DisCo: Diffusion-guided Unbiased Discriminative Learning for Unsupervised Graph Domain Adaptation

Haodong Zhang
2110496@stu.neu.edu.cn
Northeastern University
Shenyang, China

Tao Ren
rent@swc.neu.edu.cn
Northeastern University
Shenyang, China

Changhu Wang
cwang3@fredhutch.org
Fred Hutchinson Cancer Research Center
Seattle, United States

Yifan Wang*
yifanwang@uibe.edu.cn
University of International Business and Economics
Beijing, China

Wei Ju
juwei@pku.edu.cn
Peking University
Beijing, China

Huaizhi Tang
2110495@stu.neu.edu.cn
Northeastern University
Shenyang, China

Junyu Luo
luojunyu@stu.pku.edu.cn
Peking University
Beijing, China

Zimo Wang
zimo.wang@bayims.cn
Beijing Bayi School
Beijing, China

Ziyue Qiao
zyqiao@gbu.edu.cn
Great Bay University
Dongguan, China

Xian-Sheng Hua
huaxiansheng@gmail.com
Tongji University
Shanghai, China

Xiao Luo
xiao.luo@wisc.edu
University of Wisconsin-Madison
Madison, United States

Abstract

This paper investigates the task of unsupervised graph domain adaptation, which facilitates the transfer of knowledge from labeled source graphs to unlabeled target graphs. Recent approaches usually utilize graph contrastive learning and pseudo-labeling to learn from unlabeled target data, which could introduce potential biased representations and supervision of target graphs resulting from serious shifts across two domains. Towards this end, we propose a novel framework named Diffusion-guided Unbiased Discriminative Learning (DisCo) for unsupervised graph domain adaptation. The core of our DisCo is to leverage both feature disentanglement and cross-domain diffusion signals to remove the potential biases for target graphs. In particular, we first utilize adversarial feature disentanglement to extract causal features that are orthogonal to domain biases. More importantly, we retrieve the labels of cross-domain source graphs to generate the conditions, which would be utilized to optimize a diffusion model for label denoising. The consistency between pseudo-labels and denoised labels is measured to reduce the potential biases during domain alignment. Extensive experiments on several real-world benchmarks demonstrate that our proposed DisCo consistently outperforms competing state-of-the-art baselines.

*Corresponding author.

KDD '26, Jeju Island, Republic of Korea

ACM ISBN 979-8-4007-2258-5/2026/08
https://doi.org/10.1145/3770854.3780287

CCS Concepts

• **Mathematics of computing** → **Graph algorithms**; • **Computing methodologies** → **Neural networks**; **Unsupervised learning**; **Transfer learning**.

Keywords

Graph Neural Networks, Domain Adaptation, Causal Discovery, Diffusion Model

ACM Reference Format:
Haodong Zhang, Tao Ren, Changhu Wang, Yifan Wang, Wei Ju, Huaizhi Tang, Junyu Luo, Zimo Wang, Ziyue Qiao, Xian-Sheng Hua, and Xiao Luo. 2026. DisCo: Diffusion-guided Unbiased Discriminative Learning for Unsupervised Graph Domain Adaptation. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.1 (KDD '26), August 09–13, 2026, Jeju Island, Republic of Korea.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3770854.3780287

Resource Availability:
The source code of this paper has been made publicly available at https://doi.org/10.5281/zenodo.18096835.

1 Introduction

Graph Neural Networks (GNNs) have achieved remarkable success and become the *de facto* approach for learning from graph-structured data, including molecular structures [56, 70], biological networks [41], social networks [6] and citation networks [42, 55]. As a fundamental GNN-based task, graph classification endeavors to infer the characteristics of the entire graph and has attracted

considerable interest in recent years [53, 71]. The core of these methods lies in the message-passing paradigm [8, 65], which enables each node to iteratively aggregate information from its local neighborhood. Through multiple layers of aggregation and a readout operator, GNNs produce global graph representations, facilitating subsequent classification objectives.

Despite certain progress made by these methods, they often rely on the assumption that training and test graphs are drawn from the same distribution. This assumption is frequently violated in real-world scenarios, where graph data collected from different domains may exhibit significant variations [29, 61]. Meanwhile, the lack of labeled data in the target domain further exacerbates the challenge, making fully supervised learning approaches infeasible [12, 44]. Unsupervised domain adaptation (UDA) has emerged as a promising solution, aiming to leverage knowledge from a labeled source domain to facilitate effective task performance in an unlabeled target domain [20, 35, 40].

Actually, there are several approaches that apply UDA to Euclidean data (i.e., images) [59]. Among them, distance-based methods focus on minimizing statistical divergence, i.e., maximum mean discrepancy [66] and Wasserstein distance [24], between the source and target feature distributions to align two domains. Self-supervised learning methods generate pseudo-labels to facilitate knowledge transfer [14, 57]. Adversarial learning methods introduce a domain discriminator and encourage the encoder to learn domain-invariant features to confuse the discriminator [7, 28]. Drawing inspiration from these advancements, recent efforts adopt the same paradigms for graph structure data. Global alignment methods [3, 37, 67] explicitly minimize the distribution differences across the source and target domains. More recent studies [31, 46] incorporate contrastive and self-supervised learning to preserve the intrinsic topological and semantic information of graphs during adaptation.

Despite the great success of these methods, the presence of noise within target data, arising from domain shifts, poses two significant challenges: ❶ *Biased representation across domain.* Unlike Euclidean data, graph-structured inputs exhibit intricate relational dependencies that often lead to entangled node representations where task-relevant information is mixed with spurious, domain-specific correlations [9, 27]. Such a bias in the learned representations prevents the model from effectively capturing task-relevant factors and thereby limits its transferability across domains. ❷ *Limited and biased supervision in the target domain.* In typical UDA scenarios, the target domain lacks ground-truth labels, and the generated pseudo-labels are often noisy and biased due to the presence of inherent distribution shift across domains [19]. This unreliable supervision not only misguides the model during training but also reinforces incorrect domain alignment, leading to suboptimal discriminative feature learning and domain adaptation performance.

Toward this end, we propose DisCo, a **DI**ffusion-guided unbia**S**ed dis**C**riminative learning for unsupervised graph d**O**main adaptation, which facilitates unbiased feature and label learning by explicitly disentangling causal and spurious factors with adversarial training and guides the learning of domain-invariant representations with diffusion signals. On the one hand, given the encoded representation of the graph, we disentangle them into causal and spurious parts through adversarial training, where causal features are encouraged to be domain-invariant, while the spurious ones retain label-irrelevant variations. On the other hand, based on the generated pseudo-labels of the target domain, we reformulate the unbiased adaptation process from a generative-mode perspective and incorporate diffusion-guided label refinement for progressive domain alignment.

We summarize the contributions of this paper as follows:

- *New Perspective:* We highlight the spurious representation bias and label supervision bias in the target domain, proposing a framework for unbiased graph domain adaptation.
- *Novel Methodology:* We introduce DisCo, which disentangles stable causal and spurious features and leverages a generative formulation with diffusion-guided pseudo-label refinement for iterative alignment.
- *Extensive Experiments:* We perform comprehensive experiments to evaluate the efficacy of our DisCo. The empirical results confirm that the our framework consistently outperforms the baselines.

2 Related Work

2.1 Unsupervised Graph Domain Adaptation

Unsupervised graph domain adaptation (UGDA) focuses on leveraging labeled source graphs to facilitate knowledge transfer to an unlabeled target domain [61, 67, 72], yet remains a challenging task due to the non-Euclidean nature of graph data. Prevailing approaches fall into two main streams: (1) Global alignment approaches [3, 61, 63] that leverage adversarial learning to minimize domain discrepancy but risk collapsing informative graph structures while preserving spurious factors. (2) Self-training approaches [31, 46, 60] which utilize pseudo-labels to provide supervision for the target domain, yet remain vulnerable to error propagation and confirmation bias. For example, SPA [63] aligns domain graphs in eigenspaces for inter-domain transferability. CoCo [67] leverages hierarchical graph kernel to explore the topological structure in the target domain with contrastive learning paradigm. MTDF [46] aligns representations with a multi-teacher framework and bridges the domain shift with generated data. However, a central challenge of these approaches is that graph data are inherently entangled with spurious correlations, which introduces representation bias, and the reliance on pseudo-labels in the target domain, which further amplifies label bias and leads to unstable adaptation performance. Therefore, in this paper, we explicitly disentangle causal and spurious features and introduce pseudo-label refinement for unbiased graph domain adaptation.

2.2 Representation Disentanglement and Causal Discovery

Representation disentanglement and causal discovery are fundamental to stable representation learning, grounded in causal inference and invariance theory [39, 49]. Recent advances in graph learning integrate causal principles with GNNs to disentangle spurious correlations from underlying graph structures for more robust performance [2, 25, 27]. For example, RGCL [27] discovers salient graph features to create rationale-aware augmentations for effective contrastive learning. GIL [25] identifies invariant subgraphs and infers latent environments from variant ones to learn generalizable

graph representations. By leveraging representation disentanglement, causal learning is facilitated to separate domain-invariant causal factors from spurious correlations [18, 30, 69]. Inspired by causal learning, our work incorporates disentanglement learning with adversarial domain discrimination, fostering effective domain-independent causal knowledge for graph domain adaptation.

2.3 Diffusion-guided Data Recalibration

Data recalibration is a primary strategy for learning with noisy data, enhancing model robustness by progressively correcting mislabeled data or identifying clean samples [26, 52, 73]. For example, DivideMix [26] separates clean and noisy data using a Gaussian mixture model, while CC [73] relies on feature space centrality and consistency. C2D [74] further improves upon DivideMix by using self-supervised learning to enhance feature quality. A parallel research stream involves guided diffusion models, which fall into two groups. (1) Classifier guidance [4] leverages classifier gradients to steer the diffusion model's generation. (2) Classifier-free guidance, in contrast, learns the conditional distribution directly during training to improve generation quality [11, 16]. CARD [11], for example, reframes regression or classification as a conditional task, generating labels or target variables based on an image. Building on this concept, LRA [1] generates pseudo-clean labels by retrieving neighbor labels within a robust, pre-trained feature space. This approach decouples the label correction signal from the main learning process, thereby mitigating confirmation bias. In our work, we integrate this diffusion-guided label correction strategy into the pseudo-label selection process to generate stable supervisory signals for unbiased adaptation.

