LIMOS - MAAD axis, AGC theme

Thesis supervisor: Florent Foucaud (MCF), florent.foucaud@uca.fr

Title of PhD subject: Complexity measures and identification in graphs: structure and algorithms

Summary : The aim of this PhD proposal is to study data complexity measures for discrete structures (with a focus on graphs) and related problems, mainly from a theoretical point of view, from the structural and algorithmic perspectives.

Measuring the complexity of input data is crucial in data processing applications, especially as datasets grow exponentially in size, usage, and diversity. Some widely used measures include Vapnik-Chervonenkis dimension (VC-dimension for short) [1], Teaching dimension [4], and neighbourhood complexity [2]. A related field is the classic area of identification problems, where we are given a combinatorial structure, such as a graph, a set system, or a geometric configuration. The goal is to select a (small) set of certain elements of the structure, so that all other elements are uniquely determined by their relationships with that solution set. The goal of this PhD proposal is to study the rich area of graph complexity measures and its connection to identification problems.

On the structural/combinatorial side, several intriguing conjectures and open problems in the area merit further investigation. One such conjecture is the Simon-Zilles conjecture [4], about the optimal relation between recursive teaching dimension and VC-dimension, that is connected to the celebrated sample compression conjecture in theoretical machine learning [6]. Another related conjecture is the n/2-location-domination conjecture [7], a major open problem in the area of graph identification problems. These open problems are related and we expect progress can be made by studying them under a common point of view. We also wish to investigate the neighbourhood complexity of important graph classes, such as planar graphs and graphs of bounded treewidth, for which the optimal values are still unknown [2].

These structural aspects can be complemented by algorithmic aspects. Indeed, though several papers investigate algorithmic aspects of data complexity measures [1] and identification problems [3], many interesting questions remain open, such as the algorithmic complexity of recognizing graphs of n vertices with an identifying code of size exactly log2(n) (equivalently, VC-dimension exactly log2(n)), or the parameterized complexity of data complexity measures with respect to structural parameters such as the treewidth or the vertex cover number of the input graph. Another interesting research direction is to investigate algorithmic applications of neighbourhood complexity, a topic which started to be investigated recently [5] but needs to be explored more.

References

[1] C. Bazgan, F. Foucaud, and F. Sikora. Parameterized and approximation complexity of Partial VC Dimension. Theoretical Computer Science, 766 :1–15, 2019.
[2] L. Beaudou, J. Bok, F. Foucaud, D. A. Quiroz, and J. Raymond. Profile and neighbourhood complexity of graphs with excluded minors and tree-structured graphs. ArXiv preprint 2501.08895, 2025.
[3] D. Chakraborty, F. Foucaud, D. Majumdar, and P. Tale. Tight (double) exponential bounds for identification problems: Locating-dominating set and test cover. Proceedings of ISAAC 2024.
[4] T. Doliwa, G. Fan, H. U. Simon, and S. Zilles. Recursive teaching dimension, VC-dimension and sample compression. Journal of Machine Learning Research, 15(1):3107–3131, 2014.
[5] J. Dreier, I. Eleftheriadis, N. Mählmann, R. McCarty, M. Pilipczuk, and S. Torunczyk. First-order model checking on monadically stable graph classes. Proceedings of FOCS 2024.
[6] S. Floyd and M. K. Warmuth. Sample compression, learnability, and the vapnik-chervonenkis dimension. Machine Learning, 21(3):269–304, 1995.
[7] D. Garijo, A. González, and A. Márquez. The difference between the metric dimension and the determining number of a graph. Applied Mathematics and Computation, 249:487–501, 2014.