Dual Prototype-Enhanced Contrastive Framework for Class-Imbalanced Graph Domain Adaptation

Xin Ma[1*], **Yifan Wang**[2*], **Siyu Yi**[3], **Wei Ju**[1†], **Junyu Luo**[4], **Yusheng Zhao**[4], **Xiao Luo**[5], **Jiancheng Lv**[1]
[1]College of Computer Science, Sichuan University
[2]School of Information Technology & Management, University of International Business and Economics
[3]College of Mathematics, Sichuan University [4]Peking University [5]UCLA
maxin88@stu.scu.edu.cn, juwei@scu.edu.cn

Abstract

Graph transfer learning, especially in unsupervised domain adaptation, aims to transfer knowledge from a label-abundant source graph to an unlabeled target graph. However, most existing approaches overlook the common issue of label imbalance in the source domain, typically assuming a balanced label distribution that rarely holds in practice. Moreover, they face challenges arising from biased knowledge in the source graph and substantial domain distribution shifts. To remedy the above challenges, we propose a dual-branch prototype-enhanced contrastive framework for graph domain adaptation under a class-imbalanced scenario. Specifically, we introduce a dual-branch graph encoder to capture both local and global information, generating class-specific prototypes from a distilled anchor set. Then, a prototype-enhanced contrastive learning framework is introduced. On the one hand, we encourage class alignment between the two branches based on constructed prototypes to alleviate the bias introduced by class imbalance. On the other hand, we infer the pseudo-labels for the target domain and align sample pairs across domains that share similar semantics to reduce domain discrepancies. Experimental results show that our ImGDA outperforms the state-of-the-art methods across multiple datasets and settings. The code is available at: https://github.com/maxin88scu/ImGDA.

1 Introduction

Graph serves as a versatile data structure to represent complex relationships [12, 47] in a number of fields such as social networks [2], molecular biology [13, 55] and recommender systems [20, 44]. One fundamental task for graph-structured data is node classification, which endeavors to predict the category of each node within the graph and is widely applied in various applications, i.e., community detection [51], smart city [58] and knowledge graph [49]. Nevertheless, the effectiveness of this task heavily relies on label information, which is time-consuming and costly [9]. Graph transfer learning [19, 61] has emerged as an effective framework for tackling this problem by leveraging labeled information from a source graph to facilitate learning on an unlabeled target graph, significantly enhancing the model's ability to generalize on the target graph.

Actually, there are several approaches that apply graph transfer learning for domain adaptation in the node classification task. Thanks to the powerful capabilities of graph neural networks (GNNs), unsupervised graph domain adaptation focuses on reducing the distribution discrepancy between

[*]Equal Contributions
[†]Corresponding Author

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

target and source graphs within the latent representation space induced by GNNs [23, 53, 56]. Distance-based methods minimize a divergence measure between their distributions to learn invariant representations, such as maximum mean discrepancy [34] and graph subtree discrepancy [45] to align domains. Adversarial-based methods incorporate a discriminator to distinguish between the two domains and produce invariant features to confound the discriminator [4, 32].

Despite the promising performance of graph transfer learning, it often relies on the unrealistic assumption that the source domain graph exhibits a balanced label distribution. In practice, however, real-world graphs usually exhibit long-tailed structures, where most classes contain only a few labeled nodes (tail classes) and a small number of classes dominate with many labeled samples (head classes) [11, 14], resulting in severe class imbalance. For example, in the NCI dataset, which consists of graphs of chemical compounds [40], only around 5% of compounds exhibit anti-cancer activity, with the overwhelming remainder are labeled as inactive. Therefore, this class-imbalanced issue inevitably leads to the knowledge extraction bias [59] in the source graph which in turn adversely affects the graph domain adaptation. This naturally spurs a question: *How can label semantics in the target graph be effectively inferred under domain shifts and severe source imbalance?*

However, designing an effective framework for graph domain adaptation under class imbalance remains challenging due to several critical obstacles: ❶ *How to sufficiently alleviate the class imbalance for extracting knowledge from the source graph?* The latent space formed by imbalanced training data is highly skewed, with minor subspaces being compressed by the dominant ones. This imbalance forces the model to focus primarily on the head class, which results in insufficient learning of the tail class to extract unbiased knowledge. ❷ *How to effectively reduce the domain discrepancy to make accurate predictions on the target graph?* Since the target graph is entirely unlabeled, existing methods primarily focus on entire domain alignment, overlooking class-level distributions. Therefore, as the result shown in Figure 1, existing graph domain adaptation methods struggle to transfer the correct semantic knowledge, leading to poor performance, particularly for tail classes.

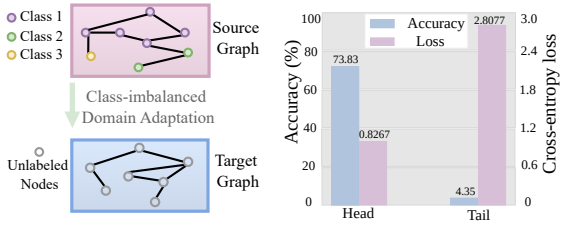


Figure 1: Illustration of class-imbalanced graph domain adaptation. The unsupervised domain adaptation method GDA-SpecReg suffers from an imbalanced class of source graphs, particularly for the tail class.

Towards this end, in this paper, we propose a holistic method termed **ImGDA**, a dual-branch prototype-enhanced contrastive framework for class-**Im**balanced **G**raph **D**omain **A**daption, which facilitates the transfer of class-imbalanced label information from the source graph to the unlabeled target graph. Specifically, we develop a dual graph encoder that jointly captures local and global information to learn generalized node representations. Building on this, a prototype for each class is generated from a distilled anchor set. Then, we introduce a prototype-aware contrastive learning framework that integrates cross-branch and cross-domain contrastive learning to enhance adaptation performance. To rebalance the feature space, the cross-branch prototype contrast encourages class alignment between the two branches. To further reduce domain discrepancy, we infer pseudo-labels of nodes from the target domain graph in a non-parametric manner and align sample pairs across domains that exhibit the same semantic meaning. Additionally, we treat temperature as a rebalancing parameter to mitigate class imbalance during training. Extensive experimental results validate the effectiveness of our proposed ImGDA for class-imbalanced graph domain adaptation.

2 Related Work

Unsupervised Domain Adaptation (UDA). UDA focuses on learning domain-invariant representations between a labeled source domain and an unlabeled target domain to transfer cross-domain information [24, 26, 43]. Recently, graph UDA methods have included adversarial learning and reducing domain discrepancies based on certain metrics (e.g., MMD [7], subtree discrepancy [45]). Among adversarial methods, RNA [25] adversarially extracts domain-invariant subgraphs to address domain shift, while leveraging spectral seriation for robust alignment under label scarcity. In the metric-based category, GDASpecReg [53] leverages spectral smoothness and maximum frequency response regularizations to enhance GNN transferability across node and link transfer scenarios,

while A2GNN [23] improves cross-domain transfer by adjusting GNN propagation layers using Lipschitz bounds. However, our ImGDA studies a novel yet practical scenario of imbalanced source data, and introduces a dual prototype-aware contrastive framework to reduce source graph bias and align the semantic space between source and target for cross-domain knowledge transfer.

Class-Imbalanced Learning on Graphs. It is well recognized that class imbalance on graphs can degrade classification performance. To address this problem, there are typically three approaches: (a) Modifying the loss function to prioritize underrepresented classes [3, 29, 36, 50]. (b) Post-hoc correction to modify logits for the tail class [10, 16]. (c) Re-sampling techniques that augment or generate tail class data [27, 31, 42, 57]. For instance, GraphSHA [22] expands tail class decision boundaries by creating more challenging samples, while ImGCL [54] addresses class imbalance by employing balanced sampling and explicitly accounting for node centrality within a contrastive learning paradigm. Extending these strategies to graph-level tasks, C^{3}GNN [14] addresses class-imbalanced graph classification by clustering majority classes and applying Mixup to learn hierarchical representations, while KDEX [28] transfers head-to-tail knowledge and trains diverse experts that are adaptively combined via a self-supervised router. Additionally, Qin et al. [33] further propose a comprehensive benchmark IGL-Bench for imbalanced graph learning, which evaluate the effectiveness, robustness, and efficiency of various algorithms within a unified framework. However, distribution shifts in graphs further complicate class-imbalanced learning on graphs. To address this, our proposed ImGDA introduces cross-branch prototype contrast to reduce class imbalance bias and cross-domain prototype contrast to align domain discrepancies, effectively generating domain-invariant representations.

3 Notations and Problem Definition

Source Domain Graph. Denote $\mathcal{G}^s = \{\mathcal{V}^s, \mathcal{E}^s, \boldsymbol{X}^s, \boldsymbol{Y}^s\}$ the source domain graph with the labeled node $\mathcal{V}^s$ and edge sets $\mathcal{E}^s$. The node feature matrix can be represented as $\boldsymbol{X}^s \in \mathbb{R}^{|\mathcal{V}^s| \times d}$, where entry $\boldsymbol{x}_v \in \mathbb{R}^d$ is associated with a feature vector of node v with dimension d. We use the adjacency matrix $\boldsymbol{A}^s \in \mathbb{R}^{|\mathcal{V}^s| \times |\mathcal{V}^s|}$ to describe the structure information of the graph, where $\boldsymbol{A}^s_{ij} = 1$ if an edge exists between v_i and v_j, i.e., $(v_i, v_j) \in \mathcal{E}^s$, otherwise, $\boldsymbol{A}^s_{ij} = 0$. The degree matrix is denoted as $\boldsymbol{D} = \text{diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_N)$ with a degree of each node $\boldsymbol{D}_i = \sum_{j=1}^{|\mathcal{V}^s|} \boldsymbol{A}^s_{ij}$. We denote the node label matrix of the source domain graph as $\boldsymbol{Y}^s \in \mathbb{R}^{|\mathcal{V}^s| \times C}$, where C corresponds to the total classes.

Target Domain Graph. The target domain graph is denoted as $\mathcal{G}^t = \{\mathcal{V}^t, \mathcal{E}^t, \boldsymbol{X}^t\}$ with unlabeled node set $\mathcal{V}^t$ and edge set $\mathcal{E}^t$. Similarly, the adjacency matrix $\boldsymbol{A}^t \in \mathbb{R}^{|\mathcal{V}^t| \times |\mathcal{V}^t|}$ indicates the node connectivity information in the target domain graph. And the feature matrix can be represented as $\boldsymbol{X}^t$. The attribute sets of the source and target domain graph could exhibit significant differences. Here, we construct a unified attribute set across both domains to align the dimensions.

Problem Definition. We consider $\mathcal{G}^s$ as a fully labeled source graph and $\mathcal{G}^t$ as an unlabeled target graph. Each c-th class in $\mathcal{G}^s$ contains N_c nodes, ordered such that $N_C \geq N_2 \geq \cdots \geq N_C$. The source domain graph is assumed to be class-imbalanced, with the degree of imbalance quantified by the factor N_1/N_C. The objective of the task is to mitigate this class imbalance while transferring knowledge from the source graph to the target domain graph to achieve accurate node label prediction. Figure 2 provides a schematic overview of our proposed framework in the following.

4 Methodology

4.1 Dual-Branch Embedding Generalization

Generalized node embeddings are crucial for effective graph domain adaptation. Therefore, we introduce a dual-branch graph encoder to fully capture both local and global structure information of the graph [30, 62].

Local Consistency Encoder. To capture local consistency knowledge (i.e., neighboring samples are prone to share the same semantics), we directly utilize the Graph Convolutional Network (GCN) [18] as the encoder. Given the adjacency matrix $\boldsymbol{A}^*$ and feature matrix $\boldsymbol{X}^*$ ($* \in \{s, t\}$) of both source and target domain graphs, the output of the l-th layer is defined as:

$$\boldsymbol{Z}^{(l)}_{*,local} = \sigma(\tilde{\boldsymbol{D}}^{*^{-\frac{1}{2}}} \tilde{\boldsymbol{A}}^* \tilde{\boldsymbol{D}}^{*^{-\frac{1}{2}}} \boldsymbol{Z}^{(l-1)}_{*,local} \boldsymbol{W}^{*(l)}), \tag{1}$$

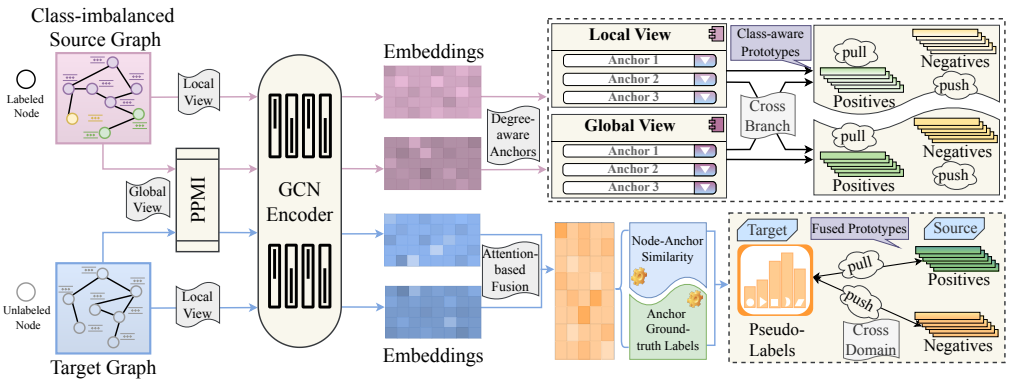


Figure 2: A schematic overview of our proposed ImGDA framework.

where $\tilde{\boldsymbol{A}}^* = \boldsymbol{I}_{|\mathcal{V}^*|} + \boldsymbol{A}^*$ is the adjacency matrix containing self-loop, $\tilde{\boldsymbol{D}}^*$ is the corresponding degree matrix accordingly. $\boldsymbol{W}^{*(l)}$ serves as the learnable filter in l-th layer and the initial feature matrix is $\boldsymbol{Z}^{(0)}_{*,local} = \boldsymbol{X}^*$. Here $\sigma(\cdot)$ denotes the activation function. By stacking L graph convolutional layers, the extracted local consistency knowledge can be expressed as $\boldsymbol{Z}^*_{local} = \boldsymbol{Z}^{(L)}_{*,local}$.

