M2 Internship / PhD proposal

Node features in Graph Machine Learning: benchmarks and new models

In the last few years, **Graph Machine Learning** [1] went from a subfield of Machine Learning (ML) to a primary citizen in the ML landscape. It encompasses many applications in chemistry, biology, or recommender systems, to name a few, and many algorithms and models, from early spectral approaches to complex Graph Neural Networks.

Most theoretical works in Graph ML still relies on classic graph theory, where the graph structure is the primary object of interest. However, it has now become clear that **node features** are an equally important part for prediction, and *the relationship between graph structure and node features* plays a key role that is still poorly understood [2]. For instance, notions such as homophily and heterophily [3] have been identified as playing a crucial role. However, the incorporation of node features in graph ML remains largely open, both in empirical benchmarks and in terms of modelization and theoretical understanding. In particular, statistical models of random graphs, which are crucial in Graph ML theory to characterize properties of learnability and generalization, often ignore node features altogether or are limited to toy models.

The goal of this Master internship, potentially followed by a PhD thesis, is to establish new models in Graph ML that will naturally incorporate node features and graph structure together. This will include a strong benchmarking of different learning algorithms and datasets to better characterize the role of node features, followed by the definition of new statistical models of graphs that include node features and their study, which will illuminate key properties of real-world datasets and provide inspiration for new GNN architectures. The balance between theoretical and empirical studies may depend on the candidate.

Infos. Location: IRISA, Rennes, France. Starting date in 2025. Funded by the MALAGA ERC Starting Grant: `https://nkeriven.github.io/malaga`

Contact. `nicolas.keriven@cnrs.fr`

References

[1] Michael M. Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv:2104.13478*, 2021.

[2] Diana Gomes, Frederik Ruelens, Kyriakos Efthymiadis, Ann Nowe, and Peter Vrancx. When are graph neural networks better than structure-agnostic methods?, 2022.

[3] Derek Lim, Felix Hohne, Xiuyu Li, Sijia Linda Huang, Vaishnavi Gupta, Omkar Bhalerao, and Ser Nam Lim. Large scale learning on non-homophilous graphs: New benchmarks and strong simple methods. *Advances in Neural Information Processing Systems*, 25:20887–20902, 2021.