Disentangled Generation-Based Prototypical Alignment for Few-Shot Unsupervised Domain Adaptation in Graph-Level Anomaly Detection

Zhibin Ni[1], **Chenghao Zhang**[1], **Hai Wan**[1,2*], **Xibin Zhao**[1*]

[1] BNRist, KLISS, and School of Software, Tsinghua University
[2] Hunan Sanyou Environmental Technology Co., Ltd., Changsha, Hunan, China
{nzb22,chenghao24}@mails.tsinghua.edu.cn, {wanhai,zxb}@tsinghua.edu.cn

Abstract

Graph-Level Anomaly Detection (GLAD) seeks to identify anomalous graphs within graph datasets, which has significant applications across diverse real-world fields. Most existing GLAD methods are trained in an unsupervised manner due to high costs for labeling, resulting in sub-optimal performance when compared to supervised methods. To fill this gap, we propose a **D**isentangled **G**eneration-Based **P**rototypical **A**lignment **(DGPA)** method that extends graph-level anomaly detection to Few-Shot Unsupervised Domain Adaptation (FUDA) setting, aiming to identify anomalous graphs from a set of unlabeled graphs (target domain) by using partially labeled graphs from a different but related domain (source domain), which fulfills the practical requirement of transferring anomaly knowledge. This is specifically achieved through a dedicated *Disentangled Sample Generation* module, which addresses ***label scarcity*** by generating faithful samples with disentangled representation learning grounded in Information Bottleneck principle, along with a *Graph-based Prototypical Self-Supervision* module, which alleviates ***domain shift*** by encoding and aligning semantic structures in the shared latent space across domains in a self-supervised manner. Extensive experiments on four benchmark datasets reveal the effectiveness of our proposed DGPA.

Introduction

Graph-level anomaly detection (GLAD) aims at capturing anomalous graphs in a graph set (Ma et al. 2021). As a critical area in data mining, GLAD finds extensive applications across diverse domains such as financial networks (Qiu et al. 2022), drug discovery (Ma et al. 2023) and malware detection (Fan et al. 2021). Despite various proposed methods, most of them are predominantly unsupervised and thus yield suboptimal detection accuracy compared to supervised approaches. Manual annotation of large-scale graph data in a single domain can be prohibitively costly and challenging in practical scenarios. However, obtaining adequate labeled data from pre-existing, related domains is often feasible and attractive (Xi et al. 2025; Shao et al. 2025a). Unsupervised Domain Adaptation (UDA) achieves this by transferring predictive knowledge from a fully-labeled source domain to an unlabeled target domain, emerging as a viable solution for GLAD and attracting significant attention in both scientific research and industrial practice (Li et al. 2024).

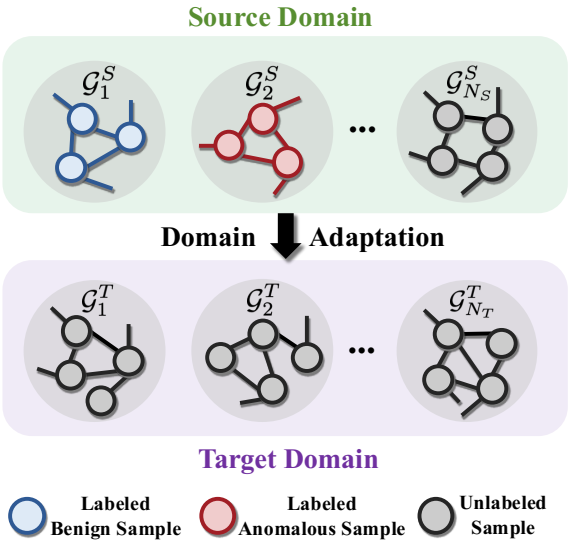

Figure 1: Illustration of Few-Shot Unsupervised Domain Adaptation (FUDA) in graph-level anomaly detection. Given a partially labeled source domain which contains a few annotated benign and anomalous samples, the task aims to generalize to a fully unlabeled target domain.

In some real-world applications, providing large-scale annotations even in the source domain is often expensive (Shao et al. 2025b). Taking molecular data for instance, it takes hours for experts to determine the properties of a molecule with merely twenty atoms via density functional theory (Wang et al. 2022b). Thus, in practice, assuming a source domain with abundant labels is overly restrictive.

In this paper, to address the costs of source domain labeling, we instead consider a few-shot unsupervised domain adaptation (FUDA) setting, where only a minimal subset of source samples are labeled while the remaining source samples and all target samples remain unlabeled. In practice, it is often feasible to acquire a small number of labeled normal and anomalous graphs at a relatively low cost. Although a few studies have explored UDA for GLAD task, FUDA remains unexplored in this context. In summary, as shown in

*Corresponding author.

Fig.1, two challenges need to be overcome.

(i) Label Scarcity. Compared to UDA, FUDA poses greater challenges for GLAD under label-scarce conditions. Prior few-shot domain adaptation techniques (Xu et al. 2022; Zhang et al. 2023a) often adopt the data augmentation mechanism, which uses generative methods to augment the data and capture the underlying data distribution. However, they struggle to effectively disentangle the two complementary components found within graphs: the label-relevant and label-irrelevant factors. This issue further amplifies the distribution shift between seen and unseen samples, causing generated samples to deviate from the true data distribution and degrading the generalization on the target domain. Here, label-relevant information pertains to a specific class and contains discriminative substructures critical for detection. Conversely, label-irrelevant information captures all label-irrelevant features, reflecting intra-class diversity. Hence, there exists a compelling need to develop an effective disentangled generation method explicitly tailored for GLAD task under FUDA setting.

(ii) Domain shift. For adapting graph-level anomaly detection to the FUDA setting, a key challenge is domain shift, specifically the distributional discrepancy between source and target domains, which manifests as variations in graph structures, sizes, and anomaly distributions. Existing FUDA methods are mostly designed for Euclidean vector-based data (e.g., images, text) that rely on distance metrics in continuous vector spaces. This makes directly applying off-the-shelf FUDA methods to GLAD impractical. While a few recent cross-domain few-shot graph anomaly detection works (Chen et al. 2024) have been proposed, they only focus on node-level tasks. However, graph-level anomaly detection presents greater complexity, as anomalies at the graph level often manifest through global patterns and semantics that cannot be captured by simply examining individual nodes.

