

# Projet de recherche doctoral

## Diversité dans les systèmes de recommandation sociaux

### Contexte

Les systèmes de recommandation, apparus dans les années 1990, visent à aider les utilisateurs d'une plateforme à identifier ce qu'ils pourraient apprécier, qu'il s'agisse de contenus culturels, d'« articles » voire de personnes. Ils facilitent la navigation et évitent aux utilisateurs d'être perdus devant une quantité d'information trop importante. Ces systèmes sont aujourd'hui devenus incontournables pour la lecture d'actualités, la musique en ligne, la vidéo à la demande, les achats en ligne, la recherche d'information, le recrutement, la recherche d'emploi, etc.

Les techniques employées ont évolué avec la forte augmentation du volume de données disponibles et des ressources de calcul exploitables. S'il est toujours possible de déceler les aspects filtrage par le contenu et filtrage collaboratif qui caractérisaient les principales approches de la recommandation, les systèmes de l'état de l'art combinent en général plusieurs sources d'information et emploient souvent des modèles décisionnels obtenus par apprentissage statistique. Des modèles reposant sur des approches orientées "graphes" ont aussi été proposées, soit comme variantes de systèmes existants [1], soit en tant que telles [6].

L'évaluation des systèmes de recommandation a aussi beaucoup évolué au cours des dernières années. Les indicateurs de performances ont longtemps été la précision et le rappel (couplé dans *l'accuracy*), ou la RMSE (l'erreur de prédiction des notes). Mais ces indicateurs, quoique efficaces, présentent des biais ou des inconvénients, à l'échelle individuelle ("bulles de filtre") ou collective (création de blockbuster, délaissement d'items). Les études de la diversité des dernières années tendent à apporter d'autres mesures [2], pour contrer ces effets. Dans une perspective plus large du domaine, il s'agit de proposer des modèles d'IA plus explicables, ce qui constitue une demande sociale importante mais a aussi fait ses preuves en terme d'efficacité [8].

Les travaux menés dans le cadre du projet ANR AlgoDiv, sur la diversité des systèmes de recommandation, ont permis d'explorer avec des modèles graphes ces indicateurs et leurs effets. Plusieurs publications sont en cours de relecture [4, 3] et la thèse s'inscrit dans la continuité directe de ces travaux.

### Description du sujet

Cette thèse s'intéresse à la question large de la diversité dans les systèmes de recommandation. L'approche de ceux-ci est centrée sur des modèles orientés "graphes", qu'il s'agisse de modèles standard ou d'*Heterogeneous Information Networks* voire de *Link streams*.

**L'objectif central est d'explorer plusieurs indicateurs de diversité et d'aboutir à une compréhension fine de l'impact de la topologie du graphe sous-jacent sur chacun de ces indicateurs.**

Le travail à mener comporte un aspect formel et un aspect expérimental, les deux étant parfaitement complémentaires. La partie théorique suppose d'acquérir une maîtrise fine de plusieurs modèles de graphes, pour choisir le plus approprié pour répondre à une question donnée. Des approches reposant sur le mélange aléatoire d'arêtes (*configuration model*) sont par exemple des pistes intéressantes. La partie expérimentale procédera par l'implémentation des modèles conçus et la confrontation de ceux-ci à des données. Ces dernières pourront être issues de jeux de données standards (MovieLens) sur lesquels des processus de recommandation

seront simulés, mais la création de données synthétiques adaptées est aussi envisageable.

En fonction des résultats sur la première partie, la thèse pourra ensuite explorer des questions complexes complémentaires du domaine :

- la *question du temps pour les données utilisées* : de nombreux algorithmes de recommandation considèrent encore de façon identique une appréciation donnée il y a 5 ans et une appréciation du mois dernier. Quelques modèles tirent réellement profit de l'incorporation d'information sur cette temporalité (voir par exemple [5]).
- le lien entre ces techniques et les approches d'apprentissage profond pour la recommandation, éventuellement reposant sur des Graph Neural Networks [7]

## Encadrement & équipe d'accueil

L'encadrement est assuré par :

- Lionel Tabourier, maître de conférences au LIP6, HDR
- Raphaël Fournier-S'niehotta, maître de conférences au Cnam Paris

L'équipe ComplexNetworks du LIP6 (<http://www.complexnetworks.fr>) est spécialisée depuis de nombreuses années dans l'étude de grands graphes de terrain, et les compétences de ses membres se sont élargies aux domaines connexes comme les systèmes de recommandation.

## Profil recherché

La candidate ou le candidat doit avoir de solides bases en mathématiques et en théorie des graphes. La maîtrise de bibliothèques de programmation permettant de traiter de grands graphes ou implémentant des systèmes de recommandation est essentielle.

## Références

- [1] Christian Desrosiers and George Karypis. A comprehensive survey of neighborhood-based recommendation methods. In *Recommender systems handbook*, pages 107–144. Springer, 2011.
- [2] Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems—a survey. *Knowledge-Based Systems*, 123 :154–162, 2017.
- [3] Pedro Ramaciotti Morales, Raphael Fournier-S'niehotta, and Lionel Tabourier. The role of the structure of heterogeneous information networks in the accuracy and diversity of recommendations, 2020. Under review.
- [4] Pedro Ramaciotti Morales, Robin Lamarche-Perrin, Raphael Fournier-S'niehotta, Remy Poulain, Lionel Tabourier, and Fabien Tarissan. Measuring diversity in heterogeneous information networks. <https://arxiv.org/abs/2001.01296>, 2020. Under review.
- [5] Peter Romov and Evgeny Sokolov. Recsys challenge 2015 : ensemble learning with categorical features. In *Proceedings of the 2015 International ACM Recommender Systems Challenge*, page 1. ACM, 2015.
- [6] Jiliang Tang, Xia Hu, and Huan Liu. Social recommendation : a review. *Social Network Analysis and Mining*, 3(4) :1113–1133, 2013.

- [7] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec. Graph convolutional neural networks for web-scale recommender systems. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Jul 2018.
- [8] Tao Zhou, Zoltán Kuscsik, Jian-Guo Liu, Matúš Medo, Joseph Rushton Wakeling, and Yi-Cheng Zhang. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, 107(10) :4511–4515, 2010.