A New Mechanism for Eliminating Implicit Conflict in Graph Contrastive Learning (Appendix)

A. Related Work

A.1 Graph Neural Networks

GNNs serve as a vital means for analyzing graph data that embodies complex relational information. In contrast to conventional neural networks, GNNs possess an additional message-passing mechanism, allowing information to propagate and aggregate within the topological structure. GCN (Kipf and Welling 2017) performs convolutional computations on the first-order neighbors of nodes in the spatial domain. Simultaneously, in the spectral domain, GCN utilizes first-order Chebyshev polynomial filters to approximate the spectral-based graph convolution. Within the field of graph self-supervised learning, GCN is the most widely used backbone. However, GCNs are limited in that the edge weights cannot be parameterized in message passing. GAT (Veličković et al. 2018) introduced an attention mechanism, assigning learnable parameters to the edges, thereby enhancing the performance of GCNs. It is noteworthy that even when GAT is used as the backbone for self-supervised training, they are still affected by the issue of EIC. Due to these conflicts, the weights of the edges in GATs tend to diminish to ensure minimal information interaction with negative nodes, fulfilling the minimization of similarity between negative pairs. This leads to an inability to effectively utilize the graph's topological structure. GraphSAGE (Hamilton, Ying, and Leskovec 2017), in comparison to GCN, has further optimized the neighbor sampling and message aggregation mechanisms.

A.2 Graph Self-Supervised Learning

Graph Self-Supervised Learning (GSSL) serves as a viable solution to the label dependency problem intrinsic to GNNs. It harnesses the innate potential of the data, mining and sampling within to identify non-manually extracted pseudo-labels, which are subsequently utilized to train the GNN encoders for the extraction of low-dimensional graph embeddings. Beyond the three primary frameworks discussed in the introduction, there exists the GBT (Bielak, Kajdanowicz, and Chawla 2022), an approach inspired by the Barlow Twins (Zbontar et al. 2021) used in CV. GBT optimizes the model by minimizing redundancy between the dimensions of the learned embeddings, thereby obviating the need for negative sampling. GVAE (Kipf and Welling 2016) employs an Auto-Encoder-based design. This design paradigm allows for training through the reconstruction of encoded data.

InfoNCE-based GCL is a mainstream method within the realm of GSSL. Inspired by visual studies such as SimCLR (Chen et al. 2020) and MoCo (He et al. 2020), works like GRACE (Zhu et al. 2020) generate augmented views and employ the InfoNCE (Oord, Li, and Vinyals 2018) loss function for training. These methods have achieved significant success and have become the paradigms. However, components derived from visual studies do not necessarily align well with graph data. In terms of data augmentation, original algorithms use random perturbations to generate augmented views. To avoid the destruction of crucial semantic information, GraphCL (You et al. 2020) explores various data augmentation techniques through experimental and theoretical investigations. GCA (Zhu et al. 2021) improves upon GRACE by perturbing less important nodes and edges with a higher probability, using centrality measures. ADGCL (Suresh et al. 2021) introduces the concept of min-max mutual information and applies learnable parametric perturbations to edges. The work (Liu et al. 2022) analyzes various data augmentation methods from a spectral perspective and proposes SpCo. AF-GCL (Wang et al. 2022) suggests eliminating data augmentation to prevent the destruction of semantic information. On the other hand, numerous studies have improved the positive and negative sampling mechanisms. LOCAL-GCL (Zhang et al. 2022) expands positive samples to the first-order neighbors of nodes, leveraging the graph homophily assumption to acquire positive samples. NeCo (He et al. 2023) discriminates homogeneity of the original graph edges, removing heterogeneous edges to obtain more accurate one-hop positive samples. ProGCL (Xia et al. 2021) uses the Beta distribution to model positive and negative examples, distinguishing the authenticity of negative samples.

Note that the occurrence of EIC is not directly related to the correct selection of positive and negative examples. While some of these sampling methods have reduced the number of negative samples within the receptive field, slightly decreasing the occurrence of EIC, many negative samples within the EIC still remain, presenting an issue that awaits resolution.