To address these two challenges, we propose a novel **D**isentangled **G**eneration-Based **P**rototypical **A**lignment **(DGPA)** model, which pioneers the extension of GLAD tasks to the FUDA setting, thereby enabling effective transfer of predictive knowledge from a sparsely labeled source domain to an unlabeled target domain. To achieve this goal, DGPA comprises two novel modules.

First, we introduce a dedicated *Disentangled Sample Generation (DSG)* module to address ***label scarcity*** problem in GLAD tasks. To achieve better representation learning, the DSG module proposes a novel disentangled information bottleneck to disentangle label-relevant information from label-irrelevant information. To improve the robustness of generated samples, the DSG module combines label-relevant information for discrimination and various label-irrelevant information for diversity. **Second**, we further propose a *Graph-based Prototypical Self-Supervision (GPS)* module to alleviate ***domain shift*** issue. To enable semantic structure learning, the GPS module introduces prototype learning to learn representative knowledge based on label-relevant information. To facilitate domain alignment, the GPS module performs a novel prototype-based domain alignment in a self-supervised manner.

In sum, our contributions can be summarized as follows:

- To the best of our knowledge, we make the first attempt to reformulate graph-level anomaly detection task as a few-shot unsupervised domain adaptation problem, thereby extending the applicability of traditional methods to cross-domain scenarios.
- We propose DGPA, a novel method which introduces a disentangled sample generation method to address the ***label scarcity*** and a graph-based prototypical self-supervision strategy to alleviate the ***domain shift***.
- Extensive experiments on four benchmark datasets demonstrate that our proposed method outperforms the state-of-the-art methods in the cross-domain few-shot graph-level anomaly detection task.

Related Work

Graph Level Anomaly Detection. Graph-level anomaly detection (GLAD) aims to find anomalous graphs within a graph set. Recent studies try to address the GLAD problem with various advanced techniques, such as knowledge distillation (Ma et al. 2022), transformation learning (Qiu et al. 2022), deep Random Walk Kernel (Zhang et al. 2022) and differential evolutionary algorithm (Ma et al. 2023). While existing methods predominantly focus on single-domain scenarios, cross-domain approaches are mainly node-level (Wang et al. 2023) and unsuitable for graph-level tasks. Conversely, our work operates GLAD in a FUDA setting, further eliminating the need for extensive annotation in the source domain, thus showing greater efficiency in cross-domain scenarios.

Graph Domain Adaptation. Recently, to address the domain shift issue caused by learning from distinct domains, increasing attention has been drawn to leveraging domain adaptation to promote graph learning. Research works combine domain adaptation with graph data by domain adversarial learning (Shen et al. 2020), knowledge selection (Zhang et al. 2023b), domain-invariant feature learning (Li et al. 2024), etc. However, most of them focus on the multi-class node classification problem rather than the GLAD task, which is more challenging due to extreme label scarcity and graph-level properties. In contrast, our model leverages DGPA for FUDA in GLAD task, effectively addressing challenges related to domain shift and label scarcity.

Problem Formulation

Denote $\mathcal{D}^S = \{(G_i^S, y_i^S)\}_{i=1}^{N_S}$ as a partially labeled graph dataset from the source domain, where each attributed graph $G_i^S = (V_i^S, E_i^S, X_i^S, A_i^S)$ consists of nodes V_i^S, edges E_i^S, node features $X_i^S \in \mathbb{R}^{|V_i^S| \times d}$ and adjacency matrix A_i^S. Unlike conventional settings where the source domain contains only normal graphs ($y_i^S = 0$), we assume a more practical scenario where $\mathcal{D}^S$ includes both normal and anomalous graphs, i.e., $y_i^S \in \{0, 1\}$ for a subset of graphs. This setting is more realistic since a small amount of labeled source domain data is easy to obtain, yet has not been well leveraged by existing works. Conversely, the target domain $\mathcal{D}^T = \{G_j^T\}_{j=1}^{N_T}$ is fully unlabeled, with no access to y_j^T. Each target graph $G_j^T = (V_j^T, E_j^T, X_j^T, A_j^T)$ shares the same label

space as the source domain but exhibits distinct distributions due to domain shift. The goal of GLAD task under FUDA setting is to learn a binary classifier $f : \mathcal{G}^T \rightarrow \{0, 1\}$ that detects anomalous graphs in $\mathcal{D}^T$ by transferring knowledge from the partially labeled source domain $\mathcal{D}^S$ while adapting to the unlabeled target domain $\mathcal{D}^T$.

The Proposed Model

Model Overview

As shown in Fig. 2, the proposed model consists of two modules: Disentangled Sample Generation (DSG) and Graph-based Prototypical Self-Supervision (GPS). **First**, a DSG module is proposed to address the label scarcity challenge under the few-shot setting. In the DSG module, a novel disentangled information bottleneck is proposed to disentangle label-relevant and label-irrelevant subgraphs for robust anomaly detection, while generating faithful unseen samples based on disentanglement. **Second**, a GPS module is proposed to alleviate domain shift problem in cross-domain scenario. In the GPS module, a novel prototype learning method on disentangled subgraphs is introduced to achieve cross-domain alignment in a self-supervised manner.

Disentangled Sample Generation

Challenge Analysis. Prior FUDA work typically employs data augmentation such as adversarial mixup (Zhang et al. 2023a) and diffusion models (Benigmim et al. 2023) to improve generalizability. However, they fail to effectively disentangle label-relevant and label-irrelevant factors. This causes generated samples to deviate from the true distribution, as entangled generators exploit spurious features.

