Complex Graph Analysis for the Detection of Corruption in Public Procurements

Laboratoire Informatique d'Avignon (LIA), France

Title: Complex graph analysis for the detection of corruption in public procurements

Main laboratory: Laboratoire Informatique d'Avignon (LIA – EA 4128), France

Duration: 3 years

Advisors:
Vincent Labatut & Rosa Figueiredo (LIA – Avignon Université, Avignon, France)
Christine Largeron (Laboratoire Hubert Curien – UJM, Saint-Étienne, France)

Keywords: Signed graphs analysis, Graph partitioning, Data analysis, Public procurements

Context

In many applications, the data to be studied is of *relational* nature, modeled in the form of a network and represented by graphs. This representation makes possible the study of interactions between people, especially in social sciences. Although network analysis is an active branch of data mining and machine learning, a majority of works focuses on homogeneous networks described by a simple graph where the nodes correspond to the entities of the network and the links (edges or arcs) to their relationships. However, in many applications, contextual information describing the relationships or the entities themselves are available and could be used to study in a more efficient way the network.

This led notably to the notions of signed graphs and attributed graphs. In a signed graph, each edge is labeled with either a negative or a positive sign, which allows representing *antagonistic* relationships [1]. In an attributed network, vertices are described by attributes which allow to take into account their *individual* characteristics, like for instance their genre or age [2].

These representations are much richer than a simple graph and permit to better model complex interaction systems. However, they require to adapt existing algorithms or to design new ones for solving efficiently all standard network visualization and analysis tasks, such as community detection, link prediction or information diffusion.

This PhD is part of the French Research project DeCoMaP (Detecting Corruption in Markets for Public Procurement) funded by the ANR (French NSF) which aims at retrieving, processing and analyzing open data related to French public procurements, in order to design a tool able to assess corruption risks between public buyers and suppliers.

Subject

Designing automatic tools for the assessment of risks of corruption in public procurements is a task called *red flagging*. It is not completely new, as some teams have been working on it for a few years [1], especially at the European level [2]. However, none of them applies to French public procurements, as they do not handle its specifics (legal framework, nature and form of the available open data). Moreover, existing approaches focus on *individual* information, which characterize buyers and suppliers independently, and ignore *relational* information, which corresponds to interactions and interdependencies between these agents.

In the context of the project DoCoMaP, we propose to design a new tool tackling these issues and limitations. More specifically, the goal of this thesis is to handle a large methodological part of this work, by representing the relationships between buyers and suppliers (normal contractual relationship or corruption) and by characterizing these agents through signed and/or attributed graphs.

As this type of graphs has been much less explored than simple graphs in the literature, the first task of this thesis consists in designing methods to extract them from the raw data, and to analyze them in the context of our application (corruption detection). In particular, it is necessary to define a corruption index that could be used to enrich the information already available in the graph at the level of its links (contracts between a public buyer and a supplier) and at the level of its vertices (characteristics of the actors).

The second task focuses on signed graph partitioning in the framework of structural balance [3]. A signed graph is said to be *balanced* if it can be partitioned into two [4] or more [5] mutually hostile subgroups, each having internal solidarity (i.e. positive edges are inside clusters, and negative ones are

between them). Some recent works started tackling this problem notably in our team [6], but there is still much to do for formalizing the problem and solving it efficiently, particularly using deep learning approaches [7]. In DeCoMap, signed graph partitioning will aim to bring out groups of agents (public buyers and supplier companies) likely to be related in fraudulent practices.

Working environment

The PhD candidate will work at the Laboratoire Informatique d'Avignon (LIA – EA 4128) under the supervision of Vincent Labatut and Rosa Figueiredo (LIA – Avignon Université, Avignon, France) and Christine Largeron, Professor in Computer Science (LabHC – Université Jean Monnet, Saint-Etienne, France).

Funding

The PhD fellowship is funded for 3 years through the ongoing DeCoMaP ANR project (ANR-19-CE38-0004) and is monthly funded about approximatively 1500 €.

Profile of the candidate

The candidate should have a master degree or equivalent in Computer Science. The subject is at the intersection of several domains: graph theory, statistics, data mining and machine learning, big data including databases (the considered networks can be huge). Thus the candidate should have strong backgrounds in several of these topics.

Other required skills:
- Good abilities in algorithm design and programming.
- Good technical skills regarding data management (databases, retrieval from web APIs)
- A very good level (written and oral) in English.
- Good communication skills (oral and written).
- The ability to work in a team with colleagues from other scientific disciplines.
- Autonomy and motivation for research.

Application instructions

Applicants are invited to contact as soon as possible.

The application file should contain the following documents:
1. a **curriculum vitæ** (CV);
2. the **official academic transcripts** of all the candidate's higher education degrees (BSc, License, MSc, Master's degree, Engineer degree, etc.). If the candidate is currently finishing a Master's degree, s/he must send the transcript of the grades obtained so far, with the rank among her/his peers, and the list of classes taken during the last year;
3. some **recommendation letters** (quality is more important than quantity, there);
4. and a motivation letter **written specifically for this position**.

Send all of these documents by email to all three advisors:
- Vincent Labatut <vincent.labatut@univ-avignon.fr>
- Rosa Figueiredo <rosa.figueiredo@univ-avignon.fr>
- Christine Largeron <christine.largeron@univ-st-etienne.fr>

The PhD thesis is expected to start in September/October 2020.

References
[1] T. Zaslavsky. *Signed graphs, Discrete Applied Mathematics*, 5:248, 1983.
[2] R. Interdonato, M. Atzmueller, S. Gaito, R. Kanawati, C. Largeron & A. Sala. *Feature-rich networks: going beyond complex network topologies*. Applied Network Science 4(1):4:1-4:13, 2019.
[3] M. Fazekas, J.T. Tóth. *New ways to measure institutionalised grand corruption in public procurement*, U4 Brief, n°9 (2014).
[4] J. Ferwerda, et I. Deleanu. *Identifying and Reducing Corruption in Public Procurement in the EU*. European Commission, OLAF, 2013.
[5] P. Doreian & A. Mrvar. *Partitioning signed social networks*. Social Networks, 31(1):1–11, 2009.
[6] D. Cartwright & F. Harary. *Structural balance: A generalization of Heider's theory*. Psychological Review, 63:277–293, 1956.
[7] J. A. Davis. *Clustering and structural balance in graphs*. Human Relations, 20(2):181–187, 1967.
[8] R. Figueiredo & Y. Frota. *The maximum balanced subgraph of a signed graph: Applications and solution approaches*. European Journal of Operational Research, 236(2):473-487, 2014.
[9] J. Huang, H. Shen, L. Hou & X. Cheng. *Signed Graph Attention Networks*, https://arxiv.org/abs/1906.10958, 2019.