Visualisation de l'activité d'un site marchand de produits illicites sur le darkweb

Encadrement: Bruno Pinaud et Guy Melançon <bruno.pinaud@u-bordeaux.fr>

Stage financé à hauteur de l'indemnité légale de stage.

Mots clés : science des réseaux, science des données, graphes, visualisation, développement orienté web

Ce sujet fait partie d'une collaboration scientifique au long cours entre le LaBRI et l'école des Sciences criminelles de l'université de Lausanne (Suisse). Notre objectif est de développer des méthodes et outils novateurs pour la construction, l'analyse, l'exploration et la fouille visuelle et interactive de réseaux (mathématiquement parlant des graphes) issus de la collecte de données sur des activités illicites. L'approche vise à soutenir des analyses de données issues de la forensique numérique (http://criminologie.com/article/science-forensique) [Rossy & Morselli 2018] [Rossy & Ribaux 2020].

Succinctement mais plus formellement, notre objectif est de développer des approches de transformation de graphes pilotées grâce à des visualisations des données – à l'image de la plateforme Porgy développé au LaBRI [Pinaud 2012] (https://porgy.labri.fr). Néanmoins, la plateforme Porgy nécessite une expertise formelle en transformations de graphe importante. Nous souhaitons aller au-delà et pouvoir rendre encore plus accessible de telles manipulations de transformations de graphes. La visualisation et l'interaction doivent ainsi totalement prendre le relais de l'aspect très conceptuel et formel des transformations de graphes.

Le stage est centré sur une prise en main des données et une familiarisation avec les questions de la forensique, avant d'aller vers des questions de recherche visant une formalisation des approches à développer.

Il s'agit d'étudier des données issues du plus important site web marchand de produits illicites accessibles uniquement depuis le darkweb via le protocole Tor (crypto-marché) pendant l'été 2020. Nos collègues suisses ont moissonné et analysé plus de 50Go de pages web librement accessibles entre juin et septembre 2020 sur le plus gros site marchand ouvert à ce moment-là.

Nous avons des données sur les annonces des produits à vendre, les produits à vendre, les vendeurs et les évaluations d'acheteurs via plusieurs millions d'avis laissés en ligne. D'un point de vue des sciences criminelles de nombreuses questions se posent telles que :

- L'identification des produits et/ou vendeurs populaires, l'évolution des prix, les produits/vendeurs qui gagnent ou perdent en popularité ;
- Faire émerger des vendeurs génériques/spécialisés dans certains produits ou type de produits (fausse monnaie, médicaments frauduleux, drogue, etc.)
- Faire émerger des comportements d'achats et ou de vente et leurs évolutions

- Caractériser le marché et les flux selon les pays d'origine et de destination des produits

Pour l'informaticien, le travail (le contenu du stage) consiste ensuite à s'approprier les données et ces questions (et d'autres que nous n'avons pas encore) pour les traduire en terme d'opérations à mener pour construire des graphes, les analyser et les visualiser tout cela interactivement [Brehmer 2013]. Ces opérations se passent dans un logiciel qui à terme pourra être mis dans les mains des experts des données. Un exemple très simple est de construire le réseau entre les vendeurs et acheteurs, où une arête indique qu'un vendeur a envoyé un produit à un acheteur. Dans ce cas, le degré du sommet vendeur ou acheteur montre immédiatement celui qui fait le plus/le moins de ventes/d'achats.

A la diversité des propriétés des entités et des liens qui en font un réseau multicouches [McGee *et al*. 2019], on peut aussi penser à construire un réseau dynamique [Beck *et al*. 2014] entre les vendeurs et les types de produits vendus. Ce réseau peut servir de base pour faire émerger des stratégies temporelles de diversification ou au contraire de spécialisation. Il faut ici étudier l'évolution de la densité des arêtes dans le temps selon différents critères.

Ensuite, il faut traduire ces opérations sur les visualisations et formuler les interactions. Par exemple,

- le degré peut être projeté sur la taille des sommets,
- des diagrammes de Sankey pour réaliser l'analyse spatiale (https://www.d3-graph-gallery.com/sankey.html)
- des cartes de chaleurs (https://www.datavis.fr/index.php?page=day-hour-heatmap) peuvent montrer des évolutions temporelles,
- des bulles pour grouper les vendeurs/acheteurs/produits selon des critères (https://www.datavis.fr/index.php?page=bubblechart)

En première intention, nous utiliserons la plateforme de visualisation Tulip [Auber *et al*. 2017] (https://tulip.labri.fr) développée au LaBRI. Les aspects visuels et interactifs sur le web pourront être traités par exemple avec la librairie D3 (https://d3js.org/).

Références

- Auber, D., et al. (2017). Tulip 5. Encyclopedia of Social Network Analysis and Mining. R. Alhajj and J. Rokne. New York, NY, Springer New York: 1-28.
- Beck, F., et al. (2014). "The state of the art in visualizing dynamic graphs." EuroVis STAR.
- Brehmer, M. and T. Munzner (2013). "A multi-level typology of abstract visualization tasks." Visualization and Computer Graphics, IEEE Transactions on 19(12): 2376-2385.
- McGee, F., et al. (2019). "The State of the Art in Multilayer Network Visualization." Computer Graphics Forum 38(6): 125-149.
- Pinaud, B., et al. (2012). "PORGY: A Visual Graph Rewriting Environment for Complex Systems." Computer Graphics Forum 31(3pt4): 1265-1274.
- Rossy, Q., & Morselli, C. (2018). The contribution of forensic science to the analysis of crime networks. The Routledge International Handbook of Forensic Intelligence and Criminology, 191-204.
- Rossy, Q., Ribaux, O. Orienting the Development of Crime Analysis Processes in Police Organisations Covering the Digital Transformations of Fraud Mechanisms. Eur J Crim Policy Res (2020). https://doi.org/10.1007/s10610-020-09438-3