Structure-Enhanced Heterogeneous Graph Contrastive Learning Supplementary Material

Yanqiao Zhu[*†‡] Yichen Xu[*§] Hejie Cui[¶] Carl Yang[¶] Qiang Liu[†‡] Shu Wu[†‡‖]

A Pseudo-code for Training STENCIL

We summarize the training procedure of the proposed STENCIL in Algorithm 1.

Algorithm 1 The STENCIL framework

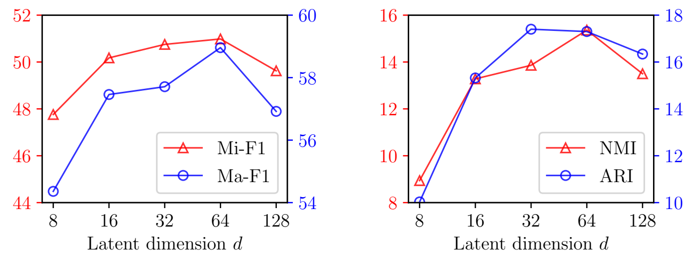
Construct multiple semantic views corresponds to metapath $p \in \mathcal{P}$
for *epoch* $\leftarrow 1, 2, \cdots$ **do**
    ▷ *Heterogeneous graph encoding* ◁
    Obtain node embeddings of each metapath-induced view $\boldsymbol{H}^p$ according to Eq. (2.1)
    Obtain aggregated embeddings $\boldsymbol{H}$ according to Eq. (2.3)
    ▷ *Structure-enhanced negative mining* ◁
    Compute hardness score $s(i, j, p)$ for each negative-anchor pair
    Sort S in ascending order
    Pick T negative nodes with the highest S in each metapath-induced view
    Synthesis M hard negative samples via Eq. (3.9)
    Update the negative bank $\mathcal{B}$ according to Eq. (3.11)
    ▷ *Model training via multiview contrastive aggregation* ◁
    Update parameters by applying stochastic gradient descent to minimize $\mathcal{J}$ as in Eq. (3.12)

B Dataset Details

For comprehensive comparison, we use three widely-used heterogeneous datasets from various domains: DBLP, ACM, and IMDb, where DBLP and ACM are two academic networks, and IMDb is a movie network.

- **DBLP** is a subset of an academic network extracted from DBLP, consisting of four kinds of nodes: authors, papers, conferences, and topics. The authors are selected from four domains: database, data mining, machine learning, and information retrieval. Each author is labeled with their research area according to the conferences they submitted, and is associated with bag-of-word features which represent keywords.
- **ACM** is an academic network extracted from papers published in KDD, SIGMOD, SIGCOMM, MobiCOMM, and VLDB. We construct a heterogeneous graph with nodes of three types: papers, authors, and subjects. Papers with bag-of-words of features are classified into three themes according to their corresponding research topic.
- **IMDb** is a subset of the movie network IMDb, where nodes represent movies, actors, or directors. We categorize movies into three classes according to their genre. Each movie node is associated with a bag-of-words feature representing plots.

[*] The first two authors contributed equally to this research.
[†] Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China.
[‡] School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing, China.
[§] School of Computer Science, Beijing University of Posts and Telecommunications, Beijing, China.
[¶] Department of Computer Science, Emory University, GA, USA.
[‖] To whom correspondence should be addressed (shu.wu@nlpr.ia.ac.cn).



(a) Node Classification (b) Node Clustering

Figure S1: Model performance with varied latent dimensions.

C Implementation Details

The proposed model is implemented using PyTorch [4], DGL [6], PyTorch Geometric [1], and PyGCL [7]. We use Adam optimizer [3] with learning rate set to 0.01, 0.0005, and 0.001 for ACM, IMDb and DBLP respectively and ℓ_2 regularization set to 10^{-5}. The model is trained for at most 3,000 epochs and is early-stopped if the training loss does not improve for 100 consecutive epochs. The dropout rate [5] is set to 0.2 on all datasets. We use 8 attention heads and the embedding size is 64 for both STENCIL and baselines for fair comparison. Furthermore, we set the temperature parameter τ to 0.9 in the contrastive objective. The number of synthesize samples M is set to 200. All parameters are initialized with Glorot initialization [2].

D Sensitivity Analysis

Additionally, we perform sensitivity study on one key hyperparameter in our proposed STENCIL model, namely the dimension of hidden representation d. Note that all other parameters described previously remain the same while we are altering a specific parameter. Two downstream tasks, node classification and node classification, are included using the corresponding evaluation metrics and the results are on the IMDb dataset.

We show the influence of varied node latent dimensions d on STENCIL in Figure S1. It is observed that at initial stages, the performance of STENCIL on both two tasks improves noticeably as the latent dimension increases. This is because that the model can encode richer information with larger dimension size, which facilitate the performance on various downstream tasks. However, as d continues to grow, the improvements in terms of Micro-F1 and ARI become less evident and finally the performance on two tasks begin to drop. The reason may be that with more latent dimensions, the model gets sized-up and harder to train, which possibly leads to under-fitting with the same amount of training samples. Therefore, we need to choose a moderate and appropriate dimension d to balance the expressiveness and the capacity of the model.

References

[1] M. Fey and J. E. Lenssen, *Fast Graph Representation Learning with PyTorch Geometric*, in RLGM@ICLR, 2019.

[2] X. Glorot and Y. Bengio, *Understanding the Difficulty of Training Deep Feedforward Neural Networks*, in AISTATS, 2010, pp. 249–256.

[3] D. P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, in ICLR, 2015.

[4] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, *PyTorch: An Imperative Style, High-Performance Deep Learning Library*, in NeurIPS, 2019, pp. 8024–8035.

[5] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, *Dropout: A Simple Way to Prevent Neural Networks From Overfitting*, JMLR, 15 (2014), pp. 1929–1958.

[6] M. Wang, D. Zheng, Z. Ye, Q. Gan, M. Li, X. Song, J. Zhou, C. Ma, L. Yu, Y. Gai, T. Xiao, T. He, G. Karypis, J. Li, and Z. Zhang, *Deep Graph Library: A Graph-Centric, Highly-Performant Package for Graph Neural Networks*, arXiv.org, (2019), https://arxiv.org/abs/1909.01315v2.

[7] Y. Zhu, Y. Xu, Q. Liu, and S. Wu, *An Empirical Study of Graph Contrastive Learning*, in NeurIPS Datasets and Benchmarks, 2021.