Non-funded internship proposals

Rémy Cazabet - LIRIS - Université Lyon 1

2020/2021

1 Adapting community detection algorithms for static networks to link streams

Many methods exist to detect communities in static networks [2]. Detecting communities in dynamic networks is already a popular topic with a growing number of publications [5]. However, very few methods exist to detect communities in *link stream*, i.e., networks which are composed of interactions without duration, such as data collected with instant messaging, contact tracing or e-mail. In a previous internship[1], we started to study the problem and developed some preliminary approaches. One of the possibilities which seem promissing and that we would like to investigate further is the adaptation of existing algorithms for static networks to link streams. In particular, Walk Trap[4] seems a very good candidate, but other possibilities could be explored. This will require both theoretical (appropraite definition of a random walk in a link stream) and practical (efficient implementation) contributions.

2 Interactions or Relations? Caracterising the nature of dynamic networks

There is a fundamental difference in the nature of dynamic networks: some represent interactions, or interaction-like phenomenon, while other represent interactions. One way to understand this difference is that the network composed of edges existing at any particular time is a well formed network in the case of relation (the state at time t of all friendship relations in Facebook is a proper network that can be analyzed with network science), and a degenerate one for interactions (the network of all Facebook messenger messages sent at a particular second t is not a proper network, even if thousands of messages have been sent at this particular second). To this day, no method exist to say if a network is rather composed of interaction or of relations. It is an important question because interactions networks are often transformed in relation networks to be analyzed, but this is done with arbitrary scale (we aggregate

[1] `http://cazabetremy.fr/rTeam/lecuyer.pdf`

edges every hour/day/month/year), without guarantee of the network to have the properties of a relation network. In a previous work, we proposed a method based on information compression to distinguish both methods, but this is not fully satisfying because this method lacks interpretability. In this internship, we will try to define notions such as the stability of edges, nodes and graphs, the synchronicity of edges, etc. as an intermediary step towards our objective.

3 Machine learning for improving null models for community detection

A popular approach for community detection (Modularity) consists in comparing, for a given partition, the number of edges found inside communities in the observed network and the number of edges found inside communities in a random network. One of the problem of this approach is that even in random networks, a *good* partition can be found that contains more edges than expected at random, due to random flucutations. We will try to solve this problem by using (classic) machine learning techniques (random forest, neural networks, etc.) to predict the number/distribution of expected edges in **the best partition** of a network of known properties for a given partition. Indeed, the problem is that there are too many parameters to find a closed form expression of this expectation. If this approach is successful, we will try to extend it by taking into account other parameters than the degree distribution, such as the clustering coefficient.

4 Graph Convolutional Neural Networks for traffic prediction

Graph Convolutional Neural Networks (GNN)[3] have been introduced a few years ago and represent a powerful technique to do machine learning on graphs. Even more recently, methods have been proposed to take into account the dynamic of what is happening on the network(e.g., [1]) to improve the learning process. In a previous internship[2], we investigated and compared existing solutions for traffic prediction using those new methods and more classic ones. In particular, we reimplemented one of the most recent approaches and used it for prediction. In this internship, we would like to improve on these results, for instance in taking into account global or local phenomenons (weather, events...). This internship would be co-supervised with a researcher expert in the domain of transportation, and benefits from large real-world data of car transit in Lyon.

[2] `http://cazabetremy.fr/rTeam/fmensi.pdf`

5 Cryptocurrency network analysis

I'm leading a research project called BITUNAM for Bitcoin User Network Analysis and Mining. The objective of this project is to analyse using machine learning and network science tools financial transactions happening in Bitcoin and other cryptocurrencies. There is a lot to do on the topic, large data, complex challenges, so there's a wide variety of topics if you're interested.

References

[1] Zulong Diao, Xin Wang, Dafang Zhang, Yingru Liu, Kun Xie, and Shaoyao He. Dynamic spatial-temporal graph convolutional neural networks for traffic forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 890–897, 2019.

[2] Santo Fortunato and Darko Hric. Community detection in networks: A user guide. *Physics reports*, 659:1–44, 2016.

[3] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[4] Pascal Pons and Matthieu Latapy. Computing communities in large networks using random walks. In *J. Graph Algorithms Appl.* Citeseer, 2006.

[5] Giulio Rossetti and Rémy Cazabet. Community discovery in dynamic networks: a survey. *ACM Computing Surveys (CSUR)*, 51(2):1–37, 2018.