TFGDA: Exploring Topology and Feature Alignment in Semi-supervised Graph Domain Adaptation through Robust Clustering

Jun Dan[1*], Weiming Liu[1*], Chunfeng Xie[2], Hua Yu[3], Shunjie Dong[4], Yanchao Tan[5,6,7†]
[1]Zhejiang University [2]Queen Mary University of London
[3]Dalian University of Technology [4]Shanghai Jiao Tong University [5]Fuzhou University
[6]Engineering Research Center of Big Data Intelligence, Ministry of Education
[7]Fujian Key Laboratory of Network Computing and Intelligent Information Processing
danjun@zju.edu.cn, 21831010@zju.edu.cn, c.xie@qmul.ac.uk
yhiccd@mail.dlut.edu.cn, sjdong@sjtu.edu.cn, yctan@fzu.edu.cn

Abstract

Semi-supervised graph domain adaptation, as a branch of graph transfer learning, aims to annotate unlabeled target graph nodes by utilizing transferable knowledge learned from a label-scarce source graph. However, most existing studies primarily concentrate on aligning feature distributions directly to extract domain-invariant features, while ignoring the utilization of the intrinsic structure information in graphs. Inspired by the significance of data structure information in enhancing models' generalization performance, this paper aims to investigate how to leverage the structure information to assist graph transfer learning. To this end, we propose an innovative framework called TFGDA. Specially, TFGDA employs a structure alignment strategy named STSA to encode graphs' topological structure information into the latent space, greatly facilitating the learning of transferable features. To achieve a stable alignment of feature distributions, we also introduce a SDA strategy to mitigate domain discrepancy on the sphere. Moreover, to address the overfitting issue caused by label scarcity, a simple but effective RNC strategy is devised to guide the discriminative clustering of unlabeled nodes. Experiments on various benchmarks demonstrate the superiority of TFGDA over SOTA methods.

1 Introduction

With the rise of deep learning, node classification techniques have made significant progress in diverse fields. However, due to the distribution shift issue, well-trained models often suffer severe performance degradation when applied directly to a new domain. Graph transfer learning (GTL) [1] has been proposed to tackle this issue by transferring domain-invariant features from a labeled source graph to an unlabeled target graph, effectively boosting the model's performance on the target graph.

Although current studies on GTL have made significant strides, they often rely on the assumption that all nodes in the source graph are labeled. However, this ideal assumption does not hold true in many scenarios, as annotating the entire source graph is time-consuming, especially for large-scale networks. Therefore, in this paper, we focus on a more realistic application scenario known as semi-supervised graph domain adaptation (SGDA) [2], where the source graph only has a few labeled nodes. Utilizing the transferable knowledge acquired from the label-scarce source graph to enhance the model's adaptation performance on the target graph is the most crucial challenge for SGDA.

[*] Equal Contribution, [†] Corresponding Author.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

However, most existing studies [3, 4, 1] tend to focus on directly aligning feature distributions across domains to extract domain-invariant node features, while overlooking the utilization of the intrinsic structure information in graphs. Notably, the recent advancements in unsupervised learning [5, 6, 7, 8, 9] have showcased the significance of data structure information in enhancing models' generalization. Considering the complex topological structure information presented in the graph, this paper seek to leverage these critical structure information to assist the transfer of shared node knowledge. To this end, we propose an innovative SGDA framework named **TFGDA** that employs a Subgraph Topological Structure Alignment (**STSA**) strategy to encode structure information into latent space. Specially, STSA utilizes persistent homology (PH) [10] to extract the topological structure information of the input and latent spaces, and then align the topological structures of two space, significantly improving the model's transfer performance. Furthermore, current works [2, 11, 4] primarily utilize adversarial training to reduce domain discrepancy. However, adversarial training is an unstable process that may destroy the discriminative details hidden in features [12], thereby affecting the transfer of shared knowledge. To remedy this issue, we propose a Sphere-guided Domain Alignment (**SDA**) strategy, aiming to achieve more stable domain alignment. Concretely, SDA initially maps node features to the spherical space. Then, geodesic projection [13] is utilized to project spherical features onto multiple great circles, where the spherical sliced-Wasserstein (SSW) distance [14] is employed to quantify the feature distributions discrepancy across domains.

More importantly, in the SGDA scenarios, due to the label scarcity of source graph, well-trained model on only a few labeled source nodes is likely to encounter overfitting. Consequently, it may make ambiguous or even incorrect predictions for certain target graph nodes located near the decision boundaries or far from their corresponding class centers. To address this overfitting issue, we devise a Robustness-guided Node Clustering (**RNC**) strategy to effectively enhance the model's robustness. RNC aims to guide the discriminative clustering of unlabeled nodes by maximizing the mutual information between the soft cluster assignment of the original node and its perturbed version, significantly improving the model's generalization performance on the target graph.

In summary, the main contributions are listed as follows:

(1) To the best of our knowledge, this is first attempt to utilize the intrinsic topological structure information hidden in graphs to assist GTL. A novel STSA strategy is proposed to preserve the topological structure information in latent space.

(2) A strategy named SDA is introduced to stably align the node feature distributions across domains.

(3) To address the overfitting issue, a simple but effective RNC strategy is devised to guide the discriminative clustering of unlabeled nodes.

(4) Experimental results show that our TFGDA outperforms SOTA methods on various benchmarks.

2 Related Works

Graph Transfer Learning. GTL [15, 16] has gained widespread attention for relieving the burden of collecting labeled data for new tasks. Early studies usually use source nodes to pre-train expressive models for related tasks in target domain [17, 18, 19, 20, 21]. To enhance model's generalization, recent studies have shifted their emphasis to domain adaptation [11, 22, 23]. There are two main ways to extract domain-invariant node features: (1) Using adversarial training to enforce domain confusion [16, 4, 2, 24]; (2) Minimizing the statistical distance between two domains [25, 26, 27, 28, 29]. However, existing works tend to focus on domain alignment while overlooking the utilization of structure information in graphs.

