Pre-Training of Graph Neural Networks: What has been achieved by now, and what can be done with it?

Jonatan Frank
jonatan.frank@uni-ulm.de
Ulm University
Ulm, Germany

1 INTRODUCTION

Pre-training has already had great success in natural language processing [2] and computer vision [3]. The idea is to use first abundant, mostly unlabeled, data for pre-training the learning model. This data differs from the downstream task data, which is usually labeled and in short supply. Afterward, one uses downstream task data for fine-tuning the pre-trained model. This usually improves performance, and one can reuse the pre-trained model for other tasks. One just fine-tunes it for a task. Hence, one can save training time by reusing it.

Intuitively, one could assume that pre-training is also beneficial for graphs. Therefore, pre-training a Graph Neural Network (GNN) should improve the performance. However, this is not always the case. There are some approaches of pre-training GNNs [4–8], but one can not simply use them and expect them to benefit the performance, because pre-training a GNN can also lead to negative transfer. Therefore, it is interesting to know when this happens because if it happens, one could simply save one's time in pre-training the model. The work by Cao et al. [1] tries to answer this by looking into the question: When to pre-train Graph Neural Networks? One of their points is that for pre-training, the pre-training data and the downstream data should be from the same domain and have topological similarities. They also deny the importance of attributes and proximity in the graphs, but one can argue that this is not always true.

Instead of simply trying to answer the question *when to pre-train*, I try to find answers to the questions *what to pre-train* and *how to pre-train* because that also answers the*when to pre-train* question. For that, I look deeper into works that are already offering approaches for pre-training GNNs. I try to answer the following questions about these approaches and works:

- How do they pre-train?
- What information do they use to pre-train?
- What are the similarities and differences between the approaches?
- How can they be combined, or what can be used for a new one?

By answering the last question, I try to develop a new pre-training approach.

Structure of the exposé.

2 METHODS [OR MODELS]

Clarify the assumptions that are made

Structure:

(1) Explain: [4–8]
(2) Answer the questions by using the explanation above and propose/explain possible experiments. (For the experiments, use at least one dataset of each work. **Limitations in answering approach**)

For all papers, except [8], the code is publically available.

- How do they pre-train?
- What information do they use to pre-train?
- What are the similarities and differences between the approaches?
- How can they be combined, or what can be used for a new one?

2.1 [Problem Statement/Problem Formalization]

2.2 Assumptions

3 DATASETS

Clarify the assumptions that are made

I looked into some datasets, and most of them are available. I have not yet downloaded anything, but in the worst case that, surprisingly, one paper has only undownloadable datasets, I try to switch to other datasets. I will try to download them this week.

4 RELATED WORK

About the five pre-training approaches ([4–8]), including their related work and separated in generative and discriminative.

REFERENCES

[1] Yuxuan Cao, Jiarong Xu, Carl Yang, Jiaan Wang, Yunchao Zhang, Chunping Wang, Lei Chen, and Yang Yang. 2023. When to Pre-Train Graph Neural Networks? An Answer from Data Generation Perspective! arXiv:2303.16458 [cs.LG]

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, Jill Burstein, Christy Doran, and Thamar Solorio (Eds.). Association for Computational Linguistics, 4171–4186. https://doi.org/10.18653/v1/n19-1423

[3] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross B. Girshick. 2020. Momentum Contrast for Unsupervised Visual Representation Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*. Computer Vision Foundation / IEEE, 9726–9735. https://doi.org/10.1109/CVPR42600.2020.00975

[4] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay S. Pande, and Jure Leskovec. 2020. Strategies for Pre-training Graph Neural Networks. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net. https://openreview.net/forum?id=HJlWWJSFDH

[5] Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. 2020. GPT-GNN: Generative Pre-Training of Graph Neural Networks. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash (Eds.). ACM, 1857–1867. https://doi.org/10.1145/3394486.3403237

[6] Yuanfu Lu, Xunqiang Jiang, Yuan Fang, and Chuan Shi. 2021. Learning to Pre-train Graph Neural Networks. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational*

Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021. AAAI Press, 4276–4284. https://ojs.aaai.org/index.php/AAAI/article/view/16552

[7] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. 2020. GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash (Eds.). ACM, 1150–1160. https://doi.org/10.1145/3394486.3403168

[8] Simiao Zuo, Haoming Jiang, Qingyu Yin, Xianfeng Tang, Bing Yin, and Tuo Zhao. 2022. DiP-GNN: Discriminative Pre-Training of Graph Neural Networks. *CoRR* abs/2209.07499 (2022). https://doi.org/10.48550/arXiv.2209.07499 arXiv:2209.07499