de centralité les plus aptes à identifier les influenceurs en les confrontant à des données réelles portant différentes catégories d’information sur un ensemble d’utilisateurs. Nous utilisons un corpus constitué dans le cadre de la compétition RepLab 2014 (Amigó et al., 2014) qui présente plus de 7000 comptes Twitter³ manuellement annotés selon qu’ils sont influenceurs ou non. Nous avons sélectionné six algorithmes qui sont le *Degré entrant* comme Baseline, l’*Intermédiation* qui mesure si un individu fait office de

1. <http://www.somaproject.eu/soma-project/>

2. « Customer Relationship Management »

3. <http://www.twitter.com>