

Extraction de connaissances dans les réseaux sociaux: Structure, Volume, Dynamique et Influence

Erick STATTNER

La capacité qu'ont aujourd'hui les médias sociaux à agréger les écrits, les opinions, les sentiments et les pensées de l'humanité en font des outils puissants pour comprendre l'inconscient collectif et identifier les tendances émergentes. En effet, il est particulièrement surprenant d'observer comment ces dernières années les données générées par les sites d'échange et de partage ont été exploitées pour aborder des phénomènes de nature très différente. De nombreux travaux ont par exemple montré qu'un réseau social tel que Twitter est un extraordinaire capteur de comportements sociaux, qui renferme en son sein des indicateurs qui, une fois collectés, traités et fouillés, permettent d'aborder des questions très larges telles que l'incidence d'une épidémie dans une population, le résultat d'une élection, l'évolution des indices de marché boursier, la prédiction du revenu d'un film au box-office, ou même la façon dont des comportements se répandent.

Si ces études frappent par la précision des résultats obtenus, elles impressionnent également par la simplicité des données qui sont utilisées pour les réaliser. La plupart de ces approches n'exploitent en effet que la structure du réseau des utilisateurs et, parfois, les informations qui sont partagées par ces derniers. Elles se fondent en réalité sur l'idée émergente, et en rupture avec les approches traditionnelles, que les interactions entre les entités sont des facteurs beaucoup plus pertinents que les entités elles-mêmes, pour aborder les phénomènes qui nous entourent. C'est donc dans ce contexte que nous avons vu émerger une multitude de travaux menés sur les réseaux sociaux, regroupés au sein d'un domaine qu'il est aujourd'hui convenu d'appeler la ``*nouvelle science des réseaux*``; "*nouvelle*" faisant ici référence au fait que les approches proposées dépassent les travaux historiques sur les réseaux qui découlaient de la théorie des graphes.

Dans ce mémoire, nous abordons plusieurs problématiques autour de la science des réseaux, et en particulier autour des problématiques de clustering.

(i) La limitation de la structure, où nous proposons une approche de clustering de liens qui tient compte de la structure du réseau, mais également des attributs des noeuds.

(ii) Le passage à l'échelle, où nous présentons une amélioration de l'algorithme d'extraction des clusters de liens.

(iii) La prise en compte de la dynamique, où nous nous intéressons aux transitions qui surviennent sur les clusters entre deux états du réseau.

(iv) L'implication des clusters, où nous étudions les comportements de diffusion sous le prisme du clustering.