Semi-supervised Learning on Graphs. Semi-supervised learning on graphs addresses node classification with a small fraction of labeled nodes. Previous works [30, 31, 32] commonly adopt the message passing paradigm to extract discriminative features. Recent studies have explored various techniques, including adversarial training [33, 34], data augmentation [35, 36], continuous graph [37], contrastive learning [27] and meta learning [38, 39] to further enhance model's generalization.

Persistent Homology (PH). PH is an essential method in topology for extracting structure information from point clouds [40, 41]. Recently, PH has shown significant advantages in various areas, including signal processing[42], shape matching[43], and design of network [44]. Some studies have investigated its differentiability [45, 46]. Recent works have also explored its potential in image segmentation [47, 48], action/image recognition [49, 50, 51, 52, 53], and evaluation of GANs [54].

3 Preliminaries: Persistent Homology (PH)

PH is a method used to capture the topological structure of complex point clouds as a scale parameter ρ is varied. In this section, we briefly introduce some key concepts. Further details on PH can be found in [10, 40].

Notation. $\mathcal{X} := \{x_i\}_{i=1}^m$ denotes a point cloud and ω is a distance metric over $\mathcal{X}$.

Vietoris-Rips (VR) Complex. The VR complex [55] is a unique simplicial complex constructed from a set of points, providing an approximation of the underlying space's topology. The VR complex of $\mathcal{X}$ at scale ρ, denoted as $\mathcal{V}_\rho(\mathcal{X})$, contains all simplices of $\mathcal{X}$. Each component of $\mathcal{X}$ satisfies the distance constraint: $\omega(x_i, x_j) \leq \rho$ for any i, j. Additionally, the VR complex exhibits a nesting property: $\mathcal{V}_{\rho_i} \subseteq \mathcal{V}_{\rho_j}$ for any $\rho_i \leq \rho_j$, which enables us to track the evolution progress of simplical complex as ρ increases.

Persistence Diagram (PD). The PD dgm is a multi-set of points (g_1, g_2) in the Cartesian plane $\mathbb{R}^2$, encoding lifespan information of topological features. Concretely, it summarizes the birth time g_1 and death time g_2 information of each topological feature with a homology group. The birth time g_1 indicates the scale of feature creation and death time g_2 refers to the scale of feature destruction.

4 Methodology

4.1 Problem Definition

Source Domain Graph: The source graph is defined as $\mathcal{G}^s = (\mathcal{V}^{s,l}, \mathcal{V}^{s,u}, Y^{s,l}, A^s, X^s)$, where $\mathcal{V}^{s,l}$ is the labeled node set, and $\mathcal{V}^{s,u}$ is the remaining unlabeled node set in $\mathcal{G}_s$. $Y^{s,l} \in \mathbb{R}^{|\mathcal{V}^{s,l}| \times K}$ denotes the label matrix of $\mathcal{V}^{s,l}$, where K is the number of node classes. If a node $n_i^s \in \mathcal{V}^{s,l}$ belongs to the k-th class, $y_{i,k}^s = 1$; otherwise, $y_{i,k}^s = 0$. $A^s \in \mathbb{R}^{N^s \times N^s}$ is an adjacency matrix, where $N^s = |\mathcal{V}^{s,l}| + |\mathcal{V}^{s,u}|$ is the number of nodes in $\mathcal{G}^s$. If there is an edge between nodes n_i and n_j, the value of A_{ij}^s is set to 1; otherwise, it is set to 0. $X^s \in \mathbb{R}^{N^s \times e}$ indicates an attribute matrix, where e is the dimension of node attributes. Notably, $|\mathcal{V}^{s,l}|$ is much smaller than $|\mathcal{V}^{s,u}|$ in the SGDA setting.

Target Domain Graph: Similarly, the target graph is represented as $\mathcal{G}^t = (\mathcal{V}^t, A^t, X^t)$, which is a completely unlabeled graph with an unlabeled node set $\mathcal{V}^t$. $A^t \in \mathbb{R}^{N^t \times N^t}$ is an adjacency matrix, and $X^t \in \mathbb{R}^{N^t \times e}$ is a node attribute matrix, where $N^t = |\mathcal{V}^t|$ denotes the number of nodes in $\mathcal{G}_t$.

Semi-Supervised Graph Domain Adaptation (SGDA): Given a partially labeled source graph $\mathcal{G}^s$ and an unlabeled target graph $\mathcal{G}^t$, the goal of SGDA is to precisely annotate target graph nodes by utilizing transferable knowledge learned from the limited labeled source nodes [2, 16].

4.2 Network Architecture

The architecture of our TFGDA model is depicted in Figure 1. It consists of two components: a graph convolutional network (GCN)-based feature extractor $\mathcal{F}$ and a node classifier $\mathcal{C}$. Mathematically, given an input graph $\mathcal{G} = (\mathcal{V}, A, X)$, the node features extracted by $\mathcal{F}$ is denoted as $H = \mathcal{F}(\mathcal{G}) \in \mathbb{R}^{|\mathcal{V}| \times d}$, and it is further normalized to map onto a spherical space $\mathbb{S}_r^{d-1}$ to obtain spherical features $Z \in \mathbb{R}^{|\mathcal{V}| \times d}$, where d is the feature dimension, r is the radius and $|\mathcal{V}|$ denotes the number of nodes in $\mathcal{G}$. The classification probability predicted by $\mathcal{C}$ is denoted as $\Psi = \mathcal{C}(Z) \in \mathbb{R}^{|\mathcal{V}| \times K}$.

