

## Modélisation de processus de diffusion avec données incomplètes

**Encadrant :** Sylvain Lamprier – Sylvain.Lamprier@lip6.fr

La diffusion d'information dans les réseaux sociaux est un sujet important pour de nombreux acteurs. La modélisation de ce genre de processus est en effet crucial pour comprendre les dynamiques d'échanges sur les réseaux et être à même d'agir sur celui-ci (tâche de recommandation de produits, maximisation d'influence, coupure de rumeur, prédiction de comportements, etc.). De nombreux travaux se sont alors intéressés à l'extraction des relations de diffusion dans les réseaux. La plupart des modèles proposés considère néanmoins que la totalité des épisodes de diffusion sont connus : si un nœud n'est pas observé comme appartenant à un épisode de diffusion, il est considéré comme n'ayant pas été impacté par le contenu correspondant. Cependant, dans de nombreux contextes il n'est pas possible d'observer la totalité de l'activité du réseau et il se peut qu'un nœud, bien qu'ayant réagi à un contenu donné, n'est pas été enregistré comme tel. Il ne faudrait pas alors le considérer comme un exemple négatif de diffusion lors de l'apprentissage du modèle. Dans ce projet, nous proposons de mettre en œuvre des modèles classiques de diffusion (modèles de dépendances bayésiennes type IC [Saito et al., 2008] ou NetRate [Gomez-Rodriguez et al., 2012]), dans lesquels on inclura un mécanisme permettant de faire des suppositions sur les nœuds absents des épisodes d'apprentissage. Quelques modèles existants ont proposé cette inclusion. Il s'agit de générer pour chaque épisode d'entraînement de multiples épisodes complétés par Gibbs Sampling (méthode d'échantillonnage type MCMC) à partir des estimations de probabilité courantes [Wu et al., 2013] pour IC et [Shaghaghian & Coates, 2016] pour NetRate. Les paramètres sont alors mis à jour selon ces épisodes complétés, puis la procédure est répétée jusqu'à convergence de l'algorithme.

Le travail à réaliser s'articule autour de différents points :

- État de l'art sur les modèles diffusion;
- Mise en œuvre des algorithmes décrits dans [Wu et al., 2013] et [Shaghaghian & Coates, 2016] pour l'apprentissage de réseaux de diffusion avec données non-observées ;
- Expérimentations et analyse des résultats ;
- Propositions et mise en œuvre éventuelles d'extensions aux algorithmes existants.

### Refs :

[Gomez-Rodriguez et al., 2012] M. Gomez-Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influence," *ACM Trans. on Knowl. Discov. from Data (TKDD)*, vol. 5, no. 4, p. 21, 2012.

[Saito et al., 2008] K. Saito, R. Nakano, and M. Kimura. *Prediction of information diffusion probabilities for independent cascade model. In Proceedings of the 12th international conference on Knowledge-Based Intelligent Information and Engineering Systems, Part III, KES '08, pages 67–75. Springer-Verlag, 2008.*

[Shaghaghian & Coates, 2016] Shaghaghian, Shohreh & Coates, Mark. (2016). *Bayesian Inference of Diffusion Networks with Unknown Infection Times.*

[Wu et al., 2013] Xiaojian Wu, Akshat Kumar, Daniel Sheldon and Shlomo Zilberstein. Parameter Learning for Latent Network Diffusion. IJCAI'13. 2013.