

# Representation and contrastive learning on graphs

**Supervisor(s):** Harry Sevi and Argyris Kalogeratos

**Contact:** <fname.sname@ens-paris-saclay.fr>

**Project type:** Research

**Theme:** Graph Machine Learning

**Group:** Machine Learning and Massive Data Analysis (MLMDA)

**Lab:** Centre Borelli, ENS Paris-Saclay, Gif-sur-Yvette

**Funding:** under the auspices of the Industrial Data Analytics & Machine Learning Chair (IdAML)

## Description

In the literature, numerous graph neural network (GNN) models have been proposed for graph-related tasks, such as node classification [1], link prediction, and graph classification. Most existing GNN approaches use semi-supervised training. For many real-world graph applications, e.g. protein analysis, they intuitively require the input of some amount of labeled data.

Alternatively, there is a series of random walk-based GNNs, including node2vec [2] and graph2vec [3], which are unsupervised. Their approach is to first learn the node embeddings, and then various supervised downstream tasks are directly applied on these node embeddings. These approaches can be considered as part of the contrastive learning framework.

Contrastive learning [4] originally aims to learn to embed each image in a self-supervised manner. Due to its impressive performance in many tasks, contrastive learning has become the hottest topic in unsupervised learning. Its motivation is to maximize the similarity of positive pairs and the distance of negative pairs. Generally speaking, the positive pairs are composed of data augmentations of the same instance, while those of different instances are regarded as negative pairs [5].

This internship aims to investigate the last findings in graph embedding and graph contrastive learning, understand, deconstruct the different steps, and make some steps towards new contrastive learning approaches for graphs.

- See references: [6, 7, 8, 9, 10]

**Keywords:** Graph neural networks, contrastive learning, random walks, graph embeddings, unsupervised learning.

## Indicative references

- [1] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.
- [2] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” in *Proc. of the ACM SIGKDD Inter. Conf. on Knowledge discovery and data mining*, 2016, pp. 855–864.
- [3] A. Narayanan, M. Chandramohan, R. Venkatesan, L. Chen, Y. Liu, and S. Jaiswal, “graph2vec: Learning distributed representations of graphs,” *arXiv preprint arXiv:1707.05005*, 2017.
- [4] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, “A simple framework for contrastive learning of visual representations,” in *Proc. of the Inter. Conf. on Machine Learning*. PMLR, 2020, pp. 1597–1607.
- [5] Z. Yang, M. Ding, C. Zhou, H. Yang, J. Zhou, and J. Tang, “Understanding negative sampling in graph representation learning,” in *Proc. of the ACM SIGKDD Inter. Conf. on Knowledge Discovery & Data Mining*, 2020, pp. 1666–1676.
- [6] U. Shaham, K. Stanton, H. Li, B. Nadler, R. Basri, and Y. Kluger, “Spectralnet: Spectral clustering using deep neural networks,” *arXiv preprint arXiv:1801.01587*, 2018.
- [7] J. Qiu, Y. Dong, H. Ma, J. Li, K. Wang, and J. Tang, “Network embedding as matrix factorization: Unifying deepwalk, line, pte, and node2vec,” in *Proc. of the ACM Inter. Conf. on web search and data mining*, 2018, pp. 459–467.

- [8] H. Zhong, J. Wu, C. Chen, J. Huang, M. Deng, L. Nie, Z. Lin, and X.-S. Hua, “Graph contrastive clustering,” *arXiv preprint arXiv:2104.01429*, 2021.
- [9] A. Nazi, W. Hang, A. Goldie, S. Ravi, and A. Mirhoseini, “Generalized clustering by learning to optimize expected normalized cuts,” *arXiv preprint arXiv:1910.07623*, 2019.
- [10] S. Chen, Z. Wang, X. Zhang, X. Zhang, and D. Peng, “Probing negative sampling for contrastive learning to learn graph representations,” in *Joint European Conf. on Machine Learning and Knowledge Discovery in Databases*. Springer, 2021, pp. 434–449.

**Doors open afternoon – 14 December 15h00-18h00:** The members of the Graph Machine Learning team will be happy to welcome the interested students at our lab. We can talk about this subject and the research opportunities at Centre Borelli. Please contact Argyris Kalogeratos to express your interest and allocate a time slot.