To capture more precise adjacency relationships of graph $\mathcal{G}$, we compute the positive point-wise mutual information (PPIM) between nodes following [2]. Specially, for a given graph $\mathcal{G} = (\mathcal{V}, A, X)$, we utilize random walk to sample a collection of paths on A and generate a frequency matrix R. Based on R, we can compute the PPIM matrix $\mathbb{P}$ as follows:

$$\begin{gathered}\mathbb{P}_{ij} = \frac{R_{ij}}{\sum_{i,j} R_{ij}},\ \mathbb{P}_{i,*} = \frac{\sum_j R_{ij}}{\sum_{i,j} R_{ij}},\ \mathbb{P}_{*,j} = \frac{\sum_i R_{ij}}{\sum_{i,j} R_{ij}}, \\ P_{ij} = \max\{\log(\frac{\mathbb{P}_{ij}}{\mathbb{P}_{i,*} \times \mathbb{P}_{*,j}}), 0\},\end{gathered} \tag{1}$$

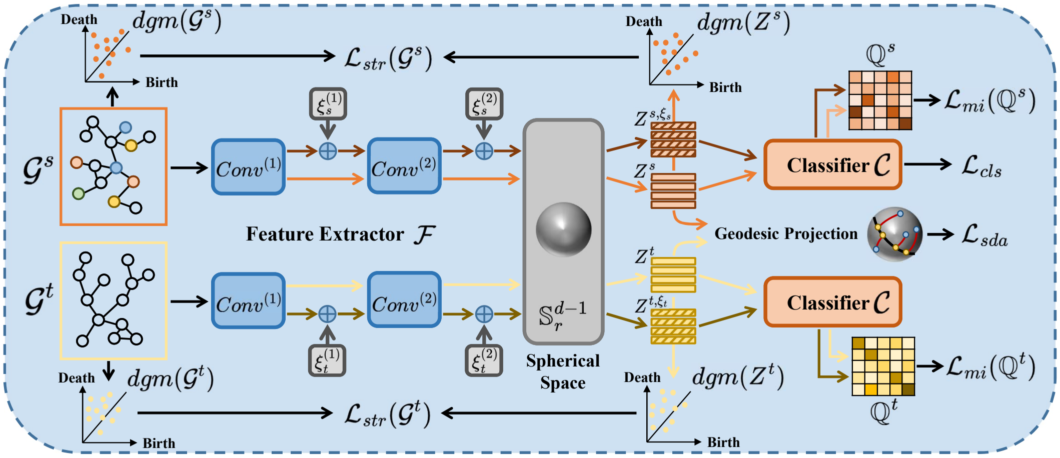


Figure 1: Global overview of the TFGDA model. STSA strategy encodes critical structure information of graphs into spherical space ($\mathcal{L}_{stsa} = \mathcal{L}_{str}(\mathcal{G}^s) + \mathcal{L}_{str}(\mathcal{G}^t)$), greatly improving the model's generalization. SDA strategy aims to extract domain-invariant node features by minimizing domain discrepancy on sphere ($\mathcal{L}_{sda}$). Moreover, to effectively solve the overfitting issue, RNC strategy is introduced to guide the discriminative clustering of unlabeled nodes ($\mathcal{L}_{rnc} = \mathcal{L}_{mi}(\mathbb{Q}^s) + \mathcal{L}_{mi}(\mathbb{Q}^t)$).

where P_{ij} denotes the positive mutual information between nodes n_i and n_j, which quantifies the topological proximity between nodes. A higher value of P_{ij} indicates a strong connection between n_i and n_j. Then, the output of the l-th GCN layer $Conv^{(l)}(\cdot)$ is defined as:

$$H^{(l)} = Conv^{(l)}(P, H^{(l-1)}) = \sigma(D^{-\frac{1}{2}} \widetilde{P} D^{-\frac{1}{2}} H^{(l-1)} W^{(l)}), \tag{2}$$

where D is the diagonal degree matrix of P, and $\widetilde{P} = P + I$ (I is a identity matrix). $W^{(l)}$ refer to the trainable parameters of the l-th layer, $\sigma(\cdot)$ is an activation function, and $H^{(0)} = X$. Thus, the feature extractor $\mathcal{F}$ can be constructed by sequentially stacking L layers of GCN $Conv^{(l)}(l = 1, 2, \cdots, L)$. Given the source labeled node set $\mathcal{V}^{s,l}$, the classification loss on the source graph $\mathcal{G}^s$ is defined as:

$$\mathcal{L}_{cls} = \frac{1}{|\mathcal{V}^{s,l}|} \sum_{n_i^s \in \mathcal{V}^{s,l}} \mathcal{L}_{ce}(\mathcal{C}(\mathcal{F}(n_i^s)), y_i^s). \tag{3}$$

where $\mathcal{L}_{ce}$ represents the standard cross-entropy loss.

4.3 Subgraph Topological Structure Alignment

As mentioned in Section 1, the graph consists of numerous nodes (e.g., *ACMv9* has over 9000 nodes), thereby containing rich structure information. Inspired by the significance of data structure information in enhancing models' generalization [56, 51, 6], we seek to leverage such critical structure information to facilitate the learning of domain-invariant features.

To achieve this goal, we treat the graph as a point cloud and attempt to directly capture its underlying topological structure using PH. However, we inevitably encounter huge computational burden due to the complex attributes and adjacency relationships of graph. Fortunately, Refs. [57, 58] have indicated that the properties of graph can be well preserved in its multiple local subgraphs. To this end, we propose a Subgraph Topological Structure Alignment (**STSA**) strategy to encode structure information of input space into the latent space. For a given graph $\mathcal{G} = (\mathcal{V}, A, X)$, STSA first sample a subgraphs $\left\{\hat{\mathcal{G}}_1, \hat{\mathcal{G}}_2, \cdots, \hat{\mathcal{G}}_a\right\}$ using random walk, and then employ PH to capture the intrinsic topological structure information of each subgraph $\hat{\mathcal{G}}_i$.