3 Preliminary

3.1 Problem Definition

We define a graph as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, with a node set $\mathcal{V}$ and an edge set $\mathcal{E}$. The topology of $\mathcal{G}$ is captured by an adjacency matrix $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, with entries $\boldsymbol{A}_{uv} = 1$ denoting an edge $(u, v) \in \mathcal{E}$, and 0 otherwise. Additionally, the graph is characterized by a feature matrix $\boldsymbol{X} \in \mathbb{R}^{N\times d}$, containing d-dimensional feature vectors $\boldsymbol{X}_v \in \mathbb{R}^d$ for each node $v \in \mathcal{V}$. We are given a labeled source domain $\mathcal{D}^{so} = \{(\mathcal{G}_i^{so}, y_i^{so})\}_{i=1}^{N_{so}}$ consisting of N_{so} graphs $\mathcal{G}_i^{so}$ with its corresponding label y_i^{so} and an unlabeled target domain $\mathcal{D}^{ta} = \{\mathcal{G}_i^{ta}\}_{i=1}^{N_{ta}}$ containing N_{ta} graphs without label. In our setting, both $\mathcal{D}^{so}$ and $\mathcal{D}^{ta}$ share an identical label space $\mathcal{Y} = \{1, \ldots, C'\}$, while exhibiting distinct data distributions. The primary objective of unsupervised graph domain adaptation is to accurately label the target domain graphs using unbiased transferable knowledge from the source domain.

3.2 Diffusion Models

Diffusion models, especially Denoising Diffusion Probabilistic Model (DDPM) [15], represent a category of generative frameworks that generate data by inverting a progressive noise addition step. Specifically, for an input sample $\boldsymbol{y}_0$ drawn from distribution $q(\boldsymbol{y}_0)$, the forward diffusion process injects Gaussian noise step-by-step via a Markov chain over T steps, producing a sequence of latent variables $\{\boldsymbol{y}_t\}_{t=1}^T$, which are governed by:

$$q(\boldsymbol{y}_t|\boldsymbol{y}_{t-1}) := \mathcal{N}(\boldsymbol{y}_t; \sqrt{1-\beta_t}\boldsymbol{y}_{t-1}, \beta_t \boldsymbol{I}), \tag{1}$$

where $\beta_t \in (0, 1)$ manages the noise level at step t and the endpoint of the process is set to $p(\boldsymbol{y}_T) := \mathcal{N}(0, I)$. In practice, the forward sampling process can be simplified by $q(\boldsymbol{y}_t|\boldsymbol{y}_0) := \mathcal{N}(\boldsymbol{y}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{y}_0, (1-\bar{\alpha}_t)\boldsymbol{I})$, where $\alpha_t := 1 - \beta_t$ and $\bar{\alpha}_t := \prod_{i=1}^t \alpha_i$. The reverse denoising process is learned via a parameterized model $p_\theta(\boldsymbol{y}_{t-1}|\boldsymbol{y}_t)$ to eliminate the added noise in the forward process:

$$p_\theta(\boldsymbol{y}_{t-1}|\boldsymbol{y}_t) := \mathcal{N}(\boldsymbol{\mu}_\theta(\boldsymbol{y}_t, t), \Sigma_\theta(\boldsymbol{y}_t, t)), \tag{2}$$

where $\boldsymbol{\mu}_\theta(\boldsymbol{y}_t, t) := \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{y}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta)$ is parameterized by a neural network for the denoising function $\boldsymbol{\epsilon}_\theta$ and $\Sigma_\theta(\boldsymbol{y}_t, t) := \tilde{\beta}_t \boldsymbol{I} = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t \boldsymbol{I}$ is often fixed. Therefore, the parameter θ can be trained by the objective:

$$\mathcal{L}_\epsilon = \mathbb{E}_{\boldsymbol{y}_0\sim q(\boldsymbol{y}_0), \boldsymbol{\epsilon}\sim\mathcal{N}(0,\boldsymbol{I}), t} ||\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boldsymbol{y}_t, t)||^2. \tag{3}$$

Note that given the condition $\boldsymbol{c}$, we can instead replace the $\boldsymbol{\epsilon}_\theta(\boldsymbol{y}_t, t)$ with $\boldsymbol{\epsilon}_\theta(\boldsymbol{y}_t, t|\boldsymbol{c})$ for the denoising function.

4 Methodology

4.1 Framework Overview

In this paper, we introduce DisCo, which utilizes diffusion guidance to facilitate unbiased discriminative learning in unsupervised graph domain adaptation tasks. The core concept of our approach is to alleviate the negative transfer knowledge effects arising from spurious correlations and biased supervision in the target graph domain. Given the labeled source domain and the unlabeled target domain, we first employ a GNN-based encoder to extract informative representations from graph-structured data. Then, we perform adversarial feature disentanglement to decompose the learned representations into two complementary parts: causal components, which capture domain-invariant semantics, and spurious components, which preserve label-irrelevant variations. Building upon this, we generate the pseudo-label for the target domain and propose a conditional diffusion model that iteratively refines the noisy pseudo-labels for progressive domain alignment. Figure 1 illustrates the overview of our proposed DisCo. More details are introduced in the following sections.

4.2 GNN-based Encoder

Generalized graph representations are crucial for effective domain adaptation. To extract transferable patterns across domains, we utilize a GNN-based encoder to extract both the topological structure and semantic attributes of each graph. For the input graph of source and target domain $\mathcal{G}_i^* = \{\mathcal{V}_i^*, \mathcal{E}_i^*\}, * \in \{so, ta\}$, the encoder aggregates node features through message passing, which can be defined as:

$$\boldsymbol{h}_v^{*,(l)} = C^{*,(l)}\left(\boldsymbol{h}_v^{*,(l-1)}, \mathcal{A}^{*,(l)}\left(\left\{\boldsymbol{h}_u^{*,(l-1)}\right\}_{u\in\mathcal{N}(v)}\right)\right), \tag{4}$$

where $\mathcal{N}(v)$ is the neighbors of node v. $\mathcal{A}(\cdot)$ and $C(\cdot)$ denote the aggregation and combination function at l-th layer. After L layers of message passing, we apply a global readout function to obtain the graph-level representation $\boldsymbol{z}$. Then, the graph prediction label

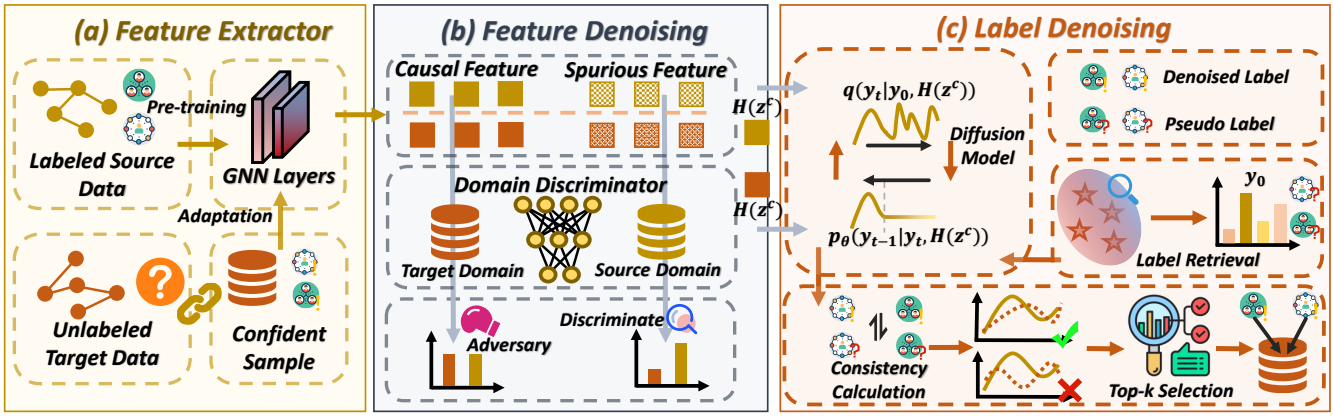


Figure 1: An overview of our DisCo. (1) A GNN-based encoder to extract informative information. (2) An adversarial feature disentanglement module to capture unbiased causal factors. (3) A diffusion-guided label denoising module to refine the noisy pseudo-labels.

is output through a multi-layer perceptron (MLP) classifier $H(\cdot)$, which can be calculated as:

$$z_i^* = \text{READOUT}\left(\left\{\boldsymbol{h}_v^{*,(L)}\right\}_{v\in\mathcal{V}_i}\right), \hat{\boldsymbol{y}}_i^* = H(z_i^*). \tag{5}$$

The model can be trained on the source domain graphs:

$$\mathcal{L}_{so} = -\frac{1}{\mathcal{D}^{so}} \sum_{\mathcal{G}_i^{so}\in\mathcal{D}^{so}} \log(\hat{\boldsymbol{y}}_i^{so}[y_i^{so}]). \tag{6}$$

4.3 Adversarial Feature Disentanglement for Unbiased Causal Discovery

To enable stable domain adaptation, we perform feature disentanglement based on the constructed causal graph, which uncovers domain-invariant causal factors and isolates domain-specific spurious signals for unbiased [32, 40].

Causal Graph Construction. As illustrated in Figure 2, we formalize dependencies between variables as a Structure Causal Model (SCM) [32, 38].