Method Rationale. To address these challenges, we propose a DSG module using disentangled graph information bottleneck to generate samples that combine source label-relevant and target label-irrelevant features, preserving discriminability while bridging the domain gap in label-scarce scenarios. The DSG objective is formally defined as follows:

$$\begin{aligned}\mathcal{L}_{DSG} = &\underbrace{I(G^{I,S};Y^S)}_{\text{Relevance}} - \underbrace{(1+\alpha)I(Y^S;G^{E,S})}_{\text{Disentanglement}} \\ &- \underbrace{\beta I(G;G^I)}_{\text{Compression}} + \underbrace{I(G;G^I,G^E)}_{\text{Reconstruction}} \\ &+ \underbrace{\lambda I(G^T;G^{I,S},G^{E,T})}_{\text{Cross-Domain}}. \end{aligned} \tag{1}$$

For the relevance term, maximizing $I(G^{I,S};Y^S)$ aims to preserve as much as possible the information on Y^S. **For the disentanglement term**, minimizing $I(Y^S;G^{E,S})$ aims to encode all label-irrelevant information into the label-irrelevant representation $G^{E,S}$ from source domain, so as to achieve disentanglement. Hyperparameter α controls the degree of disentanglement. **For the compression term**, minimizing $I(G;G^I)$ aims to find a maximally compressed representation G^I of original graphs. Hyperparameter β controls the degree of compression. **For the reconstruction term**, maximizing $I(G;G^I,G^E)$ encourages the disentangled representation pair (G^E,G^I) to be sufficient for reconstructing the input graph G. **For the cross-domain term**, maximizing $I(G^T;G^{I,S},G^{E,T})$ measures how the label-relevant representation $G^{I,S}$ from source domain collaborates with the label-irrelevant representation $G^{E,T}$ from target domain to reconstruct target domain graph G^T.

Proof Sketch. We briefly conclude the proof as follows: (i) From the variational lower bound, $I(G^{I,S};Y^S) \geq \mathbb{E}[\log q(Y^S|G^{I,S})]$ corresponds to $\mathcal{L}_{\text{cls}}$; (ii) For $I(G;G^I)$, we obtain a computable upper bound dependent on the probability and noise variance, corresponding to $\mathcal{L}_{\text{MI}}^1$ and the consistency regularization $\mathcal{L}_{\text{con}}$; (iii) For $I(G;G^I,G^E)$ and $I(G^T;G^{I,S},G^{E,T})$, we provide log-likelihood lower bounds, corresponding to $\mathcal{L}_{\text{recon}}$ and $\mathcal{L}_{\text{cro}}$; (iv) Special cases: Setting $\alpha_4{=}\alpha_5{=}0$ yields vanilla Disentangled Information Bottleneck; setting $\beta{=}0, \alpha{>}0$ and conditioning the reconstruction terms on (G^I,G^E) degrades these two terms to a conditional generative log-likelihood lower bound.

Relevance. $I(G^{I,S};Y^S)$ is realized through a noise-injected subgraph extraction mechanism. Specifically, we extract a label-relevant subgraph $G_i^{I,S}$ from the labeled source domain graph G_i^S by selectively preserving critical nodes via stochastic noise injection. For each node $v_k^S \in G_i^S$ with representation $\mathbf{h}_k^S$, we compute its preservation probability $p_k^{I,S}$ and noisy embedding $\mathbf{z}_k^{I,S}$:

$$\begin{aligned} p_k^{I,S} &= \text{Sigmoid}(\text{MLP}^I(\mathbf{h}_k^S)), \\ \mathbf{z}_k^{I,S} = \lambda_k^{I,S}\mathbf{h}_k^S + (1{-}\lambda_k^{I,S})\boldsymbol{\epsilon}_k^{I,S}, &\quad \lambda_k^{I,S} \sim \text{Bernoulli}(p_k^{I,S}), \\ \boldsymbol{\epsilon}_k^{I,S} &\sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{h}_k^{I,S}}, \boldsymbol{\sigma}_{\mathbf{h}_k^{I,S}}), \end{aligned} \tag{2}$$

where $\boldsymbol{\mu}_{\mathbf{h}_k^S}, \boldsymbol{\sigma}_{\mathbf{h}_k^S}$ are node-specific Gaussian parameters for source domain. This process adaptively compresses G_i^S into $G_i^{I,S}$ by maximizing $I(G_i^{I,S};Y_i^S)$ through high $p_k^{I,S}$ values for nodes critical to Y_i^S to retain label-relevant information.

To maximize $I(G_i^{I,S};Y_i^S)$, we employ the cross-entropy loss that encourages the extracted subgraph to preserve label-relevant information:

$$\mathcal{L}_{\text{cls}} = -\sum_{i=1}^{N}\sum_{c=1}^{K}\mathbb{I}(y_i^S = c)\log p(y_i^S = c|G_i^{I,S}), \tag{3}$$

where $p(y_i^S = c|G_i^{I,S})$ is the predicted probability of class c given the extracted subgraph $G_i^{I,S}$. Similarly, $G_i^{E,S}$ is extracted through the same noise-injected mechanism with different parameters.

Disentanglement. The disentanglement term minimizes $I(Y^S;G^{E,S})$ to eliminate label-relevant information from the label-irrelevant representation. Unfortunately, there exists a trivial solution where the model simply memorizes the distribution $p(Y^S)$ while $G^{E,S}$ is not optimized. Since direct optimization of $I(Y^S;G^{E,S})$ is infeasible, we derive an equivalent objective via conditional independencies of IB:

$$\begin{aligned} I(Y^S;G^{E,S}) = &I(G^S;G^{E,S}) + I(G^S;Y^S) \\ &- I(G^S;Y^S,G^{E,S}) \end{aligned} \tag{4}$$