Global Consistency Encoder. Furthermore, we utilize a graph encoding strategy based on positive pointwise mutual information (PPMI) [46, 62]. We represent the state at current time t as $s(t) = v_i$, with the probability transit from the node v_i to any of its neighbors v_j expressed as:

$$\boldsymbol{P}^*_{ij} = p(s(t+1) = v_j | s(t) = v_i) = \boldsymbol{A}^*_{ij} / \boldsymbol{D}_i. \tag{2}$$

We apply the random walk guided by $\boldsymbol{P}^*$ to generate a collection of node paths on $\boldsymbol{A}^*$. From these paths, we construct co-occurrence frequency matrix $\boldsymbol{F}^* \in \mathbb{R}^{|\mathcal{V}^*| \times |\mathcal{V}^*|}$, where $\boldsymbol{F}_{ij}$ records how often node v_j appears within a predefined window around v_i. The PPMI between nodes is calculated as:

$$\tilde{\boldsymbol{P}}^*_{ij} = \frac{\boldsymbol{P}^*_{ij}}{\sum_{i,j} \boldsymbol{P}^*_{ij}}, \tilde{\boldsymbol{P}}^*_i = \frac{\sum_j \boldsymbol{P}^*_{ij}}{\sum_{i,j} \boldsymbol{P}^*_{ij}}, \tilde{\boldsymbol{P}}^*_j = \frac{\sum_i \boldsymbol{P}^*_{ij}}{\sum_{i,j} \boldsymbol{P}^*_{ij}}, \quad \boldsymbol{M}^*_{ij} = \max\{\log(\frac{\tilde{\boldsymbol{P}}^*_{ij}}{\tilde{\boldsymbol{P}}^*_i \times \tilde{\boldsymbol{P}}^*_j}), 0\}, \tag{3}$$

where $\tilde{\boldsymbol{P}}^*_{ij}$ is the probability that v_j appears in the v_i's context, $\tilde{\boldsymbol{P}}^*_i$ and $\tilde{\boldsymbol{P}}^*_j$ are the estimated probability of node v_i and context v_j respectively. $\boldsymbol{M}^*_{ij}$ captures the high-order topological relationship between nodes. Thus, nodes that frequently co-occur at high frequency will have larger $\boldsymbol{M}^*_{ij}$ values compared to independent nodes. By treating the PPMI matrix $\boldsymbol{M}^*$ as a new adjacency matrix, we can effectively extract global consistency knowledge as:

$$\boldsymbol{Z}^{(l)}_{*,global} = \sigma(\boldsymbol{D}^{*-\frac{1}{2}} \boldsymbol{M}^* \boldsymbol{D}^{*-\frac{1}{2}} \boldsymbol{Z}^{(l-1)}_{*,global} \boldsymbol{W}^{*(l)}), \tag{4}$$

where $\boldsymbol{D}^*_i = \sum_j \boldsymbol{M}^*_{ij}$ and $\boldsymbol{W}^{*(l)}$ is shared learnable parameters used with the local consistency encoder. Similarly, the global structural consistency can be also extracted by stacking L graph convolutional layers, namely, $\boldsymbol{Z}^*_{global} = \boldsymbol{Z}^{(L)}_{*,global}$.

Attention-Based Consistency Fusion. To fuse the extracted local and global consistency knowledge, we leverage the attention mechanism [38] and the attention coefficients can be obtained as:

$$\zeta_{ij} = \text{softmax}(\phi(\boldsymbol{W}^* \boldsymbol{z}^*_{i,local}, \boldsymbol{W}^* \boldsymbol{z}^*_{j,global})), \tag{5}$$

where $\boldsymbol{W}^*$ is the shared parameter matrix, $\phi(\boldsymbol{z}_i, \boldsymbol{z}_j) = \text{LeakyRelu}(\boldsymbol{W}_0^{*\top}[\boldsymbol{z}_i || \boldsymbol{z}_j])$ denotes the attention function with parameter $\boldsymbol{W}^*_0$. The fused consistency knowledge of both domains can be:

$$\boldsymbol{Z}^* = \zeta_{ii} \boldsymbol{Z}^*_{local} + (1 - \zeta_{ii}) \boldsymbol{Z}^*_{global}. \tag{6}$$

4.2 Anchor-Based Prototype Construction

In contrast to directly using cross-entropy loss on class-imbalanced data, supervised contrastive learning (SCL) tends to achieve better results [15, 17, 60]. However, when the dataset is highly imbalanced, the feature space remains dominated by head class. To avoid overemphasis on head

classes, a straightforward approach is to generate a set of uniformly distributed prototypes for all classes, ensuring that each contributes approximately equally during optimization. The subsequent analysis provides theoretical guarantees that this strategy can effectively alleviate the class imbalance.

Theorem 4.1. *Let $\boldsymbol{Z} = \{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_N\}, ||\boldsymbol{z}_i|| = 1$ be the extracted consistency knowledge of N node points with label $\boldsymbol{Y} = \{y_1, \ldots, y_N\}$. The supervised contrastive loss for a class c in a batch $\mathcal{B}$ is bounded by:*

$$\mathcal{L}_{SCL}(\boldsymbol{Z}; \boldsymbol{Y}, \mathcal{B}, c) \geq \sum_{i \in \mathcal{B}_c} \log((|\mathcal{B}_c| - 1) + |\bar{\mathcal{B}_c}| \exp(\underbrace{\frac{1}{|\bar{\mathcal{B}_c}|} \sum_{k \in \bar{\mathcal{B}_c}} \boldsymbol{z}_i \cdot \boldsymbol{z}_k}_{repulsion\ term} - \underbrace{\frac{1}{|\mathcal{B}_c| - 1} \sum_{j \in \mathcal{B}_c \setminus \{i\}} \boldsymbol{z}_i \cdot \boldsymbol{z}_j}_{attraction\ term})), \quad (7)$$

where $\mathcal{B}_c$ represents the subset within $\mathcal{B}$ containing all samples of class c and $\bar{\mathcal{B}_c}$ denotes its complement set.

The lower bound of SCL above is derived from [6] and comprises two terms. The attraction term encourages intra-class instances to converge toward their prototypes regardless of the class distribution, while the repulsion term enforces uniform inter-class separation and is dominated by head classes.

Theorem 4.2. *Let $\mathcal{C}_{\mathcal{B}}, |\mathcal{C}_{\mathcal{B}}| \leq C$ denotes the set of classes that appears in $\mathcal{B}$. The supervised contrastive loss for a class c after class averaging is bounded by:*

$$\begin{aligned} \mathcal{L}_{SCL}(\boldsymbol{Z}; \boldsymbol{Y}, \mathcal{B}, c) \geq \sum_{i \in \mathcal{B}_c} & \log(1 + (|\mathcal{C}_{\mathcal{B}}| - 1) \\ & \times \exp(\underbrace{\frac{1}{|\mathcal{C}_{\mathcal{B}}| - 1} \sum_{q \in \mathcal{C}_{\mathcal{B}} \setminus \{c\}} \frac{1}{|\mathcal{B}_q|} \sum_{k \in \mathcal{B}_q} \boldsymbol{z}_i \cdot \boldsymbol{z}_k}_{repulsion\ term} - \underbrace{\frac{1}{|\mathcal{B}_c| - 1} \sum_{j \in \mathcal{B}_c \setminus \{i\}} \boldsymbol{z}_i \cdot \boldsymbol{z}_j}_{attraction\ term})), \end{aligned} \quad (8)$$

Therefore, head classes no longer dominate the repulsion term. To fully train all classes in $\mathcal{B}$, we learn the prototype of each class for prototype-aware contrastive learning. The proofs of the theorem are in Appendix A.

In practice, we distill the original graph to sample the anchor set consisting of the most important nodes. Then, we calculate the prototype based on the anchor set to facilitate effective learning. Specifically, for each class c in the source domain graph, the top k nodes ranked by the degree are selected to get the sub-graph of anchor nodes $\mathcal{G}_a^s$. We calculate the prototype as the mean vector of each class, namely, $\boldsymbol{\mu}_{c,q} = \frac{1}{k} \sum_{i=1}^{k} \boldsymbol{z}_{i,q}^s, y_i^s = c, q \in \{local, global\}$.

4.3 Prototype-Enhanced Contrastive Learning

Given the prototype sets of two branches for both source and target domain graphs, we formalize a prototype-aware contrastive learning framework, which integrates cross-branch and cross-domain prototype contrastive learning to mitigate class imbalance and effective domain adaptation.

Cross-Branch Prototype Contrast. Considering that the model extracts both the local and global consistency knowledge from both complementary branches, we contrast the consistency knowledge of the source domain graph between the two branches in a prototype manner to mitigate the imbalance effect [1]. Specifically, we calculate the prototype of both local and global consistency knowledge from two branches as $\{\boldsymbol{\mu}_{c,local}\}_{c=1}^{C}$ and $\{\boldsymbol{\mu}_{c,global}\}_{c=1}^{C}$. For each query node $v_i^s \in \mathcal{V}^s$, we use the prototype with the same class label as the positive sample and the cross-branch prototype contrastive loss can be defined as:

$$\mathcal{L}_{cb} = \frac{1}{4|\mathcal{V}^s|} \sum_{i=1}^{|\mathcal{V}_s|} \sum_{\boldsymbol{z}_+ = \boldsymbol{z}_{i,local}^s}^{\boldsymbol{z}_{i,global}^s} \sum_{\boldsymbol{\mu}_+ = \boldsymbol{\mu}_{y_i^s,local}}^{\boldsymbol{\mu}_{y_i^s,global}} \log\Big(\frac{\exp(\boldsymbol{z}_+ \cdot \boldsymbol{\mu}_+ / \tau)}{\exp(\boldsymbol{z}_+ \cdot \boldsymbol{\mu}_+ / \tau) + \sum_{\boldsymbol{\mu} \in \mathcal{P}_i^-} \exp(\boldsymbol{z}_+ \cdot \boldsymbol{\mu} / \tau)}\Big), \quad (9)$$

where $\mathcal{P}_i^-$ denote the prototype set excluding $\boldsymbol{u}_{y_i^s,q}, q \in \{local, global\}$ in two branches respectively, and τ is the temperature parameter for contrastive learning.

Cross-Domain Prototype Contrast. To alleviate the domain shift in the graph space, we seek to align the consistent knowledge of nodes in the source domain graph with nodes in the target domain graph that share the same semantics. To achieve this, we infer pseudo-labels of nodes from the target

domain graph in a non-parametric manner and employ cross-domain prototype contrastive learning between cross-domain pairs [52]. The pseudo-label for node v_j^t in the target domain graph can be:

$$\hat{p}_j^t = \sum_{(v_i^s, y_i^s) \in \mathcal{G}_a^s} \Big(\frac{\exp(\boldsymbol{z}_j^t \cdot \boldsymbol{z}_i^s / \tau)}{\sum_{(v_i^s, y_i^s) \in \mathcal{G}_a^s} \exp(\boldsymbol{z}_j^t \cdot \boldsymbol{z}_i^s / \tau)} \Big) \boldsymbol{Y}_i^s, \tag{10}$$

where $\boldsymbol{Y}_i^s$ corresponds to the one-hot label of v_i. The pseudo-label can be derived as $\hat{y}_j^t = \arg\max(\hat{p}_j^t)$. Then, for each query node $v_j^t \in \mathcal{V}^t$, we pull semantically similar prototypes close compared to those with different semantics:

$$\mathcal{L}_{cd} = \frac{1}{|\mathcal{V}^t|} \sum_{j=1}^{|\mathcal{V}^t|} \log \Big(\frac{\exp(\boldsymbol{z}_j^t \cdot \boldsymbol{\mu}_{\hat{y}_j^t})}{\exp(\boldsymbol{z}_j^t \cdot \boldsymbol{\mu}_{\hat{y}_j^t}) + \sum_{\boldsymbol{\mu} \in \mathcal{P}_j^-} \exp(\boldsymbol{z}_j^t \cdot \boldsymbol{\mu})} \Big), \tag{11}$$

where $\mathcal{P}_j^-$ denotes the prototype set of fused consistency knowledge excluding $\boldsymbol{\mu}_{\hat{y}_j^t}$ in the source domain graph. Note that the cross-domain contrastive learning process can be interpreted as an Expectation Maximization (EM) scheme, where the aligned semantics between the source and target domain nodes are inferred in the E-step, and the log-likelihood of the nodes in the target domain graph is maximized in the M-step. The proof can be seen in Appendix B.

Adaptive Temperature Formulation. The temperature parameter τ in contrastive learning controls the penalty on hard negative samples [41]. However, for tail classes, where node samples are fewer, increasing τ has a negligible impact but reduces their gradients, worsening class imbalance. We propose an adaptive mechanism in which the temperature for each class is adjusted according to the number of samples within that class:

$$\tau_c = \gamma + (1 - \gamma) \cdot N_c / N_C, \tag{12}$$

where τ_c is the temperature of class c, and γ here denotes the minimum value of temperature.

4.4 Overall Optimization

To further reduce the impact of class imbalance in the source domain graph, we employ a logit compensation strategy to correct the consistency knowledge [26, 29, 59], summarized as follows:

$$\mathcal{L}_{lc} = -\lambda_c \log \frac{\exp(\varphi_y(\boldsymbol{z}_i^s) + \delta_c)}{\sum_{c'=1}^{C} \exp(\varphi_{c'}(\boldsymbol{z}_i^s) + \delta_{c'})}, \tag{13}$$

where $\varphi(\cdot)$ is the classification function that outputs the logit for each label, λ_y represents the contribution weight for class y_i^s, and δ_c denotes the compensation value for class c. Here, we have $\lambda_c = 1$ and $\delta_c = \log N_c$, following the previous work [29]. Finally, the objective can be:

$$\mathcal{L} = \mathcal{L}_{cb} + \mathcal{L}_{cd} + \mathcal{L}_{lc}. \tag{14}$$

Typically, the dynamic weight hyperparameters can be used to balance these losses during training, but we find in practice that a simple addition already yields effective results.

5 Experiment

5.1 Experimental Settings

Benchmark Datasets. This study conducts experiments on three publicly accessible network datasets from the ArnetMiner [37]: ACMv9 (A), Citationv1 (C), and DBLPv7 (D). These datasets are derived from different sources and cover distinct time periods: ACM (post-2010), Microsoft Academic Graph (pre-2008), and DBLP (2004-2008), resulting in diverse domain characteristics. All datasets model academic papers as nodes and construct undirected edges to represent citation relationships. To realistically simulate real-world class imbalance scenarios, we employ the operation proposed by [31] for preprocessing source domain data. This involves iteratively adjusting the number of nodes within classes in the source domain to achieve specified imbalance factors (IF). The imbalance factor is defined as $\rho = N_1/N_C$, where class populations follow $N_1 \geq N_2 \geq \cdots \geq N_C$, with N_c denoting the node count for class c in the source graph. The experiments on more metrics of imbalanced distribution can be found in Appendix E.

Table 1: Results of methods across varying imbalance factors (ρ). Here A⇒C represents using A as the source graph and C as the target graph. Scores are reported as micro-average (%) and macro-average (%). The top-performing method is shown in **bold** with the runner-up is underlined.

