**[Arxiv 2020] Self-supervised Training of Graph Convolutional Networks. [paper]**

**Node/Graph Tasks:** Node classification on citation networks

**Training Type:** Pretraining by the task of graph topology reconstruction and finetuning by the task of node classification on citation networks

**Pretext task data:** Structure, node features
**The first pretext task** is to randomly removes parts of the edges inside the graph, then requires the GCN model to predict the graph link of the input graph. **The second pretext task** is to randomly cover parts of the features during training and supervise the model to predict the links between nodes.

**Initial short summary here**
The major limitation of convolutional neural networks is the massive amount of annotated data required in training. To tackle this challenge, this paper proposes two types of self-supervised learning strategies: randomly removing links (RRL) and randomly covering features (RCF) to improve the discriminating ability of the GNN based model. The graph model in both of the two pretext tasks are simply the GCN or the GAT models []. After pre-training by performing the two pretext tasks, the model is then fine-tuned by the traditional node classficiation task on citation networks.

The task RRL is to recover the graph topology (link prediction) after randomly removing edges in the graph. After we obtain the latent feature embedding for each node $\mathbf{H}_i$, the edge between any pair of nodes $i, j$ is calculated by their feature similarity as Eq. (16) and the weighted cross-entrophy loss $\mathcal{L}$ is used during training as Eq. (17).

$$A'_{ij} = \text{sigmoid}(\mathbf{H}_i(\mathbf{H}_i)^\top) \tag{16}$$

$$\mathcal{L} = -\sum_{i,j \in \mathcal{V}^{\mathcal{G}}} W(A_{ij} \log A'_{ij}) + (1 - A_{ij}) \log(1 - A'_{ij}) \tag{17}$$

where $W$ is the weight hyperparameter used for balancing two classes, node pairs with edges and node pairs without edges.

To explore the contributions of RRL and RCF to improving the model performance, the authors evaluate the performance of GCN on the Cora dataset with three SSL strategies: independent RRL and RCF, mixed RRL and RCF. The results demonsrate the outperformance of utilizing two pretext tasks.

**Bibtex:**
@articlezhu2020self, title=Self-supervised Training of Graph Convolutional Networks, author=Zhu, Qikui and Du, Bo and Yan, Pingkun, journal=arXiv preprint arXiv:2006.02380, year=2020 *copy/page the bibtex here*