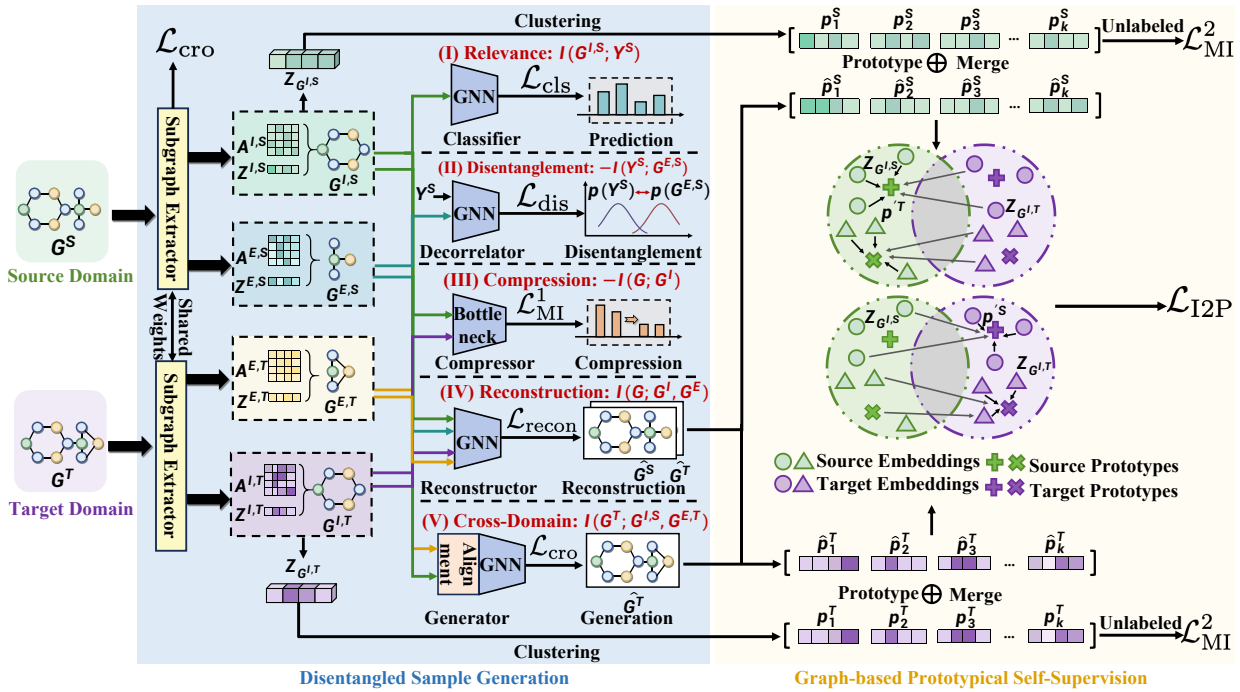

Figure 2: The main framework of the proposed DGPA for few-shot unsupervised domain adaptation in graph-level anomaly detection task. The framework consists of two modules, including a disentangled sample generation module (the blue part) and a graph-based prototypical self-supervision module (the yellow part).

where $I(G^S; Y^S)$ is a constant. $I(G^S; G^{E,S})$ is minimized for both source and target domains via its variational upper bound:

$$I(G; G^S) \leq \mathbb{E}_G \left[-\frac{1}{2} \log \mathcal{A}^S + \frac{\mathcal{A}^S + \mathcal{B}^{S^2}}{2|V_G|} \right] = \mathcal{L}^1_{\text{dis}}, \quad (5)$$

where $\mathcal{A}^S = \sum_{k=1}^{|V_G|} (1 - \lambda_k^S)^2$ and $\mathcal{B}^S = \sum_{k=1}^{|V_G|} \lambda_k^S (\mathbf{h}_k^S - \boldsymbol{\mu}_{\mathbf{h}_k^S}) / \sigma_{\mathbf{h}_k^S}$. Besides, we employ label-guided graph reconstruction of RHGC (Ni et al. 2025) during optimization of $I(G^S; Y^S, G^{E,S})$ to obtain the reconstructed adjacency matrix $\hat{A}^S$ of original labeled source graph:

$$\mathcal{L}^2_{\text{dis}} = \|\mathbf{A^S} - \hat{\mathbf{A}}^{\mathbf{S}}\|_2^2, \quad (6)$$

where A^S denotes the adjacency matrix of source graphs. In brief, the objective of disentanglement is as follows:

$$\mathcal{L}_{\text{dis}} = \mathcal{L}^1_{\text{dis}} + \mathcal{L}^2_{\text{dis}}. \quad (7)$$

Compression. Similarly, $I(G; G^I)$ is minimized via its variational upper bound:

$$\mathcal{L}^1_{\text{MI}} = \mathbb{E}_G \left[-\frac{1}{2} \log A^I + \frac{A^I + B^{I^2}}{2|V_G|} \right]. \quad (8)$$

The label-relevant embeddings $\mathbf{z}_{G^I}$ are then obtained via a graph readout function (e.g., max-pooling). To ensure stable subgraph connectivity across both domains, following PGIB (Seo, Kim, and Park 2024), we minimize the batch-wise consistency loss as follows:

$$\mathcal{L}_{\text{con}} = \|\text{Norm}(S_{\text{B}}^\top \mathbf{A}_{\text{B}} S_{\text{B}}) - I_2\|_2^2, \quad (9)$$

where S_{B} and $\mathbf{A}_{\text{B}}$ are batch-level assignments based on the preservation probability and adjacency matrices, and I_2 is the 2×2 identity matrix.

Reconstruction. The reconstruction term maximizes $I(G; G^I, G^E)$ to ensure the disentangled representations retain sufficient information for reconstructing the original graph.

We directly reconstruct the adjacency matrix using the outer product of the disentangled embeddings. The reconstruction loss optimizes the theoretical lower bound through matrix factorization:

$$\mathcal{L}_{\text{recon}} = \|\sigma(\mathbf{z}^E \mathbf{z}^{\mathbf{I}\top}) - \mathbf{A}\|_2^2, \quad (10)$$

where $\mathbf{z}^E$ and $\mathbf{z}^I$ are the label-irrelevant and label-relevant node embeddings respectively, $\sigma(\cdot)$ is the sigmoid function. This formulation ensures that maximizing the reconstruction quality directly corresponds to maximizing the mutual information lower bound.

Cross-Domain. The cross-domain term $I(G^T; G^{I,S}, G^{E,T})$ serves as a complement to the reconstruction term by introducing cross-domain collaboration, where the label-relevant representation $G^{I,S}$ from source domain collaborates with label-irrelevant representation $G^{E,T}$ from target domain for reconstructing target domain graph G^T. While the FUDA setting assumes unlabeled target domain data, our approach leverages the assumption that graphs sharing identical predicted labels exhibit similar structural patterns, enabling effective reconstruction through cross-domain alignment. We

employ soft alignment (Ling et al. 2023) to address misalignment between graphs when reconstruction. Finally, we maximize the cross-domain term through:

$$\mathcal{L}_{cro} = \|\sigma(\mathbf{z}_{G^{I,S}}\mathbf{z}_{G^{E,T}}^{\top}) - \mathbf{A}^{T}\|_2^2. \tag{11}$$

Summary. The objective of DSG is summarized as follows:

$$\begin{aligned}\mathcal{L}_{\mathrm{DSG}} = \mathcal{L}_{\mathrm{cls}} - \alpha_1\mathcal{L}_{\mathrm{dis}} - \alpha_2\mathcal{L}_{\mathrm{MI}}^{1} \\ + \alpha_3\mathcal{L}_{\mathrm{con}} + \alpha_4\mathcal{L}_{\mathrm{recon}} + \alpha_5\mathcal{L}_{\mathrm{cro}},\end{aligned} \tag{12}$$

where $\{\alpha_i\}_{i=1}^{5}$ are trade-off hyperparameters that balance the contributions of different loss components.

