

Response to Reviewers

Ziyang Liu¹, Chaokun Wang¹, Liqun Yang², Yunkai Lou¹, Hao Feng¹, Cheng Wu¹, Kai Zheng³, Yang Song³

¹Tsinghua University, Beijing, China; ²Nanhu Academy, Jiaxing, China; ³Kuaishou Technology, Beijing, China

{liu-zy21,louyk18,fh20,wuc22}@mails.tsinghua.edu.cn, chaokun@tsinghua.edu.cn,
yangliqun@cnaeit.com, zhengkai@kuaishou.com, ys@sonyis.me

Firstly, we would like to thank the meta-reviewer and reviewers for your insight and valuable comments to our ICDE 2024 submission (661). We have carefully taken your comments into consideration in preparing this revision. Below we briefly summarize how we have specifically addressed the comments of each reviewer in our revised manuscript. We believe that our response has sufficiently addressed all the issues raised by the reviewers. In our revised paper, all the modifications are marked in blue. In the responses, the deleted, updated, and added texts are marked by strike-through, orange, and blue, respectively.

I. RESPONSE TO META-REVIEWER

A. Area Chair's Summary

Comment: The reviewers found some merits of this submission. However, some concerns are also raised. We are happy to invite the authors to revise the paper and address all the comments and concerns from the reviewers.

Response: Thanks for your kind suggestion. We have replied to all the comments from the reviewers and finished all the required experiments suggested by the reviewers:

- We have addressed two reviewers' confusion regarding the connection between this paper and data management/mining problems, and added more references from major database conferences (Reviewer #1 O1, Reviewer #3 O6).
- We have incorporated the method with the same τ setting as GRACE into our ablation studies (Reviewer #2 O1).
- We have deployed GLATE with more baseline methods (DGI, SUGRL, and NCLA) in the inductive node classification task to verify its effectiveness (Reviewer #2 O2).
- We have conducted the sensitivity analysis to more hyperparameters in GLATE, including the initial momentum, momentum parameter, similarity threshold, temperature learning rate, and number of sampled nodes (Reviewer #2 O3).
- We have reviewed the entire paper and addressed the issues with the figures, e.g., Fig. 4 (Reviewer #2 O4).
- We have added more efficiency analysis of GLATE to demonstrate the high efficiency of temperature fine-tuning. Also, the newly added sensitivity analysis experiments on different types of datasets show that a general parameter configuration can achieve satisfactory performance when training our method on a new graph dataset without excessive parameter tuning time (Reviewer #3 O1).
- We have evaluated our method and demonstrated its versatility by varying the density, connectivity patterns, and node/edge attributes in graphs (Reviewer #3 O2).

- We have added the scalability analysis to demonstrate that GRACE+GLATE is scalable to large-scale graphs. The experimental results on time costs and memory peak are consistent with our complexity analysis (Reviewer #3 O3).
- We have applied our method to more downstream tasks including link prediction, graph classification, and node clustering, demonstrating its potential for diverse graph-related tasks in real-world applications (Reviewer #3 O4).
- We have added the case study to interpret the learned embeddings in molecule graphs. The analysis indicates that the embeddings learned by GRACE+GLATE align with fundamental chemical knowledge well (Reviewer #3 O5).

II. RESPONSE TO REVIEWER #1

A. O1

Comment: The paper is a machine learning paper. The main body of the paper is about how to improve the training process of graph contrastive learning. It is not specific for a data engineering or data management/mining problem. This can also be seen from the references of the paper. The majority of the cited papers are from machine learning venues. As such, it would be better for the paper to be evaluated in a machine learning venue.

Response: Thanks for your comment. In this paper, we focus on improving the training process of graph contrastive learning (GCL). It is important to note that GCL supports a variety of downstream tasks, which are integral to the field of data mining/management. For example, node classification is important for social/citation network analysis, graph classification is important for molecule property prediction, and link prediction is important for recommender systems. In the revised manuscript, we apply GLATE to four different downstream tasks, including node classification, graph classification, node clustering, and link prediction. Our proposed method GLATE not only improves the performance of the base model but also accelerates its training process, which makes the GCL model easily scalable to large-scale graph data.

Furthermore, we have investigated the works closely related to ours that were published in ICDE, SIGMOD, and VLDB, and added 17 new references to the revised manuscript. These papers cover subgraph query ([83], [84], and [85]), community search ([86], [87], and [88]), node classification or link prediction on temporal graphs ([27], [29], [30], [36] and [37]), link prediction on knowledge graphs ([35]), citation network analysis ([28], [31], and [34]), and molecule property prediction ([32] and [33]).

The main revisions for the citations in the revised manuscript are as follows:

Raw texts: On benchmark datasets, the state-of-the-art GCL models have demonstrated competitive performance against supervised learning models, e.g., graph convolutional network (GCN) [26], in various graph-related tasks such as node classification [23], [27] and graph classification [28], [29].

Revised texts: On benchmark datasets, the state-of-the-art GCL models have demonstrated competitive performance against supervised learning models, e.g., graph convolutional network (GCN) [28], in various graph-related tasks such as node classification [27-31], graph classification [32], [33], and link prediction [34-37].

Raw texts: The objective of GCL is to maximize the similarity between positives and minimize the similarity between negatives [22], [68], [69]. Recent studies [14], [16] have concluded three mainstream graph contrastive losses based on the contrastive pairs in a graph, including JSD loss [21], [27], TM loss [35]-[37], and NT-Xent loss [22], [28], [38].

Revised texts: Graph embedding converts graph structures into low-dimensional vectors, preserving key information. These vectors support a series of graph-related tasks like subgraph query [83-85] and community search [86-88]. Graph contrastive learning (GCL) is an important method for extracting graph embeddings and its objective is to maximize the similarity between positives and minimize the similarity between negatives [22], [89], [90]. Recent studies [14], [16] have concluded three mainstream graph contrastive losses based on the contrastive pairs in a graph, including JSD loss [21], [43], TM loss [44]-[46], and NT-Xent loss [22], [47], [52].

Specifically, the newly added references include:

- [27] J. Layne, J. Carpenter, E. Serra, and F. Gullo, “Temporal sir-gn: Efficient and effective structural representation learning for temporal graphs,” Proceedings of the VLDB Endowment, vol. 16, no. 9, pp. 2075–2089, 2023.
- [28] Y. Zhang and A. Kumar, “Lotan: Bridging the gap between gnns and scalable graph analytics engines,” Proceedings of the VLDB Endowment, vol. 16, no. 11, pp. 2728–2741, 2023.
- [29] F. Xiao, Y. Wu, M. Zhang, G. Chen, and B. C. Ooi, “Mint: Detecting fraudulent behaviors from time-series relational data,” Proceedings of the VLDB Endowment, vol. 16, no. 12, pp. 3610–3623, 2023.
- [30] X. Du, X. Zhang, S. Wang, and Z. Huang, “Efficient tree-svd for subset node embedding over large dynamic graphs,” Proceedings of the ACM on Management of Data, vol. 1, no. 1, pp. 1–26, 2023.
- [31] X. Zhang, Y. Shen, Y. Shao, and L. Chen, “Ducati: A dual-cache training system for graph neural networks on

giant graphs with the gpu,” Proceedings of the ACM on Management of Data, vol. 1, no. 2, pp. 1–24, 2023.