	Cora	CiteSeer	PubMed	CS	Photo	Computers
PiGCL-GT	83.71 ± 0.77	72.58 ± 0.85	86.70 ± 0.85	93.21 ± 0.85	93.10 ± 0.85	89.09 ± 0.85
PiGCL	**84.63 ± 0.78**	**73.51 ± 0.64**	**86.75 ± 0.20**	**93.30 ± 0.09**	**93.14 ± 0.30**	**89.25 ± 0.27**

Table 1: Ablation study for the component of gradient toning up. PiGCL-GT means PiGCL without gradient toning up

	Learning_Rate	Num_Hidden	Num_Proj_Hidden	Activation	Drop_Edge_Rate 1	Drop_Edge Rate_2
Cora	0.001	128	128	ReLU	0.4	0.4
CiteSeer	0.00005	1024	1024	ReLU	0.1	0.1
PubMed	0.001	512	1024	PReLU	0.4	0.4
Photo	0.00005	512	512	ReLU	0.3	0.3
CS	0.00005	512	512	ReLU	0.3	0.3
Computers	0.00001	1024	2048	ReLU	0.1	0.1

	Drop_Feature_Rate_1	Drop_Feature_Rate_2	Tau	Num_Epochs	Update_NegMatrix_Epoch	Ignore_rate
Cora	0.2	0.4	0.4	600	10	0.5
CiteSeer	0.4	0.4	1.0	200	5	0.15
PubMed	0.4	0.4	0.1	1500	10	0.05
Photo	0.2	0.2	0.4	3000	20	0.05
CS	0.2	0.2	0.4	600	20	0.05
Computers	0.3	0.3	0.05	1500	40	0.5

Table 2: Hyperparameter settings for PiGCL

B. Theoretical Supplement

Considering that the mathematical definition of EIC is intuitively difficult to understand, it is not beneficial for readers to give it before the proof part of the main text, and it is redundant to give it after the proof part. So we give here the mathematical definition of the EIC problem.

The theorem of EIC: $\forall v_a$ and its two negative samples v_b, v_c satisfied $TR(v_a) \cap TR(v_b) \neq TR(v_a) \cap TR(v_c) \neq \emptyset$. Consider a node set $C = \{v_i | v_i \in TR(v_a) \cap TR(v_b), v_i \notin TR(v_a) \cap TR(v_c)\}$ and a vector $\overrightarrow{\text{Vec}_C}$ composed of the sum of nodes representation of C. Then, $\min \mathcal{L}_{\text{InfoNCE}}(v_a, v_b) \simeq \min |\overrightarrow{\text{Vec}_C}|, \min \mathcal{L}_{\text{InfoNCE}}(v_a, v_c) \simeq \max |\overrightarrow{\text{Vec}_C}|$, where two negative samples are in conflict. $TR(v_i)$ means the topology receipted field of v_i.

C. Experiment Supplement

C.1 Datasets

We follow prior works (Zhu et al. 2020, 2021; Thakoor et al. 2021; Veličković et al. 2019) and evaluated our performance on six widely used datasets:

- **Cora**, **CiteSeer**, and **PubMed** are citation datasets from PlantoID. The nodes represent scientific papers, edges represent citation relationships, node features are bag-of-words vectors of the papers, and labels indicate the fields of the papers.
- **Photo** and **Computers** are co-purchase graphs from Amazon. The nodes represent products, edges represent frequent co-purchase relationships, node features are bag-of-words vectors of product reviews, and labels indicate the product categories.
- **CS**. is from Coauthor. The nodes represent authors, edges represent co-authorship relationships, node features are bag-of-words vectors of author's papers, and labels indicate the research domains of the authors.

The statistical information of the datasets is presented in Tab. 3. We utilized the Python library 'torch_geometric.dataset' to load the datasets. Bidirectional links were established for all edges.

Name	Nodes	Edges	Features	Classes
Cora	2708	4732	1433	7
CiteSeer	3327	5429	3703	6
PubMed	19717	44338	500	3
CS	18333	81894	6805	15
Photo	7650	119081	745	8
Computers	13752	245861	767	10

Table 3: Dataset statistics

C.2 Ablation Study for Gradient Toning Up

To further validate the effectiveness of the gradient toning up component, we conduct ablation studies as shown in Table 1. We make two findings: (1) The performance of PiGCL-GT outperforms GRACE on all datasets, which further verifies the efficacy of our partial ignore strategy. (2) PiGCL outperforms PiGCL-GT on all datasets, indicating that the gradient toning up component effectively compensates for the gradient fluctuations caused by ignoring negative nodes, thereby benefiting model training. This confirms the effectiveness of this component.