- $C \to \mathcal{G} \leftarrow S$. The observed graph data $\mathcal{G}$ is generated by both causal variables C and spurious variables S.
- $Y \leftarrow C \to PY$. The pseudo-label PY from the target domain and the true label Y from the source domain are determined by the domain-invariant causal variable C.
- $\mathcal{D}^{so} \to S \leftarrow \mathcal{D}^{ta}$. The spurious variable S are shaped by domain-specific factors.

The dashed arrow between C and S denotes the spurious correlation, which provides a backdoor path $S \leftarrow C \to Y$, making S and Y spuriously correlated.

Adversarial Feature Disentanglement for Unbiased Representations. Following the principle of causal theory [32, 62], a variable Y in an SCM has directed edges from its parent variables $PA(Y)$ to Y, if and only if there is a causal mechanism $Y = H(PA(Y), \epsilon_Y)$, where $\epsilon_Y \perp\!\!\!\perp PA(Y)$ is the exogenous noise of Y, represented as:

$$Y = H(C), Y \perp\!\!\!\perp S|C, \tag{7}$$

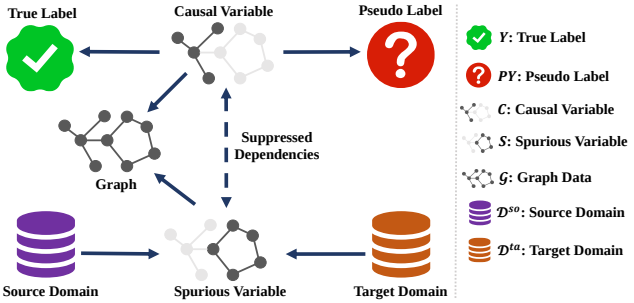


Figure 2: Causal graph for UGDA. Our objective is to train a model capable of disregarding the spurious variables S learned from the source domain $\mathcal{D}^{so}$. It is designed to focus on distilling the causal variables C that constitute the domain-invariant attributes. By leveraging the observed graph data $\mathcal{G}$, the model can ultimately achieve accurate classification within a novel target domain $\mathcal{D}^{ta}$.

where $Y \perp\!\!\!\perp S|C$ indicates that C shields effect of S on Y. In our framework, we suppress dependencies between C and S to eliminate the unintended association between the spurious features S and the label Y (or pseudo-label PY) for the stable adaptation. The objective can be formulated as:

$$\max \underbrace{I(Z^s;\mathcal{D}^*)}_{\text{Domain Specific}} - \underbrace{I(Z^c;\mathcal{D}^*)}_{\text{Domain Invariant}} - \underbrace{\beta I(Z^s;Z^c)}_{\text{Feature Disentanglement}}, \tag{8}$$

where $I(\cdot,\cdot)$ denotes the mutual information between two variables and $* \in \{so, ta\}$. In practice, we employ a domain discriminator $T(\cdot)$ with parameter ψ to identify the domain origin of the extracted features:

$$\mathcal{L}_{AL}^* = -\sum_{i=1}^{N_{so}+N_{ta}} [d_i \log T(z_i^*) + (1-d_i)\log(1-T(z_i^*))], \tag{9}$$

where d_i denotes the domain label for the graph $\mathcal{G}_i$, namely $d_i = 1$ when z_i from the source domain, otherwise $d_i = 0$. For $* \in \{s, c\}$, the first two terms of the objective can be:

$$\begin{cases} \min_{\psi} \mathcal{L}^s_{AL} + \mathcal{L}^c_{AL}, \\ \max_{\phi} \mathcal{L}^c_{AL}, \end{cases} \tag{10}$$

where ϕ denotes other parameters of the framework besides ψ. Since distance correlation is able to characterize the dependence between two variables, we here formalize the last term of the objective as:

$$\begin{aligned} \mathcal{L}_{ind} &= dCor(Z^c, Z^s) \\ dCor(Z^c, Z^s) &= \frac{dCov(Z^c, Z^s)}{\sqrt{dVar(Z^c) \cdot dVar(Z^s)}}, \end{aligned} \tag{11}$$

where $Z^* = [z^*_1, \ldots, z^*_{|\mathcal{D}^{so}|+|\mathcal{D}^{ta}|}], * \in \{s, c\}$ denotes the packed causal and spurious feature matrix. $dVar(\cdot)$ and $dCov(\cdot)$ denote the matrix variance of each matrix and the distance covariance between two matrices [51].

4.4 Diffusion-guided Learning Denoising with Cross-domain Retrieval

Since the target domain always suffers from label scarcity, we introduce an unbiased discrimination learning mechanism, which generates corresponding pseudo-labels for the target domain graphs and develops a retrieval-augmented diffusion model for progressive label denoising [1, 17, 23].

Cross-domain Retrieval Label Conditioning Generation. We combat the paucity of labels in target domain graphs by assigning pseudo-labels based on the identified causal features:

$$\boldsymbol{y}^{ta}_i = H(z^c_i),\ \mathcal{G}^{ta}_i \in \mathcal{D}^{ta}. \tag{12}$$

Then we assume that graph samples with different classes always form distinct clusters and encourage the neighbor consistency of the generated pseudo-labels. Specifically, we calculate the distance between the identified causal features from the target domain graphs and retrieve the pseudo-labels of k-nearest neighbors in source domain graphs as conditioning. This process can be formulated as:

$$\begin{aligned} r_k(z^c_i) &:= \inf\{r : |\mathcal{B}(z^c_i, r) \cap \mathcal{D}^{so}| \geq k\}, \\ \mathcal{B}(z^c_i, r) &:= \{(z^c_j, \boldsymbol{y}^{so}_j) \in \mathcal{D}^{so} : \mathrm{Dis}(z^c_i, z^c_j) \leq r\}, \end{aligned} \tag{13}$$

where $\mathrm{Dis}(\cdot, \cdot)$ denotes the distance function between two feature embeddings and $r_k(z^c_i)$ is the radius for the retrieved neighbors. The retrieved pseudo-labels can be:

$$C_k(z^c_i) := \mathcal{B}(z^c_i, r_k(z^c_i)) \cap \mathcal{D}^{so}. \tag{14}$$

We sample label $\hat{\boldsymbol{y}}^{ta}_j$ from the $\boldsymbol{y}^{ta}_i \cup \{\boldsymbol{y}^c_j | \boldsymbol{y}^c_j \in C_k(z^c_i)\}$ and convert it to the one-hot vector $\boldsymbol{y}_0$ as the conditioning label.

Diffusion-guided Learning Denoising for Domain Alignment. Since the generated pseudo-labels are often noisy and overconfident, we develop an unbiased discrimination learning mechanism, which treats the label denoising as a stochastic process of conditional label generation. Different from the vanilla diffusion model, we assume the endpoint distribution as $p(\boldsymbol{y}_T | z^c) = \mathcal{N}(H(z^c), \boldsymbol{I})$. Therefore, the conditional distribution in the forward process is:

$$\begin{aligned} q(\boldsymbol{y}_t | \boldsymbol{y}_{t-1}, H(z^c)) := \mathcal{N}(\boldsymbol{y}_t; \sqrt{1-\beta_t}\boldsymbol{y}_{t-1} + \\ (1 - \sqrt{1-\beta_t})H(z^c), \beta_t \boldsymbol{I}). \end{aligned} \tag{15}$$

The forward sampling for an arbitrary step t can then be:

$$\begin{aligned} q(\boldsymbol{y}_t | \boldsymbol{y}_0, H(z^c)) := \mathcal{N}(\boldsymbol{y}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{y}_0 + \\ (1 - \sqrt{\bar{\alpha}_t})H(z^c), (1 - \bar{\alpha}_t)\boldsymbol{I})). \end{aligned} \tag{16}$$

Thus, we can parameterize the reverse transition step:

$$p_\theta(\boldsymbol{y}_{t-1} | \boldsymbol{y}_t, H(z^c)) := \mathcal{N}(\boldsymbol{\mu}_\theta(\boldsymbol{y}_t, H(z^c), t), \tilde{\beta}_t \boldsymbol{I}), \tag{17}$$

where the mean term is constructed as $\boldsymbol{\mu}_\theta(\boldsymbol{y}_t, H(z^c), t) := \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{y}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\boldsymbol{y}_t, H(z^c), t))$. The objective can be:

$$\mathcal{L}_\epsilon = \mathbb{E}_{\boldsymbol{y}_0 \sim q(\boldsymbol{y}_0), \boldsymbol{\epsilon} \sim \mathcal{N}(0, \boldsymbol{I}), t} ||\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boldsymbol{y}_t, H(z^c), t)||^2. \tag{18}$$

Note that with the trained diffusion model, we can predict the debiased label of $\mathcal{G}^{ta}_i \in \mathcal{D}^{ta}$ with much fewer steps t' following the pre-defined sampling trajectory:

$$\begin{aligned} \tilde{\boldsymbol{y}}^{ta}_i := \tilde{\boldsymbol{y}}_0 = \frac{1}{\sqrt{\bar{\alpha}_{t'}}}[\boldsymbol{y}_\tau - (1 - \sqrt{\bar{\alpha}_{t'}})H(z^c_i) - \\ \sqrt{1 - \bar{\alpha}_{t'}}\boldsymbol{\epsilon}_\theta(\boldsymbol{y}_{t'}, H(z^c_i), \tau)], \end{aligned} \tag{19}$$

where $\boldsymbol{y}_{t'}$ is computed in the forward diffusion process. We leverage the consistency between pseudo-labels and their denoised labels to select the confident clean samples, which are subsequently trained for progressive domain alignment:

$$\mathcal{L}_{ta} = -\frac{1}{\mathcal{D}^{ta}} \sum_{\mathcal{G}^{ta}_i \in \mathcal{D}^{ta}} \mathbb{1}(S(\boldsymbol{y}^{ta}_i, \tilde{\boldsymbol{y}}^{ta}_i) > \xi) \log(\tilde{\boldsymbol{y}}^{ta}_i[y^{ta}_i]), \tag{20}$$

where $S(a, b) = CE(a, b) + CE(b, a)$ denotes the symmetric cross-entropy [54, 68] between two predictions. ξ is the threshold as the value at the τ^{th} percentile.