Specially, for each subgraph $\hat{\mathcal{G}}_i$, we represent its spherical features extracted by $\mathcal{F}$ as $\hat{Z}_i \in \mathbb{R}^{|\hat{\mathcal{V}}_i| \times d}$, which are obtained by indexing the local nodes features from the complete nodes features $Z \in \mathbb{R}^{|\mathcal{V}| \times d}$. Here $|\hat{\mathcal{V}}_i|$ denotes the number of nodes in each subgraph. Then we construct the VR complexes

$\mathcal{V}_\rho(\hat{\mathcal{G}}_i)$ and $\mathcal{V}_\rho(\hat{Z}_i)$ for point clouds $\hat{\mathcal{G}}_i$ and $\hat{Z}_i$, employ PH to extract their topological structures, and obtain their corresponding PDs $dgm(\hat{\mathcal{G}}_i)$ and $dgm(\hat{Z}_i)$ respectively. To align the topological structures of the input and latent spherical spaces, we adopt the 1-Wasserstein distance $\mathcal{W}_1$ to measure the discrepancy between two PDs (i.e., $dgm(\hat{\mathcal{G}}_i)$ for input space and $dgm(\hat{Z}_i)$ for spherical space), which aims to seek the optimal transport plan γ^* between two PDs:

$$\begin{aligned}\gamma^* &= \arg\min_{\gamma} \mathcal{W}_1(dgm(\hat{\mathcal{G}}_i), dgm(\hat{Z}_i)) \\ &= \arg\min_{\gamma} \sum_{(\alpha,\beta)\in\gamma} \|\alpha - \beta\|_\infty\end{aligned} \tag{4}$$

where $\alpha \in dgm(\hat{\mathcal{G}}_i)$, $\beta \in dgm(\hat{Z}_i)$, and $\|\cdot\|_\infty$ is the l_∞ distance. After obtaining γ^*, the local structure discrepancy $\mathcal{L}_{str}^{sub}(\hat{\mathcal{G}}_i)$ between two spaces of $\hat{\mathcal{G}}_i$ can be calculated as:

$$\mathcal{L}_{str}^{sub}(\hat{\mathcal{G}}_i) = \sum_{(\alpha,\beta)\in\gamma^*} \|\alpha - \beta\|_2^2 . \tag{5}$$

Hence, we can estimate the global structure discrepancy between the input and spherical spaces of graph $\mathcal{G}$ by aggregating the local structure discrepancy of all subgraphs:

$$\mathcal{L}_{str}(\mathcal{G}) = \sum_i \mathcal{L}_{str}^{sub}(\hat{\mathcal{G}}_i) \tag{6}$$

In the SGDA scenario, the STSA strategy will be applied to both the $\mathcal{G}^s$ and $\mathcal{G}^t$, and its loss function is defined as:

$$\mathcal{L}_{stsa} = \mathcal{L}_{str}(\mathcal{G}^s) + \mathcal{L}_{str}(\mathcal{G}^t). \tag{7}$$

Notably, preserving topological structure information has the potential to guide unlabeled nodes towards achieving discriminative clustering, thereby promoting the learning of transferable node features, as verified in Section 5.3.

Although many GCNs-based methods have already been proposed to exploit graph structure information to promote the learning of features, these methods are not effective in addressing the SGDA task. Specifically, these GCNs-based methods [59, 60, 61, 62] typically mine the graph structure information in the deep feature space by designing well-crafted GCN architectures or introducing some complex modules. However, recent studies [27, 63, 64, 16] have pointed out that GCNs are insufficient in capturing the sophisticated structure information in graph, which means that the graph structure information may be lost or destroyed after passing through the GCNs-based feature extractors. Thus, directly mining graph structure information from the deep feature space is a suboptimal way, which affects the learning of transferable node features in our SGDA setting.

The proposed STSA strategy aims to extract the graph structure information directly from the input space and encode these powerful information into the latent spherical space by aligning the topological structures of the two spaces. This method does not lose or destroy the graph structure information during training. Furthermore, our STSA strategy does not introduce any changes to the network architecture, effectively avoiding an increase in model's complexity and ensuring its adaptability to integration with other methods.

4.4 Sphere-guided Domain Alignment

As mentioned in Section 1, adversarial training has been widely adopted by existing GTL models to reduce domain discrepancy. However, it is an unstable process that may destroy the discriminative information hidden in node features, thereby impacting the learning of shared features.

To tackle this issue, we propose a Sphere-guided Domain Alignment (SDA) strategy that achieves stable alignment of cross-domain node features distributions in the spherical space. Our SDA strategy mainly comprises three steps: **(1)** Map node features onto the sphere space $\mathbb{S}_r^{d-1}$. **(2)** Use geodesic projection [13] to project node features from $\mathbb{S}_r^{d-1}$ to multiple great circles. **(3)** Compute the feature distributions discrepancy on great circles and minimize it during training.

Step (1): Motivated by the effectiveness of spherical features in improving model's transfer performance [65], we first normalize node features $H \in \mathbb{R}^{|\mathcal{V}| \times d}$ extracted by $\mathcal{F}$ with $z_i = r\frac{h_i}{\|h_i\|}$ to obtain the spherical features $Z \in \mathbb{R}^{|\mathcal{V}| \times d}$ in the sphere space $\mathbb{S}_r^{d-1} = \left\{z_i \in \mathbb{R}^d : \|z_i\|_2 = r\right\}$, where h_i and z_i is the i-th row of H and Z, respectively. Notably, Ref. [66] has proved that a proper radius r is lower bounded by parameters τ and K:

$$r \geq \frac{K-1}{K} \ln \frac{(K-1)\tau}{1-\tau} \tag{8}$$

where τ denotes the expected minimal classification probability of class center and K is the number of classes. In this work, τ is set to 0.999, and radius r is set to the lower bound.