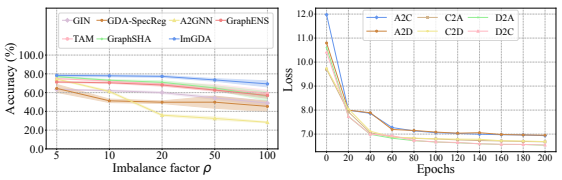
Methods	IF (ρ)	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
		Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
GCN	10	66.01$_{\pm 1.15}$	60.90$_{\pm 2.14}$	61.21$_{\pm 0.57}$	53.20$_{\pm 0.71}$	58.25$_{\pm 1.22}$	56.52$_{\pm 2.87}$	63.39$_{\pm 0.32}$	58.38$_{\pm 1.27}$	55.47$_{\pm 0.39}$	52.80$_{\pm 0.86}$	61.68$_{\pm 0.38}$	58.35$_{\pm 0.68}$
	20	62.28$_{\pm 1.17}$	55.48$_{\pm 1.79}$	60.75$_{\pm 0.69}$	51.60$_{\pm 1.48}$	51.17$_{\pm 0.58}$	43.78$_{\pm 1.54}$	59.00$_{\pm 0.88}$	48.22$_{\pm 1.57}$	50.93$_{\pm 2.70}$	45.12$_{\pm 4.44}$	54.02$_{\pm 0.27}$	46.18$_{\pm 0.55}$
	50	53.07$_{\pm 3.39}$	42.41$_{\pm 3.32}$	56.26$_{\pm 2.21}$	43.35$_{\pm 3.53}$	44.38$_{\pm 1.12}$	32.20$_{\pm 2.51}$	53.06$_{\pm 0.55}$	37.09$_{\pm 1.18}$	41.86$_{\pm 1.31}$	31.11$_{\pm 1.50}$	42.28$_{\pm 3.28}$	32.05$_{\pm 3.30}$
GAT	10	66.48$_{\pm 0.72}$	58.37$_{\pm 2.74}$	61.58$_{\pm 0.75}$	51.44$_{\pm 1.02}$	55.10$_{\pm 2.47}$	51.54$_{\pm 4.12}$	60.87$_{\pm 1.00}$	52.99$_{\pm 1.34}$	56.18$_{\pm 0.31}$	53.11$_{\pm 0.80}$	61.64$_{\pm 0.23}$	58.22$_{\pm 0.53}$
	20	61.63$_{\pm 3.34}$	50.54$_{\pm 3.72}$	61.04$_{\pm 1.87}$	50.10$_{\pm 1.46}$	47.18$_{\pm 1.04}$	37.30$_{\pm 1.61}$	55.95$_{\pm 1.26}$	43.32$_{\pm 2.17}$	49.18$_{\pm 1.59}$	42.20$_{\pm 2.78}$	53.05$_{\pm 3.25}$	45.29$_{\pm 2.63}$
	50	53.21$_{\pm 7.62}$	42.07$_{\pm 7.64}$	53.67$_{\pm 4.00}$	40.36$_{\pm 5.33}$	41.02$_{\pm 0.99}$	26.41$_{\pm 1.18}$	49.76$_{\pm 0.63}$	32.29$_{\pm 1.15}$	40.14$_{\pm 1.46}$	28.89$_{\pm 1.85}$	45.38$_{\pm 7.90}$	34.77$_{\pm 8.31}$
GIN	10	62.12$_{\pm 0.69}$	55.03$_{\pm 0.84}$	60.91$_{\pm 0.79}$	54.39$_{\pm 1.48}$	54.35$_{\pm 1.75}$	49.21$_{\pm 1.31}$	60.14$_{\pm 1.00}$	54.51$_{\pm 2.16}$	52.59$_{\pm 0.67}$	49.43$_{\pm 2.42}$	59.89$_{\pm 1.42}$	54.68$_{\pm 2.28}$
	20	60.03$_{\pm 1.03}$	50.98$_{\pm 0.94}$	59.12$_{\pm 1.05}$	48.27$_{\pm 0.92}$	50.38$_{\pm 2.75}$	42.80$_{\pm 3.88}$	55.87$_{\pm 2.47}$	46.42$_{\pm 4.71}$	47.79$_{\pm 1.49}$	41.02$_{\pm 2.77}$	53.47$_{\pm 2.49}$	44.83$_{\pm 2.61}$
	50	54.39$_{\pm 1.62}$	44.28$_{\pm 1.69}$	54.43$_{\pm 1.60}$	41.90$_{\pm 2.09}$	41.26$_{\pm 0.37}$	29.45$_{\pm 0.51}$	47.05$_{\pm 0.60}$	31.57$_{\pm 1.13}$	43.38$_{\pm 1.26}$	34.41$_{\pm 2.05}$	48.59$_{\pm 2.15}$	39.22$_{\pm 2.29}$
GDA-SpecReg	10	51.27$_{\pm 1.86}$	39.29$_{\pm 2.43}$	53.69$_{\pm 3.02}$	40.87$_{\pm 5.56}$	49.69$_{\pm 4.93}$	39.08$_{\pm 7.51}$	57.41$_{\pm 3.15}$	45.54$_{\pm 4.35}$	52.73$_{\pm 3.20}$	47.07$_{\pm 7.86}$	56.37$_{\pm 1.40}$	45.04$_{\pm 3.46}$
	20	49.68$_{\pm 1.42}$	36.44$_{\pm 1.70}$	51.34$_{\pm 3.70}$	35.82$_{\pm 3.00}$	48.11$_{\pm 4.07}$	38.20$_{\pm 6.83}$	51.64$_{\pm 1.96}$	35.15$_{\pm 3.49}$	43.84$_{\pm 3.79}$	31.89$_{\pm 3.72}$	44.53$_{\pm 6.48}$	31.21$_{\pm 6.80}$
	50	49.80$_{\pm 6.48}$	36.09$_{\pm 7.63}$	47.44$_{\pm 4.58}$	29.36$_{\pm 6.49}$	47.64$_{\pm 4.97}$	33.87$_{\pm 7.90}$	52.63$_{\pm 2.53}$	38.46$_{\pm 6.93}$	41.95$_{\pm 3.49}$	28.28$_{\pm 7.97}$	45.63$_{\pm 5.22}$	30.81$_{\pm 8.85}$
A2GNN	10	61.23$_{\pm 0.93}$	54.72$_{\pm 2.23}$	59.00$_{\pm 0.64}$	46.61$_{\pm 1.16}$	51.29$_{\pm 0.51}$	41.70$_{\pm 0.77}$	61.07$_{\pm 0.32}$	50.09$_{\pm 0.68}$	59.54$_{\pm 0.83}$	55.41$_{\pm 1.74}$	60.32$_{\pm 1.82}$	60.11$_{\pm 2.30}$
	20	35.95$_{\pm 0.99}$	25.28$_{\pm 1.45}$	49.06$_{\pm 1.28}$	35.11$_{\pm 1.11}$	46.09$_{\pm 0.19}$	33.40$_{\pm 0.39}$	56.91$_{\pm 0.24}$	42.06$_{\pm 0.38}$	41.31$_{\pm 0.79}$	31.44$_{\pm 1.20}$	41.00$_{\pm 0.70}$	31.44$_{\pm 1.04}$
	50	32.48$_{\pm 1.62}$	17.61$_{\pm 1.75}$	35.22$_{\pm 0.12}$	15.10$_{\pm 0.24}$	40.26$_{\pm 0.08}$	24.09$_{\pm 0.07}$	50.33$_{\pm 0.19}$	31.03$_{\pm 0.17}$	31.27$_{\pm 0.22}$	14.29$_{\pm 0.48}$	27.79$_{\pm 0.35}$	13.19$_{\pm 0.72}$
GraphENS	10	70.49$_{\pm 0.63}$	64.53$_{\pm 1.33}$	66.43$_{\pm 1.40}$	58.40$_{\pm 3.69}$	63.25$_{\pm 1.11}$	62.41$_{\pm 2.29}$	67.47$_{\pm 0.37}$	63.87$_{\pm 1.23}$	58.95$_{\pm 0.64}$	58.34$_{\pm 1.07}$	65.50$_{\pm 0.53}$	63.52$_{\pm 0.65}$
	20	68.12$_{\pm 1.26}$	57.59$_{\pm 1.13}$	64.41$_{\pm 0.70}$	53.09$_{\pm 1.11}$	59.61$_{\pm 1.30}$	51.96$_{\pm 1.31}$	65.58$_{\pm 1.02}$	57.31$_{\pm 3.39}$	54.72$_{\pm 0.46}$	49.54$_{\pm 1.89}$	61.11$_{\pm 0.49}$	52.47$_{\pm 1.99}$
	50	62.74$_{\pm 0.81}$	51.76$_{\pm 0.79}$	61.01$_{\pm 1.66}$	47.74$_{\pm 4.14}$	54.09$_{\pm 1.51}$	45.51$_{\pm 1.74}$	60.40$_{\pm 0.90}$	47.81$_{\pm 2.29}$	50.67$_{\pm 0.43}$	42.18$_{\pm 0.44}$	56.26$_{\pm 0.85}$	47.81$_{\pm 2.43}$
TAM	10	72.36$_{\pm 0.50}$	67.27$_{\pm 0.56}$	67.12$_{\pm 0.69}$	61.52$_{\pm 1.05}$	64.49$_{\pm 0.41}$	63.40$_{\pm 1.34}$	70.10$_{\pm 0.85}$	66.20$_{\pm 1.02}$	59.17$_{\pm 0.68}$	55.36$_{\pm 1.29}$	66.65$_{\pm 0.39}$	63.16$_{\pm 0.77}$
	20	68.90$_{\pm 1.49}$	60.01$_{\pm 1.23}$	66.11$_{\pm 0.99}$	56.88$_{\pm 1.83}$	61.36$_{\pm 0.70}$	57.64$_{\pm 2.19}$	68.11$_{\pm 0.69}$	62.57$_{\pm 1.98}$	53.81$_{\pm 1.05}$	46.92$_{\pm 1.65}$	61.23$_{\pm 1.22}$	53.28$_{\pm 1.42}$
	50	64.93$_{\pm 3.74}$	55.77$_{\pm 4.74}$	61.63$_{\pm 0.99}$	50.23$_{\pm 1.48}$	58.95$_{\pm 1.64}$	50.99$_{\pm 2.06}$	64.66$_{\pm 1.19}$	54.71$_{\pm 2.03}$	53.39$_{\pm 1.61}$	44.93$_{\pm 2.14}$	57.88$_{\pm 1.67}$	47.52$_{\pm 2.02}$
GraphSHA	10	73.20$_{\pm 0.73}$	70.43$_{\pm 1.14}$	61.79$_{\pm 0.09}$	56.09$_{\pm 1.47}$	66.07$_{\pm 1.26}$	66.24$_{\pm 1.34}$	63.05$_{\pm 0.93}$	56.69$_{\pm 4.64}$	63.28$_{\pm 0.55}$	64.06$_{\pm 0.54}$	71.11$_{\pm 0.64}$	69.03$_{\pm 0.64}$
	20	71.00$_{\pm 1.58}$	64.33$_{\pm 2.56}$	57.95$_{\pm 2.55}$	47.37$_{\pm 3.16}$	62.87$_{\pm 1.74}$	60.41$_{\pm 2.91}$	62.72$_{\pm 3.73}$	53.91$_{\pm 4.00}$	57.56$_{\pm 1.71}$	53.22$_{\pm 4.22}$	66.93$_{\pm 1.40}$	62.03$_{\pm 1.82}$
	50	64.97$_{\pm 2.30}$	53.99$_{\pm 2.49}$	65.96$_{\pm 1.44}$	54.07$_{\pm 1.91}$	56.55$_{\pm 2.73}$	49.19$_{\pm 1.58}$	56.10$_{\pm 5.25}$	44.72$_{\pm 6.29}$	50.08$_{\pm 0.95}$	41.72$_{\pm 3.42}$	57.89$_{\pm 1.59}$	50.52$_{\pm 2.26}$
ImGDA (Ours)	10	**77.93**$_{\pm 1.03}$	**73.98**$_{\pm 2.74}$	**71.34**$_{\pm 2.61}$	**66.82**$_{\pm 1.82}$	**67.79**$_{\pm 0.62}$	**68.22**$_{\pm 1.41}$	**73.06**$_{\pm 2.16}$	**69.62**$_{\pm 2.98}$	**66.64**$_{\pm 1.33}$	**66.86**$_{\pm 1.78}$	**74.72**$_{\pm 0.55}$	**72.86**$_{\pm 1.14}$
	20	**77.30**$_{\pm 1.05}$	**73.92**$_{\pm 0.83}$	**71.23**$_{\pm 1.59}$	**65.66**$_{\pm 1.25}$	**67.28**$_{\pm 0.84}$	**66.39**$_{\pm 1.33}$	**68.48**$_{\pm 1.54}$	**63.42**$_{\pm 2.03}$	**63.92**$_{\pm 0.46}$	**63.45**$_{\pm 0.74}$	**71.99**$_{\pm 1.15}$	**68.38**$_{\pm 2.50}$
	50	**73.53**$_{\pm 1.40}$	**67.42**$_{\pm 1.23}$	**69.14**$_{\pm 1.63}$	**61.27**$_{\pm 1.50}$	**66.87**$_{\pm 1.71}$	**65.45**$_{\pm 2.64}$	**70.57**$_{\pm 1.19}$	**66.32**$_{\pm 1.86}$	**60.67**$_{\pm 0.98}$	**58.96**$_{\pm 1.75}$	**66.07**$_{\pm 0.96}$	**62.77**$_{\pm 1.11}$

Compared Baselines. For our study on graph domain adaptation under class-imbalanced scenario, we employ three GNNs as baselines: GCN [18], GAT [39], and GIN [48]. Meanwhile, we incorporate two baselines that are specifically developed for domain adaptation: GDA-SpecReg [53] and A2GNN [23]. In addition, we include three methods recognized for their effectiveness in handling imbalanced graph learning: GraphENS [31], TAM [35], and GraphSHA [22].

Implementation Details. We conduct extensive experiments by alternately setting one domain as the source and the other two as targets. The source imbalance factor $\rho \in \{10, 20, 50\}$ covers mild to severe imbalance levels. For our ImGDA, we adopt GCN [18] as the backbone with a 512-dimensional feature space. The hyperparameters are set as: $\gamma = 0.5$, $K = 200$, and $\alpha = \beta = 1$. For a fair comparison, all baselines employ the same GNN encoder and are fine-tuned for optimal performance. Each method is evaluated on the target domain graph w.r.t. *micro-F1* and *macro-F1* scores, and the final results is averaged over five runs.

5.2 Performance Comparison

Table 1 presents our results for $\rho = \{10, 20, 50\}$, while additional results for $\rho = \{5, 100\}$ are provided in Appendix D.1. From the table, our ImGDA consistently achieves the best performance. And the comparison shows that our approach significantly outperforms all other methods in class-imbalanced domain adaptation tasks. Additionally, when using DBLP dataset as the source domain with fewer samples, baseline methods generally perform worse than with other datasets. However, our ImGDA mitigates this issue by effectively capturing intrinsic semantics via a prototype-enhanced contrastive framework. Furthermore, Figure 3a shows the performance trend across different ρ values in the A⇒C experiment, where ρ ranges from 5 to 100. When ρ is small (e.g., 5), performance gaps between methods are minor. As ρ increases, all methods degrade due to rising imbalance. However, our ImGDA shows the smallest decline, maintaining stable results even under severe imbalance. Notably, as ρ increases, domain adaptation-specific methods (GDA-SpecReg, A2GNN) drop sharply, while imbalance-handling methods (GraphENS, TAM, GraphSHA) outperform others. This indicates that addressing source-domain imbalance has a greater impact than mitigating domain shifts. Figure 3b also shows the rapid convergence of our method within a few epochs.



(a) Different ρ on A2C (b) Training loss ($\rho = 50$)

Figure 3: Impact of ρ and convergence analysis.

5.3 Ablation study

To evaluate the contribution of each component in our method, we conduct an ablation study. We remove three loss terms (w/o $\mathcal{L}_{lc}$, w/o $\mathcal{L}_{cb}$, w/o $\mathcal{L}_{cd}$), the global branch (w/o global, i.e., two branches are local branches), the local branch (w/o local, i.e., two branches are global branches), and the dynamic temperature (w/o γ), and assess the performance under $\rho = 50$. The results for the Micro score are shown in Figure 4, while the results for the Macro score can be found in Appendix D.2. We can clearly see that excluding any component causes the performance degradation, especially the removal of $\mathcal{L}_{cb}$ (cross-branch prototype contrastive loss). This emphasizes the importance of addressing data imbalance in the source domain and enforcing consistency. Also, eliminating $\mathcal{L}_{cd}$ (cross-domain prototype contrastive loss) leads to some decline, but the impact is less severe, supporting the argument that mitigating data imbalance is more crucial than domain adaptation. Additionally, removing the $\mathcal{L}_{lc}$, global and local branches or switching to a fixed temperature reduces performance, particularly when DBLP is the source domain. The macro score drops after fixing the temperature emphasizes the importance of dynamic temperature in contrastive learning, where adjusting it based on class distribution alleviates data imbalance. These results validate the necessity and effectiveness of each component within our framework.