4.5 Overall Optimization

In summary, the final loss objective of the proposed DisCo can be defined as follows:

$$\mathcal{L} = \mathcal{L}_{so} + \alpha(\mathcal{L}^s_{AL} - \mathcal{L}^c_{AL}) + \beta\mathcal{L}_{ind} + \gamma\mathcal{L}_{ta}. \tag{21}$$

The overall optimization process can be formulated as:

$$\begin{cases} \min_{\psi} \mathcal{L}^s_{AL} + \mathcal{L}^c_{AL}, \\ \min_{\phi} \mathcal{L}. \end{cases} \tag{22}$$

In our implementation, we first train the diffusion model on the source domain and leverage the inference output of the model as debiased labels of target domain graphs. Then, the final framework of the adaptation is optimized in an alternative manner for progressive domain alignment.

5 Theoretical Analysis from a Causal Discovery Perspective

In this section, we present a theoretical analysis of our proposed method, DisCo, from the perspective of causal discovery. This analysis justifies the use of mutual information, as defined in Equation (8), to effectively disentangle causal and spurious features.

We assume that the true feature matrix is given by $\boldsymbol{E}^* = [\boldsymbol{E}^c, \boldsymbol{E}^s] \in \mathbb{R}^d$, where $\boldsymbol{E}^c \in \mathbb{R}^{d_c}$ and $\boldsymbol{E}^s \in \mathbb{R}^{d_s}$ represent the causal and spurious features, respectively. For simplicity, we assume that $\boldsymbol{E}^*$ follows a standard multivariate Gaussian distribution, i.e.,

$$\boldsymbol{E}^* \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \tag{23}$$

where $\boldsymbol{I}$ denotes the identity matrix.

However, the observed feature matrix is a transformed version of the true features:

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{E}^*, \tag{24}$$

where $\boldsymbol{T}$ is an orthogonal transformation matrix. We model the domain-specific factors as

$$\mathcal{D}^* = \boldsymbol{B}\boldsymbol{E}^s, \tag{25}$$

where $\boldsymbol{B}$ is a transformation matrix satisfying $\boldsymbol{B}\boldsymbol{B}^\top = \boldsymbol{I}$. Our goal is to find an orthogonal transformation $\boldsymbol{F}$ such that the rotated representation

$$\boldsymbol{Z} = \boldsymbol{F}\boldsymbol{X} = \boldsymbol{F}\boldsymbol{T}\boldsymbol{E}^* = [\boldsymbol{Z}^c, \boldsymbol{Z}^s] \tag{26}$$

correctly disentangles the causal and spurious components.

Theorem 1. *The solution to the optimization problem in Equation* (8) *is given by*

$$\boldsymbol{Z}^c = \boldsymbol{E}^c, \quad \boldsymbol{Z}^s = \boldsymbol{B}\boldsymbol{E}^s. \tag{27}$$

Theorem 1 demonstrates that the optimal solution corresponds exactly to the ground-truth causal and spurious features. This result highlights the effectiveness of our proposed method in achieving feature disentanglement through the mutual information objective.

To prove Theorem 1, we first present several standard lemmas that establish fundamental properties of mutual information under the Gaussian assumption.

Lemma 2 (Non-negativity of Mutual Information). *Let X and Y be two random variables with joint distribution $p(x, y)$ and marginal distributions $p(x)$ and $p(y)$. The mutual information between X and Y, defined as*

$$I(X;Y) = \iint p(x,y) \log\left(\frac{p(x,y)}{p(x)p(y)}\right) dx\, dy, \tag{28}$$

is always non-negative:

$$I(X;Y) \geq 0, \tag{29}$$

with equality if and only if X and Y are independent, i.e., $p(x, y) = p(x)p(y)$ almost everywhere.

Lemma 3 (Zero Mutual Information Implies Independence). *Let X and Y be two random variables with joint distribution $p(x, y)$ and marginals $p(x)$, $p(y)$. Then,*

$$I(X;Y) = 0 \quad \Longleftrightarrow \quad X \perp Y. \tag{30}$$

That is, X and Y are statistically independent if and only if their mutual information is zero.

Lemma 4 (Maximum Mutual Information). *Let X be a random variable with entropy $H(X)$. Then, for any random variable Y, the mutual information satisfies:*

$$I(X;Y) \leq H(X), \tag{31}$$

with equality if and only if Y is a deterministic and invertible function of X, i.e., $Y = f(X)$ where f is bijective.

The above lemmas are classical results in information theory and can be found in textbooks; we omit their proofs for brevity.

Proof of Theorem 1. Recall the optimization problem defined in Equation (8):

$$\max_{\boldsymbol{F}\in O(d)} I(\boldsymbol{Z}^s;\mathcal{D}^*) - I(\boldsymbol{Z}^c;\mathcal{D}^*) - \beta I(\boldsymbol{Z}^s;\boldsymbol{Z}^c). \tag{32}$$

By Lemmas 2–4, all mutual information terms in the objective are non-negative and bounded above by the entropy of the involved variables.

The first term $I(\boldsymbol{Z}^s;\mathcal{D}^*)$ is maximized when $\boldsymbol{Z}^s$ fully captures the information in $\mathcal{D}^*$, which is achieved when $\boldsymbol{Z}^s = \boldsymbol{B}\boldsymbol{E}^s$.

The second term $I(\boldsymbol{Z}^c;\mathcal{D}^*)$ is minimized when $\boldsymbol{Z}^c$ is independent of $\mathcal{D}^*$, which is satisfied when $\boldsymbol{Z}^c = \boldsymbol{E}^c$, since $\boldsymbol{E}^c \perp \boldsymbol{E}^s$ and $\mathcal{D}^*$ depends only on $\boldsymbol{E}^s$.

Similarly, the third term $I(\boldsymbol{Z}^s;\boldsymbol{Z}^c)$ is minimized when $\boldsymbol{Z}^s$ and $\boldsymbol{Z}^c$ are statistically independent, i.e., $I(\boldsymbol{Z}^s;\boldsymbol{Z}^c) = 0$, which also holds under the choice $\boldsymbol{Z}^s = \boldsymbol{B}\boldsymbol{E}^s$, $\boldsymbol{Z}^c = \boldsymbol{E}^c$, due to the independence of $\boldsymbol{E}^s$ and $\boldsymbol{E}^c$.

Therefore, the choice $\boldsymbol{Z}^c = \boldsymbol{E}^c$, $\boldsymbol{Z}^s = \boldsymbol{B}\boldsymbol{E}^s$ simultaneously maximizes the first term and minimizes the latter two, thereby solving the optimization problem in Equation (8). □

6 Experiments

6.1 Experimental Settings

6.1.1 Datasets. The efficacy of our method in unsupervised graph domain adaptation is validated using two distinct experimental paradigms: cross-dataset transfer and dataset split. For the cross-dataset paradigm, we employ benchmarks including PTC [13], BZR/BZR_MD [45], and COX2/COX2_MD [45], which are chosen for their inherent impartiality across sub-domains. In the dataset split setting, we follow prior works [5, 30, 67] to partition graphs based on density, conducting experiments on TWITTER-Real-Graph-Partial [36], NCI1 [50], and Mutagenicity [21].

6.1.2 Baselines. We compare `DisCo` against 13 baselines, which can be categorized into four groups: **(1) Graph neural networks**, including GCN [22], GraphSAGE [10], GIN [64] and GAT [48]. **(2) Semi-supervised graph methods**, including InfoGraph [43] and Mean-Teacher [47]. **(3) Domain adaptation methods**, including ToAlign [58], DANN [7], DUA [33] and DARE-GRAM [34]. **(4) Graph adaptation methods**, including CoCo [67], MTDF [46] and SLOGAN [32].

6.1.3 Implementation Details. Building on the baselines, we adopt accuracy (ACC) to evaluate the performance of our model. The primary architecture is a two-layer GCN with a 128-dimensional hidden state. Using the Adam optimizer with a learning rate set to 0.001, the model is trained in the source domain for 100 epochs, followed by 30 epochs for the adaptation of the target domain. The diffusion model comprises a separate two-layer GCN encoder and a four-layer noise predictor, optimized with a learning rate of 0.0001.

6.2 Performance Comparison

In this section, we conduct an extensive performance comparison of `DisCo` against various baselines, with experimental results showcased across five benchmark datasets in Table 1, 2, 3, 4 and 5. We highlight the following observations: ❶ Semi-supervised methods, such as InfoGraph, tend to outperform standard GNNs. ❷ Classic

Table 1: PTC classification results (%) for source→target tasks.

Methods	MM→MR	MR→MM	MR→FR	MR→FM	MM→FM	MM→FR	MM→FM	FR→MR	FM→MM	FR→MM	FR→FM	FM→FR	Avg.
GIN	64.3	61.8	56.5	57.7	45.7	53.5	42.6	37.7	44.5	59.4	54.3	66.2	53.7
GCN	62.9	63.2	55.1	66.2	45.7	67.6	54.4	62.3	64.8	58.0	52.0	60.3	59.4
GAT	46.0	60.0	57.1	70.7	46.3	65.9	53.8	54.5	53.2	69.0	51.7	65.9	57.8
SAGE	55.1	58.8	56.5	67.6	47.4	66.2	48.2	48.1	45.1	58.0	52.6	63.2	55.6
MeanTeacher	61.4	61.8	60.9	73.2	52.9	50.7	44.1	65.2	35.1	66.7	42.9	55.9	55.9
InfoGraph	60.0	63.2	59.4	66.2	48.6	67.6	56.4	55.1	64.8	63.8	54.3	69.1	60.7
DANN	64.3	64.3	63.8	69.0	55.7	73.2	48.5	50.7	66.2	71.0	52.9	70.6	62.8
ToAlign	45.7	70.5	66.7	67.6	54.3	67.6	50.0	58.0	67.6	71.0	55.7	76.5	62.6
DUA	63.1	62.0	52.2	73.6	54.3	63.4	55.9	53.7	60.6	69.4	54.3	63.2	60.5
DARE-GRAM	54.3	61.8	53.6	73.3	55.7	66.2	54.1	56.4	59.4	69.6	55.7	66.2	60.5
CoCo	63.8	65.1	60.3	73.0	55.2	72.8	55.9	62.1	63.2	70.5	54.2	70.1	63.8
MTDF	64.8	65.9	61.2	**76.9**	56.0	73.9	56.8	62.6	69.0	71.1	56.0	71.6	65.5
SLOGAN	65.7	71.1	66.8	74.6	55.4	71.8	59.1	66.7	70.4	**78.3**	57.1	73.7	67.8
Ours	**67.1**	**72.8**	**68.2**	70.7	**66.9**	**74.3**	**64.1**	**67.8**	**70.6**	70.7	**61.9**	**74.5**	**69.2**

Table 2: NCI1 classification results (%) for source→target tasks, involving sub-datasets N_{D0}, N_{D1}, N_{D2}, and N_{D3}.