- [32] G. Lv and L. Chen, “On data-aware global explainability of graph neural networks,” Proceedings of the VLDB Endowment, vol. 16, no. 11, pp. 3447–3460, 2023.
- [33] G. Lv, C. J. Zhang, and L. Chen, “Hence-x: Toward heterogeneity-agnostic multi-level explainability for deep graph networks,” Proceedings of the VLDB Endowment, vol. 16, no. 11, pp. 2990–3003, 2023.
- [34] C. Yang and J. Han, “Revisiting citation prediction with cluster-aware text-enhanced heterogeneous graph neural networks,” in 39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023, pp. 682–695, IEEE, 2023.
- [35] Y. Zhang, W. Wang, H. Yin, P. Zhao, W. Chen, and L. Zhao, “Disconnected emerging knowledge graph oriented inductive link prediction,” in 39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023, pp. 381–393, IEEE, 2023.
- [36] Y. Li, Y. Shen, L. Chen, and M. Yuan, “Orca: Scalable temporal graph neural network training with theoretical guarantees,” Proceedings of the ACM on Management of Data, vol. 1, no. 1, pp. 1–27, 2023.
- [37] H. Li and L. Chen, “Early: Efficient and reliable graph neural network for dynamic graphs,” Proceedings of the ACM on Management of Data, vol. 1, no. 2, pp. 1–28, 2023.
- [83] Y. Wang, A. Khan, T. Wu, J. Jin, and H. Yan, “Semantic guided and response times bounded top-k similarity search over knowledge graphs,” in 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 445–456, IEEE, 2020.
- [84] Y. Wang, A. Khan, X. Xu, J. Jin, Q. Hong, and T. Fu, “Aggregate queries on knowledge graphs: Fast approximation with semantic-aware sampling,” in 2022 IEEE 38th International Conference on Data Engineering (ICDE), pp. 2914–2927, IEEE, 2022.
- [85] H.-W. Chen, H.-H. Shuai, D.-N. Yang, W.-C. Lee, C. Shi, S. Y. Philip, and M.-S. Chen, “Structure-aware parameter-free group query via heterogeneous information network transformer,” in 2021 IEEE 37th International Conference on Data Engineering (ICDE), pp. 2075–2080, IEEE, 2021.
- [86] X. Xu, J. Liu, Y. Wang, and X. Ke, “Academic expert finding via (k, p)-core based embedding over heterogeneous graphs,” in 2022 IEEE 38th International Conference on Data Engineering (ICDE), pp. 338–351, IEEE, 2022.
- [87] J. Chen, Y. Xia, and J. Gao, “Communityaf: An example-based community search method via autoregressive flow,” Proceedings of the VLDB Endowment, vol. 16, no. 10, pp. 2565–2577, 2023.
- [88] J. Gao, J. Chen, Z. Li, and J. Zhang, “Ics-gnn: lightweight interactive community search via graph neural network,” Proceedings of the VLDB Endowment, vol. 14, no. 6, pp. 1006–1018, 2021.

TABLE I: Accuracy results (%) under different strategies.

| Model | Cora | CiteSeer | PubMed |
|---|-----------------|-----------------|-----------------|
| GRACE+GLATE | 84.8±0.3 | 74.4±0.1 | 87.2±0.1 |
| w/ hard negative sampling | 83.1±0.4 | 67.1±0.6 | 85.3±0.1 |
| w/ static τ ($\tau=0.8$) | 84.3±0.1 | 67.8±0.2 | 85.2±0.1 |
| w/ static τ ($\tau=0.2$) | 83.1±0.3 | 64.0±0.5 | 82.1±0.1 |
| w/ static τ ($\tau=0.4/0.9/0.7$) | 80.6±0.2 | 71.1±0.7 | 83.5±0.0 |
| w/o τ_{lower}^{\dagger} | 18.4±0.1 | 12.6±0.2 | 25.6±0.0 |
| w/o momentum | 83.7±0.2 | 72.7±0.6 | 86.9±0.1 |
| TaU | 83.5±0.2 | 67.8±0.4 | 85.5±0.0 |
| iSogCLR | 83.2±0.1 | 66.3±0.4 | 84.0±0.1 |
| Adap- τ | 83.6±0.2 | 68.5±0.3 | 85.9±0.1 |

[†] The value of training loss eventually becomes “NaN”.

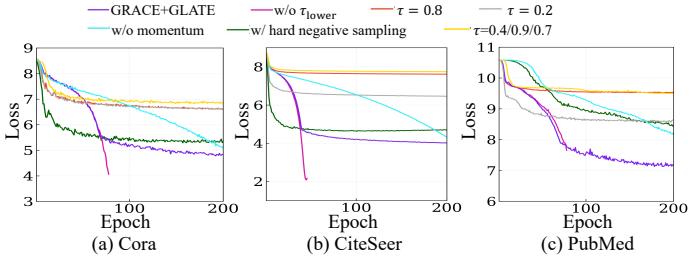


Fig. 1: Training loss curves of GRACE+GLATE and its variants.

III. RESPONSE TO REVIEWER #2

A. OI

Comment: In the ablation study, the static tau is not set the same as it is in the GRACE. In GRACE, tau is set to different values on these three datasets. Therefore, it is advised to use the same parameter settings as in the original paper.

Response: Thanks for your suggestion. We have added an experiment to use the same parameter settings for τ as reported in the original paper of GRACE. We denote this method as “w/ static τ ($\tau=0.4/0.9/0.7$)”, which means that fixing τ as the same value in GRACE (i.e., 0.4 for Cora, 0.9 for CiteSeer, and 0.7 for PubMed). We have updated this method’s accuracy results and loss curve (yellow curve) in Table I (i.e., Table XI in the revised manuscript) and Fig. 1 (i.e., Fig. 2 in the revised manuscript), respectively. The updated results still support our original experimental conclusion, i.e., GRACE+GLATE has obvious advantages over its variants with static τ either on accuracy performance or training loss.

The main revisions for the experimental analysis of ablation studies are shown as follows:

Raw texts: (ii) w/ static τ : removing dynamic temperature estimation and fixing τ as 0.8 or 0.2; (iii) w/o τ_{lower} : ... We conduct ablation studies to assess the effects of different temperature strategies concerning negatives, temperature, or momentum on model performance. ... As shown in Table VIII, by estimating temperatures dynamically, GRACE+GLATE outperforms two models with static temperatures ($\tau = 0.8$ or $\tau = 0.2$), especially on CiteSeer. ... The results demonstrate that “ $\tau = 0.2$ ”, “ $\tau = 0.8$ ”, “w/o momentum”, and GRACE+GLATE all converge rapidly, and our GRACE+GLATE model converges to the lowest loss value among them.

TABLE II: Accuracy results (%) on inductive node classification.

| Model | PPI | Model | PPI |
|----------------|----------|--------------------|------------------------------------|
| Raw model | 42.2±0.0 | DGI | 63.8±0.2 |
| GraphSAGE-GCN | 46.5±0.2 | DGI+GLATE (ours) | 67.0±0.0 |
| GraphSAGE-Mean | 48.6±0.1 | SUGRL | 66.0±0.1 |
| GraphSAGE-LSTM | 48.2±0.1 | SUGRL+GLATE (ours) | 67.8±0.1 |
| GraphSAGE-Pool | 50.2±0.2 | NCLA | 66.2±0.1 |
| GMI | 65.0±0.0 | NCLA+GLATE (ours) | 68.0±0.0 |
| GCA | 66.0±0.1 | GRACE | 66.2±0.1 |
| BGRL | 67.5±0.1 | GRACE+GLATE (ours) | 68.4±0.0 ($p=8.1e-11$) |

Revised texts: (ii) w/ static τ ($\tau=0.8$): fixing τ as 0.8; (iii) w/ static τ ($\tau=0.2$): fixing τ as 0.2; (iv) w/ static τ ($\tau=0.4/0.9/0.7$): fixing τ as the same value in GRACE (i.e., 0.4 for Cora, 0.9 for CiteSeer, and 0.7 for PubMed); (v) w/o τ_{lower} : ... We conduct ablation studies to assess the effects of different temperature strategies regarding negatives, temperature, or momentum on model performance. ... As shown in Table I, by estimating temperatures dynamically, GRACE+GLATE outperforms three models with static temperatures ($\tau=0.8$, $\tau=0.2$, and $\tau=0.4/0.9/0.7$). ... The results demonstrate that “ $\tau=0.2$ ”, “ $\tau=0.8$ ”, “ $\tau=0.4/0.9/0.7$ ”, “w/ hard negative sampling”, and GRACE+GLATE all converge rapidly, and our GRACE+GLATE model converges to the lowest loss value among them.

