Master 2 Internship Proposal (Spring 2022): Random Graphs in Machine Learning

Nicolas Keriven[1] Simone Barthelmé[1] Yohann De Castro[2]
[1]CNRS, Gipsa-lab [2]École Centrale de Lyon

Context. In the last decade, machine learning (ML) on *graph data* has known a rapid growth, with the advent of kernel methods on graphs [6], Graph Neural Networks [1, 7] (GNNs), benchmark datasets [3], and numerous applications ranging from the analysis of social networks to molecular classification or protein interface prediction.

Despite this, the study of traditional ML guarantees such as generalization bounds or sample complexities, which characterize the amount of training data needed to guarantee a low prediction error on new test data, has remained limited in the literature. This is especially true for *node-level tasks* such as Semi-Supervised Learning (SSL) [2], where the goal is to predict node labels on one or several graphs from partially labelled training nodes. This is often due to the lack *statistical modelling* of the graph structure: indeed, when no assumption is made on the graph-generating process to characterize unseen test data, the very notion of generalization becomes ill-defined.

On the other hand, *random graph* (RG) models represent a vast field of study in statistics and graph theory, but have hardly been explored in ML, despite immediate connections. In particular, *Latent Position Models* [4] (LPM: each nodes is associated to an unknown latent variable, and edges are randomly generated according to these variables) such as Stochastic Block Models, graphons or ϵ-graphs, offer striking similarities with classical ML settings, with the significant difference that latent variables are unobserved and must be indirectly deduced from the graph structure. Nevertheless, LPMs allow for instance to study some properties of GNNs in the infinite number of nodes limit [4, 5], but the generalization capacities of GNNs and other models are still mostly unknown.

Goals. The goal of this internship is the explore the use of RG models in ML, and to study quantities such as generalization bounds, sample complexities, and so on, in this context. We will compare several ML models including GNNs, and, depending on the progress of the candidate, several RG models, such as LPMs and preferential attachment models. The internship will balance between theoretical studies and validation on real data, depending on the candidate.

The internship. This internship will take place at the Gipsa-lab in Grenoble, France, and may involve a number of visits to the nearby city of Lyon. A "gratification de stage" (small wage) will be provided to the intern. The usual duration of an M2 internship is 6 months, starting in the Spring of 2022. The internship may be followed by a 3-years PhD funded by the ANR project GRandMa[1].

Application, contact. Please send a CV and short statement of interest to
`nicolas.keriven@cnrs.fr`; `simon.barthelme@grenoble-inp.fr`; `yohann.de-castro@ec-lyon.fr`
Do not hesitate to contact us if you have any question.

[1] `https://nkeriven.github.io`

References

[1] Michael M. Bronstein, Joan Bruna, Yann Lecun, Arthur Szlam, and Pierre Vandergheynst. Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

[2] Olivier Chapelle, Bernhard Schölkopf, and Alexander Zien. *Semi-Supervised Learning*. 2010.

[3] Weihua Hu, Matthias Fey, Marinka Zitnik, Yuxiao Dong, Hongyu Ren, Bowen Liu, Michele Catasta, and Jure Leskovec. Open Graph Benchmark: Datasets for Machine Learning on Graphs. *Neural Information Processing Systems (NeurIPS)*, (NeurIPS):1–34, 2020.

[4] Nicolas Keriven, Alberto Bietti, and Samuel Vaiter. Convergence and Stability of Graph Convolutional Networks on Large Random Graphs. In *Advances in Neural Information and Processing Systems (NeurIPS)*, pages 1–26, 2020.

[5] Nicolas Keriven, Alberto Bietti, and Samuel Vaiter. On the Universality of Graph Neural Networks on Large Random Graphs. In *Advances in Neural Information and Processing Systems (NeurIPS)*, 2021.

[6] Nils M. Kriege, Fredrik D. Johansson, and Christopher Morris. A Survey on Graph Kernels. 2019.

[7] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S. Yu. A Comprehensive Survey on Graph Neural Networks. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–21, 2020.