Research internship proposal: Sampling large networks

Nicolas Tremblay, Simon Barthelmé, Pierre-Olivier Amblard
CNRS, Gipsa-lab, Univ. Grenoble-Alpes, France

General context. Data are increasingly measured on irregular network structures: brain activity patterns on neuronal networks, traffic on the Internet network, load on the power grid etc. are some examples of such data. As in classical data mining, one of the first analysis step when these data and networks become too large, is to reduce the dimension of the data of interest in order to run learning algorithms in a reasonable amount of computational time. Classically, one of the main approaches to reduce dimensionality is sampling: periodical, non-uniform, random, etc. The question raised by data measured on networks is how to transpose these classical notions to very irregular network structures?

The graph sampling problem. Graphs are a central modelling tool for network-structured data. Depending on the application, the nodes of a graph may represent people in social networks, brain regions in neuronal networks, etc., basically any system made of interconnected sub-systems. Data on a graph, called graph signals [4], such as blood flow of brain regions in neuronal networks, may be represented by a scalar per node. Graph sampling thus consists in measuring an a priori smooth graph signal on a reduced set of nodes carefully chosen to enable stable reconstruction. Many strategies can be designed and discussed (for instance, see [2, 3]): greedy approaches, random approaches may they be uniform or not, correlated or not, etc.

Objective. This internship will be rooted in these problematics. We will build upon our preliminary work [5] where we identified a particular Determinantal Point Process (DPP) as optimal (in some sense) for graph sampling. The problem is that computing the parameters of this ideal DPP and sampling from it is prohibitive. An elegant first-order approximation of this ideal sampling strategy is to rely on a version of Wilson's algorithm that samples efficiently from a related DPP [1]. During this internship, we will investigate different options that take advantage of the efficiency of Wilson's algorithm while better approaching the ideal sampling strategy. Depending on the candidate, different directions of research are envisioned, between empirical and theoretical studies of this problem.

Curiosity and open-mindedness are very much welcome :-)

Environment. The internship will take place at Gipsa-lab, on the university campus of Grenoble. Partial financial support (approx. EUR 500/month) will be provided. The intern will work with Nicolas Tremblay, Pierre-Olivier Amblard and Simon Barthelmé. Research on high-dimensional statistics and data analysis is lively at Gipsa-lab, and we typically host many interns during the spring semester.
Contact: firstname.lastname@grenoble-inp.fr and 04.76.82.70.26

This is a research-oriented project for both DS and MSCI students. A PhD is not foreseen.

References [open-access versions of the articles are available on Arxiv]

[1] L. Avena and A. Gaudillière. Two Applications of Random Spanning Forests. *Journal of Theoretical Probability (arxiv 1310.1723)*, July 2017.

[2] P. Di Lorenzo, S. Barbarossa, and P. Banelli. Sampling and Recovery of Graph Signals. *arXiv:1712.09310 [eess]*, Dec. 2017. arXiv: 1712.09310.

[3] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst. Random sampling of bandlimited signals on graphs. *Applied and Computational Harmonic Analysis*, pages –, 2016.

[4] D. Shuman, S. Narang, P. Frossard, A. Ortega, and P. Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *Signal Processing Magazine, IEEE*, 30(3):83–98, May 2013.

[5] N. Tremblay, P. O. Amblard, and S. Barthelmé. Graph sampling with determinantal processes. In *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 1674–1678, Aug. 2017.