Methods	$N_{D0}\to N_{D1}$	$N_{D0}\to N_{D2}$	$N_{D0}\to N_{D3}$	$N_{D1}\to N_{D0}$	$N_{D1}\to N_{D2}$	$N_{D1}\to N_{D3}$	$N_{D2}\to N_{D0}$	$N_{D2}\to N_{D1}$	$N_{D2}\to N_{D3}$	$N_{D3}\to N_{D0}$	$N_{D3}\to N_{D1}$	$N_{D3}\to N_{D2}$	Avg.
GIN	66.0	60.6	50.3	68.0	68.4	69.9	61.0	65.6	73.1	48.3	59.4	62.9	62.8
GCN	55.8	59.1	54.0	73.3	65.0	70.7	73.5	60.7	70.2	67.8	54.5	55.1	63.3
GAT	63.4	60.0	41.7	70.1	68.2	70.1	73.2	63.1	69.3	56.6	56.3	60.5	62.7
SAGE	54.9	55.8	50.1	74.5	59.7	66.0	76.2	59.7	71.7	70.6	57.2	64.8	63.4
MeanTeacher	54.9	45.2	51.6	73.8	45.2	50.7	73.3	54.9	50.2	72.8	55.8	47.1	56.3
InfoGraph	66.5	61.0	57.6	62.7	64.6	64.1	75.7	62.6	67.1	69.9	60.7	50.2	63.6
DANN	64.1	58.7	45.6	76.2	69.8	63.6	71.3	70.9	70.0	70.4	58.3	67.5	65.5
ToAlign	65.5	61.7	47.1	73.3	69.9	59.7	71.4	69.9	69.9	68.0	59.2	63.1	64.9
DUA	69.9	60.7	58.5	71.3	69.9	68.4	67.5	68.0	70.9	56.1	50.5	66.5	64.9
DARE-GRAM	69.4	59.2	55.8	69.9	69.4	61.2	68.9	70.4	68.9	60.1	57.6	65.0	64.7
CoCo	70.9	64.0	68.7	70.0	68.5	71.2	75.1	61.2	72.8	74.6	59.6	56.4	67.7
MTDF	67.5	**70.9**	**71.8**	76.7	65.0	73.1	77.2	62.5	74.3	75.9	61.0	57.8	69.5
SLOGAN	71.4	64.1	63.1	71.8	72.3	72.8	76.7	72.5	73.3	76.3	61.7	70.9	70.6
Ours	**72.1**	66.9	66.2	**79.2**	**75.8**	**74.2**	**77.9**	**73.0**	**74.6**	**76.4**	**63.2**	**71.4**	**72.6**

Table 3: Mutagenicity classification results (%) for source→target tasks, involving sub-datasets M_{D0}, M_{D1}, M_{D2}, and M_{D3}.

Methods	$M_{D0}\to M_{D1}$	$M_{D0}\to M_{D2}$	$M_{D0}\to M_{D3}$	$M_{D1}\to M_{D0}$	$M_{D1}\to M_{D2}$	$M_{D1}\to M_{D3}$	$M_{D2}\to M_{D0}$	$M_{D2}\to M_{D1}$	$M_{D2}\to M_{D3}$	$M_{D3}\to M_{D0}$	$M_{D3}\to M_{D1}$	$M_{D3}\to M_{D2}$	Avg.
GIN	72.3	64.1	56.6	68.5	67.4	55.9	72.1	74.4	62.8	61.1	67.3	73.0	66.3
GCN	71.1	62.7	57.7	70.4	68.8	53.6	69.0	74.2	65.8	59.6	63.3	74.5	65.9
GAT	70.8	63.2	56.4	67.8	69.3	55.7	70.6	73.2	64.1	58.3	65.2	71.4	65.5
SAGE	69.7	60.7	58.4	69.6	66.2	54.2	73.8	74.0	64.6	60.7	68.5	70.9	65.9
MeanTeacher	67.3	60.4	56.6	67.4	61.4	52.0	70.4	72.8	63.9	58.6	64.7	67.8	63.6
InfoGraph	73.5	64.8	59.5	71.7	68.2	58.7	70.4	76.2	63.4	63.5	67.8	72.1	67.5
DANN	72.6	67.5	61.0	70.8	71.5	58.4	71.8	75.2	64.4	64.0	68.2	71.4	68.1
ToAlign	74.0	69.1	54.7	72.7	71.7	58.7	65.2	77.2	61.5	73.1	73.1	62.2	67.8
DUA	70.2	64.0	53.6	56.5	57.7	65.1	63.7	76.0	57.9	68.5	59.8	67.7	63.4
DARE-GRAM	71.5	67.3	60.8	62.3	70.4	57.0	70.2	73.6	61.8	62.5	70.7	72.9	66.8
CoCo	77.7	73.3	66.6	76.6	77.3	67.4	74.5	80.8	68.9	74.3	74.1	77.5	74.1
MTDF	79.3	72.4	65.5	**78.2**	78.6	67.8	72.4	82.0	70.3	75.1	76.3	77.8	74.6
SLOGAN	81.2	74.0	67.3	76.2	79.1	68.2	74.7	83.6	70.8	**75.3**	76.9	78.2	75.5
Ours	**82.6**	**75.5**	**69.4**	76.8	**80.1**	**69.1**	**76.4**	**85.7**	**72.5**	72.8	**79.2**	**79.1**	**76.6**

domain adaptation methods like DANN exhibit exceptional performance on specific tasks. However, their effectiveness can be inconsistent across the full range of scenarios due to their incapable to deal with graph-based data. ❸ Graph adaptation methods (CoCo, MTDF, and SLOGAN) show better performance, underscoring the significance of explicitly addressing the domain shift inherent in these graph classification tasks. ❹ Remarkably, DisCo delivers consistent and substantial gains in both dataset splitting and cross-dataset settings, verifying its efficacy and reliability. We attribute this superior performance to the decoupling of causal and spurious features, which effectively isolates domain-related biases. Furthermore, the incorporation of diffusion-guided pseudo-label

Table 4: TWITTER classification results (%) for source→target tasks, involving sub-datasets T_{D0}, T_{D1}, T_{D2}, and T_{D3}.

Methods	T_{D0}→T_{D1}	T_{D0}→T_{D2}	T_{D0}→T_{D3}	T_{D1}→T_{D0}	T_{D1}→T_{D2}	T_{D1}→T_{D3}	T_{D2}→T_{D0}	T_{D2}→T_{D1}	T_{D2}→T_{D3}	T_{D3}→T_{D0}	T_{D3}→T_{D1}	T_{D3}→T_{D2}	Avg.
GIN	59.7	62.8	60.4	64.2	62.2	61.3	61.7	63.2	61.0	62.3	61.8	62.4	61.9
GCN	62.0	62.9	59.7	64.1	63.4	59.8	64.2	62.8	60.5	62.7	61.4	62.7	62.2
GAT	60.6	63.2	60.0	63.1	61.6	59.8	63.5	61.6	59.5	63.4	62.1	63.7	61.9
SAGE	61.0	64.6	62.1	61.9	61.9	60.8	62.9	62.6	60.9	61.7	60.9	63.4	62.1
MeanTeacher	52.2	49.2	46.1	49.0	50.7	46.1	49.5	51.7	52.6	48.1	48.0	51.1	49.5
InfoGraph	63.9	65.1	61.6	65.6	65.0	59.2	64.3	60.8	63.3	62.4	63.3	63.2	63.2
DANN	58.4	60.0	58.0	59.0	59.4	57.4	57.7	58.1	58.4	58.2	57.9	60.4	58.6
ToAlign	58.6	59.5	55.5	57.7	58.1	56.1	56.3	57.2	57.8	57.7	57.6	60.2	57.7
DUA	64.2	64.8	62.1	65.8	56.0	62.0	65.3	63.6	60.8	64.2	63.4	64.7	63.1
DARE-GRAM	61.7	65.7	61.6	65.7	64.4	62.1	64.0	64.3	60.0	64.9	64.2	64.3	63.6
CoCo	64.5	66.1	62.1	64.0	64.6	61.6	64.5	63.5	62.1	62.8	62.5	63.5	63.5
MTDF	64.5	66.4	65.1	64.7	65.2	62.2	64.9	63.5	63.2	63.2	63.4	64.4	64.2
SLOGAN	**65.1**	**66.5**	62.3	66.2	65.4	62.4	66.6	64.4	62.2	65.8	**64.4**	65.1	64.7
Ours	63.8	66.3	**63.5**	**66.7**	**66.2**	**62.8**	**66.8**	**65.0**	**62.9**	**66.2**	64.1	**66.0**	**65.1**

denoising ensures the model benefits from more precise and refined supervision signals.

Table 5: COX2 and BZR classification results (%) for source→target tasks.