Graph-based Prototypical Self-Supervision

Challenge Analysis. Despite various graph domain adaptation methods proposed, they have some fundamental weaknesses. First, most methods adopt instance-to-instance alignment strategies that exhibit high sensitivity to outliers. Consider a case where large domain gaps exist between source and target domains, yet a single anomalous source sample is mapped closer to the target distribution than any legitimate source sample. This leads to unfavorable alignment. Second, they fail to encode the semantics of the data, resulting in suboptimal alignment where semantically similar instances are erroneously dispersed across the latent space.

Method Rationale. To address the above challenge, we propose a novel Graph-based Prototypical Self-Supervision module, which introduces prototypes to encode semantic knowledge based on label-relevant information and performs a novel prototype-based domain alignment method in a self-supervised manner.

Prototype Learning. To ensure that prototypes effectively capture the semantic structure of the graph data, k-means clustering is performed on $\boldsymbol{Z}_{\boldsymbol{G}^{I,S}}$ and $\boldsymbol{Z}_{\boldsymbol{G}^{I,T}}$ to get source clusters $\boldsymbol{C}^S = \{C_1^{(S)}, C_2^{(S)}, \ldots, C_k^{(S)}\}$ and similarly $\boldsymbol{C}^T$ with normalized source prototypes $\{p_j^S\}_{j=1}^k$ and normalized target prototypes $\{p_j^T\}_{j=1}^k$. Prototypes are maintained and updated using momentum-based exponential moving averages during training to capture evolving data distributions while ensuring stability. It should be noted that the aforementioned **Reconstruction Term** and **Cross-Domain Term** can serve as generators to augment data simultaneously, which form additional source domain dataset $\hat{\mathcal{D}^S}$ and target domain dataset $\hat{\mathcal{D}^T}$ that can be used to generate source prototypes $\{\hat{p}_j^S\}_{j=1}^k$ and target prototypes $\{\hat{p}_j^T\}_{j=1}^k$. The original prototypes and additional prototypes are then merged (i.e., averaged) to obtain final source prototypes $\{p_j^{'S}\}_{j=1}^k$ and target prototypes $\{p_j^{'T}\}_{j=1}^k$.

Instance-Prototype Matching. Given label-relevant embeddings from both domains and prototypes, we establish cross-domain alignments through bidirectional instance-prototype matching. Specifically, given a source feature $\mathbf{z}_{G_i^I}^S$ and target prototypes, we first compute the similarity distribution vector $\mathbf{P}_i^{S\to T} = [P_{i,1}^{S\to T}, \ldots, P_{i,k}^{S\to T}]$:

$$P_{i,j}^{S\to T} = \frac{\exp\left(p_j^{'T}\cdot \mathbf{z}_{G_i^{I,S}}/\tau\right)}{\sum_{r=1}^{k}\exp\left(p_r^{'T}\cdot \mathbf{z}_{G_i^{I,S}}/\tau\right)}. \tag{13}$$

Then we minimize the entropy of $\mathbf{P}_i^{S\to T}$:

$$H(\mathbf{P}_i^{S\to T}) = -\sum_{i=1}^{k} P_{i,j}^{S\to T}\log P_{i,j}^{S\to T}. \tag{14}$$

Similarly, we compute $H(\mathbf{P}_i^{T\to S})$ for target-to-source matching. The final cross-domain instance-prototype loss is:

$$\mathcal{L}_{\mathrm{I2P}} = \sum_{i=1}^{N_S} H(\mathbf{P}_i^{S\to T}) + \sum_{j=1}^{N_T} H(\mathbf{P}_j^{T\to S}), \tag{15}$$

where the N_S denotes the sample number of source domain and N_T is the sample number of target domain.

Mutual Information Maximization. For the unlabeled data, we extend the mutual information maximization objective (Liang et al. 2021) to graph data in order to encourage individually certain and globally diverse predictions on source and target domains:

$$\mathcal{L}_{\mathrm{MI}}^{2} = -\mathcal{H}\left(\sum_{i=1}^{N_u} p(y_i|\mathbf{z}_{G_i^I})\right) + \frac{1}{N_u}\sum_{i=1}^{N_u}\mathcal{H}(p(y_i|\mathbf{z}_{G_i^I})), \tag{16}$$

where the entropy metric $\mathcal{H}(p(y_i|\mathbf{z}_{G_i^I})) = -\sum_{c=1}^{K} p(y_i = c|\mathbf{z}_{G_i^I})\log p(y_i = c|\mathbf{z}_{G_i^I})$, N_u denotes the number of all unlabeled samples.

Final Loss Function. Finally, the overall learning objective can be derived as:

$$\mathcal{L} = \mathcal{L}_{\mathrm{DSG}} + \beta_1\mathcal{L}_{\mathrm{I2P}} + \beta_2\mathcal{L}_{\mathrm{MI}}^{2}, \tag{17}$$

where $\{\beta_i\}_{i=1}^{2}$ are two trade-off hyperparameters.

Experiment Setup

Datasets

Due to the lack of datasets containing ground-truth graph-level anomalies, following prior works (Niu, Pang, and Chen 2023), we use graph classification datasets for evaluation. Specifically, we use 4 datasets from **TUDataset** (Morris et al. 2020) including **Mutagenicity** (Kazius, McGuire, and Bursi 2005), **PROTEINS** (i.e., PROTEINS and DD) (Dobson and Doig 2003), **COX2** (i.e., COX2 and COX2_MD) and **BZR** (i.e., BZR and BZR_MD) (Sutherland, O'brien, and Weaver 2003). In these datasets, graphs with class 0 are treated as normal, while graphs with class 1 are considered anomalies.

