

Caractérisation de traces de navigation

Lionel Tabourier (LIP6 – CNRS / UPMC – 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. Nous chercherons à donner des éléments de réponses aux questions suivantes : Comment caractériser les traces de navigation en fonction de leur longueur, du type de pages visitées, ou de la durée passée sur chaque page ? Peut-on alors construire des classes de comportements d’utilisateurs ? Observe-t-on des modifications de comportements sur de longues périodes d’observation ? Peut-on identifier des liens de cause à effet entre les trajectoires observées et les recommandations proposées ? Dans le cadre du projet Algodiv, plusieurs jeux de données (sur des sites de médias ou commerciaux) seraient disponibles pour servir de support à cette étude.

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.