Methods	B→BM	BM→B	C→CM	CM→C	Avg.
GIN	54.3	79.1	45.9	61.7	60.3
GCN	46.8	79.0	57.4	76.6	64.9
GAT	49.4	79.5	52.6	68.1	62.4
SAGE	48.1	65.4	58.5	59.2	57.8
MeanTeacher	54.8	76.5	49.2	76.7	64.3
InfoGraph	56.5	77.8	57.4	72.3	66.0
DANN	54.8	**84.0**	52.6	67.0	67.0
ToAlign	54.9	79.0	57.6	76.6	64.4
DUA	54.8	78.8	57.3	76.6	66.9
DARE-GRAM	53.2	79.0	52.9	77.1	65.6
CoCo	57.2	79.1	57.4	76.3	67.5
MTDF	58.1	79.8	59.0	77.0	68.5
SLOGAN	58.6	80.2	59.0	77.7	68.9
Ours	**61.7**	81.3	**60.0**	**78.2**	**70.3**

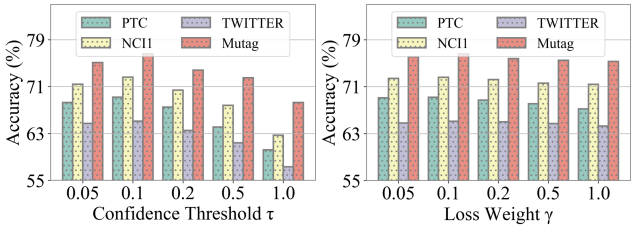


Figure 3: Comparison of performance with respect to varying confidence thresholds τ (left) and loss weight γ (right).

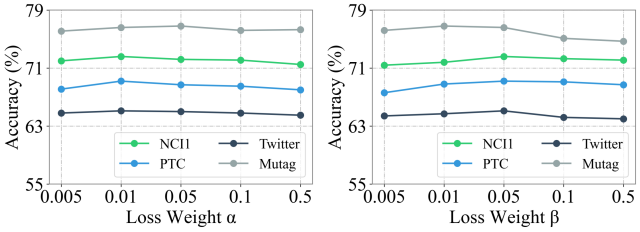


Figure 4: Comparison of performance with respect to varying loss weights α (left) and β (right).

6.3 Scalability Analysis

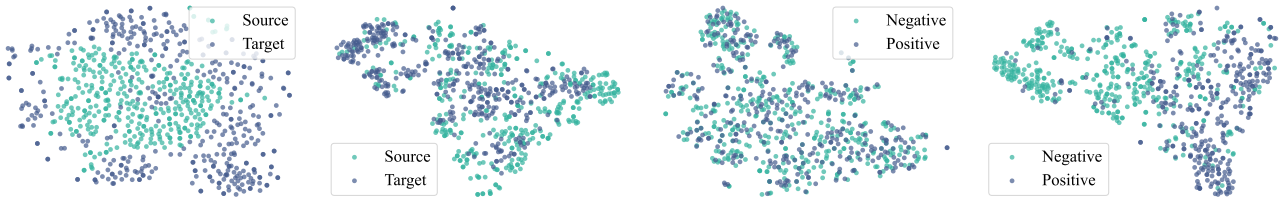
This section evaluates the overall complexity of DisCo and subsequently discuss its scalability. Let L be the number of GNN layers, d be the hidden dimension, N_{ta} and N_{so} be the number of graphs in the target and source domains, and $|V_{so}|$ and $|V_{ta}|$ be the average number of nodes in the source and target domains. The computational consumption of our DisCo is mainly composed of three parts: **(1)** pre-training phase in the source domain, **(2)** pre-training and pseudo-labeling phases with diffusion-based label denoising module, and **(3)** adaptation phase in the target domain. For **(1)**, the complexity of supervised pre-training of the GNN classifier is $O(N_{so}L|V_{so}|d^2)$. For **(2)**, the complexity of the pre-training of diffusion-based label denoising module is $O(N_{so}L|V_{so}|d^2)$, and the complexity of pseudo-labeling is $O(N_{ta}L|V_{ta}|d^2)$. For **(3)**, the complexity of adaptation training is $O(N_{ta}L|V_{ta}|d^2)$. In total, the overall complexity of our model can be calculated as $O(2Ld^2(N_{so}|V_{so}| + N_{ta}|V_{ta}|)$. This complexity is linearly related to the size of the dataset, which is comparable with other most recent domain adaptation works. We also provide a deeper discussion to show the comparison of training/adaptation time and parameter counts in Table 8 in the Appendix A.2.

6.4 Sensitivity Analysis

Confidence threshold τ. We vary $\tau \in \{5\%, 10\%, 20\%, 30\%, 50\%\}$. From the results shown in Figure 3 (left), we conclude that: ❶ Optimal accuracy is consistently achieved at $\tau = 10\%$ across all datasets. ❷ A lower threshold, such as $\tau = 5\%$ appears to select an insufficient number of pseudo-labels for effective adaptation, while

Table 6: The ablation study on NCI1 for source→target tasks, involving sub-datasets N_{D0}, N_{D1}, N_{D2}, and N_{D3}.

Methods	$N_{D0}\to N_{D1}$	$N_{D0}\to N_{D2}$	$N_{D0}\to N_{D3}$	$N_{D1}\to N_{D0}$	$N_{D1}\to N_{D2}$	$N_{D1}\to N_{D3}$	$N_{D2}\to N_{D0}$	$N_{D2}\to N_{D1}$	$N_{D2}\to N_{D3}$	$N_{D3}\to N_{D0}$	$N_{D3}\to N_{D1}$	$N_{D3}\to N_{D2}$	Avg.
Variant 1	68.9	59.6	57.6	75.0	68.4	70.6	73.5	69.8	71.4	71.2	59.8	64.7	67.5
Variant 2	70.4	61.5	58.2	75.2	67.3	71.4	74.1	70.4	72.4	72.8	61.3	66.7	68.5
w/o $\mathcal{L}^*_{AL}$	71.3	63.9	64.1	76.8	72.4	72.5	74.8	72.6	74.3	73.9	60.8	69.0	70.5
w/o $\mathcal{L}_{ind}$	71.1	64.7	65.3	78.0	74.8	73.7	76.3	71.4	73.8	72.7	61.2	70.8	71.2
Ours	**72.1**	**66.9**	**66.2**	**79.2**	**75.8**	**74.2**	**77.9**	**73.0**	**74.6**	**76.4**	**63.2**	**71.4**	**72.6**



(a) Spurious Feature **(b) Causal Feature** **(c) Spurious Feature** **(d) Causal Feature**

Figure 5: t-SNE visualization of spurious and causal features, grouped by domain and label.

higher thresholds ($\tau > 10\%$) likely introduce excessive noise from lower-quality samples, leading to performance degradation. Based on these findings, we set $\tau = 10\%$ as the default value.

Loss weight α, β, γ. We investigate the sensitivity to loss weights by varying $\alpha \in \{0.005, 0.01, 0.03, 0.05, 0.1\}$, $\beta \in \{0.005, 0.01, 0.05, 0.1, 0.5\}$, and $\gamma \in \{0.05, 0.1, 0.2, 0.5, 1\}$. From the results illustrated in Figure 3 (right) and 4, we can conclude that: ❶ Our model is not sensitive to γ. ❷ Excessively low values of α and β fail to adequately disentangle causal and spurious factors, while excessively high values may over-penalize the causal representation, diminishing its quality and causing a slight performance decline. Based on the experimental results, we recommend that setting α and β to 0.01 and 0.05, respectively, are suitable for most datasets.

6.5 Ablation Study

Ablation study is conducted to investigate the impact of our two primary components: ❶ the diffusion-guided label denoising strategy and ❷ feature disentanglement for causal discovery. To evaluate the first component, we adopt two simpler pseudo-labeling variants using a direct confidence threshold on predicted probabilities: **Variant 1** ($\xi = 0.90$) and **Variant 2** ($\xi = 0.98$). For the second component, we assess the performance by individually ablating the adversarial loss (**w/o** $\mathcal{L}^*_{AL}$) and the orthogonality loss (**w/o** $\mathcal{L}_{ind}$).

As shown in Table 6, our ablation study on the NCII dataset highlights the essential contributions of our models' components. ❶ The marked performance drop in both **Variant 1** and **Variant 2** underscores the importance of the diffusion-guided label denoising module for enhancing pseudo-label quality in the target domain. This indicates that diffusion-guided pseudo-labeling mitigate the confirmation bias by providing more precise supervisory signals for adaptation. ❷ The accuracy decrease observed when ablating either the adversarial or orthogonality loss confirms their synergistic role in disentangling domain-specific and domain-invariant factors, validating that each component is integral to the overall effectiveness. To more comprehensively investigate the efficacy of diffusion-guided label denoising module, deeper discussion can be found in Table 7 in the Appendix A.1.

6.6 Visualization

To intuitively illustrate the effectiveness of our DisCo in disentangling representations, we visualize the learned spurious and causal features using t-SNE, where features are distinguished by domain and label. The observations can be summarized as follows: ❶ In Figure 5(a) and 5(b), the spurious features exhibit a clear boundary between the source and target domains, while the causal features maintain a consistent distribution across them. ❷ Figure 5(c) and 5(d) show that causal features form distinct clusters for positive and negative samples, demonstrating a high correlation with the semantic labels. Conversely, spurious features retain variations that are irrelevant to the labels. These provide evidence that DisCo effectively extracts domain-invariant, semantically meaningful information into causal features, thereby enhancing the overall adaptation.

7 Conclusion

In this paper, we propose a diffusion-guided unbiased discriminative learning framework for unsupervised graph domain adaptation termed DisCo. Specifically, we leverage adversarial feature disentanglement to decouple the representations into causal and spurious parts, where causal features are encouraged to be domain-invariant, while the spurious features retain label-irrelevant variations. Furthermore, we retrieve the labels of cross-domain source graphs to generate the conditions, which would be utilized to optimize a diffusion model for label denoising. Building upon this, we measure the consistency between pseudo-labels and denoised labels to reduce the bias during domain adaptation. Comprehensive experiments across various real-world benchmarks demonstrates the effectiveness of DisCo. In our future work, we plan to harness the generative potential of diffusion models to address source-free unsupervised graph domain adaptation.

