[ICML 2020] When Does Self-Supervision Help Graph Convolutional Networks? [paper] [code]

Node/Graph Tasks: Node classification

Training Type: pre-training and fine-tuning, self-training, multi-task learning (also known as joint training). Specifically, self-training starts by pretraining a model over the labeled data, then assigning pseudo-labels to highly confident unlabled samples and including them into the labeled data for the next round of training. The pretext tasks are divided into follows:
Node clustering: node features

The pretext task here is to train the GNN model to extract features that encode node clustering information. Node clusters can be obtained by running any clustering-related algorithms based on feature similarity and we want the embeddings extracted by the learnt GNN model could encode such similarity equivalence.
Graph partitioning: graph topology

The pretext task here is to train the GNN model to extract features that encode node partition information. Node partitioning can be obtained by minimizing the edge cut. Note that graph partition provide prior regularization based on graph topology rather than node features as what node clustering does.
Graph completion: graph topology, node features

The pretext task here is to train the GNN model to extract features that encode node feature information. The graph completion first masks target nodes by removing their features and then aims at recovering masked node features by feeding to GCNs unmasked node features.

Initial short summary here
This paper provides a systematic research on incorporating the self-supervision into GNNs, which integrates the study of three training schemes and three pretext tasks, and examine the roles of SSL in gaining robustness against various graph adversarial attacks. Since the three training scheme and pretext tasks have been covered previously, we only focus on the role of robustness gaining by SSL and the results part.

The result that pretraining-finetuning provides some performance improvement for the small dataset Cora but does not do so for the larger datasets Citeseer and Pubmed suggests that information learnt in SSL may be largely lost during finetuning. The difference between self-training and the multi-task training is that self-training generate peudo-labels based on node proximity in embedding space while the multi-task training generates pseudo-labels not only based on proximity in embedding space but also based on graph structure and node features. The self-training benefits more in the few-shot learning. Therefore, multi-task learning can be more general. The numerous results also demonstrate that different SSL tasks could benefit different network architectures on different datasets to different extents.

Bibtex:

@inproceedingsyou2020does, title=When does self-supervision help graph convolutional networks?, author=You, Yuning and Chen, Tianlong and Wang, Zhangyang and Shen, Yang, booktitle=International Conference on Machine Learning, pages=10871–10880, year=2020, organization=PMLR