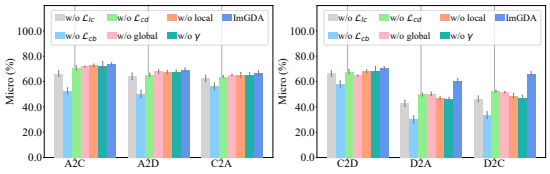
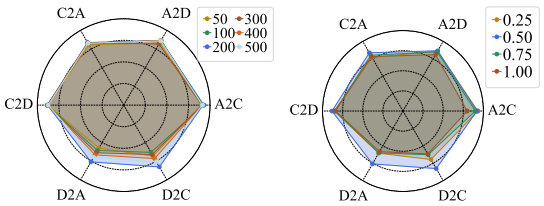


Figure 4: Results of ablation study on all data pairs ($\rho = 50$, Micro score (%) with standard deviation).

5.4 Sensitivity Study

Effect of k. We here study the impact of the number of anchor nodes k, where k takes values from {50, 100, 200, 300, 400, 500}, as shown in Figure 5a. Both small and large values of k degrade performance, with the impact more pronounced for small k. A small k samples too few nodes, ignoring useful information and hindering learning. On the other hand, for a large k, while the number of head class nodes sampled remains unaffected, the number of tail class nodes sampled decreases due to the limited number of such nodes, leading to a data imbalance phenomenon similar to that in the source domain, which also drops performance. Hence, selecting an appropriate number of anchor nodes k is critical for effective prototype learning.



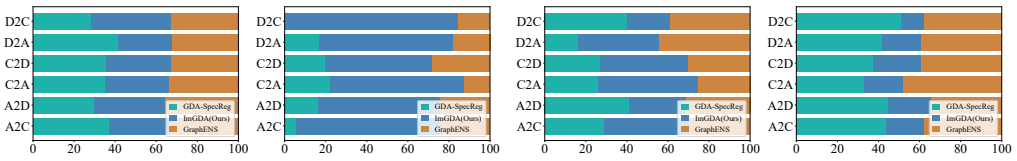
(a) k (b) γ

Figure 5: Impact of k and γ ($\rho = 50$). The distance between the points on each line and the center point represents the magnitude of the Micro score (%).

Effect of γ. We explore the impact of dynamic temperature γ in Eq. (12), where γ represents the minimum temperature associated with the tail class. It can be observed that as γ decreases, the tail temperature lowers and the gradients increase, making γ a weight parameter related to the gradients. We experiment with γ in {0.25, 0.50, 0.75, 1.00}, as shown in Figure 5b. Compared to the default $\gamma = 0.5$, both high and low temperatures degrade performance, with lower temperatures slightly better. This highlights that smaller temperatures help alleviate data imbalance, but overly small values disproportionately increase the gradient of tail class, harming the training of other classes.

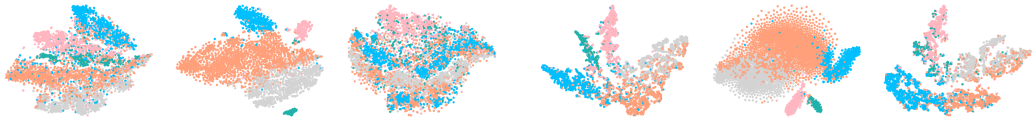
5.5 Capability to Mitigate the Imbalance Issue

We evaluate our ImGDA's capability to handle data imbalance by examining cross-entropy loss and accuracy across tail and head classes under ρ= 20. To enhance comparability, we normalize these metrics for each method and visualize the results in Figure 6. The results show that for head classes, all three methods achieve similar accuracy and cross-entropy loss ratios, indicating comparable effectiveness when abundant training data is available. However, for tail classes, the other two methods degrade significantly, revealing difficulty in handling imbalance. It can be observed that GraphENS performs better than GDA-SpecReg, likely due to its design advantages in dealing with imbalanced data. In comparison, our ImGDA consistently performs well in both the head and tail classes, underscoring its ability to mitigate class imbalance while ensuring robust generalization.



(a) Accuracy (Head) (b) Accuracy (Tail) (c) CE Loss (Head) (d) CE Loss (Tail)

Figure 6: Comparison of head and tail class performance in terms of accuracy and cross-entropy (CE) loss. The horizontal axis here indicates the relative contribution of each compared method, normalized across all the three methods.



(a) GDA-SpecReg (b) ImGDA (Ours)

Figure 8: *t*-SNE visualization of node embeddings from the baseline GDA-SpecReg and our proposed ImGDA. In each panel we sequentially present the balanced source domain graph, the imbalanced source domain graph, and the target domain graph, respectively.

5.6 Visualization

To further evaluate our ImGDA from a qualitative perspective, we compare its t-SNE visualizations with GDA-SpecReg. With ρ=20, node features learned by the encoder are projected into a 2D space to visualize three cases: a balanced source domain graph, an imbalanced source domain graph, and a target domain graph (Figure 8). As for the baseline, we can find that class boundaries appear blurred in both source domains, and the target domain shows a clear shift, indicating challenges in transferring domain-invariant features. In contrast, our ImGDA maintains well-separated class boundaries even though the source domain suffers from severe imbalance, and its target domain closely resembles that of the balanced source domain. This indicates that ImGDA effectively learns domain-invariant class structures and remains robust in the presence of imbalanced training data.

5.7 Different Sampling Strategies

To examine the impact of different anchor node sampling strategies on prototype computation, we compare three strategies: sampling based on node degree (Deg.), sampling based on inverse degree probability (Inv.), and random sampling (Rand.). We visualize the results for $\rho = 50$ in Figure 7, with additional results available in the Appendix D.4. From the figure, it can be observed that random sampling performs worse compared to the graph-structure-based sampling strategies, while the degree-based strategy employed by our method achieves the best performance. Additionally, our experiments reveal that inverse degree sampling significantly reduces training speed, further demonstrating the effectiveness of our selected sampling strategy.

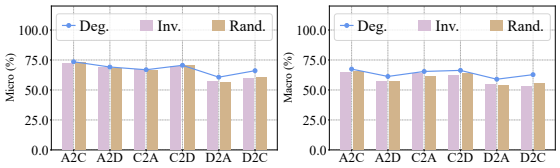


Figure 7: Results of sampling strategies (ρ = 50).

6 Conclusion

This paper tackles the challenge of class-imbalanced graphs by introducing a dual prototype-enhanced contrastive framework for graph domain adaptation. We use a dual graph encoder to capture local and global information and generate class-specific prototypes from a distilled anchor set. A prototype-aware contrastive learning module is introduced, combining cross-branch and cross-domain contrast. It mitigates source-domain imbalance via class alignment and reduces domain discrepancy by generating pseudo-labels to align semantically similar cross-domain pairs. Extensive experiments demonstrate the effectiveness of ImGDA compared to state-of-the-art methods.

Acknowledgements

Jiancheng Lv and Xin Ma are supported in part by the National Major Scientific Instruments and Equipments Development Project of National Natural Science Foundation of China under Grant 62427820, the Fundamental Research Funds for the Central Universities under Grant 1082204112364. Wei Ju is supported by the National Natural Science Foundation of China under Grant 62306014, the Postdoctoral Fellowship Program (Grade A) of CPSF under Grant BX20250376, the Sichuan Science and Technology Program under Grant 2025ZNSFSC1506, the Fundamental Research Funds for the Central Universities under Grant 1082204112K97, and the Sichuan University Interdisciplinary Innovation Fund 1082204112J74. Siyu Yi is supported by the National Natural Science Foundation of China under Grant 12501344, the Postdoctoral Fellowship Program (Grade A) of CPSF under Grant BX20240239, the China Postdoctoral Science Foundation under Grant No. 2024M762201, and the Sichuan Science and Technology Program under Grant 2025ZNSFSC0808.

References

[1] Thong Bach, Anh Tong, Truong Son Hy, Vu Nguyen, and Thanh Nguyen-Tang. Global contrastive learning for long-tailed classification. *Transactions on Machine Learning Research*, 2023.

[2] Lars Backstrom and Jure Leskovec. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the International ACM Conference on Web Search & Data Mining*, pages 635–644, 2011.

[3] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9268–9277, 2019.

[4] Quanyu Dai, Xiao-Ming Wu, Jiaren Xiao, Xiao Shen, and Dan Wang. Graph transfer learning via adversarial domain adaptation with graph convolution. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4908–4922, 2022.

[5] Corrado Gini. *Memorie di metodologia statistica*, volume 1. EV Veschi, 1955.

[6] Florian Graf, Christoph Hofer, Marc Niethammer, and Roland Kwitt. Dissecting supervised contrastive learning. In *Proceedings of the International Conference on Machine Learning*, pages 3821–3830, 2021.

[7] Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012.

[8] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 2017.

[9] Zhongkai Hao, Chengqiang Lu, Zhenya Huang, Hao Wang, Zheyuan Hu, Qi Liu, Enhong Chen, and Cheekong Lee. Asgn: An active semi-supervised graph neural network for molecular property prediction. In *Proceedings of the International ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 731–752, 2020.

[10] Youngkyu Hong, Seungju Han, Kwanghee Choi, Seokjun Seo, Beomsu Kim, and Buru Chang. Disentangling label distribution for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6626–6636, 2021.

[11] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5375–5384, 2016.

[12] Wei Ju, Zheng Fang, Yiyang Gu, Zequn Liu, Qingqing Long, Ziyue Qiao, Yifang Qin, Jianhao Shen, Fang Sun, Zhiping Xiao, et al. A comprehensive survey on deep graph representation learning. *Neural Networks*, 173:106207, 2024.

[13] Wei Ju, Zequn Liu, Yifang Qin, Bin Feng, Chen Wang, Zhihui Guo, Xiao Luo, and Ming Zhang. Few-shot molecular property prediction via hierarchically structured learning on relation graphs. *Neural Networks*, 163:122–131, 2023.

[14] Wei Ju, Zhengyang Mao, Siyu Yi, Yifang Qin, Yiyang Gu, Zhiping Xiao, Jianhao Shen, Ziyue Qiao, and Ming Zhang. Cluster-guided contrastive class-imbalanced graph classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2025.

[15] Wei Ju, Yifan Wang, Yifang Qin, Zhengyang Mao, Zhiping Xiao, Junyu Luo, Junwei Yang, Yiyang Gu, Dongjie Wang, Qingqing Long, et al. Towards graph contrastive learning: A survey and beyond. *arXiv preprint arXiv:2405.11868*, 2024.

[16] Bingyi Kang, Saining Xie, Marcus Rohrbach, Zhicheng Yan, Albert Gordo, Jiashi Feng, and Yannis Kalantidis. Decoupling representation and classifier for long-tailed recognition. In *International Conference on Learning Representations*, 2020.

[17] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

[18] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. 2016.

[19] Jaekoo Lee, Hyunjae Kim, Jongsun Lee, and Sungroh Yoon. Transfer learning for deep learning on graph-structured data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

[20] Hourun Li, Yifan Wang, Zhiping Xiao, Jia Yang, Changling Zhou, Ming Zhang, and Wei Ju. Disco: graph-based disentangled contrastive learning for cold-start cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 12049–12057, 2025.

[21] Junnan Li, Pan Zhou, Caiming Xiong, and Steven CH Hoi. Prototypical contrastive learning of unsupervised representations. *arXiv preprint arXiv:2005.04966*, 2020.

[22] Wen-Zhi Li, Chang-Dong Wang, Hui Xiong, and Jian-Huang Lai. Graphsha: Synthesizing harder samples for class-imbalanced node classification. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1328–1340, 2023.

[23] Meihan Liu, Zeyu Fang, Zhen Zhang, Ming Gu, Sheng Zhou, Xin Wang, and Jiajun Bu. Rethinking propagation for unsupervised graph domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 13963–13971, 2024.

[24] Junyu Luo, Yiyang Gu, Xiao Luo, Wei Ju, Zhiping Xiao, Yusheng Zhao, Jingyang Yuan, and Ming Zhang. Gala: Graph diffusion-based alignment with jigsaw for source-free domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9038–9051, 2024.

[25] Junyu Luo, Zhiping Xiao, Yifan Wang, Xiao Luo, Jingyang Yuan, Wei Ju, Langechuan Liu, and Ming Zhang. Rank and align: Towards effective source-free graph domain adaptation. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 4706–4714, 2024.

[26] Xin Ma, Yifan Wang, Siyu Yi, Wei Ju, Bei Wu, Ziyue Qiao, Chenwei Tang, and Jiancheng Lv. Pala: Class-imbalanced graph domain adaptation via prototype-anchored learning and alignment. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 3198–3207, 2025.

[27] Zhengyang Mao, Wei Ju, Yifang Qin, Xiao Luo, and Ming Zhang. Rahnet: Retrieval augmented hybrid network for long-tailed graph classification. In *Proceedings of the 31st ACM international conference on multimedia*, pages 3817–3826, 2023.

[28] Zhengyang Mao, Wei Ju, Siyu Yi, Yifan Wang, Zhiping Xiao, Qingqing Long, Nan Yin, Xinwang Liu, and Ming Zhang. Learning knowledge-diverse experts for long-tailed graph classification. *ACM Transactions on Knowledge Discovery from Data*, 19(2):1–24, 2025.

[29] Aditya Krishna Menon, Sadeep Jayasumana, Ankit Singh Rawat, Himanshu Jain, Andreas Veit, and Sanjiv Kumar. Long-tail learning via logit adjustment. In *Proceedings of the International Conference on Learning Representations*, 2021.

[30] Jinhui Pang, Zixuan Wang, Jiliang Tang, Mingyan Xiao, and Nan Yin. Sa-gda: Spectral augmentation for graph domain adaptation. In *Proceedings of the 31st ACM international conference on multimedia*, pages 309–318, 2023.

[31] Joonhyung Park, Jaeyun Song, and Eunho Yang. Graphens: Neighbor-aware ego network synthesis for class-imbalanced node classification. In *International Conference on Learning Representations*, 2021.

[32] Ziyue Qiao, Xiao Luo, Meng Xiao, Hao Dong, Yuanchun Zhou, and Hui Xiong. Semi-supervised domain adaptation in graph transfer learning. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pages 2279–2287, 2023.

[33] Jiawen Qin, Haonan Yuan, Qingyun Sun, Lyujin Xu, Jiaqi Yuan, Pengfeng Huang, Zhaonan Wang, Xingcheng Fu, Hao Peng, Jianxin Li, et al. Igl-bench: Establishing the comprehensive benchmark for imbalanced graph learning. *arXiv preprint arXiv:2406.09870*, 2024.

[34] Xiao Shen, Quanyu Dai, Sitong Mao, Fu-lai Chung, and Kup-Sze Choi. Network together: Node classification via cross-network deep network embedding. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5):1935–1948, 2020.

[35] Jaeyun Song, Joonhyung Park, and Eunho Yang. Tam: topology-aware margin loss for class-imbalanced node classification. In *International Conference on Machine Learning*, pages 20369–20383. PMLR, 2022.

[36] Jingru Tan, Changbao Wang, Buyu Li, Quanquan Li, Wanli Ouyang, Changqing Yin, and Junjie Yan. Equalization loss for long-tailed object recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11662–11671, 2020.