8 Acknowledgments

Tao Ren is supported by the National Natural Science Foundation of China (62276058, 41774063), the Fundamental Research Funds for the Central Universities (N25GFZ011). Yifan Wang is supported by the Fundamental Research Funds for the Central Universities in UIBE (Grant No. 23QN02).

References

[1] Jian Chen, Ruiyi Zhang, Tong Yu, Rohan Sharma, Zhiqiang Xu, Tong Sun, and Changyou Chen. 2023. Label-retrieval-augmented diffusion models for learning from noisy labels. In *Proceedings of the Conference on Neural Information Processing Systems*. 66499–66517.

[2] Debo Cheng, Jiuyong Li, Lin Liu, Jixue Liu, and Thuc Duy Le. 2024. Data-driven causal effect estimation based on graphical causal modelling: A survey. *Comput. Surveys* 56, 5 (2024), 1–37.

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

[4] Prafulla Dhariwal and Alexander Nichol. 2021. Diffusion models beat gans on image synthesis. In *Proceedings of the Conference on Neural Information Processing Systems*. 8780–8794.

[5] Ming Ding, Jie Tang, and Jie Zhang. 2018. Semi-supervised learning on graphs with generative adversarial nets. In *Proceedings of the International Conference on Information and Knowledge Management*. 913–922.

[6] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph neural networks for social recommendation. In *Proceedings of the Web Conference*. 417–426.

[7] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. 2016. Domain-adversarial training of neural networks. *Journal of Machine Learning Research* 17, 59 (2016), 1–35.

[8] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. In *Proceedings of the International Conference on Machine Learning*. 1263–1272.

[9] Shurui Gui, Meng Liu, Xiner Li, Youzhi Luo, and Shuiwang Ji. 2023. Joint learning of label and environment causal independence for graph out-of-distribution generalization. In *Proceedings of the Conference on Neural Information Processing Systems*. 3945–3978.

[10] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In *Proceedings of the Conference on Neural Information Processing Systems*. 1025–1035.

[11] Xizewen Han, Huangjie Zheng, and Mingyuan Zhou. 2022. Card: Classification and regression diffusion models. In *Proceedings of the Conference on Neural Information Processing Systems*. 18100–18115.

[12] Zhongkai Hao, Chengqiang Lu, Zhenya Huang, Hao Wang, Zheyuan Hu, Qi Liu, Enhong Chen, and Cheekong Lee. 2020. 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*. 731–752.

[13] Christoph Helma, Ross D. King, Stefan Kramer, and Ashwin Srinivasan. 2001. The predictive toxicology challenge 2000–2001. *Bioinformatics* 17, 1 (2001), 107–108.

[14] Samitha Herath, Basura Fernando, Ehsan Abbasnejad, Munawar Hayat, Shahram Khadivi, Mehrtash Harandi, Hamid Rezatofighi, and Gholamreza Haffari. 2023. Energy-based self-training and normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 11653–11662.

[15] Jonathan Ho, Ajay Jain, and Pieter Abbeel. 2020. Denoising diffusion probabilistic models. In *Proceedings of the Conference on Neural Information Processing Systems*. 6840–6851.

[16] Jonathan Ho and Tim Salimans. 2022. Classifier-free diffusion guidance. *arXiv* (2022).

[17] Senyu Hou, Gaoxia Jiang, Jia Zhang, Shangrong Yang, Husheng Guo, Yaqing Guo, and Wenjian Wang. 2025. Directional Label Diffusion Model for Learning from Noisy Labels. In *Proceedings of the Computer Vision and Pattern Recognition Conference*. 25738–25748.

[18] Binbin Hu, Zhicheng An, Zhengwei Wu, Ke Tu, Ziqi Liu, Zhiqiang Zhang, Jun Zhou, Yufei Feng, and Jiawei Chen. 2025. Graph Disentangle Causal Model: Enhancing Causal Inference in Networked Observational Data. In *Proceedings of the International ACM Conference on Web Search & Data Mining*. 1–9.

[19] Wei Ju, Siyu Yi, Yifan Wang, Zhiping Xiao, Zhengyang Mao, Hourun Li, Yiyang Gu, Yifang Qin, Nan Yin, Senzhang Wang, et al. 2024. A survey of graph neural networks in real world: Imbalance, noise, privacy and ood challenges. *arXiv preprint arXiv:2403.04468* (2024).

[20] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. 2019. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 4893–4902.

[21] Jeroen Kazius, Ross McGuire, and Roberta Bursi. 2005. Derivation and validation of toxicophores for mutagenicity prediction. *Journal of medicinal chemistry* 48, 1 (2005), 312–320.

[22] Thomas N Kipf and Max Welling. 2017. Semi-supervised classification with graph convolutional networks. In *Proceedings of the International Conference on Learning Representations*.

[23] Claude Knaus and Matthias Zwicker. 2014. Progressive image denoising. *IEEE transactions on image processing* 23, 7 (2014), 3114–3125.

[24] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. 2019. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 10285–10295.

[25] Haoyang Li, Ziwei Zhang, Xin Wang, and Wenwu Zhu. 2022. Learning invariant graph representations for out-of-distribution generalization. In *Proceedings of the Conference on Neural Information Processing Systems*. 11828–11841.

[26] Junnan Li, Richard Socher, and Steven CH Hoi. 2020. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv* (2020).

[27] Sihang Li, Xiang Wang, An Zhang, Yingxin Wu, Xiangnan He, and Tat-Seng Chua. 2022. Let invariant rationale discovery inspire graph contrastive learning. In *Proceedings of the International Conference on Machine Learning*. 13052–13065.

[28] Xiaofeng Liu, Zhenhua Guo, Fangxu Xing, Jane You, C-C Jay Kuo, Georges El Fakhri, and Jonghye Woo. 2021. Adversarial unsupervised domain adaptation with conditional and label shift: Infer, align and iterate. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 10367–10376.

[29] Yixin Liu, Kaize Ding, Huan Liu, and Shirui Pan. 2023. Good-d: On unsupervised graph out-of-distribution detection. In *Proceedings of the International ACM Conference on Web Search & Data Mining*. 339–347.

[30] Bin Lu, Ze Zhao, Xiaoying Gan, Shiyu Liang, Luoyi Fu, Xinbing Wang, and Chenghu Zhou. 2024. Graph Out-of-Distribution Generalization With Controllable Data Augmentation. *IEEE Transactions on Knowledge and Data Engineering* 36, 11 (2024), 6317–6329.

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

[32] Junyu Luo, Yuhao Tang, Xiao Luo, Zhizhuo KOU, Zhiping Xiao, Wei Ju, Wentao Zhang, Ming Zhang, et al. 2025. Sparse Causal Discovery with Generative Intervention for Unsupervised Graph Domain Adaptation. In *Proceedings of the International Conference on Machine Learning*.

[33] M Jehanzeb Mirza, Jakub Micorek, Horst Possegger, and Horst Bischof. 2022. The norm must go on: Dynamic unsupervised domain adaptation by normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14765–14775.

[34] Ismail Nejjar, Qin Wang, and Olga Fink. 2023. Dare-gram: Unsupervised domain adaptation regression by aligning inverse gram matrices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 11744–11754.

[35] Poojan Oza, Vishwanath A Sindagi, Vishal M Patel, et al. 2023. Unsupervised domain adaptation of object detectors: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 6 (2023), 4018–4040.

[36] Shirui Pan, Jia Wu, and Xingquan Zhu. 2015. CogBoost: Boosting for fast cost-sensitive graph classification. *IEEE Transactions on Knowledge and Data Engineering* 27, 11 (2015), 2933–2946.

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

[38] Judea Pearl, Madelyn Glymour, and Nicholas P Jewell. 2016. *Causal inference in statistics: A primer*. John Wiley & Sons.

[39] Jonas Peters, Peter Bühlmann, and Nicolai Meinshausen. 2016. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 78, 5 (2016), 947–1012.

[40] Sanqing Qu, Tianpei Zou, Lianghua He, Florian Röhrbein, Alois Knoll, Guang Chen, and Changjun Jiang. 2024. LEAD: Learning Decomposition for Source-free Universal Domain Adaptation. In *CVPR*.

[41] Manon Réau, Nicolas Renaud, Li C Xue, and Alexandre MJJ Bonvin. 2023. DeepRank-GNN: a graph neural network framework to learn patterns in protein–protein interfaces. *Bioinformatics* 39, 1 (2023), btac759.

[42] Tao Ren, Haodong Zhang, Yifan Wang, Wei Ju, Chengwu Liu, Fanchun Meng, Siyu Yi, and Xiao Luo. 2025. MHGC: Multi-scale hard sample mining for contrastive deep graph clustering. *Information Processing & Management* 62, 4 (2025), 104084.

[43] Fan-Yun Sun, Jordon Hoffman, Vikas Verma, and Jian Tang. 2020. InfoGraph: Unsupervised and Semi-supervised Graph-Level Representation Learning via Mutual Information Maximization. In *Proceedings of the International Conference on Learning Representations*.

[44] Susheel Suresh, Pan Li, Cong Hao, and Jennifer Neville. 2021. Adversarial graph augmentation to improve graph contrastive learning. In *Proceedings of the Conference on Neural Information Processing Systems*. 15920–15933.

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

[46] Yuhao Tang, Junyu Luo, Ling Yang, Xiao Luo, Wentao Zhang, and Bin Cui. 2024. Multi-view teacher with curriculum data fusion for robust unsupervised domain adaptation. In *Proceedings of the IEEE Conference on Data Engineering*. 2598–2611.

[47] Antti Tarvainen and Harri Valpola. 2017. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *Proceedings of the Conference on Neural Information Processing Systems*. 1195–1204.

[48] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph attention networks. In *Proceedings of the International Conference on Learning Representations*.

