

Internship proposal: Sampling large networks

Nicolas Tremblay, CR CNRS, Gipsa-lab, Grenoble

General context. In today’s context, data are increasingly measured on irregular network structures: opinions on social networks, brain activity patterns on neuronal networks, traffic on the Internet network, 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 machine learning algorithms in a reasonable amount of computational time. Classically, one of the main approaches to reduce dimensions is sampling: may it be 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 individual hobbies, or blood flow of brain regions, 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 [3, 1]): greedy approaches, random approaches may they be uniform or not, correlated or not, etc.

Objective. This internship will be rooted in these problematics. The actual directions of research will in practice depend on the candidate. It could range from a tentative classification of existing methods from a complexity and algorithmic point-of-view, to theoretical questions on particular random walks on graphs that have been shown recently to sample from a determinantal process [2], and also to provide an approximate graph sampling strategy [5].

A basic knowledge in linear algebra is needed; coding skills in Python (or Matlab) are preferred. Curiosity and open-mindedness are very much welcome :-)

Environment. The internship will take place at Gipsa-lab, on the university campus of Grenoble. A “gratification de stage” (small wage) will be provided to the intern. **Contact:** firstname.lastname@gipsa-lab.grenoble-inp.fr

References [only open-access versions of the articles are provided]

- [1] A. Anis, A. Gadde, and A. Ortega. Efficient sampling set selection for bandlimited graph signals using graph spectral proxies. *arXiv:1510.00297*, 2015.
- [2] L. Avena and A. Gaudillière. Random spanning forests, markov matrix spectra and well distributed points. *arXiv:1310.1723*, 2013.
- [3] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst. Random sampling of bandlimited signals on graphs. *arXiv:1511.05118*, 2015.
- [4] D. I. Shuman, S. K. 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. *arXiv:1211.0053*, 2012.
- [5] N. Tremblay, P.-O. Amblard, and S. Barthelmé. Graph sampling with determinantal processes. *arXiv:1703.01594*, 2017.