

Caractérisation de traces de navigation

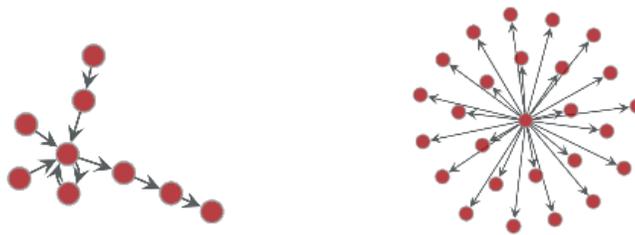
encadrement : Lionel Tabourier, Pedro Ramaciotti (LIP6 – CNRS / SU)
Christophe Prieur (Télécom Paris Tech)

lieu : Sorbonne Université – 4 place Jussieu, Paris

stages@complexnetworks.fr, <http://complexnetworks.fr>

Le stage s’inscrit dans le cadre du projet ANR Algodiv, réunissant informaticiens et sociologues. L’objectif d’Algodiv est de définir et mesurer la notion de diversité dans les algorithmes de filtrage et de classement de l’information sur le web. En effet, ces algorithmes visent en premier lieu à répondre le plus efficacement possible aux besoins immédiats des utilisateurs. Cependant, leurs éventuels effets collatéraux sont depuis longtemps l’objet de débats, comme en témoigne la large couverture médiatique, dans les derniers mois, de l’effet dit de bulle informationnelle (*filter bubble*, [Par11]), désignant une possible tendance à confiner un individu dans une gamme restreinte de choix, déterminée par ses choix passés. Que cette tendance soit ou non effective, par exemple sur des plateformes comme Facebook ou les moteurs de recherche qui pratiquent la personnalisation des résultats, ou que le confinement informationnel soit simplement un effet de l’homophilie, mesurée dans de nombreux contextes ([MSLC01]), faire une place significative à la diversité des horizons est un enjeu majeur des évolutions des algorithmes, et pris au sérieux par les concepteurs de plateformes ([BMA15]).

L’objet du stage est d’étudier les traces de navigation sur des pages web et de caractériser les trajectoires des utilisateurs sur le site Melty (www.melty.fr). Il fait suite à un premier travail de recherche au cours duquel nous cherchions à caractériser les sessions de navigation en fonction de leurs caractéristiques à la fois structurelles (chaîne, étoile) et temporelles (durée, temps inter-requêtes, ...) – cf figure ci-dessous.



Graphes représentant la structure de sessions variées d’utilisateurs observées sur Melty.

On observe que certains types de sessions sont associées à des consultations de contenus sur une faible variété de thématiques. L’hypothèse la plus vraisemblable est que la structuration du site, ainsi que la nature des liens de recommandation, conduirait à un enfermement dans certaines thématiques, et ainsi à une faible diversité de la consommation des utilisateurs. Pour tester cette hypothèse, nous souhaitons explorer des données de navigation sur ce site correspondant à différents algorithmes de recommandation. En particulier, nous chercherons

à mesurer si les recommandations visant à amener l'utilisateur vers des contenus de nature différente sont effectivement suivis et peuvent alors améliorer la diversité de la consommation sur ce site.

Le stage s'adresse à des étudiants issus de formations variées (réseaux complexes, fouille de données, apprentissage ...). Les principales compétences recherchées sont une connaissance générale de l'algorithmique, une certaine capacité d'adaptation à des méthodes nouvelles et de l'intérêt pour les thématiques interdisciplinaires. Du point de vue des langages et outils utilisés, le/la stagiaire bénéficiera d'une grande liberté.

Références

- [BMA15] Eytan Bakshy, Solomon Messing, and Lada A Adamic. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239) :1130–1132, 2015.
- [MSLC01] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather : Homophily in social networks. *Annual review of sociology*, 27(1) :415–444, 2001.
- [Par11] Eli Pariser. *The filter bubble : What the Internet is hiding from you*. Penguin UK, 2011.