Anomaly detection in link streams

**Supervision:** Matthieu Latapy (LIP6 - CNRS / SU), Esteban Bautista (LIP6 / SU) and Mehdi Naima (LIP6 / SU)

**Location:** Campus Pierre et Marie Curie, Sorbonne Université - 4 place Jussieu, 75005 Paris, France

**Topic:** Algorithms in complex networks

---

**Context**

In many domains, data are endowed with a temporal-relational structure naturally modeled by a link stream [1]: a sequence of triplets \((t, u, v)\), indicating that \(u\) and \(v\) interacted at time \(t\). Link streams capture a wealth of information that bears great potential for progress in high-impact areas. For instance, frauds or thefts in monetary transactions may leave signatures expressed as sub-streams that heavily interact in a short span of time [2]. Another example are network attacks which may be characterized as repetitive bursts of links that deviate from normal activity [3]. Consequently, it is crucial to develop algorithms that allow us to efficiently process and extract relevant information from link streams.

**Goal**

A recent work [4] proposed an on-line algorithm to rank interactions in link streams and detect microcluster anomalies: suddenly arriving groups of suspiciously similar edges. However, the algorithm of [4] ranks triplet \((u, v, t)\) solely based on how frequently \(u\) and \(v\) interacted in the recent past, meaning that rankings depend only on the time properties and not on the structural properties of link streams. We therefore aim to develop ranking algorithms that take into account both the time and structural properties of link streams.

Numerous recent works have generalized graph theoretical concepts to link streams [1, 5, 6], providing a solid foundation to analyze their structural properties. It is thus a timely challenge to leverage these concepts in the context of ranking algorithms. We are particularly interested in exploring the recent definitions of temporal random walks [7, 8, 9]. Our interest in random walks is four-fold: (i) they have already been successfully used to extend the PageRank algorithm [10]; (ii) their behavior is highly sensitive to structural properties; (iii) they are algorithmically efficient; and (iv) they allow to derive strong analytic results. We are also open to explore other centrality metrics, such as betweenness [11], which raise important algorithmic challenges but have great interpretability.

**Requested profile**

This internship is directed at students with various background (complex networks, algorithmic, graph theory) with a strong interest in graph algorithmics and/or graph theory and its applications.

The intern will be part of the Complex Networks team of the LIP6 (SU-CNRS), located in Paris on Jussieu Campus. The internship will be supervised by Matthieu Latapy.
References