[37] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 990–998, 2008.

[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Proceedings of the Conference on Neural Information Processing Systems*, pages 5998–6008, 2017.

[39] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, and Yoshua Bengio. Graph attention networks. In *International Conference on Learning Representations*, 2018.

[40] Nikil Wale, Ian A Watson, and George Karypis. Comparison of descriptor spaces for chemical compound retrieval and classification. *Knowledge and Information Systems*, 14:347–375, 2008.

[41] Feng Wang and Huaping Liu. Understanding the behaviour of contrastive loss. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2495–2504, 2021.

[42] Jianfeng Wang, Thomas Lukasiewicz, Xiaolin Hu, Jianfei Cai, and Zhenghua Xu. Rsg: A simple but effective module for learning imbalanced datasets. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3784–3793, 2021.

[43] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.

[44] Yifan Wang, Suyao Tang, Yuntong Lei, Weiping Song, Sheng Wang, and Ming Zhang. Disenhan: Disentangled heterogeneous graph attention network for recommendation. In *Proceedings of the International Conference on Information and Knowledge Management*, pages 1605–1614, 2020.

[45] Jun Wu, Jingrui He, and Elizabeth Ainsworth. Non-iid transfer learning on graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10342–10350, 2023.

[46] Man Wu, Shirui Pan, Chuan Zhou, Xiaojun Chang, and Xingquan Zhu. Unsupervised domain adaptive graph convolutional networks. In *Proceedings of the web conference 2020*, pages 1457–1467, 2020.

[47] Zonghan Wu, Shirui Pan, Fengwen Chen, Guodong Long, Chengqi Zhang, and Philip S Yu. A comprehensive survey on graph neural networks. *IEEE transactions on neural networks and learning systems*, 32(1):4–24, 2020.

[48] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2018.

[49] Junwei Yang, Hanwen Xu, Srbuhi Mirzoyan, Tong Chen, Zixuan Liu, Zequn Liu, Wei Ju, Luchen Liu, Zhiping Xiao, Ming Zhang, et al. Poisoning medical knowledge using large language models. *Nature Machine Intelligence*, 6(10):1156–1168, 2024.

[50] Si-Yu Yi, Zhengyang Mao, Wei Ju, Yong-Dao Zhou, Luchen Liu, Xiao Luo, and Ming Zhang. Towards long-tailed recognition for graph classification via collaborative experts. *IEEE Transactions on Big Data*, 9(6):1683–1696, 2023.

[51] Siyu Yi, Wei Ju, Yifang Qin, Xiao Luo, Luchen Liu, Yongdao Zhou, and Ming Zhang. Redundancy-free self-supervised relational learning for graph clustering. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[52] Nan Yin, Li Shen, Mengzhu Wang, Long Lan, Zeyu Ma, Chong Chen, Xian-Sheng Hua, and Xiao Luo. Coco: A coupled contrastive framework for unsupervised domain adaptive graph classification. In *International Conference on Machine Learning*, pages 40040–40053. PMLR, 2023.

[53] Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. Graph domain adaptation via theory-grounded spectral regularization. In *The eleventh International Conference on Learning Representations*, 2023.

[54] Liang Zeng, Lanqing Li, Ziqi Gao, Peilin Zhao, and Jian Li. Imgcl: Revisiting graph contrastive learning on imbalanced node classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 11138–11146, 2023.

[55] Haodong Zhang, Tao Ren, Yifan Wang, Fanchun Meng, Wei Ju, and Ying Tian. Cluster-aware few-shot molecular property prediction with factor disentanglement. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

[56] Xiaowen Zhang, Yuntao Du, Rongbiao Xie, and Chongjun Wang. Adversarial separation network for cross-network node classification. In *Proceedings of the International Conference on Information and Knowledge Management*, pages 2618–2626, 2021.

[57] Tianxiang Zhao, Xiang Zhang, and Suhang Wang. Graphsmote: Imbalanced node classification on graphs with graph neural networks. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 833–841, 2021.

[58] Yusheng Zhao, Xiao Luo, Wei Ju, Chong Chen, Xian-Sheng Hua, and Ming Zhang. Dynamic hypergraph structure learning for traffic flow forecasting. In *2023 IEEE 39th International Conference on Data Engineering (ICDE)*, pages 2303–2316. IEEE, 2023.

[59] Boyan Zhou, Quan Cui, Xiu-Shen Wei, and Zhao-Min Chen. Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9719–9728, 2020.

[60] Jianggang Zhu, Zheng Wang, Jingjing Chen, Yi-Ping Phoebe Chen, and Yu-Gang Jiang. Balanced contrastive learning for long-tailed visual recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6908–6917, 2022.

[61] Yun Zhu, Yaoke Wang, Haizhou Shi, Zhenshuo Zhang, Dian Jiao, and Siliang Tang. Graphcontrol: Adding conditional control to universal graph pre-trained models for graph domain transfer learning. In *Proceedings of the ACM on Web Conference 2024*, pages 539–550, 2024.

[62] Chenyi Zhuang and Qiang Ma. Dual graph convolutional networks for graph-based semi-supervised classification. In *Proceedings of the Web Conference*, pages 499–508, 2018.

A Proof of Theorem 4.2

The supervised contrastive loss for a class c in a batch $\mathcal{B}$ can be re-written as the following form [60]:

$$\begin{aligned}\mathcal{L}_{SCL}(\boldsymbol{Z};\boldsymbol{Y},\mathcal{B},c) &= \sum_{i\in\mathcal{B}_c} -\frac{1}{|\mathcal{B}_c|-1}\sum_{p\in\mathcal{B}_c\setminus\{i\}}\log\frac{\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_p)}{\sum_{j\in\mathcal{Y}_B}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k)} \\ &= \sum_{i\in\mathcal{B}_c}\log\left(\frac{\sum_{j\in\mathcal{C}_\mathcal{B}}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k)}{\prod_{p\in\mathcal{B}_c\setminus\{i\}}\exp(\boldsymbol{z}_i,\boldsymbol{z}_p)^{1/|\mathcal{B}_c|-1}}\right) \\ &= \sum_{i\in\mathcal{B}_c}\log\left(\frac{\sum_{j\in\mathcal{C}_\mathcal{B}}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k)}{\exp(\frac{1}{|\mathcal{B}_c|-1}\sum_{p\in\mathcal{B}_c\setminus\{i\}}\boldsymbol{z}_i\cdot\boldsymbol{z}_p)}\right).\end{aligned} \tag{15}$$

We divide the sum in the numerator into the positive and negative terms and since the exponential function is convex, we apply Jensen's inequality to get the lower bound of two terms as follows:

$$\begin{aligned}&\sum_{j\in\mathcal{C}_\mathcal{B}}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k) = \underbrace{\frac{1}{|\mathcal{B}_c|-1}\sum_{k\in\mathcal{B}_c\setminus\{i\}}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k)}_{positive\ term} + \underbrace{\sum_{\substack{j\in\mathcal{C}_\mathcal{B}\\ j\neq c}}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k)}_{negative\ terms}, \\ &\frac{1}{|\mathcal{B}_c|-1}\sum_{k\in\mathcal{B}_c\setminus\{i\}}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k)\geq\exp\Big(\frac{1}{|\mathcal{B}_c|-1}\sum_{k\in\mathcal{B}_c\setminus\{i\}}\boldsymbol{z}_i\cdot\boldsymbol{z}_k\Big), \\ &\sum_{\substack{j\in\mathcal{C}_\mathcal{B}\\ j\neq c}}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k)\geq\sum_{\substack{j\in\mathcal{C}_\mathcal{B}\\ j\neq c}}\exp\Big(\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\boldsymbol{z}_i\cdot\boldsymbol{z}_k\Big)\geq(|\mathcal{C}_\mathcal{B}|-1)\exp\Big(\frac{1}{|\mathcal{C}_\mathcal{B}|-1}\sum_{\substack{j\in\mathcal{C}_\mathcal{B}\\ j\neq c}}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\boldsymbol{z}_i\cdot\boldsymbol{z}_k\Big).\end{aligned} \tag{16}$$

And the lower bound of the numerator can be written as:

$$\sum_{j\in\mathcal{C}_\mathcal{B}}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\exp(\boldsymbol{z}_i\cdot\boldsymbol{z}_k)\geq\exp\Big(\frac{1}{|\mathcal{B}_c|-1}\sum_{k\in\mathcal{B}_c\setminus\{i\}}\boldsymbol{z}_i\cdot\boldsymbol{z}_k\Big)+(|\mathcal{C}_\mathcal{B}|-1)\exp\Big(\frac{1}{|\mathcal{C}_\mathcal{B}|-1}\sum_{\substack{j\in\mathcal{C}_\mathcal{B}\\ j\neq c}}\frac{1}{|\mathcal{B}_j|}\sum_{k\in\mathcal{B}_j}\boldsymbol{z}_i\cdot\boldsymbol{z}_k\Big). \tag{17}$$

Thus, the lower bound of supervised contrastive loss can be written as:

$$\mathcal{L}_{SCL}(\boldsymbol{Z};\boldsymbol{Y},\mathcal{B},c)\geq\sum_{i\in\mathcal{B}_c}\log\Big(1+(|\mathcal{C}_\mathcal{B}|-1)\exp\Big(\underbrace{\frac{1}{|\mathcal{C}_\mathcal{B}|-1}\sum_{q\in\mathcal{C}_\mathcal{B}\setminus\{c\}}\frac{1}{|\mathcal{B}_q|}\sum_{k\in\mathcal{B}_q}\boldsymbol{z}_i\cdot\boldsymbol{z}_k}_{repulsion\ term}-\underbrace{\frac{1}{|\mathcal{B}_c|-1}\sum_{j\in\mathcal{B}_c\setminus\{i\}}\boldsymbol{z}_i\cdot\boldsymbol{z}_j}_{attraction\ term}\Big)\Big). \tag{18}$$

Thus, the proof of Theorem 4.2 is complete.

B Proof of Expectation-Maximization Perspective

In unsupervised graph domain adaptation, we aim to learn the graph encoder with parameter θ to maximize the log-likelihood of nodes in target graph $\mathcal{G}^t$ with source graphs $\mathcal{G}^s$, written as:

$$\theta^* = \arg\max_\theta\sum_{v_j\in\mathcal{V}^t}\log\sum_{v_i=1}^{\mathcal{V}^s}p(v_j^t,v_i^s;\theta). \tag{19}$$

We introduce a surrogate function $Q(v_i^s)$ ($\sum_{i=1}^{\mathcal{V}^s}Q(v_i^s)=1$) to estimate the lower-bound via Jensen's inequality [21, 52]:

$$\sum_{v_j^t\in\mathcal{V}^t}\log\sum_{i=1}^{\mathcal{V}^s}p(v_j^t,v_i^s;\theta)=\sum_{v_j^t\in\mathcal{V}^t}\log\sum_{i=1}^{\mathcal{V}^s}Q(v_i^s)\frac{p(v_j^t,v_i^s;\theta)}{Q(v_i^s)}\geq\sum_{v_j^t\in\mathcal{V}^t}\sum_{i=1}^{\mathcal{V}^s}Q(v_i^s)\log\frac{p(v_j^t,v_i^s;\theta)}{Q(v_i^s)}. \tag{20}$$

Note that the equality holds when $Q(v_i^s)/p(v_i^s, v_j^t; \theta)$ is constant. Thus, we have $Q(v_i^s) = p(v_i^s; v_j^t, \theta)$. Since $-\sum_{v_j^t \in \mathcal{G}^t} \sum_{i=1}^{\mathcal{G}^s} Q(v_i^s) \log Q(v_i^s)$ does not influence the optimization process, we objective can be re-written as:

$$\mathcal{L}_{cd} \geq \sum_{v_j^t \in \mathcal{V}^t} \sum_{i=1}^{\mathcal{V}^s} p(v_i^s; v_j^t, \theta) \log p(v_j^t, v_i^s; \theta). \tag{21}$$

Here we optimize the objective via an EM algorithm. In the E-step, we infer the posterior probability $p(v_i^s; v_j^t, \theta) = \frac{1}{|\Pi(j)|}\mathbb{I}(v_j^t, v_i^s)$ where indicator $\mathbb{I}(v_j^t, v_i^s) = 1$ if they has the same label and $|\Pi(j)| = \sum_{i=1}^{\mathcal{V}^s} \mathbb{I}(v_j^t, v_i^s)$. In the M-step, we aim to optimize the lower-bound, which can be defined as:

$$\theta = \arg\max_{\theta} \sum_{v_j^t \in \mathcal{V}^t} \frac{1}{|\Pi(j)|} \sum_{i=1}^{\mathcal{V}^s} \mathbb{I}(v_j^t, v_i^s) \log\Big(\frac{\exp(\boldsymbol{z}_j^t \cdot \boldsymbol{\mu}_{\hat{y}_j^t})}{\exp(\boldsymbol{z}_j^t \cdot \boldsymbol{\mu}_{\hat{y}_j^t}) + \sum_{\boldsymbol{\mu} \in \mathcal{P}_j^-} \exp(\boldsymbol{z}_j^t \cdot \boldsymbol{\mu})}\Big). \tag{22}$$

which is equivalent to our cross-domain contrastive learning objective in Eq. 11.

C Complexity Analysis

For the time complexity, let N^s and $|\mathcal{E}^s|$ represent the number of nodes and edges in the source domain graph, and $|N^t|$ and $|\mathcal{E}^t|$ represent the number of nodes and edges in the target domain graph. Assuming an L-layer GCN encoder with a feature dimension of d, the computational complexity of feature encoding is $\mathcal{O}(L|\mathcal{E}|^s d + LN^s d^2)$ for the source domain, and $\mathcal{O}(L|\mathcal{E}|^t d + LN^t d^2)$ for the target domain. When $|\mathcal{E}|^s \gg n$, this simplifies to $\mathcal{O}(|\mathcal{E}|^s d)$ and $\mathcal{O}(|\mathcal{E}|^t d)$, respectively. For anchor node sampling, we sort nodes based on their degree and select m nodes per class. Since $m \ll N^s$, the complexity is $\mathcal{O}(N^s \log N^s)$. For prototype computation, the Mean prototype computation per class takes $\mathcal{O}(md)$, the Cross-branch prototype contrastive loss takes $\mathcal{O}(md)$, the Cross-domain prototype contrastive loss takes $\mathcal{O}(N^t d)$ and the Supervised loss computation in the source domain takes $\mathcal{O}(N^s d)$. Thus, the total time complexity is $\mathcal{O}(|\mathcal{E}^s|d + |\mathcal{E}^t|d + N^s \log N^s + md + N^t d + N^s d)$, which can be approximated as: $\mathcal{O}(\max(|\mathcal{E}^s| + |\mathcal{E}^t|, N^s \log N^s, N^t d))$.

As for the space complexity: storing node features incurs a complexity of $\mathcal{O}(Nd)$; storing the sparse adjacency matrix requires $\mathcal{O}(|\mathcal{E}|)$; the PPMI matrix has a space complexity of $\mathcal{O}(N^2)$; the model parameters require $\mathcal{O}(D^2)$; and storing the prototypes and pseudo-labels requires $\mathcal{O}(Cd)$ and $\mathcal{O}(NC)$ respectively, and C denotes the number of classes. In summary, the main space cost of our method arises from the PPMI matrix, which is $\mathcal{O}(N^2)$.