C.3 Implementation Details

Consistent with other GSSL methods (Zhu et al. 2020, 2021; Zhang et al. 2022), we adopt a two-layer GCN (Kipf and

Welling 2017) as the encoder $f(\cdot)$. The projector $g(\cdot)$ is implemented using a two-layer multilayer perceptron (MLP). We introduce a hyperparameter une (Update NegMatrix Epoch in Table 2) to control the frequency of iterations for capturing EIC negative pairs and updating the ignored set IgS. IgS is initially empty at the first iteration of training. The overall model parameters are summarized in Table 2.

References

Bielak, P.; Kajdanowicz, T.; and Chawla, N. V. 2022. Graph barlow twins: A self-supervised representation learning framework for graphs. *Knowledge-Based Systems*, 256: 109631.

Chen, T.; Kornblith, S.; Norouzi, M.; and Hinton, G. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, 1597–1607.

Hamilton, W.; Ying, Z.; and Leskovec, J. 2017. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30: 1024–1034.

He, D.; Zhao, J.; Guo, R.; Feng, Z.; Jin, D.; Huang, Y.; Wang, Z.; and Zhang, W. 2023. Contrastive Learning Meets Homophily: Two Birds with One Stone. In *International Conference on Machine Learning*.

He, K.; Fan, H.; Wu, Y.; Xie, S.; and Girshick, R. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9729–9738.

Kipf, T. N.; and Welling, M. 2016. Variational graph auto-encoders. *arXiv preprint arXiv:1611.07308*.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *International Conference on Learning Representations*.

Liu, N.; Wang, X.; Bo, D.; Shi, C.; and Pei, J. 2022. Revisiting graph contrastive learning from the perspective of graph spectrum. *Advances in Neural Information Processing Systems*, 35: 2972–2983.

Oord, A. v. d.; Li, Y.; and Vinyals, O. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*.

Suresh, S.; Li, P.; Hao, C.; and Neville, J. 2021. Adversarial graph augmentation to improve graph contrastive learning. *Advances in Neural Information Processing Systems*, 34: 15920–15933.

Thakoor, S.; Tallec, C.; Azar, M. G.; Munos, R.; Veličković, P.; and Valko, M. 2021. Bootstrapped representation learning on graphs. In *ICLR 2021 Workshop on Geometrical and Topological Representation Learning*.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2018. Graph Attention Networks. In *International Conference on Learning Representations*.

Veličković, P.; Fedus, W.; Hamilton, W. L.; Liò, P.; Bengio, Y.; and Hjelm, R. D. 2019. Deep Graph Infomax. In *International Conference on Learning Representations*.

Wang, H.; Zhang, J.; Zhu, Q.; and Huang, W. 2022. Augmentation-free graph contrastive learning with performance guarantee. *arXiv preprint arXiv:2204.04874*.

Xia, J.; Wu, L.; Wang, G.; Chen, J.; and Li, S. Z. 2021. Progcl: Rethinking hard negative mining in graph contrastive learning. In *International conference on machine learning*, 24332–24346.

You, Y.; Chen, T.; Sui, Y.; Chen, T.; Wang, Z.; and Shen, Y. 2020. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33: 5812–5823.

Zbontar, J.; Jing, L.; Misra, I.; LeCun, Y.; and Deny, S. 2021. Barlow twins: Self-supervised learning via redundancy reduction. In *International Conference on Machine Learning*, 12310–12320.

Zhang, H.; Wu, Q.; Wang, Y.; Zhang, S.; Yan, J.; and Yu, P. S. 2022. Localized Contrastive Learning on Graphs. *arXiv preprint arXiv:2212.04604*.

Zhu, Y.; Xu, Y.; Yu, F.; Liu, Q.; Wu, S.; and Wang, L. 2020. Deep graph contrastive representation learning. *arXiv preprint arXiv:2006.04131*.

Zhu, Y.; Xu, Y.; Yu, F.; Liu, Q.; Wu, S.; and Wang, L. 2021. Graph contrastive learning with adaptive augmentation. In *Proceedings of the Web Conference 2021*, 2069–2080.