

Investigation visuelle d'événements dans un grand flot de liens

Sébastien Heymann*, Bénédicte Le Grand**

*Université Pierre et Marie Curie, LIP6-CNRS
sebastien.heyman@lip6.fr

**Université Paris 1 Panthéon-Sorbonne, CRI
benedicte.le-grand@univ-paris1.fr

Résumé. Nous présentons une nouvelle méthode d'analyse exploratoire de grands flots de liens que nous appliquons à la détection d'événements significatifs dans plus de 2 millions d'interactions (pendant 4 mois) entre utilisateurs du réseau social en ligne Github. Nous combinons une méthode statistique de détection automatique d'événements dans une série temporelle, *Outskewer*, avec un système de visualisation de graphes. *Outskewer* identifie des instants de l'évolution du graphe d'interactions méritant d'être étudiés, et un analyste peut valider et interpréter ces événements par la visualisation de motifs anormaux dans les sous-graphes correspondants. Nous montrons par de multiples exemples que cette approche 1) permet de détecter des événements pertinents et de rejeter ceux qui ne le sont pas, 2) est adaptée à une démarche exploratoire car elle ne nécessite pas de connaissance a priori sur les données.

1 Introduction

L'activité d'un réseau social en ligne, d'un réseau de télécommunication, de requêtes sur un moteur de recherche, ou encore d'un système de paiement par carte bancaire, peut être modélisé par un grand réseau d'interactions ; chaque interaction peut en effet être vue comme un lien entre deux entités, apparaissant au déclenchement de l'interaction. Le réseau est alors représenté par un **flot de liens** ordonnés chronologiquement.

Malgré la diversité des systèmes modélisables par des flots de liens, la dynamique de ces flots de liens reste peu explorée. La dynamique des graphes étant elle-même un sujet d'étude scientifique récent, il a fallu attendre ces dernières années et la publication de grands jeux de données à ce niveau de précision pour que l'étude des flots de liens puisse émerger. La plupart des graphes dynamiques étudiés jusqu'ici sont en effet constitués d'instantanés capturés à une certaine fréquence (ex. un graphe par jour). L'enjeu principal est de caractériser l'évolution de ces systèmes pour mieux les comprendre et en particulier différencier une dynamique normale de comportements anormaux. Pour un analyste surveillant un système et levant des alertes en cas d'anomalies, comme une fraude à la carte bancaire ou une intrusion dans un intranet, il est crucial de pouvoir identifier et valider de tels événements rapidement. En d'autres termes, il s'agit de déterminer *où* et *quand* la structure du réseau d'interactions est anormalement altérée.