D Supplement of Experiments

D.1 Performance Experiment

To provide a more comprehensive evaluation of our ImGDA on class-imbalanced domain adaptation tasks, we consider additional scenarios beyond the main experiments presented in the body of the paper. Specifically, we explore more extreme settings with ρ values in $\{5, 100\}$. All results are presented in Table 1 in the Appendix. At $\rho = 5$, the source graph is relatively less imbalanced, and the performance of all methods is similar, with generally good results. As ρ increases, the advantages of our ImGDA become more pronounced. Even when $\rho = 100$, where the data is highly imbalanced, our ImGDA still maintains stable performance and significantly outperforms other competitive baselines. This highlights the robustness and effectiveness of our proposed ImGDA in addressing class-imbalanced domain adaptation tasks.

D.2 Ablation Study

To better validate which parts of our method ImGDA are effective and how much they contribute to the improvement in model performance, we conduct ablation experiments based on the full experiments. Specifically, we evaluate the effects of the three loss terms ($\mathcal{L}_{lc}, \mathcal{L}_{cb}, \mathcal{L}_{cd}$), the two branches (local and global), and the dynamic temperature (γ). Figure 1 and Table 2 in the Appendix show the macro score performance from the ablation study, which follows a trend similar to that of the micro scores.

Table 1: The complete experiment results of approaches on class-imbalanced domain adaptation with ρ range from 5 to 100. The best are highlighted in **bold** and the second-best are underlined.

Methods	IF(ρ)	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
		Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
GCN	5	69.51$_{\pm 0.81}$	66.61$_{\pm 0.64}$	64.53$_{\pm 0.92}$	60.20$_{\pm 0.93}$	64.07$_{\pm 0.36}$	64.65$_{\pm 0.50}$	68.52$_{\pm 0.31}$	65.56$_{\pm 0.52}$	60.04$_{\pm 0.17}$	60.10$_{\pm 0.33}$	66.88$_{\pm 0.20}$	65.18$_{\pm 0.33}$
	10	66.01$_{\pm 1.15}$	60.90$_{\pm 2.14}$	61.21$_{\pm 0.57}$	53.20$_{\pm 0.71}$	58.25$_{\pm 1.22}$	56.52$_{\pm 2.87}$	63.39$_{\pm 0.32}$	58.38$_{\pm 1.27}$	55.47$_{\pm 0.39}$	52.80$_{\pm 0.86}$	61.68$_{\pm 0.38}$	58.35$_{\pm 0.68}$
	20	62.28$_{\pm 1.17}$	55.48$_{\pm 1.79}$	60.75$_{\pm 0.69}$	51.60$_{\pm 1.48}$	51.17$_{\pm 0.58}$	43.78$_{\pm 1.54}$	59.00$_{\pm 0.88}$	48.22$_{\pm 1.57}$	50.93$_{\pm 2.70}$	45.12$_{\pm 4.44}$	54.02$_{\pm 0.27}$	46.18$_{\pm 0.55}$
	50	53.07$_{\pm 3.39}$	42.41$_{\pm 3.32}$	56.26$_{\pm 2.21}$	43.35$_{\pm 3.53}$	44.38$_{\pm 1.12}$	32.20$_{\pm 2.51}$	53.06$_{\pm 0.55}$	37.09$_{\pm 1.18}$	41.86$_{\pm 1.31}$	31.11$_{\pm 1.50}$	42.28$_{\pm 3.28}$	32.05$_{\pm 3.30}$
	100	45.54$_{\pm 2.24}$	34.20$_{\pm 1.17}$	51.34$_{\pm 2.44}$	36.95$_{\pm 2.66}$	41.54$_{\pm 0.72}$	26.75$_{\pm 1.41}$	49.76$_{\pm 0.93}$	31.25$_{\pm 1.61}$	36.37$_{\pm 0.17}$	22.72$_{\pm 0.76}$	38.40$_{\pm 3.73}$	25.99$_{\pm 4.22}$
GAT	5	69.96$_{\pm 0.95}$	66.17$_{\pm 1.03}$	65.53$_{\pm 0.30}$	61.43$_{\pm 0.46}$	63.34$_{\pm 0.92}$	63.74$_{\pm 1.02}$	67.20$_{\pm 0.91}$	64.02$_{\pm 1.53}$	60.67$_{\pm 0.28}$	60.74$_{\pm 0.46}$	68.13$_{\pm 0.85}$	65.58$_{\pm 1.17}$
	10	66.48$_{\pm 0.72}$	58.37$_{\pm 2.74}$	61.58$_{\pm 0.75}$	51.44$_{\pm 1.02}$	55.10$_{\pm 2.47}$	51.54$_{\pm 4.12}$	60.87$_{\pm 1.00}$	52.99$_{\pm 1.34}$	56.18$_{\pm 0.31}$	53.11$_{\pm 0.80}$	61.64$_{\pm 0.23}$	58.22$_{\pm 0.53}$
	20	61.63$_{\pm 3.34}$	50.54$_{\pm 3.72}$	61.04$_{\pm 1.87}$	50.10$_{\pm 1.46}$	47.18$_{\pm 1.04}$	37.30$_{\pm 1.61}$	55.95$_{\pm 1.26}$	43.32$_{\pm 2.17}$	49.18$_{\pm 1.59}$	42.20$_{\pm 2.78}$	53.05$_{\pm 3.25}$	45.29$_{\pm 2.63}$
	50	53.21$_{\pm 7.62}$	42.07$_{\pm 7.64}$	53.67$_{\pm 4.00}$	40.36$_{\pm 5.33}$	41.02$_{\pm 0.99}$	26.41$_{\pm 1.18}$	49.76$_{\pm 0.63}$	32.29$_{\pm 1.15}$	40.14$_{\pm 1.46}$	28.89$_{\pm 1.85}$	45.38$_{\pm 7.90}$	34.77$_{\pm 8.31}$
	100	38.52$_{\pm 4.23}$	25.13$_{\pm 4.65}$	45.19$_{\pm 5.76}$	29.49$_{\pm 7.86}$	38.35$_{\pm 1.13}$	22.45$_{\pm 1.76}$	46.25$_{\pm 0.47}$	26.89$_{\pm 0.49}$	35.31$_{\pm 0.56}$	20.88$_{\pm 0.82}$	31.91$_{\pm 1.78}$	18.58$_{\pm 1.90}$
GIN	5	63.36$_{\pm 0.37}$	58.32$_{\pm 0.89}$	61.11$_{\pm 0.35}$	54.84$_{\pm 0.67}$	58.76$_{\pm 0.47}$	57.47$_{\pm 0.76}$	64.52$_{\pm 0.39}$	60.99$_{\pm 0.45}$	56.69$_{\pm 0.36}$	55.89$_{\pm 0.53}$	64.51$_{\pm 0.42}$	62.02$_{\pm 0.52}$
	10	62.12$_{\pm 0.69}$	55.03$_{\pm 0.84}$	60.91$_{\pm 0.79}$	54.39$_{\pm 1.48}$	54.35$_{\pm 1.75}$	49.21$_{\pm 1.31}$	60.14$_{\pm 1.00}$	54.51$_{\pm 2.16}$	52.59$_{\pm 0.67}$	49.43$_{\pm 2.42}$	59.89$_{\pm 1.42}$	54.68$_{\pm 2.28}$
	20	60.03$_{\pm 1.03}$	50.98$_{\pm 0.94}$	59.12$_{\pm 1.05}$	48.27$_{\pm 0.92}$	50.38$_{\pm 2.75}$	42.80$_{\pm 3.88}$	55.87$_{\pm 2.47}$	46.42$_{\pm 4.71}$	47.79$_{\pm 1.49}$	41.02$_{\pm 2.77}$	53.47$_{\pm 2.49}$	44.83$_{\pm 2.61}$
	50	54.39$_{\pm 1.62}$	44.28$_{\pm 1.69}$	54.43$_{\pm 1.60}$	41.90$_{\pm 2.09}$	41.26$_{\pm 0.37}$	29.45$_{\pm 0.51}$	47.05$_{\pm 0.60}$	31.57$_{\pm 1.13}$	43.38$_{\pm 1.26}$	34.41$_{\pm 2.05}$	48.59$_{\pm 2.15}$	39.22$_{\pm 2.29}$
	100	48.82$_{\pm 2.59}$	37.42$_{\pm 2.28}$	50.15$_{\pm 1.22}$	35.91$_{\pm 2.19}$	36.97$_{\pm 0.24}$	22.75$_{\pm 0.45}$	43.65$_{\pm 0.43}$	25.90$_{\pm 1.58}$	37.83$_{\pm 2.14}$	24.27$_{\pm 4.59}$	40.01$_{\pm 4.39}$	26.99$_{\pm 5.83}$
A2GNN	5	74.43$_{\pm 0.46}$	70.85$_{\pm 0.77}$	65.08$_{\pm 0.59}$	59.76$_{\pm 1.52}$	64.54$_{\pm 1.12}$	64.36$_{\pm 1.42}$	69.25$_{\pm 0.71}$	65.81$_{\pm 1.07}$	**68.82**$_{\pm 0.35}$	**70.45**$_{\pm 0.44}$	**78.28**$_{\pm 0.21}$	**76.70**$_{\pm 0.36}$
	10	61.23$_{\pm 0.93}$	54.72$_{\pm 2.23}$	59.00$_{\pm 0.64}$	46.61$_{\pm 1.16}$	51.29$_{\pm 0.51}$	41.70$_{\pm 0.77}$	61.07$_{\pm 0.32}$	50.09$_{\pm 0.68}$	59.54$_{\pm 0.83}$	55.41$_{\pm 1.74}$	60.32$_{\pm 1.82}$	60.11$_{\pm 2.30}$
	20	35.95$_{\pm 0.99}$	25.28$_{\pm 1.45}$	49.06$_{\pm 1.28}$	35.11$_{\pm 1.11}$	46.09$_{\pm 0.19}$	33.40$_{\pm 0.39}$	56.91$_{\pm 0.24}$	42.06$_{\pm 0.38}$	41.31$_{\pm 0.79}$	31.44$_{\pm 1.20}$	41.00$_{\pm 0.70}$	31.44$_{\pm 1.04}$
	50	32.48$_{\pm 1.62}$	17.61$_{\pm 1.75}$	35.22$_{\pm 0.12}$	15.10$_{\pm 0.24}$	40.26$_{\pm 0.08}$	24.09$_{\pm 0.07}$	50.33$_{\pm 0.19}$	31.03$_{\pm 0.17}$	31.27$_{\pm 0.22}$	14.29$_{\pm 0.48}$	27.79$_{\pm 0.35}$	13.19$_{\pm 0.72}$
	100	28.37$_{\pm 0.61}$	12.57$_{\pm 0.84}$	34.13$_{\pm 0.08}$	12.35$_{\pm 0.17}$	38.58$_{\pm 0.18}$	22.08$_{\pm 0.16}$	46.16$_{\pm 0.62}$	25.98$_{\pm 0.49}$	29.79$_{\pm 0.08}$	10.45$_{\pm 0.34}$	28.31$_{\pm 0.19}$	13.76$_{\pm 0.34}$
GDA-SpecReg	5	64.51$_{\pm 4.22}$	53.23$_{\pm 5.35}$	59.49$_{\pm 5.16}$	48.61$_{\pm 6.38}$	57.66$_{\pm 5.69}$	55.46$_{\pm 7.39}$	64.04$_{\pm 3.87}$	55.30$_{\pm 4.54}$	53.42$_{\pm 4.48}$	45.43$_{\pm 6.85}$	61.75$_{\pm 2.95}$	52.49$_{\pm 4.53}$
	10	51.27$_{\pm 1.86}$	39.29$_{\pm 2.43}$	53.69$_{\pm 3.02}$	40.87$_{\pm 5.56}$	49.69$_{\pm 4.93}$	39.08$_{\pm 7.51}$	57.41$_{\pm 3.15}$	45.54$_{\pm 4.35}$	52.73$_{\pm 3.20}$	47.07$_{\pm 7.86}$	56.37$_{\pm 1.40}$	45.04$_{\pm 3.46}$
	20	49.68$_{\pm 1.42}$	36.44$_{\pm 1.70}$	51.34$_{\pm 3.70}$	35.82$_{\pm 3.00}$	48.11$_{\pm 4.07}$	38.20$_{\pm 6.83}$	51.64$_{\pm 1.96}$	35.15$_{\pm 3.49}$	43.84$_{\pm 3.79}$	31.89$_{\pm 3.72}$	44.53$_{\pm 6.48}$	31.21$_{\pm 6.80}$
	50	49.80$_{\pm 6.48}$	36.09$_{\pm 7.63}$	47.44$_{\pm 4.58}$	29.36$_{\pm 6.49}$	47.64$_{\pm 4.97}$	33.87$_{\pm 7.90}$	52.63$_{\pm 2.53}$	38.46$_{\pm 6.93}$	41.95$_{\pm 3.49}$	28.28$_{\pm 7.97}$	45.63$_{\pm 5.22}$	30.81$_{\pm 8.85}$
	100	45.35$_{\pm 5.42}$	30.65$_{\pm 5.54}$	46.44$_{\pm 3.33}$	27.08$_{\pm 5.21}$	44.66$_{\pm 2.65}$	31.91$_{\pm 6.15}$	50.88$_{\pm 2.84}$	34.10$_{\pm 7.70}$	38.93$_{\pm 2.76}$	23.07$_{\pm 4.25}$	42.88$_{\pm 3.62}$	27.97$_{\pm 3.76}$
GraphSHA	5	77.01$_{\pm 0.37}$	**75.26**$_{\pm 0.40}$	67.36$_{\pm 1.33}$	62.90$_{\pm 0.49}$	**69.15**$_{\pm 0.46}$	**69.91**$_{\pm 0.46}$	68.03$_{\pm 0.26}$	65.54$_{\pm 0.34}$	65.69$_{\pm 0.19}$	66.61$_{\pm 0.35}$	74.10$_{\pm 0.20}$	72.89$_{\pm 0.35}$
	10	73.20$_{\pm 0.73}$	70.43$_{\pm 1.14}$	61.79$_{\pm 0.09}$	56.09$_{\pm 1.47}$	66.07$_{\pm 1.26}$	66.24$_{\pm 1.34}$	63.05$_{\pm 0.93}$	56.69$_{\pm 4.64}$	63.28$_{\pm 0.55}$	64.06$_{\pm 0.54}$	71.11$_{\pm 0.64}$	69.03$_{\pm 0.64}$
	20	71.00$_{\pm 1.58}$	64.33$_{\pm 2.56}$	57.95$_{\pm 2.55}$	47.37$_{\pm 3.16}$	62.87$_{\pm 1.74}$	60.41$_{\pm 2.91}$	62.72$_{\pm 3.73}$	53.91$_{\pm 4.00}$	57.56$_{\pm 1.71}$	53.22$_{\pm 4.22}$	66.93$_{\pm 1.40}$	62.03$_{\pm 1.82}$
	50	64.97$_{\pm 2.30}$	53.99$_{\pm 2.49}$	65.96$_{\pm 1.44}$	54.07$_{\pm 1.91}$	56.55$_{\pm 2.73}$	49.19$_{\pm 1.58}$	56.10$_{\pm 5.25}$	44.72$_{\pm 6.29}$	50.08$_{\pm 0.95}$	41.72$_{\pm 3.42}$	57.89$_{\pm 1.59}$	50.52$_{\pm 2.26}$
	100	56.13$_{\pm 4.86}$	44.73$_{\pm 4.47}$	60.87$_{\pm 3.26}$	47.25$_{\pm 3.79}$	51.82$_{\pm 3.13}$	42.68$_{\pm 2.95}$	56.63$_{\pm 5.80}$	45.16$_{\pm 8.37}$	43.47$_{\pm 2.99}$	32.60$_{\pm 4.18}$	47.84$_{\pm 4.22}$	37.33$_{\pm 4.40}$
GraphENS	5	71.26$_{\pm 0.61}$	68.65$_{\pm 0.59}$	67.26$_{\pm 0.79}$	63.25$_{\pm 0.86}$	65.35$_{\pm 0.55}$	65.92$_{\pm 0.53}$	69.23$_{\pm 0.42}$	66.72$_{\pm 0.74}$	61.91$_{\pm 0.32}$	62.92$_{\pm 0.25}$	68.04$_{\pm 0.51}$	64.91$_{\pm 2.10}$
	10	70.49$_{\pm 0.63}$	64.53$_{\pm 1.33}$	66.43$_{\pm 1.40}$	58.40$_{\pm 3.69}$	63.25$_{\pm 1.11}$	62.41$_{\pm 2.29}$	67.47$_{\pm 0.37}$	63.87$_{\pm 1.23}$	58.95$_{\pm 0.64}$	58.34$_{\pm 1.07}$	65.50$_{\pm 0.53}$	63.52$_{\pm 0.65}$
	20	68.12$_{\pm 1.26}$	57.59$_{\pm 1.13}$	64.41$_{\pm 0.70}$	53.09$_{\pm 1.11}$	59.61$_{\pm 1.30}$	51.96$_{\pm 1.31}$	65.58$_{\pm 1.02}$	57.31$_{\pm 3.39}$	54.72$_{\pm 0.46}$	49.54$_{\pm 1.89}$	61.11$_{\pm 0.49}$	52.47$_{\pm 1.99}$
	50	62.74$_{\pm 0.81}$	51.76$_{\pm 0.79}$	61.01$_{\pm 1.66}$	47.74$_{\pm 4.14}$	54.09$_{\pm 1.51}$	45.51$_{\pm 1.74}$	60.40$_{\pm 0.90}$	47.81$_{\pm 2.29}$	50.67$_{\pm 0.43}$	42.18$_{\pm 0.44}$	56.26$_{\pm 0.85}$	47.81$_{\pm 2.43}$
	100	56.88$_{\pm 2.99}$	44.33$_{\pm 4.53}$	57.12$_{\pm 1.91}$	41.54$_{\pm 3.48}$	52.12$_{\pm 3.56}$	43.76$_{\pm 3.10}$	57.25$_{\pm 1.97}$	43.10$_{\pm 4.01}$	44.39$_{\pm 1.02}$	32.31$_{\pm 1.04}$	48.91$_{\pm 1.90}$	35.50$_{\pm 1.41}$
TAM	5	74.91$_{\pm 0.25}$	72.17$_{\pm 0.59}$	68.20$_{\pm 0.50}$	64.37$_{\pm 0.46}$	67.21$_{\pm 0.31}$	67.78$_{\pm 0.32}$	71.83$_{\pm 0.29}$	**68.92**$_{\pm 0.35}$	63.39$_{\pm 0.12}$	63.19$_{\pm 0.25}$	71.45$_{\pm 0.16}$	69.83$_{\pm 0.23}$
	10	72.36$_{\pm 0.50}$	67.27$_{\pm 0.56}$	67.12$_{\pm 0.69}$	61.52$_{\pm 1.05}$	64.49$_{\pm 0.41}$	63.40$_{\pm 1.34}$	70.10$_{\pm 0.85}$	66.20$_{\pm 1.02}$	59.17$_{\pm 0.68}$	55.36$_{\pm 1.29}$	66.65$_{\pm 0.39}$	63.16$_{\pm 0.77}$
	20	68.90$_{\pm 1.49}$	60.01$_{\pm 1.23}$	66.11$_{\pm 0.99}$	56.88$_{\pm 1.83}$	61.36$_{\pm 0.70}$	57.64$_{\pm 2.19}$	68.11$_{\pm 0.69}$	62.57$_{\pm 1.98}$	53.81$_{\pm 1.05}$	46.92$_{\pm 1.65}$	61.23$_{\pm 1.22}$	53.28$_{\pm 1.42}$
	50	64.93$_{\pm 3.74}$	55.77$_{\pm 4.74}$	61.63$_{\pm 0.99}$	50.23$_{\pm 1.48}$	58.95$_{\pm 1.64}$	50.99$_{\pm 2.06}$	64.66$_{\pm 1.19}$	54.71$_{\pm 2.03}$	53.39$_{\pm 1.61}$	44.93$_{\pm 2.14}$	57.88$_{\pm 1.67}$	47.52$_{\pm 2.02}$
	100	59.08$_{\pm 3.61}$	48.70$_{\pm 3.60}$	59.81$_{\pm 1.83}$	47.58$_{\pm 2.18}$	55.54$_{\pm 1.43}$	46.82$_{\pm 1.57}$	63.03$_{\pm 0.95}$	51.54$_{\pm 1.58}$	46.92$_{\pm 3.00}$	36.33$_{\pm 3.89}$	52.51$_{\pm 3.57}$	40.28$_{\pm 4.04}$
ImGDA (Ours)	5	**78.35**$_{\pm 1.48}$	75.02$_{\pm 2.21}$	**72.08**$_{\pm 1.86}$	**67.73**$_{\pm 2.88}$	67.49$_{\pm 2.19}$	67.66$_{\pm 2.35}$	**72.14**$_{\pm 1.74}$	68.66$_{\pm 2.83}$	67.65$_{\pm 0.97}$	68.43$_{\pm 1.30}$	76.06$_{\pm 0.47}$	74.50$_{\pm 0.43}$
	10	**77.93**$_{\pm 1.03}$	**73.98**$_{\pm 2.74}$	**71.34**$_{\pm 2.61}$	**66.82**$_{\pm 1.82}$	**67.79**$_{\pm 0.62}$	**68.22**$_{\pm 1.41}$	**73.06**$_{\pm 2.16}$	**69.62**$_{\pm 2.98}$	**66.64**$_{\pm 1.33}$	**66.86**$_{\pm 1.78}$	**74.72**$_{\pm 0.55}$	**72.86**$_{\pm 1.14}$
	20	**77.30**$_{\pm 1.05}$	**73.92**$_{\pm 0.83}$	**71.23**$_{\pm 1.59}$	**65.66**$_{\pm 1.25}$	**67.28**$_{\pm 0.84}$	**66.39**$_{\pm 1.33}$	**68.48**$_{\pm 1.54}$	**63.42**$_{\pm 2.03}$	**63.92**$_{\pm 0.46}$	**63.45**$_{\pm 0.74}$	**71.99**$_{\pm 1.15}$	**68.38**$_{\pm 2.50}$
	50	**73.53**$_{\pm 1.40}$	**67.42**$_{\pm 1.23}$	**69.14**$_{\pm 1.63}$	**61.27**$_{\pm 1.50}$	**66.87**$_{\pm 1.71}$	**65.45**$_{\pm 2.64}$	**70.57**$_{\pm 1.19}$	**66.32**$_{\pm 1.86}$	**60.67**$_{\pm 0.98}$	**58.96**$_{\pm 1.75}$	**66.07**$_{\pm 0.96}$	**62.77**$_{\pm 1.11}$
	100	**69.30**$_{\pm 3.20}$	**58.80**$_{\pm 1.87}$	**68.61**$_{\pm 2.13}$	**56.79**$_{\pm 1.75}$	**60.22**$_{\pm 1.24}$	**51.79**$_{\pm 1.97}$	**64.74**$_{\pm 1.78}$	**52.82**$_{\pm 2.60}$	**52.98**$_{\pm 3.42}$	**45.64**$_{\pm 2.63}$	**58.79**$_{\pm 2.23}$	**48.52**$_{\pm 0.86}$

