Anomaly Detection in Link Streams

Internship Research

Supervision	Maroua BAHRI (RO)
	Mehdi Naima (ComplexNetworks)
Mail	maroua.bahri@lip6.fr, mehdi.naima@lip6.fr
Address	LIP6, Campus Pierre et Marie Curie, Sorbonne Université

Keywords	Anomaly Detection, Link Streams, AutoML, Meta-Learning, Hyperparameter Optimization

Context

Anomaly detection of link streams refers to identifying unusual patterns, structures, or suspicious behaviors, in a sequence of edges that arrive continuously over time, see Figure 1 for an example. This research area is crucial in cybersecurity, system failures, social network analysis, or network attacks. Detecting anomalies in link streams presents several challenges, *inter alia*, (i) *dynamic nature*: since the graph evolves over time, algorithms need to operate in an online and real-time manner to process the continuous flow of new edges and nodes, (ii) *scalability*: link streams can be large-scale with billions of edges requiring highly scalable algorithms with low computational and memory overhead, (iii) *lack of labeled data*: in many real-world settings, it is hard to obtain labeled data (ground truth), as result, researchers concentrate on identifying randomly injected links considered as anomalies.

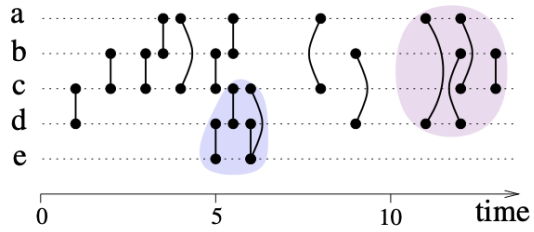

Figure 1: From [1]. A link stream: temporal interactions between *a, b, c, d, e* over time range $[0, 15]$. Interactions in shaded areas may be traces of frauds or attacks.

However, a critical limitation exists: the selection and configuration of anomaly detection methods remain largely manual and depend on human expertise. This project aims to address this gap by focusing specifically on the AutoML dimension.

1 Objectives

In this project, the student will focus on developing an **intelligent AutoML system** for anomaly detection in link streams. This innovation will make detectors accessible to non-experts while achieving state-of-the-art performance. We propose two main components:

The first component automatically selects the most appropriate detection methods for a given link stream. We will extract topological and temporal features from the stream (average degree, density, clustering coefficient, link arrival rate, ...) and develop a meta-learning system [2, 3] that learns to associate these characteristics with the best-performing detection methods. This system will automatically switch methods when the graph characteristics evolve.

Second, to tackle dynamic hyperparameter optimization in real-time, we would develop adaptive optimization mechanisms using optimization techniques (e.g., Bayesian or random optimization) tailored to stream constraints. The system will balance exploration (testing new configurations) and exploitation (using good configurations). This builds upon recent work in online hyperparameter optimization [4].

Expected outcomes and contributions:

- The proposed solution will be rigorously documented, including theoretical analysis, algorithmic details, and an extensive experimental study to validate the proposed solution. The work will also feature a comparative evaluation against current state-of-the-art methods, showcasing the performance gains achieved

in terms of detection accuracy, scalability, and robustness to different types of anomalies.

- To support transparency and further research, a fully documented open-source implementation will be provided. This will include all source code, datasets, and experimental configurations used, allowing researchers and practitioners to reproduce and build upon our work easily.

Internship

The project is dedicated to a Master's thesis (or equivalent) student, dedicated to the objectives of the project. The student will do his/her master's thesis in either of the teams concerned with the project, namely ComplexNetworks and RO.

Skills

- Master level research internship M2 or equivalent (stage de fin d'études ingénieur).
- Strong programming skills in Python.
- Familiarity with reading, understanding, and building on academic publications.
- Sound knowledge in graphs and machine learning.

Gratification

According to current regulations.

Contact to apply

Send the following documents (exclusively in PDF format) to maroua.bahri@lip6.fr and mehdi.naima@lip6.fr:

- A cover letter explaining your qualifications, experiences, and motivation for this topic.
- Curriculum vitae.
- Transcripts of grades from the third year of your bachelor's degree, the first year of your master's degree, and any available grades from the second year of your master's degree (or equivalent for engineering schools).
- If possible, recommendation letters.
- If possible, a link to repositories of personal projects (e.g., GitHub).

References

[1] "Labcom fit." `https://www-complexnetworks.lip6.fr/~latapy/fit/`. Accessed: 2024-10-30.

[2] T. Lacombe, Y. S. Koh, G. Dobbie, and O. Wu, "A meta-learning approach for automated hyperparameter tuning in evolving data streams," in *2021 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, IEEE, 2021.

[3] M. Feurer, K. Eggensperger, S. Falkner, M. Lindauer, and F. Hutter, "Auto-sklearn 2.0: Hands-free automl via meta-learning," *Journal of Machine Learning Research*, vol. 23, no. 261, pp. 1–61, 2022.

[4] N. Verma, A. Bifet, B. Pfahringer, and M. Bahri, "Bayesian stream tuner: Dynamic hyperparameter optimization for real-time data streams," in *31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 2871–2882, 2025.