B. O2

Comment: GLATE is only deployed with GRACE in the inductive node classification task. It would be beneficial to extend the application of GLATE to more baseline methods to demonstrate the effectiveness of the proposed approach.

Response: Thanks for your suggestion. We have conducted additional experiments by applying GLATE to other graph contrastive learning methods, including DGI, SUGRL, and NCLA, in the inductive node classification task. The results of these additional experiments have been included in Table II (i.e., Table VII in the revised manuscript) and show that GLATE consistently improves the performance of these baseline methods. Specifically, GLATE improves the test accuracy of DGI, SUGRL, and NCLA on PPI by 3.2, 1.8, and 1.8 percentage points, respectively. The updated results further validate the versatility and effectiveness of GLATE in the inductive node classification task.

The main revisions for the experimental analysis in the node classification task are as follows:

Raw texts: Table V shows the testing accuracy results on the task of transductive node classification. Firstly, we observe that several GCL models outperform the supervised GCN model on certain datasets, such as Coauthor-CS and Coauthor-Physics. ... Given that GRACE+GLATE consistently outperforms other models across all datasets, we select it as the representative model for GLATE in subsequent experiments.

The comparison results on the inductive node classification task are presented in Table VI. In addition, GRACE+GLATE significantly outperforms all self-supervised baselines, highlighting the effectiveness of dynamic temperature estimation in inductive learning.

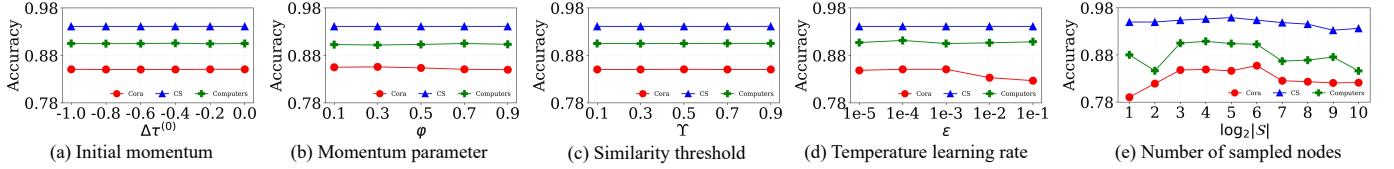


Fig. 2: Performance of GRACE+GLATE under different values of $\Delta\tau^{(0)}$, φ , Υ , ε , and $|\mathcal{S}|$ on Cora, CS, and Computers.

Revised texts: Table V shows the testing accuracy results on transductive node classification. Firstly, we observe that several GCL models outperform the supervised GCN model on certain datasets, such as **CS and Physics**. ... Given that GRACE+GLATE consistently outperforms other models **on all** datasets, we select it as the representative model for GLATE in subsequent experiments.

The comparison results on inductive node classification are presented in Table VI. ... In addition, **GLATE improves the test accuracy of DGI, SUGRL, NCLA, and GRACE by 3.2, 1.8, 1.8, and 2.2 percentage points, respectively**, highlighting the effectiveness of dynamic temperature estimation in inductive learning.

C. O3

Comment: More sensitivity analysis experiments are expected. There are several hyperparameters in the temperature definition, such as the initial momentum, the momentum parameter, the similarity threshold gamma and so on. However, the authors do not conduct enough parameter analysis in Section IV-4.

Response: As you suggest, we have expanded our sensitivity analysis to include additional experiments examining the effects of the initial momentum $\Delta\tau^{(0)}$, momentum parameter φ , similarity threshold Υ , temperature learning rate ε , and number $|\mathcal{S}|$ of sampled nodes on the performance of GRACE+GLATE. The experimental results on three different types of graph datasets (Cora: citation network, CS: co-authorship network, and Computers: co-purchase network) are shown in Fig. 2 (i.e., Fig. 4 in the revised manuscript). It verifies that the performance of GRACE+GLATE is robust to the changes in $\Delta\tau^{(0)}$, φ , and Υ , but sensitive to ε and $|\mathcal{S}|$. When $\varepsilon \in [10^{-4}, 10^{-3}]$ and $|\mathcal{S}| \in [8, 64]$, our method can achieve good performance. This general parameter configuration can achieve satisfactory performance when training our method on a new graph dataset without excessive parameter tuning time.

The main revisions for the experimental analysis of parameter sensitivity are as follows:

Raw texts: We conduct sensitivity analysis to assess the impact of three crucial hyperparameters, namely edge removing ratio p_r , attribute masking ratio p_m , and initial temperature $\tau^{(0)}$, on the performance of GRACE+GLATE. ... Furthermore, we observe that the setting of $\tau^{(0)}=0.8$ yields the highest accuracy result, reaching 84.5% on the test set. Given this finding, we recommend setting $\tau^{(0)}=0.8$ **on Cora**. These conclusions are also applicable to the other datasets.

Revised texts: We conduct sensitivity analysis to assess the impact of edge removing ratio p_r , attribute masking ratio p_m , and initial temperature $\tau^{(0)}$, on the performance of GRACE+GLATE. ... **This conclusion is also applicable to the other datasets**. Furthermore, we observe that the setting of $\tau^{(0)}=0.8$ yields the highest accuracy result, reaching 84.5% on the test set. Given this finding, we recommend setting $\tau^{(0)}=0.8$. **We also analyze the sensitivity of the initial momentum $\Delta\tau^{(0)}$, momentum parameter φ , similarity threshold Υ , temperature learning rate ε , and number $|\mathcal{S}|$ of sampled nodes on Cora (citation network), CS (co-authorship network), and Computers (co-purchase network)**. Based on the results in Fig. 4(a), (b), and (c), we can see that GRACE+GLATE is robust to the changes of $\Delta\tau^{(0)}$, φ , and Υ . Also, the results in Fig. 4(d) and (e) reveal that $\varepsilon \in [10^{-4}, 10^{-3}]$ and $|\mathcal{S}| \in [8, 64]$ enable our method to achieve high test accuracy. This general parameter configuration can achieve satisfactory performance when training our method on a new graph dataset without excessive parameter tuning time.

D. O4

Comment: There are some minor issues in this paper. For example, the figures in Fig. 4 are not clear enough. The authors should ensure a clear separation between the points and text in each figure to avoid overlap. Furthermore, a clearer representation of how ‘local’ and ‘global’ change with increasing training epochs should be provided.

Response: Thanks for your kind reminder. We have revised the scatter plot (shown in Fig. 3, i.e., Fig. 5 in the revised manuscript) in Section IV-B-7 to ensure a clearer separation between points and text, thereby avoiding any overlap. Also, to better illustrate how “local” and “global” change with increasing training epochs, we use sampled epochs (i.e., epoch 0, 5, 10, ..., 100) and employ a color gradient from red to blue to differentiate training epochs, where red represents “Epoch 0” (i.e., initial state) and blue represents “Epoch 100”.