Baselines

We select baselines from three related tasks: graph-level anomaly detection methods, graph few-shot learning methods and cross-domain graph classification methods. For graph-level anomaly detection methods, we select three

Methods	M0 → M1		M1 → M0		M0 → M2		M2 → M0		M1 → M2		M2 → M1	
	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot
ARMET	51.23	64.71	63.78	66.97	53.35	55.43	52.43	56.08	54.59	57.81	58.21	63.43
iGAD	54.97	63.89	53.15	57.43	53.11	55.96	55.78	59.01	53.60	55.42	59.74	62.83
GmapAD	50.76	61.87	54.76	58.87	54.89	56.20	55.71	56.89	56.82	58.94	56.89	60.98
MGNN	54.71	55.64	53.69	55.56	52.89	54.96	53.82	55.63	51.83	52.52	53.81	59.45
PAR	40.77	44.28	46.90	67.72	49.17	52.03	55.35	58.25	56.41	58.71	61.11	62.72
FAITH	53.01	59.38	54.22	60.40	55.64	56.17	54.54	59.67	50.56	51.02	60.26	63.06
CDTC	40.34	49.74	62.27	64.78	49.24	53.20	56.44	59.88	53.63	55.66	49.52	57.24
DEAL	54.81	61.10	54.22	60.61	51.73	53.24	52.72	57.91	55.12	56.73	52.11	59.72
CoCo	54.32	63.96	55.45	63.31	53.83	56.74	55.98	60.91	51.92	58.39	54.11	60.48
MTDF	53.50	62.74	55.60	65.87	55.61	57.84	53.83	61.90	51.73	56.40	54.05	61.45
Ours	**64.83**	**69.34**	**67.58**	**69.56**	**58.22**	**63.12**	**63.67**	**65.48**	**60.33**	**63.87**	**65.92**	**68.34**

Table 1: Overall performance (AUROC scores) of graph-level anomaly detection under few-shot unsupervised domain adaptation setting on Mutagenicity (source → target). The best results are shown in **bold type** and the runner-ups are underlined.

Methods	P → D		D → P		C → CM		CM → C		B → BM		BM → B	
	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot	5-shot	10-shot
ARMET	45.76	47.02	42.42	59.37	42.16	43.09	40.56	41.73	33.98	39.24	48.63	51.46
iGAD	45.12	47.45	41.68	55.20	41.44	43.68	38.80	40.11	32.08	34.01	48.67	50.89
GmapAD	45.78	46.05	46.56	58.12	40.35	44.61	43.76	45.85	37.24	41.99	44.66	49.57
MGNN	51.10	52.90	48.27	52.69	50.38	52.74	50.45	51.01	50.99	51.99	52.13	56.52
PAR	61.24	64.42	53.98	57.32	51.22	53.42	38.68	42.21	50.31	50.69	51.52	53.43
FAITH	49.41	50.90	40.99	58.56	47.84	49.98	39.23	41.18	31.88	39.90	59.72	62.73
CDTC	57.79	58.54	46.68	52.71	50.11	51.37	42.76	49.14	52.27	52.56	52.54	55.29
DEAL	46.75	47.02	39.31	57.89	42.72	44.03	35.92	38.34	30.53	32.11	41.14	49.13
CoCo	43.23	46.34	40.62	55.03	42.01	44.24	39.32	40.81	30.03	34.72	46.72	50.14
MTDF	47.52	50.13	40.24	56.31	43.12	45.63	40.51	42.42	31.83	35.92	47.31	52.52
Ours	**66.63**	**70.44**	**65.19**	**68.91**	**58.61**	**62.80**	**56.66**	**59.34**	**55.12**	**58.81**	**62.75**	**65.51**

Table 2: Overall performance (AUROC scores) of graph-level anomaly detection under few-shot unsupervised domain adaptation setting on PROTEINS, COX2, and BZR (source → target). Best and runner-up results are showed with **bold** and underline.

baselines including **ARMET** (Li et al. 2024), **GmapAD** (Ma et al. 2023) and **iGAD** (Zhang et al. 2022). For graph few-shot learning methods, we select three models including **MGNN** (Guo et al. 2021), **PAR** (Wang et al. 2021) and **FAITH** (Wang et al. 2022a). For cross-domain graph classification methods, we adopt four baselines including **CDTC** (Zhang et al. 2023b), **DEAL** (Yin et al. 2022), **CoCo** (Yin et al. 2023) and **MTDF** (Tang et al. 2024).

Evaluation Metrics and Settings

Follwing convention (Kim et al. 2024), for GLAD performance, we report the AUROC w.r.t. anomaly scores and labels. We evaluate our model by conducting GLAD task for a variety of transfer learning tasks (i.e., source → target). M, P, D, C, CM, B, and BM denote Mutagenicity, PROTEINS, DD, COX2, COX2_MD, BZR, and BZR_MD, respectively. Mutagenicity dataset is separated into three parts according to the density of edges. We execute experiments under the 2-way 5-shot and 2-way 10-shot settings.

Implementation Details

DGPA is implemented with *PyTorch* 2.1 framework on Ubuntu 22.04. For our model, we adopt GIN as the GNN backbone. We train the model using Adam optimizer with the learning rate set to 0.005. For a fair comparison, we tune the hyper-parameters of both our model and baselines using grid-search. To mitigate the effects of random noise, we report the results from 5 runs with different random seeds.

Results and Analysis

Comparison with State-of-the-arts

We evaluate model performance on GLAD against state-of-the-art baselines. Results are concluded in Table 1 and 2. Here we have the following observations:

(1) DGPA consistently outperforms the baselines on different transfer tasks. Compared with the second-best base-

Methods	M0 → M2		M2 → M0	
	5-shot	10-shot	5-shot	10-shot
Baseline	47.05	50.49	46.43	51.65
DSG	51.74	53.04	54.15	59.36
GPS	54.56	57.16	52.21	58.01
DSG + GPS	**58.22**	**63.12**	**63.67**	**65.48**

Table 3: Ablation study results for M0 → M2 and M2 → M0. The best results are in **bold**, runner-ups are underlined.