Step (2): Previous studies [67, 68, 69, 70] have shown the superiority of optimal transport in aligning feature distributions. Let Ω be a probability space and μ, ν be two probability measures in $\mathcal{P}(\Omega)$. For any $q \geq 1$, the q-Wasserstein distance between μ and ν is defined as:

$$\mathcal{W}_q^q(\mu, \nu) = \inf_{\gamma \in \Pi(\mu,\nu)} \int_{\Omega \times \Omega} \mathcal{M}^q(z, g) d\gamma(z, g) \tag{9}$$

where $\Pi(\mu, \nu) = \{\gamma \in \mathcal{P}(\Omega \times \Omega) | \pi_{1\#}\gamma = \mu, \pi_{2\#}\gamma = \nu\}$ is the set of couplings, π_1 and π_2 denote the two marginal projections of $\Omega \times \Omega$ to Ω, $\#$ denotes the push-forward operator, and $\mathcal{M} : \Omega \times \Omega \to \mathbb{R}^+$ is a geodesic metric. However, we find that directly using the classical Wasserstein distance to compute features distributions discrepancy on the sphere $\mathbb{S}_r^{d-1}$ is computationally expensive, due to numerous nodes in the graph.

For more efficient calculations, we utilize geodesic projection I^U to project node features lying on $\mathbb{S}_r^{d-1}$ to b great circles $\{\mathbb{C}_1, \mathbb{C}_2, \cdots, \mathbb{C}_b\}$. On the hypersphere, great circles [71] are circles whose diameter is equal to that of the sphere, and they correspond to the geodesics. Specially, the geodesic projection I^U is determined by U:

$$I^U(z) = U^\top \underset{g \in \mathrm{span(UU^\top)} \cap \mathbb{S}_\mathrm{r}^{\mathrm{d}-1}}{\arg\min} \mathcal{M}_{\mathbb{S}_r^{d-1}}(z, g) = \underset{\varrho \in \mathbb{S}_r^1}{\arg\min} \mathcal{M}_{\mathbb{S}_r^{d-1}}(z, U\varrho), \forall U \in \mathbb{V}_{d,2}, \forall z \in \mathbb{S}_r^{d-1}. \tag{10}$$

where $\mathcal{M}_{\mathbb{S}_r^{d-1}}(z, g) = \arccos(\langle z, g \rangle)$, and $\mathbb{V}_{d,2} = \left\{U \in \mathbb{R}^{d \times d}, U^\top U = I_2\right\}$ is the Stiefel manifold [72].

Step (3): Next, we utilize the spherical sliced-Wasserstein (SSW) distance [14] to measure the feature distributions discrepancy of two domain on multiple great circles, which can be formulated as:

$$SSW_q^q(u, v) = \int_{\mathbb{V}_{d,2}} \mathcal{W}_q^q(I_\#^U \mu, I_\#^U \nu) \mathrm{d}\sigma(\mathrm{U}), \tag{11}$$

where σ is the uniform distribution over $\mathbb{V}_{d,2}$. In SSW, the geodesic metric $\mathcal{M}^q(z, g)$ in Wasserstein distance $\mathcal{W}_q^q(\mu, \nu)$ is defined as the geodesic distance [71] $\mathcal{M}_{\mathbb{S}_r^1}(z, g) = \min(|z - g|, r - |z - g|)$.

In practice, it's common to approximate the source and target feature distributions using samples $(z_i^s)_{i=1}^{N^s}$ and $(z_j^t)_{j=1}^{N^t}$ (i.e., through the empirical approximations $\tilde{\mu} = \frac{1}{N^s} \sum_{i=1}^{N^s} \delta_{z_i^s}$ and $\tilde{\nu} = \frac{1}{N^t} \sum_{j=1}^{N^t} \delta_{z_j^t}$), where δ is the Dirac function. As a result, the node features distributions discrepancy between two domains can be measured as:

$$\mathcal{L}_{sda} = SSW_p^p(\tilde{\mu}, \tilde{\nu}) \approx \frac{1}{b} \textstyle\sum_{m=1}^b \mathcal{W}_q^q(\tilde{\mu}, \tilde{\nu}) \tag{12}$$

where b is the number of projections. As training progresses, SDA strategy gradually reduces domain discrepancy, making the learning of domain-invariant features easier.

4.5 Robustness-guided Node Clustering

Due to the label scarcity in $\mathcal{G}^s$, the model is prone to overfitting when solely relying on $\mathcal{L}_{cls}$ for optimization, severely degrading the model's generalization performance on $\mathcal{G}^t$. To alleviate this overfitting issue, we devise a novel Robustness-guided Node Clustering (**RNC**) strategy to enhance the model's robustness by guiding the discriminative clustering of unlabeled nodes. Specially, RNC first introduces trainable shift parameters $\xi_s = \left\{\xi_s^{(1)}, \xi_s^{(2)}, \cdots, \xi_s^{(L)}\right\}$ and $\xi_t = \left\{\xi_t^{(1)}, \xi_t^{(2)}, \cdots, \xi_t^{(L)}\right\}$ to perturbs the source and target node features respectively, at each layer of $\mathcal{F}$:

$$H^{s/t, \xi_{s/t}, (l)} = \begin{cases} Conv^{(l)}(P^{s/t}, X^{s/t}) + \xi_{s/t}^{(l)}, & l = 1 \\ Conv^{(l)}(P^{s/t}, H^{s/t,(l-1)}) + \xi_{s/t}^{(l)}, & 1 < l \leq L \end{cases} \tag{13}$$

where $H^{s,\xi_s,(l)}$ and $H^{t,\xi_t,(l)}$ denotes the perturbed source and target node features encoded by $Conv^{(l)}$ respectively, and each $\xi^{(i)}_{s/t}$ is specific to the output of $Conv^{(l)}$. The shift parameters ξ_s and ξ_t are defined as randomly initialized multi-layer parameter matrices. After perturbation, based on $H^{s,\xi_s,(L)}$ and $H^{t,\xi_t,(L)}$, we can obtain the perturbed source node spherical features Z^{s,ξ_s} and the perturbed target node spherical features Z^{t,ξ_t} respectively.

For brevity, we will omit domain-specific notations in the following text. Ideally, regardless of how the node feature is perturbed, the model's prediction for it should remain unchanged because its class label has not changed. To achieve this goal, RNC aims to maximize the mutual information between the soft cluster assignment (i.e., classification prediction of classifier $\mathcal{C}$) of the spherical features Z and its perturbed version Z^{ξ}, capturing their intrinsic invariant information between Z and Z^{ξ}.