The main revisions for the experimental analysis of embedding quality are as follows:

Raw texts: We find that the global separation and local separation of the CNEs learned from GRACE+GLATE increase gradually **in the training process on the three datasets**, ... Notably, due to the different characteristics **among the three graph datasets**, GLATE enhances the quality of learned CNEs to varying extents. For instance, the observed increases in the values of $I_{\text{global}}^{(t)}$ and $I_{\text{local}}^{(t)}$ on Cora and CiteSeer are higher than those on PubMed. ...

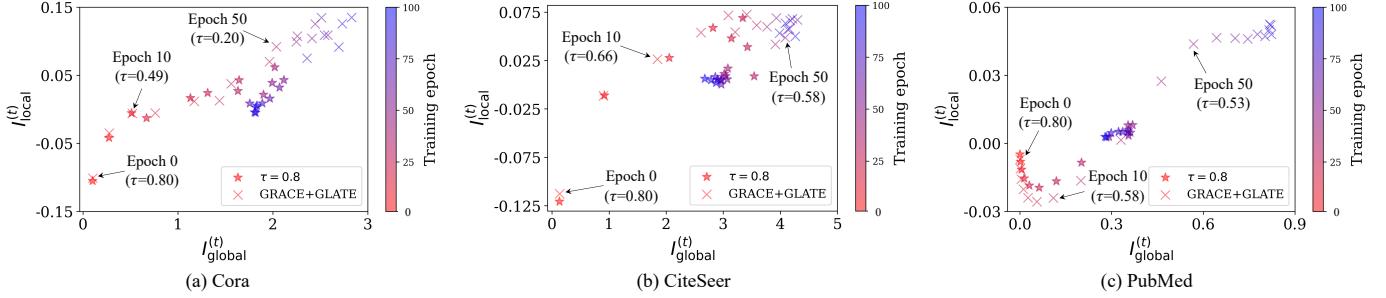


Fig. 3: Evolution trend of $I_{\text{global}}^{(t)}$ and $I_{\text{local}}^{(t)}$ during training. Red and blue represent early and late training epochs, respectively. The dots closer to the top right corner correspond to more reasonable CNEs.

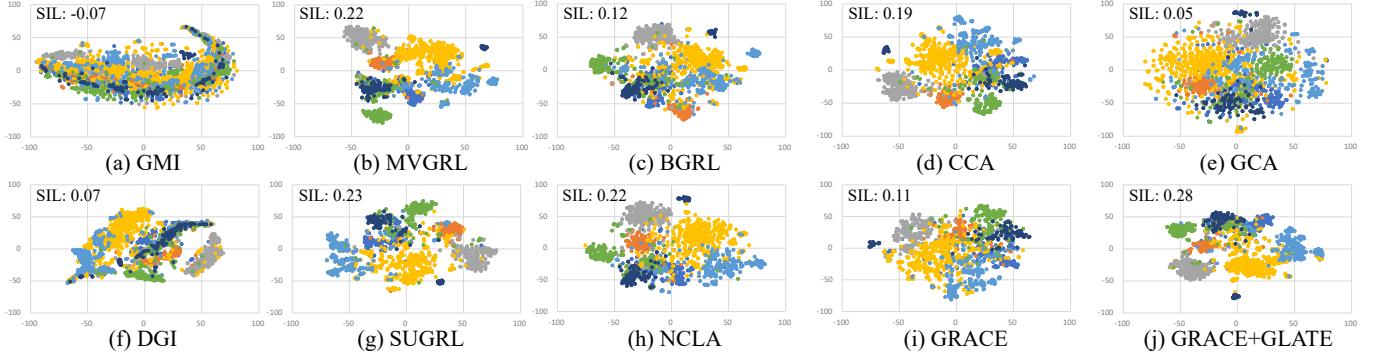


Fig. 4: Visualization of CNEs using t-SNE on Cora. Different colors represent different classes.

Revised texts: From Fig. 5, we find that the global separation and local separation of the CNEs learned from GRACE+GLATE increase gradually during training on these datasets, ... Due to the different characteristics of datasets, GLATE enhances the quality of learned CNEs to varying extents: the observed increases in $I_{\text{global}}^{(t)}$ and $I_{\text{local}}^{(t)}$ on Cora and CiteSeer are higher than those on PubMed; ...

Furthermore, to present more clearly, we have checked the whole paper and adjusted other figure presentations and writing statements in the paper. For example:

(i) We adjust the text in Fig. 4 (i.e., Fig. 10 in the revised manuscript) so that the scatter symbols and the text for the SIL score in each subfigure do not overlap and add subfigure number from (a) to (j) in Fig. 4.

(ii) The lemmas in the paper are concisely expressed, e.g., the revision for Lemma 1 is as follows:

Raw texts:

Lemma 1. When temperature τ is assigned a low value, the optimizer using \mathcal{L}_{JSD} imposes a higher penalty on easy negatives in contrast to hard negatives. Conversely, for high values of τ , the optimizer applies similar penalties to all negatives.

Revised texts:

Lemma 1. A low τ in the optimizer using \mathcal{L}_{JSD} results in a higher penalty on easy negatives than hard negatives, while a high τ leads to similar penalties for all negatives.

IV. RESPONSE TO REVIEWER #3

A. O1

Comment: The innovative approach of dynamically estimating temperatures within contrastive losses, while theoretically sound, may pose challenges for practical implementation. This complexity stems from the need to fine-tune the dynamic temperature parameters, which requires a deep understanding of the model's workings and the specific characteristics of the dataset being used. Such fine-tuning can be time-consuming and may necessitate extensive experimental trials to identify optimal settings.

Response: Thanks for your comment. We have expanded the efficiency analysis of GLATE to provide a more comprehensive verification of the efficiency of temperature fine-tuning. Specifically: (i) We plot the time ratio of different stages in the overall training of GRACE+GLATE in Fig. 5. The results show that the time consumption of temperature fine-tuning is small in the overall training of GRACE+GLATE, not exceeding 1.2% on each dataset. (ii) Moreover, we have conducted comprehensive sensitivity analysis experiments on Cora (citation network), CS (co-authorship network), and Computers (co-purchase network). The results in Fig. 2 (i.e., Fig. 4 in the revised manuscript) show that GRACE+GLATE is robust to the changes of most hyperparameters, e.g., initial momentum and momentum parameter. Although GRACE+GLATE is sensitive to the temperature learning rate ε and number $|\mathcal{S}|$ of sampled nodes, the setting of $\varepsilon \in [10^{-4}, 10^{-3}]$ and $|\mathcal{S}| \in [8, 64]$ enables our method to achieve good results. This general parameter configuration can achieve satisfactory performance

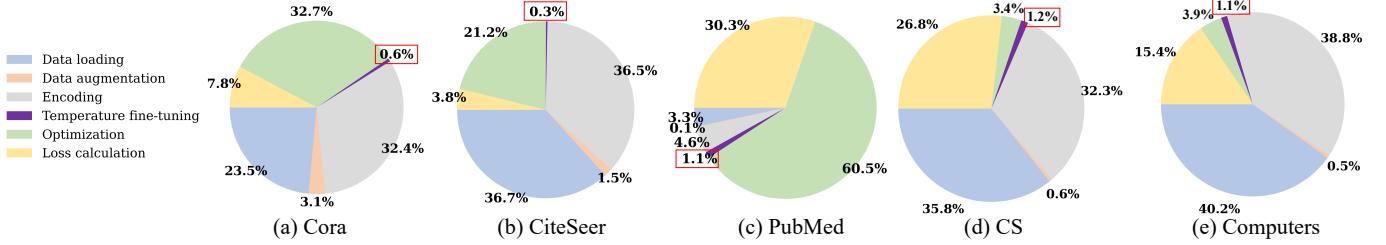


Fig. 5: Time ratio of different stages in the overall training of GRACE+GLATE. The temperature fine-tuning time ratio is marked with a red box. Different from the entire graph training on Cora, CiteSeer, CS, and Computers, GRACE+GLATE uses batch training on PubMed, so the loss calculation and optimization are relatively large.

