Master's Research Internship 2025

Title:	Change Point Detection in Temporal Graphs
Supervisors:	Esteban Bautista (ULCO), Matthieu Puigt (ULCO), Laurent Brisson (IMT Atlantique)
Internship duration:	6 months
Place:	LISIC laboratory - Université du Littoral Côte d'Opale (St-Omer)

Context

Numerous datasets, such as interactions between people, traffic between computers, or transactions between bank accounts can be very well modeled as a temporal graph: a set of triplets (u, v, t) indicating that u and v interacted at time t. The proper analysis of temporal graphs is a task of utmost importance, with numerous applications in Industry 4.0, cyber-security, or social network analysis [1]. Temporal graphs often go through different regimes in which interactions behave according to a distinctive pattern. For instance, card transactions do not occur at the same rate nor between the same individuals during week-days than during week-ends, etc. Detecting these changes — a task known change point (CP) detection — is important to improve fault detection, do better predictions, or filter corrupted/noisy interactions, among other applications. However, CP detection is a challenging task, as the highly sparse and erratic nature of real-world temporal graphs makes it very hard to extract regular patterns that accurately describe the data. Thus, while numerous algorithms for CP detection have been developed in recent years [2, 3], existing solutions still perform poorly in practice.

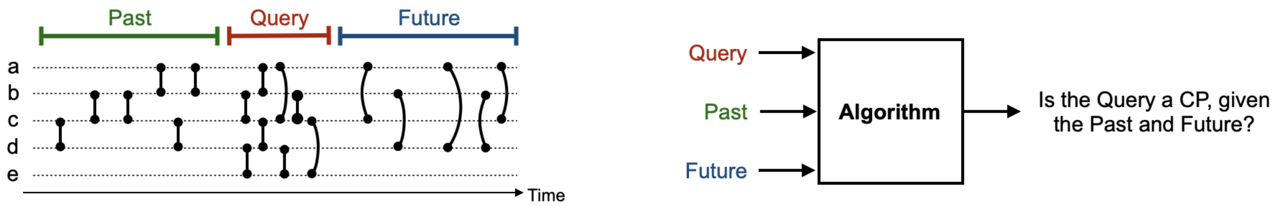

Figure 1: Interactions in a temporal graph exhibit different activity regimes (left-panel). Detectors can be seen as black boxes whose role is to spot transition periods by analyzing past and future interactions (top-right). The main difference between algorithms lies in how they extract patterns from the inputs in order to compare them and detect transitions.

CP detection algorithms can be seen as black-boxes that aim to answer to the question: is a query group of interactions a CP according to observed past and future interactions? The difference between algorithms is how they compare the inputs, which is often done through ad-hoc descriptors that struggle with data sparsity and that cannot account for all situations. Recently, [4] laid the foundations of an improved methodology to analyze and compare temporal graphs: it shows that time-series and graph dictionaries can be used to map any temporal graph (including sparse ones), without any information loss, to a matrix of coefficients that reveals all their dynamical and structural information. Namely, if a clique structure appears periodically in the data, then such matrix reveals it, provided that the used dictionaries contain a clique and a periodic signal. Thus, the analysis of [4] is a promising approach to tackle CP detection. Yet, one fundamental question must be addressed: how to select the dictionaries so as to make CPs visible in the coefficient matrices?

Goal

The goal of this internship is to build upon [4] and to construct graph and signal dictionaries that allow us to *(i)* maximize CP detection accuracy; and *(ii)* minimize CP false detection rate. While we are open to tackle the problem through an end-to-end dictionary learning pipeleline [5], we would like to first explore the potential of well-established analytic dictionaries from which it is possible to establish theoretically bounds on the detection rates. Since most temporal graphs are binary, we aim to carry an in depth comparison between the Haar [6], Walsh [7] and boolean-based [6] dictionaries as time-series ones. For the case of graphs, we aim to leverage the methodology proposed in [8] to build custom dictionaries with user-defined motifs. We aim to develop a methodology that identifies the best motifs to add into the dictionary in a data-driven way.

Requested profile

This internship is directed at students with various backgrounds (computer science, data science, signal processing, complex systems) but with a strong interest in data science and graphs. Interest in the theoretical aspects of machine learning and in Python development will be a plus.

To apply, please **send an e-mail** to {esteban.bautista, matthieu.puigt} [at] univ-littoral.fr and laurent.brisson [at] imt-atlantique.fr while attaching the documents that can support your application:

- your resume;
- a cover letter;
- your transcripts from the last year of B.Sc to the last year of M.Sc. (if the latter is already available);
- two reference letters or the names and means of contact of two academic advisers.

Applications will be reviewed on a rolling basis until the position is filled.

References

[1] J. Tang, I. Leontiadis, S. Scellato, V. Nicosia, C. Mascolo, M. Musolesi, and V. Latora, "Applications of temporal graph metrics to real-world networks," *Temporal Networks*, pp. 135–159, 2013.

[2] D. Sulem, H. Kenlay, M. Cucuringu, and X. Dong, "Graph similarity learning for change-point detection in dynamic networks," *Machine Learning*, vol. 113, no. 1, pp. 1–44, 2024.

[3] S. Huang, Y. Hitti, G. Rabusseau, and R. Rabbany, "Laplacian change point detection for dynamic graphs," in *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 349–358, 2020.

[4] E. Bautista and M. Latapy, "A frequency-structure approach for link stream analysis," in *Temporal Network Theory*, pp. 449–482, Springer, 2023.

[5] I. Tošić and P. Frossard, "Dictionary learning," *IEEE Signal Processing Magazine*, vol. 28, no. 2, pp. 27–38, 2011.

[6] T. Sasao and J. Butler, *Progress in Applications of Boolean Functions*. Springer Nature, 2022.

[7] R. O'Donnell, "Analysis of boolean functions," *arXiv preprint arXiv:2105.10386*, 2021.

[8] E. Bautista, L. Brisson, C. Bothorel, and G. Smits, "Mad: Multi-scale anomaly detection in link streams," in *Proceedings of the 17th ACM International Conference on Web Search and Data Mining*, pp. 38–46, 2024.