Concretely, given a node spherical feature z_i, the classification probability predicted by $\mathcal{C}$ is denoted as $\Psi(z_i) \in \mathbb{R}^K$, that can be viewed as the distribution of a discrete random variable ϕ over K classes: $\mathit{\Gamma}(\phi = k|z) = \Psi_k(z_i)$. Let ϕ and ϕ^{ξ} denote the cluster assignment variables of z_i and z_i^{ξ}, respectively. Then, their conditional joint distribution is defined as: $\mathit{\Gamma}(\phi = k, \phi^{\xi} = k^{\xi}|z_i, z_i^{\xi}) = \Psi_k(z_i) \cdot \Psi_{k^{\xi}}(z_i^{\xi})$. After marginalization, the joint probability distribution can be formulated as a matrix $\mathbb{Q} \in \mathbb{R}^{K \times K}$:

$$\mathbb{Q} = \frac{1}{|\mathcal{V}|} \textstyle\sum_{i=1}^{|\mathcal{V}|} \Psi(z_i) \cdot \Psi(z_i^{\xi})^{\top}, \tag{14}$$

where $\mathbb{Q}_{kk^{\xi}} = \mathit{\Gamma}(\phi = k, \phi^{\xi} = k^{\xi})$, $\mathbb{Q}_k = \mathit{\Gamma}(\phi = k)$, and $\mathbb{Q}_{k^{\xi}} = \mathit{\Gamma}(\phi^{\xi} = k^{\xi})$. To preserve the equivalence between pairs (z_i, z_i^{ξ}) and (z_i^{ξ}, z_i), matrix $\mathbb{Q}$ is typically symmetrized using $(\mathbb{Q}+\mathbb{Q}^{\top})/2$. In this way, the mutual information [73, 74] between the soft cluster assignment of Z and Z^{ξ} can be computed as:

$$\mathcal{L}_{mi}(\mathbb{Q}) = \sum_{k=1}^{K} \sum_{k^{\xi}=1}^{K} \mathbb{Q}_{kk^{\xi}} \cdot \ln \frac{\mathbb{Q}_{kk^{\xi}}}{\mathbb{Q}_k \cdot \mathbb{Q}_{k^{\xi}}}, s.t., ||\xi^{(l)}||_F \leq \varpi, \forall \xi^{(l)} \in \xi. \tag{15}$$

where ϖ is a coefficient that controls the scale of feature perturbation.

In the SGDA scenario, all target domain nodes $\mathcal{V}_t$ and source domain nodes $\mathcal{V}^{s,l} \cup \mathcal{V}^{s,u}$ are used to calculate matrices $\mathbb{Q}^t$ and $\mathbb{Q}^s$, respectively. Therefore, the objective function $\mathcal{L}_{rnc}$ of RNC can be expressed as:

$$\begin{gathered} \mathcal{L}_{rnc} = \mathcal{L}_{mi}(\mathbb{Q}^s) + \mathcal{L}_{mi}(\mathbb{Q}^t), \\ s.t., ||\xi^{s,(l)}||_F \leq \varpi, \forall \xi^{s,(l)} \in \xi^s, ||\xi^{t,(l)}||_F \leq \varpi, \forall \xi^{t,(l)} \in \xi^t. \end{gathered} \tag{16}$$

Notably, we leverage source labeled nodes $\mathcal{V}^{s,l}$ in RNC as they can significantly guide the discriminative clustering of unlabeled nodes in the right direction. With the help of RNC, more and more intrinsic invariant features of nodes are extracted, which greatly promotes the learning of transferable features. Moreover, unlike previous studies that employ pseudo-labels strategy [2] or conditional entropy term [75] to guide the learning of unlabeled nodes, our RNC strategy does not involve any pseudo-labels and naturally avoids degenerate clustering solutions (see Figure 3 for further analysis).

4.6 Model Optimization

In summary, the total objective of TFGDA can be expressed as follows:

$$\begin{gathered} \min_{\mathcal{F},\mathcal{C},\xi^s,\xi^t} \mathcal{L}_{cls} + \eta\mathcal{L}_{sda} + \varepsilon\mathcal{L}_{stsa} - \lambda\mathcal{L}_{rnc} \\ s.t., ||\xi^{s,(l)}||_F \leq \varpi, \forall \xi^{s,(l)} \in \xi^s, ||\xi^{t,(l)}||_F \leq \varpi, \forall \xi^{t,(l)} \in \xi^t. \end{gathered} \tag{17}$$

where hyper-parameters η, ε and λ are used to balance the contributions of the corresponding term.

4.7 Theoretical Analysis

The theoretical analysis of our method is based on the theory of domain adaptation (DA) [76, 77].

Formally, let $\mathcal{H}$ be the hypothesis space. Given two domains $\mathcal{S}$ and $\mathcal{T}$, the probabilistic bound of error of hypothesis h on the target domain is defined as: $\psi_{\mathcal{T}}(h) \leq \psi_{\mathcal{S}}(h) + \frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S}, \mathcal{T}) + \mu^*$, where the expected error on the target domain $\psi_{\mathcal{T}}(h)$ are bounded by three terms: **(1)** the expected error on source domain $\psi_{\mathcal{S}}(h)$; **(2)** the $\mathcal{H}\Delta\mathcal{H}$-divergence between the source and target domains $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S},\mathcal{T})$; **(3)** the combined error of ideal joint hypothesis $\mu^* = \min_{h' \in \mathcal{H}} \psi_{\mathcal{S}}(h') + \psi_{\mathcal{T}}(h')$.

