

Détection de Fraudes dans les Transactions Financières

– offre de postdoc –

Matthieu Latapy

LIP6 – CNRS et Sorbonne Université – Paris

matthieu.latapy@lip6.fr

Détecter les fraudes dans les transactions financières est un sujet applicatif crucial : les pertes dues aux fraudes sont colossales, et l'argent souvent utilisé pour financer des activités nocives. C'est aussi un sujet qui nécessite des travaux de recherche fondamentale importants, par exemple concernant l'analyse de dynamiques de graphes, l'algorithmique en *streaming* ou la définition de métriques capturant la nature à la fois structurelle et temporelle des données.

Le plus souvent, l'information disponible consiste en un enregistrement de transferts d'argent, comme des virements ou des paiements par carte. On connaît l'origine de l'argent, sa destination, le moment de la transaction et son montant, mais relativement peu d'autres informations (parfois le type de compte ou de carte, le pays de la transaction, etc). Les données sont alors très bien modélisées par des flots de liens (*link streams*)¹. Voir Figure 1.

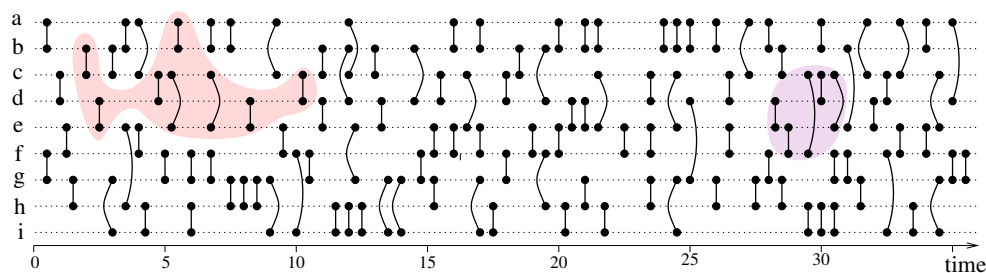


FIGURE 1 – Un exemple de flot de liens représentant des transactions financières entre a, b, \dots, i (axe vertical à gauche) de l'instant 0 à l'instant 36 (axe horizontal en bas). Une transaction a lieu entre a et b à l'instant 0.5, ainsi qu'entre f et g . Une autre a lieu entre c et d à l'instant 1, et ainsi de suite. Des groupes de liens spécifiques, comme par exemple ceux mis en évidence par des zones colorées ici, peuvent être des traces d'activités illicites.

Avec cette vision, les fraudes sont des sous-séquences de liens dans un flot de liens, dont on s'attend à ce qu'elles aient des caractéristiques particulières. Tout l'enjeu consiste à identifier ces caractéristiques, à les modéliser par des métriques calculables en temps et espace raisonnables, et enfin à s'en servir pour détecter les fraudes. L'évaluation des méthodes pose également question, face au manque de vérités de terrain, c'est-à-dire de données dans lesquelles les fraudes réelles sont connues.

Ce postdoc propose une approche orientée graphes et données pour répondre à ces problématiques et améliorer drastiquement l'état-de-l'art sur ces sujets.

1. *Stream Graphs and Link Streams for the Modeling of Interactions over Time et Weighted, Bipartite, or Directed Stream Graphs for the Modeling of Temporal Networks*. Matthieu Latapy, Clémence Magnien, Tiphaine Viard. <https://arxiv.org/abs/1710.04073> et <https://arxiv.org/abs/1906.04840>

La personne recrutée nous rejoindra en particulier sur deux travaux en cours :

1. L'extension d'une méthode de détection basée sur des propriétés simples de graphes, pour laquelle un premier prototype est déjà opérationnel et donne d'excellents résultats ; l'objectif est maintenant d'enrichir l'ensemble de propriétés de graphes considérées, d'ajouter des propriétés temporelles et des propriétés de flots de liens, d'intégrer les notions de graphes valués ou orientés, et de tenter de détecter des anomalies plus complexes ² ;

2. La conception d'un générateur de flots de liens artificiels avec certaines caractéristiques prescrites, et l'injection d'anomalies dans ces flots de liens ; en effet, de telles données artificielles sont essentielles pour tester les méthodes de détection dans différents scénarios et évaluer leur pertinence. Nous utilisons le fait qu'il existe une correspondance entre flots de liens et graphes bipartis (Figure 2), ce qui permet d'utiliser l'état-de-l'art disponible pour la génération de tels graphes ³.

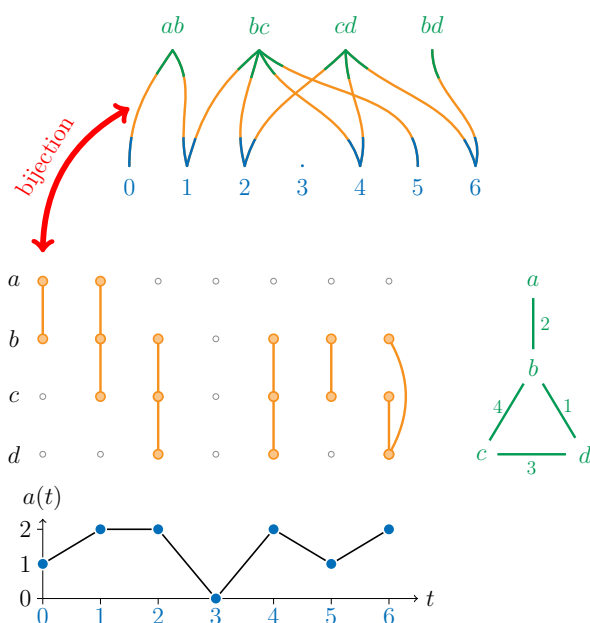


FIGURE 2 – Un graphe biparti (en haut) est équivalent à un flot de liens (au milieu). Les poids sur les liens du graphe induit par ce flot de liens (à droite) correspondent aux degrés d'une des parties des nœuds du graphe biparti (en vert) ; le nombre de liens à chaque instant dans ce flot de liens (en bas) correspond aux degrés d'une de l'autre partie des nœuds du graphe biparti (en bleu). Le tirage d'un graphe biparti aléatoire ayant ces deux distributions de degrés correspond à un flot de liens (en orange).

À titre indicatif, **nous sommes particulièrement intéressés par des compétences en** analyse de grands graphes et de réseaux, utilisation de machine learning (en particulier scikit-learn), programmation python, graphes aléatoires, manipulation de grandes données en streaming, expérimentations à grande échelle, cybersécurité et traitement du signal. Une expérience de rédaction d'articles pour des conférences internationales sélectives est également souhaitée. Enfin, un équilibre entre les aspects théoriques et pratiques de la recherche, très présent dans le sujet, serait un plus.

2. *Anomaly detection in dynamic networks : a survey*. Stephen Ranshous, Shitian Shen, Danai Koutra, Steve Harenberg, Christos Faloutsos, Nagiza F. Samatova. <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wics.1347>

3. *Bipartite Graphs as Models of Complex Networks*. Jean-Loup Guillaume, Matthieu Latapy. <https://arxiv.org/abs/cond-mat/0307095>