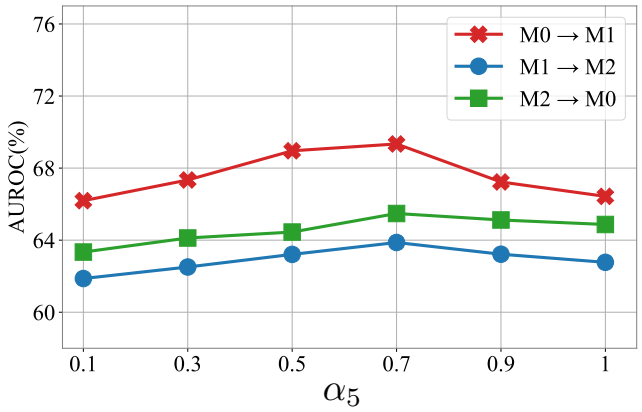

Figure 3: Hyper-parameter analysis results for different values of hyper-parameter α_5 on all datasets.

line, the proposed method achieves 5.72 % improvement of AUROC on average. The results demonstrate that DGPA can effectively address the domain shift challenges.

(2) Single-domain GLAD and few-shot learning methods suffer significant performance degradation in FUDA scenarios due to domain shift challenges. Besides, cross-domain graph learning methods underperform in label-scarce conditions, while DGPA leverages disentangled sample generation to effectively handle few-shot scenarios.

Ablation Study

In this section, an ablation study is conducted on representative transfer tasks (i.e., M0 → M2 and M2 → M0) with different combinations of the key modules. Specifically, 4 combinations of key modules are compared in the ablation study as follows:

- **Baseline:** The basic model of DGPA backbone.
- **DSG:** The baseline model with DSG module.
- **GPS:** The baseline model with GPS module.
- **DSG+GPS:** The proposed DGPA model.

As shown in Table 3, the baseline performs the worst because it cannot cope with domain shift and label scarcity. Besides, the baseline fails to achieve robust representation learning, leading to poor performance on datasets like **MUTAG**, where irrelevant noisy features prevail. Adding **DSG** markedly improves performance by disentangling label-relevant and label-irrelevant factors. Incorporating **GPS** further boosts results via prototypical alignment, enhancing discriminability and transferability. The proposed DGPA, which combines DSG and GPS, fully exploits these advantages and achieves the best performance.

Hyper-parameter Sensitivity

In this section, the hyper-parameter analysis is performed on representative transfer tasks with α_5 set to $\{0.1, 0.3, 0.5, 0.7, 0.9, 1.0\}$, as depicted in Fig. 3. It is observed that our method achieves the best performance when $\alpha_5 = 0.7$.

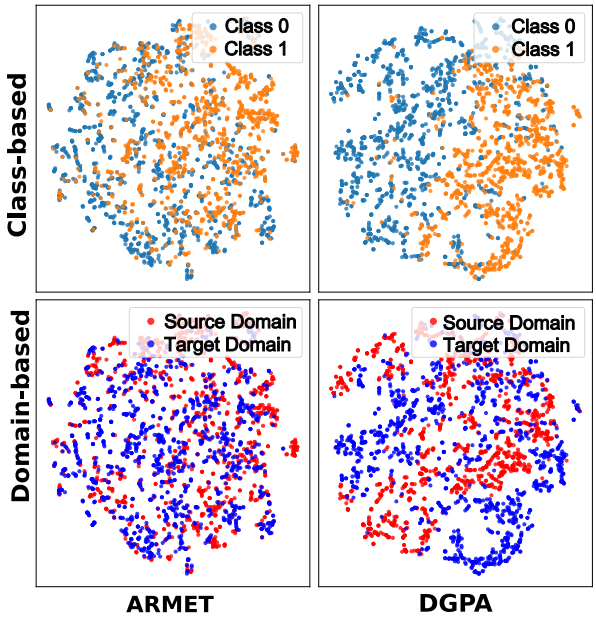

Figure 4: t-SNE visualization results of ours and baselines on M1→M0.

Visualization

We visualize the learned features using t-SNE (Maaten and Hinton 2008) on the M1→M0, as shown in Fig. 4. In the top row, the color represents the class of each sample; while in the bottom row, red represents source samples and blue represents target samples. Compared to ARMET, it qualitatively shows that DGPA favors more discriminative features. Also, the features from DGPA are more closely aggregated than ARMET, which demonstrates that DGPA learns a better semantic structure of the datasets.

Conclusion

This paper presents a novel few-shot unsupervised domain adaptation method named DGPA for graph-level anomaly detection. We introduce DSG and GPS modules to address ***label scarcity*** and ***domain shift*** through disentangled generation and prototype-based cross-domain alignment. Extensive experiments on four benchmarks demonstrate DGPA's significant superiority over state-of-the-art baselines in GLAD under FUDA settings. In the future, we plan to explore various advanced techniques to further improve model performance, such as prompt learning and adversarial learning.

Acknowledgments

This work was partially supported by the Science and Technology Innovation Project of Hunan Province (No. 2023RC4014) and the NSFC (No. 6212780016).

References

Benigmim, Y.; Roy, S.; Essid, S.; Kalogeiton, V.; and Lathuilière, S. 2023. One-shot unsupervised domain adaptation with personalized diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 698–708.

Chen, J.; Fu, S.; Zhang, Z.; Ma, Z.; Feng, M.; Wirjanto, T.; and Peng, Q. 2024. Towards Cross-Domain Few-Shot Graph Anomaly Detection. In *ICDM*, 51–60.

Dobson, P. D.; and Doig, A. J. 2003. Distinguishing enzyme structures from non-enzymes without alignments. *Journal of molecular biology*, 330(4): 771–783.

Fan, Y.; Ju, M.; Hou, S.; Ye, Y.; Wan, W.; Wang, K.; Mei, Y.; and Xiong, Q. 2021. Heterogeneous temporal graph transformer: An intelligent system for evolving android malware detection. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, 2831–2839.

Guo, Z.; Zhang, C.; Yu, W.; Herr, J.; Wiest, O.; Jiang, M.; and Chawla, N. V. 2021. Few-shot graph learning for molecular property prediction. In *Proceedings of the web conference 2021*, 2559–2567.

Kazius, J.; McGuire, R.; and Bursi, R. 2005. Derivation and validation of toxicophores for mutagenicity prediction. *Journal of medicinal chemistry*, 48(1): 312–320.

Kim, S.; Lee, S. Y.; Bu, F.; Kang, S.; Kim, K.; Yoo, J.; and Shin, K. 2024. Rethinking reconstruction-based graph-level anomaly detection: Limitations and a simple remedy. *Advances in Neural Information Processing Systems*, 37: 95931–95962.

Li, Z.; Liang, S.; Shi, J.; and van Leeuwen, M. 2024. Cross-domain graph level anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*.

