[Arxiv 2020] Self-Supervised Graph Representation Learning via Global Context Prediction. [paper]

Node/Graph Tasks: Node classification, link prediction and clustering

Training Type: pre-training by pretext tasks and then the resulted representations are feed into the logistic regression for classification purpose

Pretext task data: graph topology, node features
The pretext task here is to train a GNN model to extract node embeddings that could be used to reconstruct the topological distance (hop count) in the original graph between two nodes. Suppose that $\mathcal{C}_i$ includes a set of nodes that can reach node v_i through the shortest path with different hop counts, i.e., $\mathcal{C}_i$ is composed of multiple specific k-hop context $\mathcal{C}_i^K$ which only contains nodes within k hops:

$$\mathcal{C}_i = \mathcal{C}_i^1 \cup \mathcal{C}_i^2 \cup \cdots \cup \mathcal{C}_i^{\delta_i}, \quad \mathcal{C}_i^k = \{v_j | d_{ij} = k\}, \quad k = 1, 2, \cdots, \delta_i, \tag{49}$$

where δ_i is the upper bound of the hop count from other nodes to v_i in the graph and d_{ij} is the length of path p_{ij}. To guide our GNN model to extract node embeddings that encode the node topological distance information, we optimize the following objective:

$$\min_{\omega,\theta} \sum_{v_i \in \mathcal{V}} \sum_{v_j \in \mathcal{C}_i} \mathcal{L}(\tilde{Y}_j, h_\theta(< f_\omega(v_i), f_\omega(v_j) >)), \tag{50}$$

where f_ω can be any GNN-based encoder parametrized by ω, h_θ is a classifier to predict the pseudo-labels with θ as the parameter. $<,>$ is used to measure the interaction between pairs of nodes and should be symmetric and we take element-wise distance, i.e., $< \mathbf{z}_i, \mathbf{z}_j >= abs(\mathbf{z}_i - \mathbf{z}_j)$. $\tilde{Y}_j$ is the category of node j set up based on its topological distance to node i.

Initial short summary here:
Following the traditional idea of incorporating SSL into GNNs, this paper provides a self-supervised graph representation learning framework S^{2}GRL involving predicting relative contextual position for a pair of nodes in a graph. By optimizing the Eq. (50), the features extracted from our GNN encoder could be used to predict the node relative contextual position.

Since the upper bound of hop count (topological distance) for different target nodes varies and precisely determining this upper bound is time-consuming for a big graph, we assume that the number of hops (distance) is under control based on small-world phenomenon and further divide the distance into several major categories that clearly discriminating the dissimilarity and partly tolerating the similarity. Another problem is that nodes that are closer to one node is far less than nodes that are far away from one node. To circumvent this imbalance problem, node pairs are sampled with adaptive ratio.

In node classification, S^{2}GRL outperforms all other unsupervised algorithms on all datasets. Besides, S^{2}GRL exhibits comparable results to some supervised

models like GCN and GWNN. In clustering, although DGI achieves the best performance on Cora and Citeseer, S^2GRL also exhibits competitive performance. In link prediction, S^2GRL consistently outperforms DGI and node2vec under different edge removal rates. Besides, the work also investigates how the quality of the self-supervised learned embeddings depends on the construction of major categories of distance. Distinguishing 1, 2, 3-hop contexts into 3 distinct major categories benefits the node representations while further differentiating 4-hop and higher-hop contexts would degrade the performance.

Bibtex:

@articlepeng2020self, title=Self-supervised graph representation learning via global context prediction, author=Peng, Zhen and Dong, Yixiang and Luo, Minnan and Wu, Xiao-Ming and Zheng, Qinghua, journal=arXiv preprint arXiv:2003.01604, year=2020