[49] Petar Veličković, William Fedus, William L Hamilton, Pietro Liò, Yoshua Bengio, and R Devon Hjelm. 2019. Deep Graph Infomax. In *Proceedings of the International Conference on Machine Learning*.

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

[51] Xiang Wang, Hongye Jin, An Zhang, Xiangnan He, Tong Xu, and Tat-Seng Chua. 2020. Disentangled Graph Collaborative Filtering. In *Proceedings of the International ACM SIGIR Conference on Research & Development in Information Retrieval*.

[52] Yisen Wang, Weiyang Liu, Xingjun Ma, James Bailey, Hongyuan Zha, Le Song, and Shu-Tao Xia. 2018. Iterative learning with open-set noisy labels. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 8688–8696.

[53] Yifan Wang, Xiao Luo, Chong Chen, Xian-Sheng Hua, Ming Zhang, and Wei Ju. 2024. DisenSemi: Semi-supervised graph classification via disentangled representation learning. *IEEE Transactions on Neural Networks and Learning Systems* (2024).

[54] Yisen Wang, Xingjun Ma, Zaiyi Chen, Yuan Luo, Jinfeng Yi, and James Bailey. 2019. Symmetric cross entropy for robust learning with noisy labels. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 322–330.

[55] Yifan Wang, Yiping Song, Shuai Li, Chaoran Cheng, Wei Ju, Ming Zhang, and Sheng Wang. 2022. Disencite: Graph-based disentangled representation learning for context-specific citation generation. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 11449–11458.

[56] Yuyang Wang, Jianren Wang, Zhonglin Cao, and Amir Barati Farimani. 2022. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence* 4, 3 (2022), 279–287.

[57] Guoqiang Wei, Cuiling Lan, Wenjun Zeng, and Zhibo Chen. 2021. Metaalign: Coordinating domain alignment and classification for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 16643–16653.

[58] Guoqiang Wei, Cuiling Lan, Wenjun Zeng, Zhizheng Zhang, and Zhibo Chen. 2021. ToAlign: task-oriented alignment for unsupervised domain adaptation. In *Proceedings of the Conference on Neural Information Processing Systems*. 13834–13846.

[59] Garrett Wilson and Diane J Cook. 2020. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology (TIST)* 11, 5 (2020), 1–46.

[60] Jun Wu, Jingrui He, and Elizabeth Ainsworth. 2023. Non-IID Transfer Learning on Graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 10342–10350.

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

[62] Ying-Xin Wu, Xiang Wang, An Zhang, Xiangnan He, and Tat-Seng Chua. 2022. Discovering invariant rationales for graph neural networks. In *Proceedings of the International Conference on Learning Representations*.

[63] Zhiqing Xiao, Haobo Wang, Ying Jin, Lei Feng, Gang Chen, Fei Huang, and Junbo Zhao. 2023. SPA: a graph spectral alignment perspective for domain adaptation. In *Proceedings of the Conference on Neural Information Processing Systems*. 37252–37272.

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

[65] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How powerful are graph neural networks?. In *Proceedings of the International Conference on Learning Representations*.

[66] Hongliang Yan, Yukang Ding, Peihua Li, Qilong Wang, Yong Xu, and Wangmeng Zuo. 2017. Mind the class weight bias: Weighted maximum mean discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2272–2281.

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

[68] Yeonguk Yu, Sungho Shin, Seunghyeok Back, Mihwan Ko, Sangjun Noh, and Kyoobin Lee. 2024. Domain-Specific Block Selection and Paired-View Pseudo-Labeling for Online Test-Time Adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 22723–22732.

[69] An Zhang, Jingnan Zheng, Xiang Wang, Yancheng Yuan, and Tat-Seng Chua. 2023. Invariant collaborative filtering to popularity distribution shift. In *Proceedings of the Web Conference*. 1240–1251.

[70] Haodong Zhang, Tao Ren, Yifan Wang, Fanchun Meng, Wei Ju, and Ying Tian. 2025. Cluster-Aware Few-Shot Molecular Property Prediction With Factor Disentanglement. *IEEE Transactions on Neural Networks and Learning Systems* 36 (2025), 19644–19656.

[71] Muhan Zhang, Zhicheng Cui, Marion Neumann, and Yixin Chen. 2018. An end-to-end deep learning architecture for graph classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*. 4438–4445.

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

[73] Ganlong Zhao, Guanbin Li, Yipeng Qin, Feng Liu, and Yizhou Yu. 2022. Centrality and consistency: two-stage clean samples identification for learning with instance-dependent noisy labels. In *Proceedings of the European Conference on Computer Vision*. 21–37.

[74] Evgenii Zheltonozhskii, Chaim Baskin, Avi Mendelson, Alex M Bronstein, and Or Litany. 2022. Contrast to divide: Self-supervised pre-training for learning with noisy labels. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1657–1667.

Table 7: A statistical comparison of pseudo-label selection methods on the target domain on the NCI1 dataset. Metrics include the total number of confident samples, the count of correctly labeled samples, and the selection accuracy for four approaches: direct confidence thresholding (Variant 1, $\xi = 0.90$; Variant 2, $\xi = 0.98$) and our diffusion-based denoising module (Diffusion 1, $\tau = 10\%$; Diffusion 2, $\tau = 100\%$).

		$N_{D0}\to N_{D1}$	$N_{D0}\to N_{D2}$	$N_{D0}\to N_{D3}$	$N_{D1}\to N_{D0}$	$N_{D1}\to N_{D2}$	$N_{D1}\to N_{D3}$	$N_{D2}\to N_{D0}$	$N_{D2}\to N_{D1}$	$N_{D2}\to N_{D3}$	$N_{D3}\to N_{D0}$	$N_{D3}\to N_{D1}$	$N_{D3}\to N_{D2}$
Variant 1	Total	163	123	104	80	195	252	279	184	174	211	393	221
	Correct	119	83	52	65	149	167	195	136	145	156	219	144
	Ratio	73.4	67.5	50.0	78.5	76.2	66.3	70.0	73.9	83.3	74.2	55.7	65.2
Variant 2	Total	31	3	13	24	13	24	15	26	19	56	22	24
	Correct	26	2	6	20	11	17	11	20	15	40	14	18
	Ratio	84.2	66.7	46.2	83.3	84.6	70.8	73.3	76.9	78.9	71.4	63.6	75.0
Diffusion 1	Total	82	82	82	82	82	82	82	82	82	82	82	82
	Correct	71	62	49	74	65	69	77	65	71	72	53	59
	Ratio	86.6	75.6	60.4	90.2	79.3	84.2	93.9	79.3	86.6	87.8	64.6	72.0
Diffusion 2	Total	821	821	821	821	821	821	821	821	821	821	821	821
	Correct	524	494	363	609	565	555	586	535	564	423	447	511
	Ratio	63.8	60.2	44.2	74.3	68.8	67.6	71.4	65.2	68.7	51.5	54.5	62.2

Table 8: Comparison of computational efficiency. The table details the number of parameters and the pre-training/adaptation times (in seconds) for our method and the latest baselines across six datasets (P: PTC, N: NCI1, T: TWITTER-Real-Graph-Partial, M: Mutageniciy, B: BZR, C: COX2).

		P	N	T	M	B	C	# of Parameters
CoCo	Pre-training time (classifier)	5.60	12.71	327.52	12.68	5.81	5.26	0.23 M
	Adaptation time	0.91	0.96	16.80	1.08	1.35	0.97	
MTDF	Pre-training time (classifier)	5.80	13.67	382.45	12.37	4.87	5.02	0.67 M
	Adaptation time	1.45	2.14	14.40	2.55	1.14	1.20	
SLOGAN	Pre-training time (classifier)	4.93	12.12	303.93	11.65	5.27	5.30	0.19 M
	Adaptation time	1.22	1.51	11.31	2.18	1.36	1.04	
Ours	Pre-training time (classifier)	5.07	11.36	284.81	12.49	5.03	5.19	0.87 M
	Pre-training time (diffusion)	4.80	10.24	186.64	10.20	5.24	4.86	
	Adaptation time	0.92	0.98	10.70	1.29	1.08	0.89	

A Extra Experiments

A.1 Extra Analysis on Diffusion Model

In this section, we analyze the effectiveness of different pseudo-label selection strategies by examining the quantity and quality of the confident samples from the target domain on the NCI1 dataset. From the results shown in the Table 7, we can draw the following conclusions: ❶ The direct confidence thresholding methods, **Variant 1** and **Variant 2**, exhibit a precision-recall trade-off. The stricter threshold in **Variant 2** may yield higher selection accuracy (e.g. 84.6% in the $N_{D1}\to N_{D2}$ task), but at the cost of selecting a very limited number of samples. Conversely, **Variant 1** selects a large pool of pseudo-labels, yet suffers from lower and inconsistent accuracy. ❷ In contrast, our diffusion-based denoising module (**Diffusion 1**, $\tau = 10\%$) demonstrates a better balance in both count and accuracy. It consistently identifies a stable number of samples (82 across all tasks) while achieving relatively higher selection accuracy. ❸ The **Diffusion 2** variant, which straightly applies the diffusion model pre-trained on our source domain to our target domain without subsequent adaptation, yields the lowest accuracy across all tasks. This indicates that the subsequent domain adaptation process is essential to refine these labels and achieve optimal performance.

A.2 Complexity and Scalability Analysis

To more intuitively analyze the computational complexity and efficiency of our model, we compare with the latest baselines in Table 8, focusing on parameter counts and training time. While `DisCo` exhibits the largest number of parameters by incorporating a diffusion-based label denoising module, it maintains exceptional efficiency during the adaptation stage. This efficiency stems from a high-quality pseudo-label selection process and the straightforward yet potent feature disentangling strategy, in contrast to other methods that incur greater computational costs for adaptation refinement. Overall, our method is comparable with previous approaches, demonstrating a compelling trade-off between its superior performance and its efficiency.