Liang, J.; Hu, D.; Wang, Y.; He, R.; and Feng, J. 2021. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE Trans. Pattern Anal. Mach. Intell.*

Ling, H.; Jiang, Z.; Liu, M.; Ji, S.; and Zou, N. 2023. Graph mixup with soft alignments. In *International Conference on Machine Learning*, 21335–21349. PMLR.

Ma, R.; Pang, G.; Chen, L.; and van den Hengel, A. 2022. Deep graph-level anomaly detection by glocal knowledge distillation. In *Proceedings of the fifteenth ACM international conference on web search and data mining*, 704–714.

Ma, X.; Wu, J.; Xue, S.; Yang, J.; Zhou, C.; Sheng, Q. Z.; Xiong, H.; and Akoglu, L. 2021. A comprehensive survey on graph anomaly detection with deep learning. *IEEE transactions on knowledge and data engineering*, 35(12): 12012–12038.

Ma, X.; Wu, J.; Yang, J.; and Sheng, Q. Z. 2023. Towards graph-level anomaly detection via deep evolutionary mapping. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 1631–1642.

Maaten, L. v. d.; and Hinton, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research*, 9(Nov): 2579–2605.

Morris, C.; Kriege, N. M.; Bause, F.; Kersting, K.; Mutzel, P.; and Neumann, M. 2020. TUDataset: A collection of benchmark datasets for learning with graphs. In *ICML 2020 Workshop on Graph Representation Learning and Beyond (GRL+ 2020)*.

Ni, Z.; Liu, C.; Wan, H.; and Zhao, X. 2025. Robust Heterogeneous Graph Classification for Molecular Property Prediction with Information Bottleneck. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 640–648.

Niu, C.; Pang, G.; and Chen, L. 2023. Graph-level anomaly detection via hierarchical memory networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 201–218. Springer.

Qiu, C.; Kloft, M.; Mandt, S.; and Rudolph, M. 2022. Raising the bar in graph-level anomaly detection. *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*.

Seo, S.; Kim, S.; and Park, C. 2024. Interpretable prototype-based graph information bottleneck. *Advances in Neural Information Processing Systems*, 36.

Shao, Z.; Xi, H.; Hensher, D. A.; Wang, Z.; Gong, X.; and Gao, J. 2025a. A spatial–temporal dynamic attention-based Mamba model for multi-type passenger demand prediction in multimodal public transit systems. *Transportation Research Part E: Logistics and Transportation Review*, 202: 104282.

Shao, Z.; Xi, H.; Lu, H.; Wang, Z.; Bell, M. G.; and Gao, J. 2025b. A spatial–Temporal Large Language Model with Denoising Diffusion Implicit for predictions in centralized multimodal transport systems. *Transportation Research Part C: Emerging Technologies*, 179: 105249.

Shen, X.; Dai, Q.; Chung, F.-l.; Lu, W.; and Choi, K.-S. 2020. Adversarial deep network embedding for cross-network node classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, 2991–2999.

Sutherland, J. J.; O'brien, L. A.; and Weaver, D. F. 2003. Spline-fitting with a genetic algorithm: A method for developing classification structure- activity relationships. *Journal of chemical information and computer sciences*, 43(6): 1906–1915.

Tang, Y.; Luo, J.; Yang, L.; Luo, X.; Zhang, W.; and Cui, B. 2024. Multi-view teacher with curriculum data fusion for robust unsupervised domain adaptation. In *2024 IEEE 40th International Conference on Data Engineering (ICDE)*, 2598–2611. IEEE.

Wang, Q.; Pang, G.; Salehi, M.; Buntine, W.; and Leckie, C. 2023. Cross-domain graph anomaly detection via anomaly-aware contrastive alignment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 4676–4684.

Wang, S.; Dong, Y.; Huang, X.; Chen, C.; and Li, J. 2022a. FAITH: Few-Shot Graph Classification with Hierarchical Task Graphs. In *IJCAI*.

Wang, Y.; Abuduweili, A.; Yao, Q.; and Dou, D. 2021. Property-aware relation networks for few-shot molecular property prediction. *Advances in Neural Information Processing Systems*, 34: 17441–17454.

Wang, Y.; Wang, J.; Cao, Z.; and Barati Farimani, A. 2022b. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4(3): 279–287.

Xi, H.; Shao, Z.; Hensher, D. A.; Nelson, J. D.; Chen, H.; and Wijayaratna, K. 2025. A multi-task Transformer with mixture-of-experts for personalized periodic predictions of individual travel behavior in multimodal public transport. *Transportation Research Part C: Emerging Technologies*, 179: 105287.

Xu, B.; Zeng, Z.; Lian, C.; and Ding, Z. 2022. Few-shot domain adaptation via mixup optimal transport. *IEEE Transactions on Image Processing*, 31: 2518–2528.

Yin, N.; Shen, L.; Li, B.; Wang, M.; Luo, X.; Chen, C.; Luo, Z.; and Hua, X.-S. 2022. Deal: An unsupervised domain adaptive framework for graph-level classification. In *Proceedings of the 30th ACM International Conference on Multimedia*, 3470–3479.

Yin, N.; Shen, L.; Wang, M.; Lan, L.; Ma, Z.; Chen, C.; Hua, X.-S.; and Luo, X. 2023. CoCo: A coupled contrastive framework for unsupervised domain adaptive graph classification. In *International Conference on Machine Learning*, 40040–40053. PMLR.

Zhang, G.; Yang, Z.; Wu, J.; Yang, J.; Xue, S.; Peng, H.; Su, J.; Zhou, C.; Sheng, Q. Z.; Akoglu, L.; et al. 2022. Dual-discriminative graph neural network for imbalanced graph-level anomaly detection. *Advances in Neural Information Processing Systems*, 35: 24144–24157.

Zhang, J.; Chao, H.; Dhurandhar, A.; Chen, P.-Y.; Tajer, A.; Xu, Y.; and Yan, P. 2023a. Spectral adversarial mixup for few-shot unsupervised domain adaptation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 728–738. Springer.

Zhang, Q.; Pei, S.; Yang, Q.; Zhang, C.; Chawla, N. V.; and Zhang, X. 2023b. Cross-domain few-shot graph classification with a reinforced task coordinator. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, 4893–4901.