Research Project

Graph operators and graph representation learning

Theme: Mathematics for Artificial Intelligence

Responsible: Argyris Kalogeratos | `<argyris.kalogeratosens-paris-saclay.fr>`
Permanent researcher, co-ordinator of the theme *ML on Graphs*

Institution: Centre Borelli, ENS Paris-Saclay | MLMDA research group

Keywords: machine learning on graphs, graph learning, graph representations, embeddings, diffusion operators, graph neural networks, multiview learning

Context and motivation

Graph-based modeling can deal with complex problems involving several interrelated entities, or generally graph-structured data. There are several ways to represent a networked system by a graph, and on the top of that many alternative graph operators (i.e. matrices) for the same input graph. Understandably, the haunting question for a practitioner is: *Which is a good representation for the information the original graph conveys?* On the other side of the same coin, there is the question: *How can I learn and represent a graphical structure from available observed data*. Naturally, both these questions may seek task-dependent answers to maximize the utility of their solutions for the associated real-world problems. These questions, about graph operators and graph learning from data, are the two main directions of this research proposal.

The introduction of Graph Neural Networks (GNNs), which are able to handle graph-structured multidimensional data (i.e. each node carries a feature vector), has brought a lot of attention around graph operators. The interest concerns building efficient operators that go further than the standard adjacency matrix or typical off-the-shelf variants of the Laplacian matrix. Diffusion maps and a series of more sophisticated parametrized random walk-based operators, have started appearing in the the literature for different predictive tasks.

The momentum of GNNs is partially due to the accumulation of techniques and experience coming from the earlier-started deep learning era. However, the research at the level of neural network architectures seems to be saturating around modifications of existing frameworks, usually aiming to respond to the needs of specific applications. Looking for ways to further improve the state-of the-art in the field, we propose to study the role of graph operators as part of a number of modern machine learning frameworks. Moreover, we would like to address a number of challenges that have been highlighted in the literature, such as dealing efficiently with heterogeneous and multiview data, graph directionality, and model downsizing.

Concerning the second direction, the one about learning graphs from data, we should note that this is a problem that is rapidly concentrating a lot interest and attention in the academic and industrial world. The reason is that, when there are available data but no specific idea about a graph structure governing the inherent data correlations or interactions, it is not feasible to exploit the power of graph-based machine learning approaches. There are several problems where experts agree that graph-based modeling and learning is suitable to the nature of their problem of interest, while at the same time they cannot figure out themselves a precise structure (although they may be in position to validate if an inferred graph is meaningful). There is a vast range of applications that could benefit from advances in graph learning.

Scientific objectives

▪ One of the main ambitions is to find principled ways for defining sophisticated graph operators for common predictive tasks, such as node classification, regression, etc, or cluster analysis and data visualization.

▪ Among the intentions it is to develop approaches leading to smaller and more economic models (e.g. networks that require less layers or less training time), which still produce competitive results compared to typical approaches. This would be in contrast to most of the current approaches that leave the graph representation aspect underexploited.

▪ Another direction is the extension of diffusion-based operators to heterogeneous and multi-view data. This latter is particularly interesting in the case where the graph nodes carry vector data with features.

▪ Devising novel graph embedding procedures is also a fundamental axis of the project. Those can be produced: i) either by existing trajectory sampling mechanisms (e.g. node2vec, deepwalk), yet over more sophisticated graph representations; ii) or by devising new mechanisms that can sample trajectories from a graph.

▪ Learning graph from data is the second main ambition of the project. Although in the literature there has been a great deal of work seeing the problem from a graph signal processing viewpoint, the ambition of the project is to go beyond that and use advanced non-parametric statistical machine learning approaches.

A fitting candidate should have an excellent background in general machine learning and in graph-based models. He/She should be eager to do research in a dynamic team where the expected contribution is both at the theoretical and the implementation level. The research will be in the frame of the *Machine Learning on Graphs* research theme of Centre Borelli. The expected output is the production of a number of high-quality articles to be published in ML conferences and journals. In addition, this research will be put in relation to the AI Chair of Centre Borelli, ENS Paris-Saclay, which could offer opportunities to apply our methodologies to real industrial application in collaboration with its partners.

Indicative references

[1] H. Sevi, M. Jonckheere, and A. Kalogeratos. (2022). Clustering for directed graphs using parametrized random walk diffusion kernels. Preprint arXiv:2210.00310.

[2] H. Sevi, M. Jonckheere, and A. Kalogeratos. (2022). Generalized spectral clustering for directed and undirected graphs. Preprint arXiv:2203.03221.

[3] G. Dasoulas, J. Lutzeyer, and M. Vazirgiannis. (2021). Learning parametrised graph shift operators. International Conference on Learning Representations.

[4] T.N. Kipf and M. Welling. (2017). Semi-supervised classification with graph convolutional networks. International Conference on Learning Representations.

[5] A. Grover and J. Leskovec, node2vec: Scalable feature learning for networks. (2016). ACM SIGKDD International Conferene on Knowledge discovery and data mining.

[6] A. Narayanan, M. Chandramohan, R. Venkatesan, L. Chen, Y. Liu, and S. Jaiswal. (2017). graph2vec: Learning distributed representations of graphs. Preprint arXiv:1707.05005.

[7] Boaz Nadler, S. Lafon, R.R. Coifman, I.G. Kevrekidis. (2005). Diffusion maps, spectral clustering and eigenfunctions of Fokker-Planck operators. Advances in Neural Information Processing Systems.

[8] R.R. Coifman and S. Lafon. (2006). Diffusion maps. Applied and Computational Harmonic Analysis 21,1.

[9] O. Lindenbaum, A. Yeredor, M. Salhov, and A. Averbuch. Multi-view diffusion maps. Information Fusion 55.

[10] A. Iravanizad, E.I.S Medina, M. Stoll. (2021) RaWaNet: Enriching Graph Neural Network Input via Random Walks on Graphs. Preprint arxiv:2109.07555.

[11] S. Jeong and C. Donnat. (2022). Tuning the geometry of Graph Neural Networks. Preprint arXiv:2207.05887.

[12] Z. Yang, W. Cohen, and R.Salakhudinov. (2016). Revisiting semi-supervised learning with graph embeddings. International Conference on Machine Learning.

[13] A. Baptista, A. Gonzalez, and A. Baudot. (2022). Universal multilayer network exploration by random walk with restart. Communications Physics 5, 170.

[14] A. Valdeolivas, L. Tichit, C. Navarro, S. Perrin, G. Odelin, N. Levy, P. Cau, E. Remy, and A. Baudot. (2019). Random walk with restart on multiplex and heterogeneous biological networks. Bioinformatics, 35, 3.

[15] M. De Domenico, A. Solé-Ribalta, S. Gómez, and A. Arenas. (2014). Random walks on multiplex networks: Supplementary information for "Navigability of interconnected networks under random failures". PNAS.

[16] P. Humbert, B. Le Bars, L. Oudre, and A. Kalogeratos. Learning Laplacian Matrix from Bandlimited Graph Signals, B. Le Bars, P. Humbert, L. Oudre, and A. Kalogeratos, IEEE International Conference on Acoustics, Speech, and Signal Processing.

[17] P. Humbert, B. Le Bars, L. Oudre, A. Kalogeratos, and N. Vayatis. (2021). Learning Laplacian Matrix from Graph Signals with Sparse Spectral Representation. Journal of Machine Learning Research.

[18] A. de la Concha, N. Vayatis, and A. Kalogeratos. (2022). Collaborative likelihood-ratio estimation over graphs, arxiv preprint (under journal review).