Generally, removing any part of the method leads to a certain degree of performance degradation. However, removing $\mathcal{L}_{cb}$ leads to significant drops, which clearly demonstrates the importance of utilizing source domain information and mitigating source domain data imbalance.

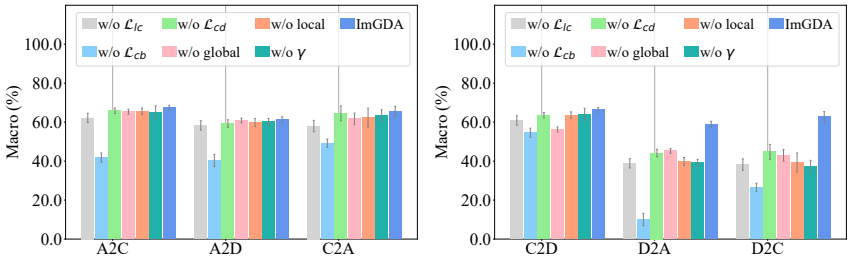
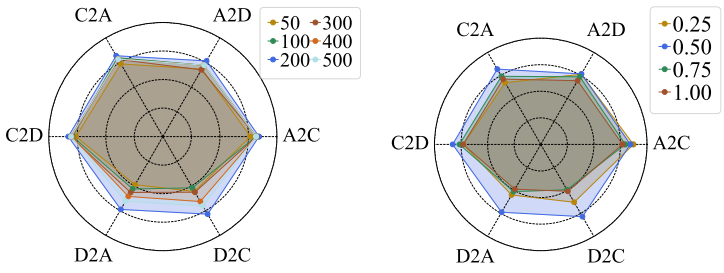


Figure 1: The results of ablation study ($\rho = 50$, Macro score (%) with standard deviation).

Table 2: The complete results of ablation study ($\rho = 50$).

Methods	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
w/o $\mathcal{L}_{lc}$	66.28$_{\pm 2.27}$	62.17$_{\pm 2.38}$	64.16$_{\pm 2.59}$	58.33$_{\pm 2.38}$	62.58$_{\pm 2.38}$	58.00$_{\pm 2.92}$	66.35$_{\pm 2.04}$	60.89$_{\pm 2.71}$	42.79$_{\pm 2.44}$	38.88$_{\pm 2.10}$	46.27$_{\pm 2.73}$	38.23$_{\pm 2.22}$
w/o $\mathcal{L}_{cb}$	52.63$_{\pm 2.40}$	41.84$_{\pm 2.35}$	50.33$_{\pm 2.83}$	40.36$_{\pm 3.12}$	56.12$_{\pm 2.70}$	49.23$_{\pm 2.14}$	58.09$_{\pm 2.78}$	54.56$_{\pm 2.37}$	30.04$_{\pm 0.96}$	10.06$_{\pm 1.86}$	33.43$_{\pm 1.90}$	26.54$_{\pm 1.01}$
w/o $\mathcal{L}_{cd}$	70.81$_{\pm 1.71}$	65.87$_{\pm 1.40}$	65.44$_{\pm 1.22}$	59.27$_{\pm 1.94}$	63.90$_{\pm 0.87}$	64.50$_{\pm 3.83}$	67.81$_{\pm 0.74}$	63.54$_{\pm 3.06}$	49.75$_{\pm 1.94}$	44.17$_{\pm 2.53}$	52.22$_{\pm 2.06}$	44.75$_{\pm 2.53}$
w/o global	71.93$_{\pm 0.36}$	65.34$_{\pm 1.27}$	67.96$_{\pm 1.40}$	60.84$_{\pm 1.23}$	65.30$_{\pm 0.64}$	61.69$_{\pm 2.93}$	64.84$_{\pm 0.43}$	56.30$_{\pm 1.50}$	50.26$_{\pm 1.87}$	45.27$_{\pm 2.42}$	51.56$_{\pm 1.60}$	43.00$_{\pm 1.87}$
w/o local	72.69$_{\pm 1.17}$	65.63$_{\pm 1.67}$	67.40$_{\pm 1.33}$	59.80$_{\pm 2.04}$	65.15$_{\pm 1.95}$	62.30$_{\pm 4.84}$	68.14$_{\pm 0.96}$	63.60$_{\pm 2.60}$	46.69$_{\pm 1.43}$	39.82$_{\pm 1.98}$	48.84$_{\pm 1.52}$	39.34$_{\pm 1.19}$
w/o γ	72.07$_{\pm 3.68}$	65.14$_{\pm 3.26}$	67.59$_{\pm 1.39}$	60.13$_{\pm 1.78}$	65.15$_{\pm 1.89}$	63.34$_{\pm 3.04}$	68.33$_{\pm 1.01}$	63.85$_{\pm 2.63}$	46.09$_{\pm 2.16}$	39.16$_{\pm 4.44}$	47.36$_{\pm 2.81}$	37.37$_{\pm 4.41}$
ImGDA	73.53$_{\pm 1.40}$	67.42$_{\pm 1.23}$	69.14$_{\pm 1.63}$	61.27$_{\pm 1.50}$	66.87$_{\pm 1.71}$	65.45$_{\pm 2.64}$	70.57$_{\pm 1.19}$	66.32$_{\pm 1.86}$	60.67$_{\pm 0.98}$	58.96$_{\pm 1.75}$	66.07$_{\pm 0.96}$	62.77$_{\pm 1.11}$



(a) k (b) γ

Figure 2: Visualization of the sensitivity analysis of (a) γ and (b) k. (Macro score (%))

D.3 Sensitivity Study

In this section, we further present the sensitivity analysis results of two key hyperparameters concerning the Macro score, along with the complete numerical results for both metrics, to provide a clearer understanding of their impact on model performance. The experiments are conducted under the setting of $\rho = 50$.

Table 3: The impact of different values of k on performance.

k	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
50	$72.11_{\pm 1.44}$	$61.40_{\pm 1.56}$	$66.38_{\pm 2.86}$	$54.64_{\pm 2.61}$	$64.42_{\pm 1.30}$	$58.83_{\pm 2.80}$	$70.45_{\pm 1.55}$	$60.99_{\pm 3.10}$	$46.00_{\pm 3.17}$	$38.97_{\pm 2.33}$	$54.34_{\pm 3.25}$	$44.39_{\pm 2.50}$
100	$74.18_{\pm 1.72}$	$65.36_{\pm 1.59}$	$68.89_{\pm 1.33}$	$57.77_{\pm 0.68}$	$66.10_{\pm 1.96}$	$63.84_{\pm 2.38}$	$70.79_{\pm 0.72}$	$63.23_{\pm 2.96}$	$48.55_{\pm 2.76}$	$42.40_{\pm 2.07}$	$50.53_{\pm 2.55}$	$41.50_{\pm 3.08}$
200	$73.53_{\pm 1.40}$	$67.42_{\pm 1.23}$	$69.14_{\pm 1.63}$	$61.27_{\pm 1.50}$	$66.87_{\pm 1.71}$	$65.45_{\pm 2.64}$	$70.57_{\pm 1.19}$	$66.32_{\pm 1.86}$	$60.67_{\pm 0.98}$	$58.96_{\pm 1.75}$	$66.07_{\pm 0.96}$	$62.77_{\pm 1.11}$
300	$72.44_{\pm 1.58}$	$65.63_{\pm 3.14}$	$65.35_{\pm 3.02}$	$54.08_{\pm 2.52}$	$66.72_{\pm 1.23}$	$62.44_{\pm 2.24}$	$70.50_{\pm 1.00}$	$64.62_{\pm 3.86}$	$50.89_{\pm 1.62}$	$45.24_{\pm 2.84}$	$52.84_{\pm 1.70}$	$45.00_{\pm 2.67}$
400	$72.33_{\pm 1.43}$	$65.45_{\pm 0.97}$	$69.32_{\pm 0.93}$	$57.78_{\pm 2.30}$	$65.93_{\pm 2.20}$	$62.83_{\pm 2.09}$	$70.31_{\pm 0.77}$	$63.70_{\pm 3.86}$	$53.18_{\pm 1.82}$	$48.45_{\pm 2.99}$	$57.24_{\pm 3.66}$	$52.56_{\pm 2.11}$
500	$72.96_{\pm 2.02}$	$65.35_{\pm 0.78}$	$68.64_{\pm 1.67}$	$57.85_{\pm 1.34}$	$66.12_{\pm 1.65}$	$62.68_{\pm 2.37}$	$69.92_{\pm 0.95}$	$64.39_{\pm 3.44}$	$55.80_{\pm 2.44}$	$52.32_{\pm 3.77}$	$60.59_{\pm 3.02}$	$56.12_{\pm 2.74}$

Effect of k. To explore the impact of different values of k (the number of anchor nodes) for our ImGDA, we conduct a sensitivity analysis of k. The results are shown in Table 3, and Figure 2a provides a visual analysis of the macro scores. It can be observed that both excessively small and large values of k affect the model's performance. As discussed in Section 5.4 in the body of the paper, a too-small value of k results in insufficient utilization of node information from the source domain, limiting the exploration of graph structural semantics. On the other hand, an excessively large k may introduce data imbalance, leading to negative feedback that ultimately affects model performance.

