Graph Contrastive Learning Automated

Yuning You¹, Tianlong Chen², Yang Shen¹, Zhangyang Wang²

¹Texas A&M University, ²University of Texas at Austin

➢ Motivations

- Unlike images, graph data are inherently heterogeneous (e.g. pandemics, product co-purchase relation, molecules).
- The SOTA GraphCL [1] handles the heterogeneity with ad-hoc choices of augmentation for every datasets.
- The rules ofthumb for selection are derived from tedious tuning.

Question: Given a new and unseen graph dataset, can GraphCL select augmentations in a principled and automated fashion?

Method. Joint Augmentation Optimization

- We propose the unified joint augmentation optimization framework (JOAO) for automatic augmentation selection (specifically for the sampling distribution \( \mathbb{P}(\mathbf{A}_1, \mathbf{A}_2) \)), as a bi-level optimization (Eq. (2)).

\[
\min_{q(G, A_1, A_2, \theta)} \mathbb{L}(q(G, A_1, A_2, \theta)) \quad \text{s.t.} \quad \mathbb{P}(A_1, A_2) \in \arg \max_{q(G, A_1, A_2, \theta)} \mathbb{E}[\log p(G, \mathbf{A}_1, \mathbf{A}_2, \theta)],
\]

\[
\min_{q(G, A_1, A_2, \theta)} \mathbb{L}(q(G, A_1, A_2, \theta)) \quad \text{s.t.} \quad \mathbb{P}(A_1, A_2) \in \arg \max_{q(G, A_1, A_2, \theta)} \mathbb{E}[\log p(G, \mathbf{A}_1, \mathbf{A}_2, \theta)],
\]

- Motivated from model-based adversarial training, we propose a min-max Instantiation (Eq. (3)), that in the lower-level optimization:
  - It always tries to exploit the most challenging augmentation;
  - We regularize it with prior distribution.

We solve Eq. (3) with alternating gradient descent (AGD), with the empirical convergence demonstrated.

Method. Augmentation-Aware Multi-Projection

- JOAO makes automatic and dynamic selection, but also aggressive and diverse, potentially leading to training distortion [2,3]. We introduce multiple projection heads with augmentation-aware selection scheme to address it, referred as JOAOv2.

Figure 4: An overview of GraphCL with multiple augmentation-aware projection heads where \( p_{\text{aug-avg}} \) refers to \( \mathbb{P}(A_1, A_2) \).

Selection of JOAO aligns with previous “best practices”.

- How reasonable are the JOAO-selected augmentation pairs per dataset?
- We observe that the selection of JOAO is generally consistent with previous “best practices”.
- While JOAO is free from manual tuning as in GraphCL

Experiments

- We numerically show that JOAO is comparable with SOTA competitors on datasets from diverse sources, and generalizes better than GraphCL on domain specific / large-scale datasets.

Table 4: Semi-supervised learning on TUDataset. Shown in \( \text{red} \) are the best accuracy (%) and those within the standard deviation of the best accuracy or the best average ranks. - indicates that label rate is too low for a given dataset size. L.R. and A.R. are short for label rate and average rank, respectively. The compared results except those for ConvNet/Préd are as published under the same experiment setting.

Table 7: Semi-supervised learning on large-scale OGB datasets. Bold numbers indicate the top-3 performances (accuracy in % on ogbg-ppa, F1 score in % on ogbg-code).