The goal of DA is to lower the upper bound of the expected target domain error $\psi_{\mathcal{T}}(h)$. Note that in unsupervised domain adaptation (UDA), minimizing $\psi_{\mathcal{S}}(h)$ can be easily achieved with source label information, as source domain samples are completely annotated. However, in our SGDA setting, due to the label scarcity of source domain, the model is prone to overfitting when solely relying on the source domain classification loss $\mathcal{L}_{cls}$ for optimization. Therefore, we introduce the **RNC** strategy ($\mathcal{L}_{rnc}$) to address the overfitting issue, with the aim of guiding $\psi_{\mathcal{S}}(h)$ towards further minimization.

Most DA methods mainly focus on reducing the domain discrepancy $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S},\mathcal{T})$, such as utilizing techniques like adversarial learning, MMD, optimal transport [78, 79, 80, 81], and CORAL. In comparison to these methods, our **SDA** strategy ($\mathcal{L}_{sda}$) effectively eliminates the feature norm discrepancy in spherical space $\mathbb{S}_r^{d-1}$ and guide a more stable alignment of feature distributions. Furthermore, considering that graph data contains rich structure information that encodes complex relationships among nodes and edges, and existing GTL methods usually adopt GCNs-based feature extractors to learn domain-invariant node features. However, recent studies [27, 63, 64, 16] have pointed out that GCNs are insufficient in capturing the sophisticated structure information in graph, which seriously affects the transfer of domain-invariant knowledge and consequently limits the model's generalization ability. To solve this problem, we thus propose the **STSA** strategy ($\mathcal{L}_{stsa}$) to align the topological structures of the input space and the spherical space, in order to facilitate the GCNs-based feature extractors to capture more domain-invariant node features. Consequently, the combination of **SDA** ($\mathcal{L}_{sda}$) and **STSA** ($\mathcal{L}_{stsa}$) strategies further promotes the minimization of the domain discrepancy $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S},\mathcal{T})$.

Notably, μ^* is expected to be extremely small, and therefore it is often neglected by previous methods. However, it is possible that μ^* tends to be large when the cross-domain category distributions are not well aligned [82, 83]. In this paper, we leverage the **RNC** strategy ($\mathcal{L}_{rnc}$) to guide both labeled nodes and unlabeled nodes toward achieving robust clustering, effectively promoting the fine-grained alignment of category distributions and ensuring that μ^* remains at a relatively small value.

In summary, our proposed method not only minimizes the source expected error $\psi_{\mathcal{S}}(h)$ and domain discrepancy $d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{S},\mathcal{T})$, but also keeps μ^* at a small value, thereby ensuring a low upper bound.

5 Experiments

5.1 Setup

Datasets. Our experiments involve three real-world graphs: *ACMv9* (**A**), *Citationv1* (**C**), and *DBLPv7* (**D**), obtained from ArnetMiner [84]. Since these graphs have varying sets of node attributes, we union their attribute sets and adjust the attribute dimension to 6775 following [2]. Each node is assigned a five-class label, determined by its relevant research areas. Six typical transfer tasks are considered in our experiments: **A**→**C**, **A**→**D**, **C**→**A**, **C**→**D**, **D**→**A** and **D**→**C**. Due to the page size limitation, more settings and implementation details are placed on **Appendix**.

Compared Methods. We compare TFGDA with several SOTA **(1)** graph semi-supervised learning methods and **(2)** graph domain adaptation methods as Ref.[2]: **(1) GCN** [30], **GSAGE** [31], **GAT**[32], **GIN** [85], **(2) DANN** [86], **CDAN** [12], **UDA-GCN** [75], **AdaGCN** [1], **CoCo** [27], **StruRW** [22] and **SGDA** [2]. DANN$_{GCN}$ and CDAN$_{GCN}$ are two variants that adopt GCN-based feature extractor.

5.2 Results and Discussion

Following [2], to showcase the superiority of our TFGDA, we report its performance on the challenging scenario, where only **5%** of the nodes in the source graph are labeled. **Micro-F1** and **Macro-F1** are employed as evaluation metrics, and the classification results on the target graph are gathered in Table 1. For all transfer tasks, we run each experiment 5 times and record the average accuracy with standard deviation. And we sample different label sets each time to mitigate the randomness. As can be seen, our model obtains the overall best results on all transfer tasks. Specially, TFGDA greatly surpasses the SOTA method SGDA [2] by $+7.3\%$ and $+10.0\%$ on "Micro-F1" and "Macro-F1" respectively for the **C**→**A** task, implying the superiority in extracting domain-invariant features. Notably, TFGDA enhances performance substantially on two hard transfer tasks, **C**→**A** and **D**→**A**,

Table 1: Transfer performance (%) on six transfer tasks with a source graph label rate of 5% for semi-supervised graph domain adaptation.