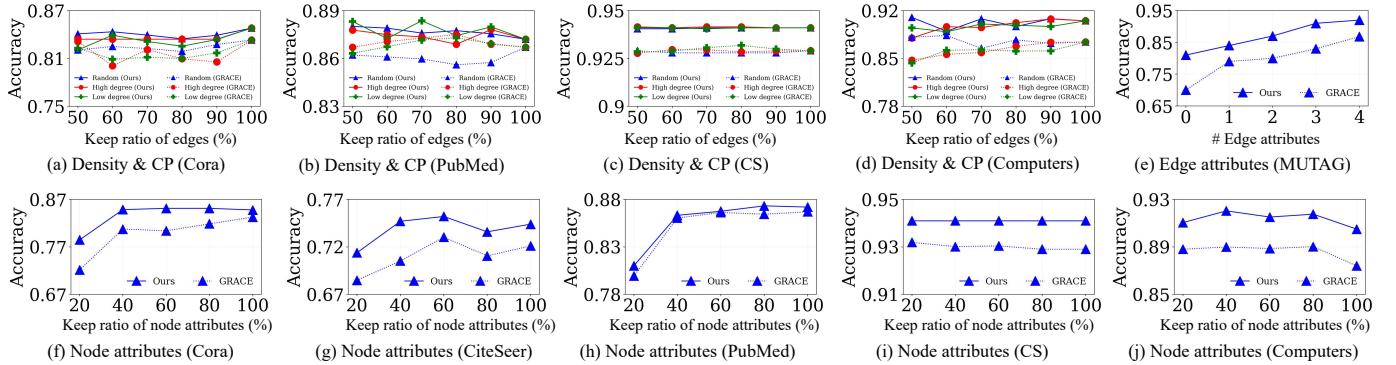


Fig. 6: Performance of GRACE+GLATE (Ours) and GRACE under different graph structures and sizes. “CP” denotes connectivity pattern. “Keep ratio” refers to retaining a certain proportion of the edges or attributes from the original graph.

when training our method on a new graph dataset without excessive parameter tuning time.

The main revisions in the experimental analysis of model efficiency are as follows:

Raw texts: Based on the above analysis, we conclude that GRACE+GLATE has higher computational efficiency compared with the other GCL models.

Revised texts: Based on the above analysis, we conclude that GRACE+GLATE has higher computational efficiency compared with the other GCL models. In Fig. 7, we plot the time ratio of different stages during the training of GRACE+GLATE. We observe that the time consumption of temperature fine-tuning is small, not exceeding 1.2% on each dataset. Since GLATE updates $\tau^{(t)}$ based on sampled nodes at intervals of a few epochs during training, it does not impose a significant time burden on the original GCL model training.

B. O2

Comment: The paper’s limited discussion on how different graph structures and sizes affect the model’s performance leaves a gap in understanding its applicability across diverse real-world scenarios. Graphs can vary widely in terms of density, connectivity patterns, and node and edge attributes, each of which could significantly impact the effectiveness of the proposed method. A more comprehensive analysis, including a variety of graph types, would provide clearer insights into the model’s versatility and limitations.

Response: Thanks for your suggestion. We have conducted additional experiments on a broader range of graph types with varying density, connectivity patterns, node attributes, and edge attributes. These experiments include:

- (i) Varying graph density by removing different ratios of edges.
- (ii) Varying graph connectivity pattern by setting different edge deletion priorities on different datasets.
- (iii) Varying node attributes by masking different ratios of node attributes.
- (iv) Varying edge attributes by masking different numbers of edge attributes.

The results of these experiments are shown in Fig. 6 (i.e., Fig. 6 in the revised manuscript). The results reveal that compared to the base model GRACE, GRACE+GLATE exhibits better overall performance when changes occur in the graph structure and size or attribute information. For example, on CS, when different ratios of edges are removed from the graph (whether it is “Random”, “High degree”, or “Low degree”), GRACE+GLATE shows better robustness than its base model GRACE.

We have added a new subsection to our experiment, titled “Performance on Graphs with Different Structures, Sizes, and Attributes”. The added content is as follows:

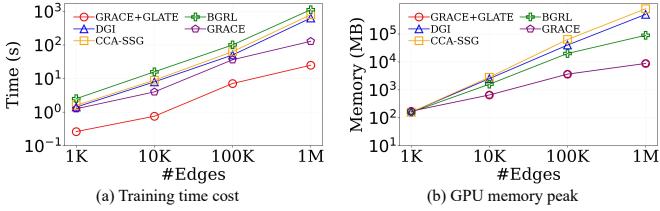


Fig. 7: Results of the scalability experiments.

Revised texts: *Performance on graphs with different structures, sizes, and attributes:* Next, we analyze GRACE+GLATE’s performance across different graph structures, graph sizes, and node (or edge) attributes, validated through density, connectivity patterns, node attributes, and edge attributes. Firstly, we validate density and connectivity patterns on Cora, PubMed, CS, and Computers, each with different connectivity patterns. We perform edge removing operations on each dataset, categorized into three types: random edge removing (“Random”), preferential removing of edges corresponding to high-degree nodes (“High degree”), and preferential removing of edges corresponding to low-degree nodes (“Low degree”). These operations result in three different connectivity patterns for each dataset. From Fig. 6(a-d), we observe that compared to the base model GRACE, GRACE+GLATE exhibits better robustness on these datasets. It indicates that the temperature fine-tuning technique in GLATE enables the GCL model to learn good node embeddings even when the graph structure information is incomplete.

Secondly, since the MUTAG dataset contains edge attributes (i.e., four chemical bond types: aromatic, single, double, and triple bonds), we randomly mask edge attributes on it and observe GRACE+GLATE’s performance as the scale of edge attributes changes. From Fig. 6(e), we see that complete edge attribute information allows the model to achieve high accuracy, and at the same time, GRACE+GLATE is more robust than GRACE when the number of edge attributes is reduced.

Thirdly, we randomly mask node attributes on five datasets and observe the performance of GRACE+GLATE as well as GRACE as the scale of node attributes changes. From Fig. 6(f-j), as node attributes are missing, we observe that both models exhibit a trend of overall performance degradation in citation networks (Cora, CiteSeer, and PubMed); while in the co-authorship network (CS) or co-purchase network (Computers), both models exhibit a trend of stable or improving performance. Moreover, GRACE+GLATE consistently outperforms GRACE, indicating the effectiveness of GLATE.

C. O3

Comment: Given the computational complexity associated with dynamic temperature adjustments and the evaluation of embeddings, there could be scalability issues when applying the method to very large graph datasets. The paper does not fully address these potential challenges, particularly in

TABLE III: Comparison results on running time (seconds).

| Model | Cora | CiteSeer | PubMed | Model | Cora | CiteSeer | PubMed |
|---------|-------|----------|--------|-------------|------------|------------|-------------|
| GMI | 100.5 | 24.8 | >500.0 | DGI | 10.5 | 3.9 | 158.2 |
| MVGRL | 21.6 | 13.5 | >500.0 | SUGRL | 17.2 | 17.8 | >500.0 |
| BGRL | 6.3 | 6.4 | 149.7 | NCLA | >500.0 | >500.0 | >500.0 |
| CCA-SSG | 12.7 | 21.3 | 89.9 | GRACE | 4.2 | 4.5 | 349.9 |
| GCA | 11.4 | 4.2 | 315.4 | GRACE+GLATE | 3.5 | 1.8 | 28.6 |

contexts where computational resources are limited or where real-time processing is required. Scalability is a critical factor for many applications, and without a detailed examination of this aspect, it’s difficult to assess the feasibility of deploying the proposed method in large-scale graph analytics tasks.

