Rethinking Feature Augmentation In Graph Contrastive Learning

Congfeng Cao[1], Pengyu Zhang[2], Zeyu Zhang[2], and Jelke Bloem[1]

[1] Institute for Logic, Language and Computation, University of Amsterdam
[2] Informatics Institute, University of Amsterdam
{C.Cao, P.Zhang, Z.Zhang2, J.Bloem}@uva.nl

Abstract. Graph Contrastive Learning (GCL) has emerged as a powerful framework for graph representation learning. GCL typically employs separate masking strategies for edges and node features. However, the stochastic Masking Node Feature (MF) method, which masks a portion of the columns in the node feature matrix, results in irrecoverable feature information loss at high masking rates. In other words, MF harms the uniformity of representations. To address this, we introduce a novel augmentation strategy called Random Feature Masking (RFM) for GCL. Unlike MF, RFM applies random masking across the entire set of node features for each individual node. Experiments on three widely used datasets for node classification demonstrate that RFM enables GCL to outperform the MF method, achieving higher accuracy, and greater robustness, even at high masking rates (e.g., 0.7, 0.8, and 0.9). Since RFM does not mask a fixed fraction of the entire node feature matrix, it inherently preserves more feature information. To our best knowledge, this is the first study to introduce and comprehensively evaluate Random Feature Masking in GCL.

Keywords: Graph Contrastive Learning · Augmentation Strategies · Random Feature Masking (RFM).

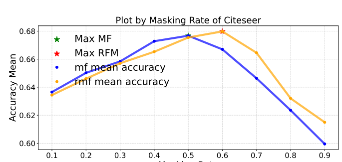
1 Introduction

Graph Contrastive Learning (GCL) learns graph representations by maximizing the agreement between augmented views of a graph through contrastive objectives like InfoNCE, by leveraging graph encoders and augmentations for tasks such as node and graph classification [1, 5, 7]. GCL follows a specific pipeline. First, it generates multiple views of a graph using various augmentation strategies. Next, two views derived from the same node are treated as a positive pair, while views derived from different nodes are treated as negative pairs. The optimization objective of contrastive learning is to maximize the agreement between jointly sampled positive pairs while minimizing the agreement between independently sampled negative pairs.

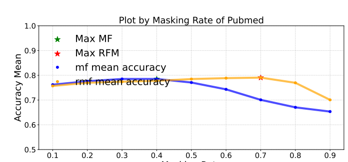
Augmentation strategies are crucial in GCL. Feature-based augmentation modifies the node feature matrix using techniques such as masking [4, 8] and shuffling [3]. Typically, these augmentations rely on stochastic Masking Node Feature (MF) methods, which mask a fraction of attributes in the node feature matrix, as seen in approaches like GRACE [10] and CCA-SSG [9]. However, MF suffers from inherent drawbacks: masking a portion of the feature matrix results in irreversible feature information loss at high masking rates, leading to performance degradation, which is not compatible with conclusions from contrastive learning in other modalities [2, 6].

To address these challenges, we introduce a feature-based augmentation method in GCL: Random Feature Masking (RFM). RFM is a simple yet effective strategy which is widely used in graph generation models but not used in graph contrastive learning. Unlike MF, RFM applies random masking across the entire set of node features for each individual node. Theoretically, $1 - \left(\prod_{i=1}^{N} p_i\right)^e$ of the feature information can be preserved in one view during RFM, where N denotes the number of nodes, p denotes the mask rate and e denotes the number of epochs.

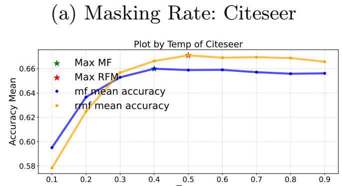
We compare our method to the Masking Node Feature (MF) method on the widely adopted GCL task of node classification, using three benchmark datasets. As can be seen from Figure 1: **1.** It shows that RFM outperforms existing MF methods, achieving superior performance at high masking rates. Notably, our approach attains state-of-the-art results on the PubMed dataset for node classification when mask rate reaches 0.7. **2.** As the masking rate increases, the performance of GCL with MF initially improves at low masking rates but decreased after mask rate larger than 0.4. In contrast, the performance of GCL with RFM consistently improves, even at masking rates as high as 0.7. Additionally, RFM consistently achieves the highest results across all scenarios.**3.** RFM is more tolerant to variations in temperature, with RFM generally achieving higher performance than MF across different temperature settings.



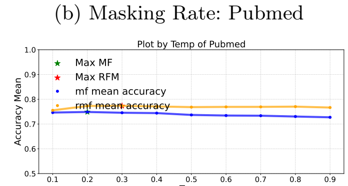
(a) Masking Rate: Citeseer



(b) Masking Rate: Pubmed



(c) Temperature: Citeseer



(d) Temperature: Pubmed

Fig. 1: Comparison of Masking Rate (Top Row) and Temperature (Bottom Row) Across Datasets (Citeseer and Pubmed).

References

1. Hassani, K., Khasahmadi, A.H.: Contrastive multi-view representation learning on graphs. In: International conference on machine learning. pp. 4116–4126. PMLR (2020)
2. Huang, Z., Jin, X., Lu, C., Hou, Q., Cheng, M.M., Fu, D., Shen, X., Feng, J.: Contrastive masked autoencoders are stronger vision learners. IEEE Transactions on Pattern Analysis and Machine Intelligence (2023)
3. Jing, B., Park, C., Tong, H.: Hdmi: High-order deep multiplex infomax. In: Proceedings of the Web Conference 2021. pp. 2414–2424 (2021)
4. Thakoor, S., Tallec, C., Azar, M.G., Munos, R., Veličković, P., Valko, M.: Bootstrapped representation learning on graphs. In: ICLR 2021 Workshop on Geometrical and Topological Representation Learning (2021)
5. Velickovic, P., Fedus, W., Hamilton, W.L., Liò, P., Bengio, Y., Hjelm, R.D.: Deep graph infomax. ICLR (Poster) **2**(3), 4 (2019)
6. Wettig, A., Gao, T., Zhong, Z., Chen, D.: Should you mask 15% in masked language modeling? In: Vlachos, A., Augenstein, I. (eds.) Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. pp. 2985–3000. Association for Computational Linguistics, Dubrovnik, Croatia (May 2023). https://doi.org/10.18653/v1/2023.eacl-main.217, https://aclanthology.org/2023.eacl-main.217/
7. Xu, D., Cheng, W., Luo, D., Chen, H., Zhang, X.: Infogcl: Information-aware graph contrastive learning. Advances in Neural Information Processing Systems **34**, 30414–30425 (2021)
8. You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z., Shen, Y.: Graph contrastive learning with augmentations. Advances in neural information processing systems **33**, 5812–5823 (2020)
9. Zhang, H., Wu, Q., Yan, J., Wipf, D., Yu, P.S.: From canonical correlation analysis to self-supervised graph neural networks. Advances in Neural Information Processing Systems **34**, 76–89 (2021)
10. Zhu, Y., Xu, Y., Yu, F., Liu, Q., Wu, S., Wang, L.: Deep Graph Contrastive Representation Learning. In: ICML Workshop on Graph Representation Learning and Beyond (2020), http://arxiv.org/abs/2006.04131