

Étude et modélisation de la propagation d'une rumeur dans un réseau social

Les réseaux sociaux jouent un rôle majeur dans la diffusion de l'information. Ainsi les « fake news », omniprésentes sur les réseaux, peuvent avoir d'importantes conséquences sur la société. J'ai donc décidé de m'intéresser à la manière dont ces rumeurs se propagent afin de mieux comprendre comment les neutraliser.

Une rumeur qui se propage sur un réseau social est une information qui est transmise entre les utilisateurs du réseau : cette étude concerne donc le transport de l'information.

Positionnement thématique (phase 2)

INFORMATIQUE (Informatique Théorique), INFORMATIQUE (Informatique pratique), MATHEMATIQUES (Mathématiques Appliquées).

Mots-clés (phase 2)

Mots-Clés (en français)	Mots-Clés (en anglais)
<i>Diffusion d'une rumeur</i>	<i>Rumor propagation</i>
<i>Maximisation d'influence</i>	<i>Influence maximization</i>
<i>Fonctions sous-modulaires</i>	<i>Submodular functions</i>
<i>Optimisation combinatoire</i>	<i>Combinatorial optimization</i>
<i>Bouche à oreille</i>	<i>Word of mouth</i>

Bibliographie commentée

La communication est un des piliers de notre société et elle s'est grandement développée avec l'apparition d'Internet qui nous a fait entrer dans ce que l'on appelle l'"ère de l'information".

L'information se transmet de diverses manières comme par la télévision ou les journaux, ou bien plus directement par le "bouche à oreille" qui reste le moyen le plus efficace pour faire connaître une idée ou un produit [4] , ce qui en fait un outil puissant.

Des modèles ont donc été créés afin de mieux comprendre comment fonctionne ce phénomène et comment prédire ses effets.

De nos jours on dit qu'une information devient "virale" lorsque elle est très largement diffusée, de la même manière qu'une maladie se propage. Cette expression est utilisée à juste titre puisque c'est sur cette idée que se sont basées les premières études sur le sujet. Ainsi en 1964 Goffman et Newill établissent une analogie entre une épidémie et la propagation d'une idée au sein d'un réseau social [1] et proposent d'adapter au problème le modèle épidémiologique SIR caractérisant l'évolution de la population infectée par des équations différentielles.

Notons que le concept de réseau social se réfère à toute structure reliant les personnes entre elles, que ce soit sur un site Internet (Twitter, Facebook...) ou autres (relations d'amitié, famille, collègues de travail etc).

Un inconvénient du modèle SIR est qu'il ne prend pas en compte la structure des relations entre les individus, ce qui a poussé des chercheurs à proposer ensuite des alternatives modélisant un réseau social par un graphe. Le fait de considérer les relations entre chaque personne fait apparaître d'autres problématiques et notamment celle de la maximisation d'influence qui consiste à trouver le groupe de personnes qui va mener à la plus grande diffusion de l'information. Plus formellement, on définit la fonction d'influence comme la fonction qui à un ensemble de diffuseurs initiaux associe le nombre total de personnes atteintes. L'objectif devient alors de trouver un ensemble de personnes maximisant cette fonction. Se pose alors la question du modèle à utiliser : les chercheurs D.Kempe, J.Kleinberg et E.Tardos s'appuient sur les modèles IC (Independent Cascade) et LT (Linear Threshold) dans leur étude du problème de maximisation d'influence [2]. Ils prouvent alors qu'il est NP-difficile en montrant que les problèmes de couverture par sommets [5] et de couverture maximale [6] en sont des cas particuliers.

Il devient alors nécessaire de disposer d'algorithmes approchant la solution optimale et pour cela un algorithme glouton est une bonne solution. En effet, la fonction d'influence possède une propriété intéressante nommée sous-modularité [2] qui permet d'appliquer des résultats d'optimisation combinatoire garantissant une certaine efficacité de cet algorithme [3].

Problématique retenue

Il s'agit ici d'étudier le problème de maximisation de l'influence, c'est à dire la question suivante : Quelles personnes doivent lancer une rumeur pour que celle-ci ait le plus d'impact possible ? Plus formellement, comment trouver un ensemble qui maximise la fonction d'influence ?

Objectifs du TIPE

- 1) Modéliser informatiquement les mécanismes de diffusion d'une rumeur sur un réseau social à l'aide de graphes
- 2) Étudier un algorithme de recherche des noeuds (personnes) optimaux et notamment établir des propriétés d'optimisation combinatoire en utilisant le caractère sous-modulaire de la fonction d'influence
- 3) Implémenter cet algorithme sur des réseaux réels afin de comparer la théorie à l'expérience

Références bibliographiques (phase 2)

- [1] W. GOFFMAN, V. A. NEWILL : Generalization of epidemic theory: an application to the transmission of ideas : *Nature*, vol. 204, no. 4955, pp. 225–228, 1964.
- [2] D. KEMPE, J. KLEINBERG, É. TARDOS : Maximizing the Spread of Influence through a Social Network : <https://theoryofcomputing.org/articles/v011a004/v011a004.pdf>
- [3] D. ROLNICK, J. WEED : Greedy Maximization of Submodular Functions : <https://pdfs.semanticscholar.org/e047/d5ba349e746f5d8b38f2e3b4b187ba82cc21.pdf>
- [4] <http://www.leparisien.fr/economie/consommation-le-bouche-a-oreille-plus-fort-que-les-reseaux-sociaux-10-09-2015-5078693.php>
- [5] https://fr.wikipedia.org/wiki/Probl%C3%A8me_de_couverture_par_sommets
- [6] https://en.wikipedia.org/wiki/Maximum_coverage_problem