Effect of γ. As discussed in Section 5.4 in the body of the paper, the hyperparameter γ can be viewed as a weight parameter associated with the gradient of the tail class. To evaluate the impact of different values of γ on model performance, we conduct a sensitivity analysis of γ. The results are shown in Table 4 in the Appendix, and Figure 2b provides a visual analysis of the macro scores. Both excessively small and large values of γ lead to a decline in model performance, indicating that the gradient of the tail class should maintain a moderate weight during model training (i.e., too small a value affects the classification performance of other classes, while too large a value negatively impacts the classification of the tail class itself).

D.4 Analysis of Different Sampling Strategies

Different sampling strategies affect the quality of the sampled anchor nodes. To investigate the impact of different sampling strategies on model performance, we conduct a comparative analysis of three different sampling strategies. Table 5 in the Appendix presents the experiment results with $\rho = \{20, 50, 100\}$. It can be observed that graph-structure-based sampling strategies (Deg. and Inv.) significantly outperform random sampling (Rand.), highlighting the importance of structural information in learning meaningful node representations. Furthermore, compared to the inverse degree-based strategy, our degree-based sampling approach achieves better performance, possibly because a node's local subgraph structure plays a crucial role in its representation within the entire graph. While the inverse degree-based strategy balances sampling bias, it inevitably leads to greater structural semantic loss by disregarding key graph information. Additionally, our experiments show

Table 4: The impact of different values of γ on performance.

γ	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
0.25	$74.47_{\pm 2.25}$	$69.70_{\pm 1.58}$	$68.59_{\pm 1.59}$	$60.95_{\pm 1.66}$	$61.46_{\pm 3.43}$	$53.64_{\pm 2.06}$	$68.38_{\pm 2.55}$	$58.34_{\pm 2.44}$	$48.40_{\pm 3.12}$	$43.94_{\pm 3.75}$	$55.64_{\pm 2.64}$	$50.26_{\pm 2.69}$
0.50	$73.53_{\pm 1.40}$	$67.42_{\pm 1.23}$	$69.14_{\pm 1.63}$	$61.27_{\pm 1.50}$	$66.87_{\pm 1.71}$	$65.45_{\pm 2.64}$	$70.57_{\pm 1.19}$	$66.32_{\pm 1.86}$	$60.67_{\pm 0.98}$	$58.96_{\pm 1.75}$	$66.07_{\pm 0.96}$	$62.77_{\pm 1.11}$
0.75	$72.06_{\pm 3.40}$	$63.59_{\pm 3.07}$	$67.97_{\pm 3.31}$	$58.79_{\pm 2.52}$	$64.00_{\pm 1.58}$	$59.17_{\pm 3.26}$	$67.60_{\pm 1.68}$	$61.21_{\pm 3.69}$	$48.05_{\pm 2.90}$	$41.01_{\pm 4.15}$	$50.12_{\pm 2.21}$	$39.42_{\pm 3.41}$
1.00	$63.76_{\pm 2.19}$	$61.46_{\pm 2.23}$	$64.76_{\pm 3.17}$	$55.37_{\pm 2.38}$	$62.48_{\pm 2.07}$	$56.74_{\pm 3.06}$	$67.18_{\pm 1.57}$	$58.46_{\pm 2.25}$	$46.55_{\pm 2.33}$	$38.92_{\pm 2.91}$	$49.28_{\pm 3.50}$	$40.73_{\pm 3.52}$

that the inverse degree-based sampling strategy results in significantly longer training times. These factors collectively justify our choice of the degree-based sampling strategy.

Table 5: The results of different sampling strategies.

Sampling Strategy	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
$\rho = 20$	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
Deg.	$77.30_{\pm 1.05}$	$73.92_{\pm 0.83}$	$71.23_{\pm 1.59}$	$65.66_{\pm 1.25}$	$67.28_{\pm 0.84}$	$66.39_{\pm 1.33}$	$68.48_{\pm 1.54}$	$63.42_{\pm 2.03}$	$63.92_{\pm 0.46}$	$63.45_{\pm 0.74}$	$71.99_{\pm 1.15}$	$68.38_{\pm 2.50}$
Inv.	$77.57_{\pm 1.17}$	$74.15_{\pm 1.16}$	$71.27_{\pm 0.36}$	$65.24_{\pm 0.97}$	$65.90_{\pm 0.81}$	$61.14_{\pm 3.46}$	$69.80_{\pm 1.10}$	$61.28_{\pm 3.15}$	$64.47_{\pm 0.54}$	$64.26_{\pm 0.54}$	$71.40_{\pm 0.76}$	$69.18_{\pm 0.94}$
Rand.	$76.86_{\pm 0.39}$	$72.71_{\pm 0.67}$	$71.44_{\pm 0.57}$	$65.11_{\pm 0.67}$	$66.15_{\pm 1.22}$	$61.51_{\pm 3.77}$	$69.67_{\pm 2.25}$	$62.13_{\pm 2.40}$	$63.68_{\pm 1.51}$	$63.17_{\pm 1.84}$	$71.70_{\pm 1.11}$	$68.93_{\pm 0.80}$
sampling method	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
$\rho = 50$	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
Deg.	$73.53_{\pm 1.40}$	$67.42_{\pm 1.23}$	$69.14_{\pm 1.63}$	$61.27_{\pm 1.50}$	$66.87_{\pm 1.71}$	$65.45_{\pm 2.64}$	$70.57_{\pm 1.19}$	$66.32_{\pm 1.86}$	$60.67_{\pm 0.98}$	$58.96_{\pm 1.75}$	$66.07_{\pm 0.96}$	$62.77_{\pm 1.11}$
Inv.	$72.40_{\pm 1.82}$	$64.91_{\pm 1.62}$	$68.39_{\pm 1.91}$	$56.99_{\pm 3.48}$	$66.99_{\pm 1.52}$	$63.81_{\pm 2.64}$	$69.89_{\pm 1.01}$	$62.22_{\pm 2.24}$	$57.09_{\pm 1.90}$	$54.46_{\pm 3.45}$	$59.86_{\pm 2.07}$	$53.23_{\pm 3.37}$
Rand.	$73.15_{\pm 1.12}$	$65.87_{\pm 0.17}$	$68.04_{\pm 1.18}$	$57.38_{\pm 1.86}$	$66.39_{\pm 1.16}$	$61.04_{\pm 2.93}$	$70.31_{\pm 0.83}$	$63.83_{\pm 3.25}$	$56.65_{\pm 1.54}$	$53.41_{\pm 2.10}$	$60.68_{\pm 2.66}$	$55.12_{\pm 3.77}$
sampling method	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
$\rho = 100$	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
Deg.	$69.30_{\pm 3.20}$	$58.80_{\pm 1.87}$	$68.61_{\pm 2.13}$	$56.79_{\pm 1.75}$	$60.22_{\pm 1.24}$	$51.79_{\pm 1.97}$	$64.74_{\pm 1.78}$	$52.82_{\pm 2.60}$	$52.98_{\pm 3.42}$	$45.64_{\pm 2.63}$	$58.79_{\pm 2.23}$	$48.52_{\pm 0.86}$
Inv.	$69.25_{\pm 4.37}$	$60.36_{\pm 4.11}$	$67.38_{\pm 2.62}$	$55.61_{\pm 2.39}$	$58.94_{\pm 1.90}$	$48.19_{\pm 2.55}$	$64.36_{\pm 1.21}$	$50.98_{\pm 2.90}$	$54.66_{\pm 1.71}$	$47.59_{\pm 1.42}$	$58.08_{\pm 4.11}$	$48.20_{\pm 2.45}$
Rand.	$69.25_{\pm 4.68}$	$57.04_{\pm 3.45}$	$67.54_{\pm 3.00}$	$56.54_{\pm 2.11}$	$59.32_{\pm 1.96}$	$48.61_{\pm 2.63}$	$64.82_{\pm 1.53}$	$51.93_{\pm 2.93}$	$54.88_{\pm 1.97}$	$47.61_{\pm 1.47}$	$57.30_{\pm 2.14}$	$46.23_{\pm 2.36}$

D.5 Analysis of Different GNN Encoders

To explore the effect of different encoders in our method, we replace our encoder from GCN to GraphSAGE [8], a widely used classic GNN encoder, and conduct experiments under $\rho = 50$. The results are shown in Table 6. It can be observed that in our method, GCN outperforms GraphSAGE, possibly because GraphSAGE samples only a subset of a target node's neighbors, which limits the exploration of local graph structures and leads to a performance drop. This highlights the importance of selecting an appropriate backbone encoder.

Table 6: The results of different GNN encoders.

Encoder	A⇒C		A⇒D		C⇒A		C⇒D		D⇒A		D⇒C	
	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro	Micro	Macro
GCN	$73.53_{\pm 1.40}$	$67.42_{\pm 1.23}$	$69.14_{\pm 1.63}$	$61.27_{\pm 1.50}$	$66.87_{\pm 1.71}$	$65.45_{\pm 2.64}$	$70.57_{\pm 1.19}$	$66.32_{\pm 1.86}$	$60.67_{\pm 0.98}$	$58.96_{\pm 1.75}$	$66.07_{\pm 0.96}$	$62.77_{\pm 1.11}$
GraphSAGE	$62.28_{\pm 1.77}$	$56.77_{\pm 2.12}$	$60.13_{\pm 1.21}$	$50.03_{\pm 2.14}$	$53.06_{\pm 1.44}$	$50.80_{\pm 0.91}$	$59.68_{\pm 1.36}$	$50.61_{\pm 1.40}$	$48.22_{\pm 1.85}$	$44.50_{\pm 2.03}$	$52.59_{\pm 1.44}$	$46.38_{\pm 1.98}$

Table 7: Detailed values of Standard Deviation, Mean/Median Ratio, Gini Coefficient with different imbalance factors of different Datasets.

Methods	IF(ρ)	σ	ϵ	δ
ACMv9	5	628.79	1.17	0.31
	10	714.46	1.36	0.41
	20	762.39	1.66	0.49
	50	798.08	2.25	0.57
	100	814.47	2.92	0.62
Citationv1	5	525.18	1.17	0.31
	10	596.31	1.36	0.41
	20	636.40	1.66	0.39
	50	666.23	2.25	0.57
	100	679.81	2.92	0.62
DBLPv7	5	420.95	1.17	0.31
	10	471.26	1.36	0.41
	20	502.90	1.66	0.49
	50	526.27	2.25	0.57
	100	537.29	2.92	0.62

E Additional Metrics of Imbalanced Distributions

In our experiments, we employ the imbalance factor (ρ) to quantify the degree of class imbalance in the dataset, defined as:

$$\rho = \max\{N_1, N_2, \cdots, N_C\} / \min\{N_1, N_2, \cdots, N_C\}, \tag{23}$$

where N_i denotes the number of nodes belonging to class i in the source graph. To provide a more comprehensive assessment of imbalance across different datasets, we further introduce several additional metrics to evaluate the distributional skewness of the data.

- Standard Deviation (σ): It is widely used in probability and statistics to measure statistical dispersion, and in some cases it can reflect sampling uncertainty, which is defined as:

$$\sigma = \sqrt{\frac{1}{C}\sum_{i=1}^{C}(n_i - \bar{n}_i)^2}, \tag{24}$$

  where C represents the number of classes, n_i represents the instance number of class i, and $\bar{n}_i$ represents the average number of instances.

- Mean/Median Ratio (ϵ): The median is a fundamental statistical measure widely applied in fields such as economics, sociology, and medicine. Unlike the mean, it is less sensitive to extreme values and thus provides a more robust representation of the data distribution. Consequently, the mean-to-median ratio serves as an indicator of data skewness, defined as:

$$\epsilon = \frac{\text{mean}(N_1, N_2, \cdots, N_C)}{\text{median}(N_1, N_2, \cdots, N_C)}. \tag{25}$$

- Gini Coefficient (δ): It was originally introduced by Italian economist Gini [5], measures distributional equality based on the Lorenz curve. Commonly used to quantify income or wealth inequality, it can similarly be applied as a metric of imbalanced distribution, as class imbalance parallels inequality across categories.

And we provide the corresponding Standard Deviation, Mean/Median Ratio, and Gini Coefficient values for each dataset under different imbalance factors in Table 7.

F Broader Impact and Limitations

Our method addresses class imbalance in graph domain adaptation and has the potential to benefit various applications such as scientific discovery, recommender systems, and bioinformatics by improving the representation of underrepresented classes. However, it also has limitations: the performance depends on the quality of the selected anchor set, the dual-branch design introduces additional computational overhead, and the pseudo-labeling process may suffer from noise under large domain shifts. Moreover, the framework requires the same label space across source and target domains, which could not be applied to open-set scenarios.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The main claims in the abstract and introduction accurately reflect the paper's contributions and scope by proposing a dual-branch prototype-enhanced contrastive framework that effectively addresses class imbalance and domain shifts in graph transfer learning, supported by extensive experiments across multiple datasets.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The appendix (Section F) states that the performance depends on the quality of the selected anchor set, the dual-branch design introduces additional computational overhead, and the pseudo-labeling process may suffer from noise under large domain shifts. Moreover, the framework assumes shared label space between source and target domains, which may limit its applicability in open-set scenarios.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: We have provided the theoretical proof in detail in the appendix.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: In Section 5.1, "Experimental Settings", we provide detailed explanations of the datasets used and implementation details for our experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: In the abstract, we provide a link to the code.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: In Section 5.1, "Experimental Settings", we provide detailed explanations of the datasets used and implementation details for our experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We precisely defined and reported the error bars.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: In Section 5.1, "Experimental Settings", we provide detailed explanations of the datasets used and implementation details for our experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper complies with the NeurIPS Code of Ethics in every respect.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: This paper discusses potential positive impacts such as improving fairness and robustness in graph learning by addressing class imbalance and domain shifts, as well as limitations that could lead to negative effects like noise from pseudo-labeling.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: This paper poses no such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We explicitly cited the sources of the relevant data and other materials used in the paper.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: The link provided in the abstract contains well-documented related materials.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: This paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [NA]

    Justification: This paper does not involve LLMs as a core component of its methodology.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.