Response: Thanks for your comment. The findings presented in Fig. 5 (i.e., Fig. 7 in the revised manuscript) illustrate that the fine-tuning of temperature incurs minimal additional training time, suggesting that GLATE does not significantly increase the training load of the original GCL model. Furthermore, the application of a dynamic temperature enables the acquisition of high-quality embeddings more rapidly than a static approach, allowing GRACE+GLATE to achieve its objectives with fewer training epochs compared to GRACE alone, thereby shortening the overall training time, as demonstrated in Table III (i.e., Table XII in the revised manuscript). In addition, to examine the scalability of GLATE, we use the citation network of ogbn-arxiv to generate a group of graphs in different scales by sampling edges from it. The comparison of training time and peak GPU memory usage for different models is depicted in Fig. 7 (i.e., Fig. 8 in the revised manuscript). These findings reveal that the training time for GRACE+GLATE increases at a nearly linear rate as graph size expands, enhancing the time efficiency compared to GRACE. Moreover, both GRACE+GLATE and GRACE maintain a similar peak memory usage, surpassing other GCL baselines in terms of memory efficiency as the number of edges grows. The observations in Fig. 7 align with the insights from our complexity analysis in Section III-D, affirming that GRACE+GLATE is scalable to large-scale graphs.

We have added a new subsection to our experiment, titled “Scalability”. The added content is as follows:

Revised texts: *Scalability:* To examine the scalability of GLATE, we use the citation network of ogbn-arxiv to generate a group of graphs in different scales by sampling edges from it. The training time cost and GPU memory peak of different models are plotted in Fig. 8(a) and (b), respectively. We can see that with the increase in graph size, the time cost of GRACE+GLATE grows at a near-linear speed. Also, GRACE+GLATE and GRACE have almost the same memory peak, outperforming other GCL baselines in memory efficiency as the number of edges increases. Therefore, we conclude that GRACE+GLATE is scalable to large-scale graphs.

D. O4

Comment: The paper did not sufficiently explore the model’s generalization capabilities across different graph-related tasks,

TABLE IV: Node clustering performance.

| Model | Cora | | | CiteSeer | | | PubMed | | |
|---------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| | Accuracy | NMI | ARI | Accuracy | NMI | ARI | Accuracy | NMI | ARI |
| Spectral clustering | 36.7 | 12.6 | 3.1 | 23.8 | 5.5 | 1.0 | 52.8 | 9.7 | 6.2 |
| <i>k</i> -means | 49.2 | 32.1 | 22.9 | 54.0 | 30.5 | 27.8 | 59.5 | 31.5 | 28.1 |
| GAE | 59.6 | 42.9 | 34.7 | 40.8 | 17.6 | 12.4 | 67.2 | 27.7 | 27.9 |
| VGAE | 50.2 | 32.9 | 25.4 | 46.7 | 26.0 | 20.5 | 63.0 | 22.9 | 21.3 |
| DGI | 55.4 | 41.1 | 32.7 | 51.4 | 31.5 | 32.6 | 58.9 | 27.7 | 31.5 |
| GRACE | 65.3 | 47.3 | 38.8 | 67.1 | 38.9 | 37.4 | 56.2 | 20.2 | 19.9 |
| GRACE+GLATE | 70.8 | 52.4 | 45.4 | 70.2 | 41.9 | 41.8 | 67.9 | 31.9 | 32.9 |

such as link prediction, graph classification, or node clustering. Understanding how the proposed method performs in a variety of tasks is crucial for assessing its utility in real-world applications.

Response: Thanks for your suggestion. We have extended our evaluation to include additional graph-related tasks: in addition to the experiments on node classification and graph classification in the original manuscript, we have applied GLATE to node clustering and link prediction. In node clustering, we evaluated different methods on Cora, CiteSeer, and PubMed using the metrics of accuracy, normalized mutual information (NMI), and adjusted rand index (ARI). In link prediction, we evaluated different methods on UCI, Last.fm, and Kuaishou using the metrics of Recall@50 and mean reciprocal rank (MRR). The results of node clustering in Table IV (i.e., Table IX in the revised manuscript) show that GRACE+GLATE surpasses its base model GRACE by 6.7%, 6.6%, and 8.0% averagely in terms of accuracy, NMI, and ARI, respectively. The results of link prediction in Table V (i.e., Table X in the revised manuscript) show that GRACE+GLATE outperforms its base model GRACE by 4.3% and 1.2% averagely in terms of Recall@50 and MRR, respectively. The expanded evaluation highlights the potential of our proposed method for a variety of graph-related tasks in real-world applications.

We have added two new subsections to our experiment, titled “Performance on node clustering” and “Performance on link prediction”. The added content is as follows:

Revised texts: *Performance on node clustering:* We further evaluate the performance of GLATE on node clustering using the metrics of accuracy, normalized mutual information (NMI), and adjusted rand index (ARI). Table IX shows that GRACE+GLATE surpasses its base model GRACE by 6.7%, 6.6%, and 8.0% averagely in terms of accuracy, NMI, and ARI, respectively, highlighting the benefits of dynamic temperature estimation to GCL in the node clustering task.

Revised texts: *Performance on link prediction:* We also evaluate the performance of GLATE on link prediction using the metrics of Recall@50 and mean reciprocal rank (MRR). Based on the results presented in Table X, we observe that GRACE+GLATE achieves superior performance in the link prediction task on all datasets. Specifically, GRACE+GLATE outperforms its base model GRACE by 4.3% and 1.2% averagely in terms of Recall@50 and MRR, respectively. These results demonstrate the effectiveness of incorporating dynamic temperature estimation into the GCL model for link prediction.

TABLE V: Link prediction performance.

| Model | UCI | | Last.fm | | Kuaishou | |
|-------------|--------------|-------------|--------------|-------------|-------------|-------------|
| | Recall@50 | MRR | Recall@50 | MRR | Recall@50 | MRR |
| LINE | 10.86 | 2.49 | 8.57 | 1.12 | 2.23 | 0.31 |
| DGI | 25.49 | 5.74 | 5.81 | 1.00 | 5.11 | 0.60 |
| GraphCL | 19.72 | 3.10 | 9.72 | 1.51 | 4.85 | 0.60 |
| AD-GCL | 18.19 | 3.23 | 8.22 | 1.11 | 1.84 | 0.24 |
| GraphMAE | 1.70 | 0.52 | 3.07 | 0.43 | 2.06 | 0.31 |
| GRACE | 16.69 | 2.91 | 10.12 | 1.45 | 4.68 | 0.61 |
| GRACE+GLATE | 27.22 | 6.40 | 11.88 | 1.65 | 5.34 | 0.66 |

Additionally, we have supplemented the dataset description and baseline description with relevant information on the datasets and baselines used in node clustering and link prediction.

Raw texts: For the node classification task, ... and node **features** represent paper keywords for each author’s papers. In a co-purchase network, each node represents a product whose **features** are bag-of-words from product reviews or product descriptions, and each edge indicates that two products are frequently bought together. ... Additionally, we present the statistics of the above 10 datasets in Table III.

For the graph classification task, we use three datasets from TUDataset [50], including NCI1, MUTAG, and PROTEINS. NCI1 or MUTAG is grouped by small molecules, while PROTEINS is grouped by macromolecules. In NCI1 or MUTAG, a small molecule corresponds to a graph where nodes and edges are atoms and chemical bonds, respectively. In PROTEINS, a macromolecule corresponds to a graph where nodes are secondary structure elements. If two nodes are connected, they are neighbors along the amino-acid sequence or one of the three nearest neighbors in space. Furthermore, we present the statistics of the three datasets in Table IV.

