

## ÉLÉMENT DE PORTFOLIO 02



### Publication

## 1 DÉFINITION DE CET ÉLÉMENT

### Titre de l'élément :

An Homophily-based Approach for Fast Post Recommendation in Microblogging Systems

### Fichier de élément :

<https://hal.sorbonne-universite.fr/hal-01679120v1>

## 2 MOTIVATIONS DU CHOIX DE CET ÉLÉMENT

Cet travail est le résultat d'une collaboration avec le laboratoire CEDRIC du CNAM Paris dans le cadre de la thèse de Quentin Grossetti [2]. Il a donné lieu à une publication dans la conférence EDBT 2018 [3]. Ce travail a été poursuivi avec la thèse de Jonathan Debure [1], et celle de Yifan Li [4], également co-encadrées avec le CEDRIC, pour analyser le comportement des utilisateurs dans les réseaux sociaux et réaliser le passage à l'échelle d'algorithmes de calculs utilisés dans les systèmes de recommandation.

## 3 PRÉSENTATION DE CET ÉLÉMENT

Les services de micro-blogging sont devenus récemment une source d'information importante. Cependant, victimes de leur succès, ils doivent actuellement gérer une quantité sans précédent d'informations générées par les utilisateurs. Il devient par conséquent difficile pour les utilisateurs de trouver dans ces services des contenus proches de leurs intérêts. Nous nous sommes intéressés dans ce travail au problème de la recommandation des utilisateurs à suivre sur un sujet donné. Pour cela nous avons défini des scores basés sur la topologie du graphe social et sur le contenu textuel des micro-blogs. Après une étude approfondie d'un large jeu de donnée issu de Twitter, nous nous appuyons sur ces conclusions concernant le fonctionnement et les usages de Twitter pour définir un modèle de propagation qui repose sur l'homophilie présente dans le réseau pour recommander des messages aux utilisateurs. Nous avons montré que la durée de vie des messages était extrêmement courte et qu'il était par conséquent nécessaire de recommander des messages rapidement après leur publication. Notre approche s'appuie sur la construction d'un graphe de similarités lié aux interactions des utilisateurs. Ces similarités sont cependant assez faibles compte tenu du faible nombre d'interactions en commun entre utilisateurs, il est donc crucial d'être en mesure de surmonter ce manque d'information. Nous construisons un réseau de similarités permettant de mettre à profit facilement la transitivité entre utilisateurs. Cette transitivité est la clé pour recommander des messages sur lesquels on ne dispose que de peu d'interactions. Notre méthode de recommandation est capable de recommander rapidement des messages après leur publication et de passer à l'échelle. Pour ce faire, nous nous appuyons sur le pré-calcul de recommandations pour certains comptes choisis dans le graphe.

Nous présentons plusieurs expérimentations pour démontrer la qualité de prédiction de notre modèle et sa capacité à passer à l'échelle. En comparant les résultats avec d'autres modèles de recommandation nous pouvons analyser le fonctionnement standard de ces méthodes avec leurs forces et leurs faiblesses. Certaines méthodes ne demandent aucun travail d'initialisation et peuvent facilement se mettre à jour mais ont tendance à recommander des messages populaires du fait des marches aléatoires. Certaines méthodes sont capables de fournir un nombre très élevé de recommandations aux utilisateurs mais ne disposent pas d'un excellent ratio précision/rappel, signe que les messages recommandés ne sont pas toujours pertinents. En revanche, le coût de calcul d'initialisation est trop élevé et ne permet pas d'envisager ce modèle de façon réaliste pour une plateforme comme Twitter. D'autres méthodes ont des coûts d'initialisation sont négligeables tandis que le coût de propagation est très élevé, ce qui limite également l'utilisation de ce type de modèle avec de larges plateformes de micro-blogging. Nous montrons également que les scores calculés par notre méthode peuvent être mis à jour à faible coût tout en proposant des recommandations pertinentes.

## 4 RÉFÉRENCES BIBLIOGRAPHIQUES

- [1] Jonathan Debure. *Détection de comportements et identification de rôles dans les réseaux sociaux*. Theses, Conservatoire national des arts et metiers - CNAM, July 2021.
- [2] Quentin Grossetti. *Système de recommandation sur les plateformes de micro-blogging et bulles filtrantes*. Theses, Sorbonne Université, November 2018.
- [3] Quentin Grossetti, Camelia Constantin, Cédric Du Mouza, and Nicolas Travers. An Homophily-based Approach for Fast Post Recommendation in Microblogging Systems. In *21st International Conference on Extending Database Technology (EDBT 2018)*, pages 229–240, Vienne, Austria, March 2018.
- [4] Yifan Li. *Edge partitioning of large graphs*. Theses, Université Pierre et Marie Curie - Paris VI, December 2017.