

PhD. position in Computer Science

Machine learning and data mining with incomplete or uncertain relational data Laboratoire Hubert Curien (LabHC), France

Title: Complex data analysis with incomplete or uncertain data

Main laboratory: Laboratoire Hubert Curien (LabHC UMR 5516), France

Duration: 3 years

Advisors:

[Christophe Gravier](#) (Pr, [Laboratoire Hubert Curien](#) – [UJM](#), Saint-Étienne, France)

[Christine Largeton](#) (Pr, [Laboratoire Hubert Curien](#) – [UJM](#), Saint-Étienne, France)

Keywords: Data mining, Machine learning, Social network, Graphs analysis, Uncomplete data

Context

In many applications, the data to be studied is *relational*, 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 in other fields focuses also graphs where the nodes correspond to the entities of the network and the links (edges or arcs) to their relationships, for instance in biology or image processing.

Among the main tasks studied on this data, one can mention supervised and unsupervised clustering, link prediction or, recommendation.

Subject

One of the main limitations of the existing methods allowing to solve these issues is that they usually consider the network to be fully known so that it can be perfectly defined by a graph or by a sequence of graphs for a dynamic network. However, in practice, the graph is often only partially observable, or the data can be observed with a degree of uncertainty. This uncertainty is derived from the presence or absence of a node (or link) at a given moment and, if applicable, the lack of weight or orientation characterizing the link.

Although it is very widespread in practice, this uncertainty is not without effect on the outcomes resulting from classical graph analysis but it is understudied in the literature. Concerning the classification of entities, one can cite the works of Sevon et al. or Pfeiffer et al.. For the prediction of links, Mallek et al., Ahmed et al., Murata et al. Finally for the detection of communities, we can mention Dahlin et al., Liu et al., Martin et al., Zhang et al. as well as our works (Benyahia et al.). In the frame of Graph Neural Networks, only few works focus on learning attribute-missing graph embeddings (Chen et al.). The applications on many downstream information communication, science and technology domains are nonetheless important, such as in databases (Banerjee et al.) and computer networks (Sasaki et al.). Thus, in this project our objectives are twofold. First, we will study the impact of the incompleteness of the relational data on network algorithms. Second, we will develop methods able to factor nodes and links incompleteness in graph processing algorithms as well as machine learning techniques based on graphs.

Working environment

The PhD candidate will work at the Laboratoire Hubert Curien (UMR 5516) under the supervision of Christophe Gravier and [Christine Largeton](#), Professor in Computer Science ([LabHC](#) – [Université Jean Monnet](#), Saint-Etienne, France).

Funding

The PhD fellowship is funded for 3 years through and is monthly funded about approximately 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. 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 mining, machine learning and data management
- 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 vitae** (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 the advisors:

- Christophe Gravier christophe.gravier@univ-st-etienne.fr
- Christine Largeron christine.largeron@univ-st-etienne.fr

References

- S. Banerjee and B. Pal, DySky: Dynamic Skyline Queries on Uncertain Graphs, WISE 2020: Web Information Systems Engineering – WISE 2020 pp 242-254
- Y. Sasaki, Y. Fujiwara, and M. Onizuka, Efficient network reliability computation in uncertain graphs. 2020 arXiv preprint arXiv:2009.03158.
- P. Sevón, L. Eronen, P. Hintsanen, K. Kuloovesi, H. Toivonen, Link Discovery in Graphs Derived from Biological Databases, Data Integration in the Life Science, 2006, Springer Berlin Heidelberg, 35-49
- J. Pfeiffer, J. Neville, Methods to Determine Node Centrality and Clustering in Graphs with Uncertain Structure, CoRR abs/1104.0319, 2011, arXiv 1104.0319
- X. Chen, S. Chen, J. Yao, H. Zheng, Y. Zhang, I. W Tsang, to appear in IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE (TPAMI), 2020 -<https://arxiv.org/abs/2011.01623>
- S. Mallek, I. Boukhris, Z. Elouedi, E. Lefevre, Evidential Missing Link Prediction in Uncertain Social Networks, Information Processing and Management of Uncertainty in Knowledge-Based Systems - 16th International Conference, 274—285, 2016
- N. M. Ahmed and L. Chen, An efficient algorithm for link prediction in temporal uncertain social networks, Information Sciences, vol 331, 120—136, 2016
- T. Murata, S. Moriyasu, Link Prediction of Social Networks Based on Weighted Proximity Measures, Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence, WI, 85-88, 2007

J. Dahlin, S. Pontus, A Method for Community Detection in Uncertain Networks, European Intelligence and Security Informatics Conference, 155--162, 2011
L. Liu, R. Jin, C. Aggrawal, Y. Shen, Reliable Clustering on Uncertain Graphs Proceedings of the 2012 IEEE 12th International Conference on Data Mining, 459—468, 2012
T. Martin, B. Ball, M.E.J. Newman, Structural inference for uncertain networks, Phys. Rev. E, Vol 93, 1, 012306, 2016