Revised texts: **(i)** For the node classification task, ... and node **attributes** represent paper keywords for each author’s papers. In a co-purchase network, each node represents a product whose **attributes** are bag-of-words from product reviews or product descriptions, and each edge indicates that two products are frequently bought together. ... **(ii)** For the graph classification task, we use three datasets from TUDataset [62], including NCI1, MUTAG, and PROTEINS. NCI1 or MUTAG is grouped by small molecules, while PROTEINS is grouped by macromolecules. **(iii)** For the node clustering task, we use the datasets of Cora and CiteSeer as benchmarks. **(iv)** For the link prediction task, we use three datasets: UCI (message communications among University of California, Irvine students in an online community), Last.fm (<user, artist, song> tuples from the Last.fm API, representing the listening habits of nearly 1,000 users), and Kuaishou (interactions between 6,840 users and 131,972 videos from the Kuaishou online video-watching platform). We present the statistics of the above datasets in Tables III, IV, and V.

TABLE VI: Statistics of the datasets for link prediction.

| Dataset | # Node | # Edge | Link |
|----------|---------|-----------|---|
| UCI | 1,677 | 56,617 | http://konect.unikoblenz.de/networks/ |
| Last.fm | 127,786 | 720,537 | http://www.last.fm/ |
| Kuaishou | 138,812 | 1,779,639 | http://www.kuaishou.com/ |

Raw texts: For example, a common READOUT function is $\mathcal{R}(U) = \sigma\left(\frac{1}{N} \sum_{i=1}^N u_i\right)$ where $\sigma(\cdot)$ is the Sigmoid function.

Revised texts: For example, a common READOUT function is $\mathcal{R}(U) = \sigma\left(\frac{1}{N} \sum_{i=1}^N u_i\right)$ where $\sigma(\cdot)$ is the Sigmoid function. **(iii) For the node clustering task, we compare GLATE with two clustering methods (spectral clustering and k -means), two graph auto-encoder models (GAE [74] and VGAE [74]), and two GCL models (DGI [21] and GRACE [22]). (iv) For the link prediction task, we compare GLATE with one graph representation model (LINE [75]), one graph autoencoder model (GraphMAE [76]), and four GCL models (DGI [21], GraphCL [52], AD-GCL [77], and GRACE [22]).**

The references of newly added baselines (i.e., GAE [74], VGAE [74], LINE [75], GraphMAE [76], and AD-GCL [77]) are included in the revised manuscript:

- [74] T. N. Kipf and M. Welling, “Variational graph auto-encoders,” NeurIPS, 2016.
- [75] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, “Line: Large-scale information network embedding,” in Proceedings of the 24th international conference on world wide web, pp. 1067–1077, 2015.
- [76] Z. Hou, X. Liu, Y. Cen, Y. Dong, H. Yang, C. Wang, and J. Tang, “Graphmae: Self-supervised masked graph autoencoders,” in Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp. 594–604, 2022.
- [77] S. Suresh, P. Li, C. Hao, and J. Neville, “Adversarial graph augmentation to improve graph contrastive learning,” Advances in Neural Information Processing Systems, vol. 34, pp. 15920–15933, 2021.

We also added the statistics of the used datasets for link prediction in Table VI (i.e., Table V in the revised manuscript).

Accordingly, we have revised the descriptions of the experiments in the abstract, contributions and conclusion:

Raw texts: Finally, the extensive experiments on 13 benchmark datasets demonstrate that GLATE consistently outperforms the state-of-the-art graph contrastive learning models **on either classification accuracy or training efficiency.**

Revised texts: Finally, the extensive experiments on 16 benchmark datasets demonstrate that GLATE consistently outperforms the state-of-the-art graph contrastive learning models **in terms of both model performance and training efficiency.**

Raw texts: Through the extensive experiments on 13 benchmark datasets, we demonstrate that GLATE significantly improves the performance of five representative GCL base models on downstream tasks, including node classification and graph classification (Section IV).

Revised texts: Through the extensive experiments on 16 benchmark datasets, we demonstrate that GLATE significantly improves the performance of five representative GCL base models on downstream tasks, including node classification, graph classification, **node clustering, and link prediction** (Section IV).

Raw texts: In this paper, our primary focus is on tackling the challenge of generating high-quality contrastive node embeddings in graph contrastive learning. We propose GLATE, a novel and effective method that dynamically estimates the temperatures in three mainstream graph contrastive losses to improve the quality of contrastive node embeddings. ... Our experiments on **two important graph-related tasks, node classification and graph classification**, demonstrate the significant improvements of GLATE in **classification accuracy** and training efficiency compared with the state-of-the-art graph contrastive learning models. Furthermore, we validate the high-quality contrastive node embeddings learned by GLATE using the indicators of **global separation and local separation**. In future work, we plan to extend the GLATE method to distributed environments, exploring strategies to leverage parallel processing and efficient communication protocols.

Revised texts: In this paper, we propose GLATE, a novel and effective method that dynamically estimates the temperatures in three mainstream graph contrastive losses to improve the quality of contrastive node embeddings. ... Our experiments on **four different graph-related tasks** demonstrate the significant improvements of GLATE in **model performance** and training efficiency compared with the state-of-the-art graph contrastive learning models. In future work, we plan to extend the GLATE method to distributed environments, exploring strategies to leverage parallel processing and efficient communication protocols.

E. O5

Comment: While the method focuses on improving the quality of contrastive node embeddings, there is limited discussion on the interpretability of these embeddings. For practical applications, especially in domains like bioinformatics or social network analysis, the ability to interpret and understand the embeddings can be as important as their performance.

Response: Thanks for your comment. We have added a new Section IV-B-11 to the revised manuscript that discusses the interpretability of the embeddings generated by our method. Specifically, for a query molecule in MUTAG, we identified five example molecules and calculated their embedding distances and RDKit fingerprint (RDKitFP) similarities with

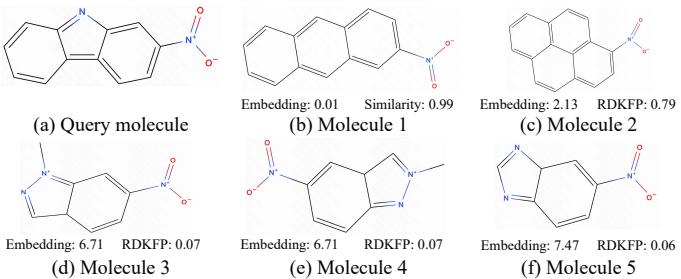


Fig. 8: Query molecule and five example molecules in MUTAG. “Embedding” and “RDKFP” denote embedding distance and RDKit fingerprint similarity, respectively.

the query molecule. The results in Fig. 8 (i.e., Fig. 9 in the revised manuscript) show that the embedding distance is closely related to the RDKFP similarity between chemical compounds, which indicates that the embeddings learned by GRACE+GLATE match the basic chemical knowledge of molecules. Specifically:

(i) The query molecule and molecule 1 have an embedding distance of 0.01 and an RDKFP similarity of 0.99 because both have two benzene rings and three rings in total.

(ii) The query molecule and molecule 2 have an embedding distance of 2.13 and an RDKFP similarity of 0.79. Although both of them have at least one benzene ring, molecule 2 has more rings than the query molecule, resulting in a greater embedding distance and lower RDKFP similarity compared to molecule 1.

(iii) The query molecule and molecules 3, 4, and 5 have embedding distances of 6.71, 6.71, and 7.47, respectively, and RDKFP similarities of 0.07, 0.07, and 0.06, respectively. Compared to molecules 1 and 2, these three molecules contain only one benzene ring, so their similarity to the query molecule is smaller and their embedding distance is farther. Furthermore, since molecules 3 and 4 have the same number of atoms and chemical bonds, their values are the same in terms of embedding distance and RDKFP similarity.

We have added a new subsection to our experiment, titled “Case study”. The added content is as follows:

Revised texts: *Case study:* To interpret the learned embeddings, we identify five example molecules about a query molecule in MUTAG. From Fig. 9, we find that the distance between a pair of embeddings closely corresponds to the computed similarity between chemical compounds based on their RDKit fingerprints (RDKFP). For example, since the query molecule and molecule 1 both have two benzene rings and three rings in total, they have closer embedding distance and higher RDKFP similarity than the other molecules. The case study demonstrates that the learned embeddings match the basic chemical knowledge of molecules.

F. 06

Comment: This work appears more aligned with AI/ML research themes, as indicated by its heavy citation of sources

from AI/ML conferences and journals, which may not perfectly fit the traditional focus areas of ICDE.

Response: Thanks for your comment. In this paper, we focus on improving the training process of graph contrastive learning (GCL). It is important to note that GCL supports a variety of downstream tasks, which are integral to the field of data mining/management. For example, node classification is important for social/citation network analysis, graph classification is important for molecule property prediction, and link prediction is important for recommender systems. In the revised manuscript, we apply GLATE to four different downstream tasks, including node classification, graph classification, node clustering, and link prediction. Our proposed method GLATE not only improves the performance of the base model but also accelerates its training process, which makes the GCL model easily scalable to large-scale graph data.

Furthermore, we have investigated the works closely related to ours that were published in ICDE, SIGMOD, and VLDB, and added 17 new references to the revised manuscript. These papers cover subgraph query ([83], [84], and [85]), community search ([86], [87], and [88]), node classification or link prediction on temporal graphs ([27], [29], [30], [36] and [37]), link prediction on knowledge graphs ([35]), citation network analysis ([28], [31], and [34]), and molecule property prediction ([32] and [33]).

The main revisions for the citations in the revised manuscript are as follows:

Raw texts: On benchmark datasets, the state-of-the-art GCL models have demonstrated competitive performance against supervised learning models, e.g., graph convolutional network (GCN) [26], in various graph-related tasks such as node classification [23], [27] and graph classification [28], [29].

Revised texts: On benchmark datasets, the state-of-the-art GCL models have demonstrated competitive performance against supervised learning models, e.g., graph convolutional network (GCN) [28], in various graph-related tasks such as node classification [27-31], graph classification [32], [33], and link prediction [34-37].

Raw texts: The objective of GCL is to maximize the similarity between positives and minimize the similarity between negatives [22], [68], [69].

Revised texts: Graph embedding converts graph structures into low-dimensional vectors, preserving key information. These vectors support a series of graph-related tasks like subgraph query [83-85] and community search [86-88]. Graph contrastive learning (GCL) is an important method for extracting graph embeddings and its objective is to maximize the similarity between positives and minimize the similarity between negatives [22], [89], [90].

Specifically, the newly added references include:
[27] J. Layne, J. Carpenter, E. Serra, and F. Gullo, “Temporal sir-gn: Efficient and effective structural representation learning

- for temporal graphs,” Proceedings of the VLDB Endowment, vol. 16, no. 9, pp. 2075–2089, 2023.
- [28] Y. Zhang and A. Kumar, “Lotan: Bridging the gap between gnns and scalable graph analytics engines,” Proceedings of the VLDB Endowment, vol. 16, no. 11, pp. 2728–2741, 2023.
- [29] F. Xiao, Y. Wu, M. Zhang, G. Chen, and B. C. Ooi, “Mint: Detecting fraudulent behaviors from time-series relational data,” Proceedings of the VLDB Endowment, vol. 16, no. 12, pp. 3610–3623, 2023.
- [30] X. Du, X. Zhang, S. Wang, and Z. Huang, “Efficient tree-svd for subset node embedding over large dynamic graphs,” Proceedings of the ACM on Management of Data, vol. 1, no. 1, pp. 1–26, 2023.
- [31] X. Zhang, Y. Shen, Y. Shao, and L. Chen, “Ducati: A dual-cache training system for graph neural networks on giant graphs with the gpu,” Proceedings of the ACM on Management of Data, vol. 1, no. 2, pp. 1–24, 2023.
- [32] G. Lv and L. Chen, “On data-aware global explainability of graph neural networks,” Proceedings of the VLDB Endowment, vol. 16, no. 11, pp. 3447–3460, 2023.
- [33] G. Lv, C. J. Zhang, and L. Chen, “Hence-x: Toward heterogeneity-agnostic multi-level explainability for deep graph networks,” Proceedings of the VLDB Endowment, vol. 16, no. 11, pp. 2990–3003, 2023.
- [34] C. Yang and J. Han, “Revisiting citation prediction with cluster-aware text-enhanced heterogeneous graph neural networks,” in 39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023, pp. 682–695, IEEE, 2023.
- [35] Y. Zhang, W. Wang, H. Yin, P. Zhao, W. Chen, and L. Zhao, “Disconnected emerging knowledge graph oriented inductive link prediction,” in 39th IEEE International Conference on Data Engineering, ICDE 2023, Anaheim, CA, USA, April 3-7, 2023, pp. 381–393, IEEE, 2023.
- [36] Y. Li, Y. Shen, L. Chen, and M. Yuan, “Orca: Scalable temporal graph neural network training with theoretical guarantees,” Proceedings of the ACM on Management of Data, vol. 1, no. 1, pp. 1–27, 2023.
- [37] H. Li and L. Chen, “Early: Efficient and reliable graph neural network for dynamic graphs,” Proceedings of the ACM on Management of Data, vol. 1, no. 2, pp. 1–28, 2023.
- [83] Y. Wang, A. Khan, T. Wu, J. Jin, and H. Yan, “Semantic guided and response times bounded top-k similarity search over knowledge graphs,” in 2020 IEEE 36th International Conference on Data Engineering (ICDE), pp. 445–456, IEEE, 2020.
- [84] Y. Wang, A. Khan, X. Xu, J. Jin, Q. Hong, and T. Fu, “Aggregate queries on knowledge graphs: Fast approximation with semantic-aware sampling,” in 2022 IEEE 38th International Conference on Data Engineering (ICDE), pp. 2914–2927, IEEE, 2022.
- [85] H.-W. Chen, H.-H. Shuai, D.-N. Yang, W.-C. Lee, C. Shi, S. Y. Philip, and M.-S. Chen, “Structure-aware parameter-free group query via heterogeneous information network trans-
- former,” in 2021 IEEE 37th International Conference on Data Engineering (ICDE), pp. 2075–2080, IEEE, 2021.
- [86] X. Xu, J. Liu, Y. Wang, and X. Ke, “Academic expert finding via (k, p)-core based embedding over heterogeneous graphs,” in 2022 IEEE 38th International Conference on Data Engineering (ICDE), pp. 338–351, IEEE, 2022.
- [87] J. Chen, Y. Xia, and J. Gao, “Communityaf: An example-based community search method via autoregressive flow,” Proceedings of the VLDB Endowment, vol. 16, no. 10, pp. 2565–2577, 2023.
- [88] J. Gao, J. Chen, Z. Li, and J. Zhang, “Ics-gnn: lightweight interactive community search via graph neural network,” Proceedings of the VLDB Endowment, vol. 14, no. 6, pp. 1006–1018, 2021.