Methods	**A→C**		**A→D**		**C→A**		**C→D**		**D→A**		**D→C**	
	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1	Micro-F1	Macro-F1
MLP [2]	41.3$_{\pm1.15}$	35.8$_{\pm0.72}$	42.8$_{\pm0.88}$	36.3$_{\pm0.77}$	39.4$_{\pm0.57}$	33.7$_{\pm0.58}$	43.7$_{\pm0.69}$	36.7$_{\pm0.55}$	37.3$_{\pm0.32}$	30.8$_{\pm0.37}$	39.4$_{\pm0.99}$	32.8$_{\pm0.99}$
GCN [30]	54.4$_{\pm1.52}$	52.0$_{\pm1.62}$	56.9$_{\pm2.33}$	53.4$_{\pm2.81}$	54.1$_{\pm1.40}$	52.3$_{\pm1.98}$	58.9$_{\pm0.99}$	54.5$_{\pm1.55}$	50.1$_{\pm2.14}$	48.0$_{\pm3.28}$	56.0$_{\pm1.24}$	51.9$_{\pm1.49}$
GSAGE [31]	49.3$_{\pm2.18}$	46.4$_{\pm2.06}$	51.8$_{\pm1.35}$	47.4$_{\pm1.62}$	46.8$_{\pm2.56}$	45.0$_{\pm2.78}$	51.7$_{\pm1.95}$	48.1$_{\pm1.97}$	41.7$_{\pm2.17}$	37.4$_{\pm4.59}$	45.4$_{\pm2.11}$	39.3$_{\pm3.45}$
GAT [32]	55.1$_{\pm3.22}$	50.8$_{\pm1.45}$	55.3$_{\pm2.52}$	51.8$_{\pm2.60}$	50.0$_{\pm1.20}$	45.6$_{\pm2.36}$	55.4$_{\pm2.73}$	49.2$_{\pm2.59}$	44.8$_{\pm2.74}$	38.3$_{\pm4.84}$	50.4$_{\pm3.35}$	42.0$_{\pm4.46}$
GIN [85]	64.6$_{\pm2.47}$	56.0$_{\pm2.73}$	60.0$_{\pm2.09}$	51.3$_{\pm3.99}$	57.1$_{\pm1.19}$	54.4$_{\pm2.57}$	62.0$_{\pm1.05}$	56.8$_{\pm1.40}$	51.9$_{\pm2.00}$	45.4$_{\pm2.16}$	60.2$_{\pm3.05}$	53.0$_{\pm2.10}$
DANN [86]	44.3$_{\pm2.03}$	39.3$_{\pm1.86}$	44.0$_{\pm1.42}$	38.7$_{\pm1.47}$	41.8$_{\pm1.95}$	37.6$_{\pm1.24}$	45.5$_{\pm0.71}$	39.6$_{\pm1.55}$	37.8$_{\pm3.66}$	33.2$_{\pm2.23}$	41.7$_{\pm2.32}$	35.6$_{\pm2.55}$
CDAN [12]	44.6$_{\pm1.30}$	38.6$_{\pm1.07}$	45.5$_{\pm0.85}$	38.0$_{\pm0.86}$	42.4$_{\pm0.64}$	36.2$_{\pm1.17}$	46.7$_{\pm1.17}$	39.2$_{\pm0.96}$	39.0$_{\pm1.08}$	32.3$_{\pm1.09}$	41.7$_{\pm1.55}$	34.8$_{\pm1.56}$
DANN$_{GCN}$ [2]	63.0$_{\pm6.75}$	59.6$_{\pm6.02}$	62.2$_{\pm1.90}$	57.7$_{\pm3.16}$	56.7$_{\pm0.38}$	55.2$_{\pm1.03}$	65.3$_{\pm2.04}$	59.0$_{\pm2.39}$	52.3$_{\pm2.59}$	48.6$_{\pm4.52}$	58.1$_{\pm2.78}$	52.4$_{\pm3.81}$
CDAN$_{GCN}$ [2]	70.3$_{\pm0.84}$	66.5$_{\pm0.66}$	65.0$_{\pm1.00}$	61.3$_{\pm0.96}$	56.3$_{\pm1.78}$	53.6$_{\pm2.70}$	65.2$_{\pm2.19}$	58.8$_{\pm2.38}$	53.0$_{\pm1.34}$	48.7$_{\pm3.51}$	59.0$_{\pm1.52}$	53.3$_{\pm1.99}$
UDA-GCN [75]	72.4$_{\pm2.75}$	65.2$_{\pm6.51}$	68.0$_{\pm6.38}$	64.3$_{\pm7.12}$	62.9$_{\pm0.33}$	62.2$_{\pm1.44}$	71.4$_{\pm2.56}$	67.5$_{\pm2.25}$	55.8$_{\pm3.50}$	52.4$_{\pm2.68}$	65.2$_{\pm4.41}$	60.7$_{\pm6.84}$
AdaGCN [1]	70.8$_{\pm0.95}$	68.5$_{\pm0.73}$	68.2$_{\pm3.84}$	64.2$_{\pm3.91}$	61.5$_{\pm2.20}$	60.4$_{\pm3.15}$	69.1$_{\pm1.96}$	65.8$_{\pm2.87}$	56.1$_{\pm1.75}$	53.8$_{\pm2.95}$	64.1$_{\pm0.91}$	62.8$_{\pm1.56}$
CoCo [27]	72.7$_{\pm1.36}$	66.8$_{\pm1.15}$	68.3$_{\pm2.31}$	64.1$_{\pm2.68}$	62.7$_{\pm0.95}$	61.5$_{\pm1.18}$	71.6$_{\pm1.76}$	67.3$_{\pm1.93}$	56.7$_{\pm1.47}$	54.1$_{\pm1.29}$	66.0$_{\pm0.88}$	64.4$_{\pm1.13}$
StruRW [22]	72.9$_{\pm1.21}$	67.1$_{\pm1.07}$	68.5$_{\pm0.94}$	64.4$_{\pm1.03}$	63.6$_{\pm1.05}$	61.9$_{\pm1.19}$	71.8$_{\pm2.06}$	67.6$_{\pm2.45}$	57.0$_{\pm1.72}$	54.2$_{\pm1.38}$	65.7$_{\pm0.96}$	63.1$_{\pm1.25}$
SGDA [2]	75.6$_{\pm0.57}$	71.4$_{\pm0.82}$	69.2$_{\pm0.73}$	64.7$_{\pm2.36}$	66.3$_{\pm0.68}$	62.3$_{\pm0.96}$	72.9$_{\pm1.26}$	68.9$_{\pm1.83}$	60.6$_{\pm0.86}$	56.0$_{\pm0.90}$	73.2$_{\pm0.59}$	69.3$_{\pm1.01}$
TFGDA-S	55.8$_{\pm1.76}$	53.6$_{\pm1.84}$	54.2$_{\pm2.11}$	44.9$_{\pm2.04}$	58.2$_{\pm1.52}$	48.9$_{\pm1.94}$	57.0$_{\pm1.03}$	46.3$_{\pm1.60}$	49.8$_{\pm2.33}$	41.0$_{\pm3.46}$	55.9$_{\pm1.41}$	45.2$_{\pm1.72}$
TFGDA-T	72.6$_{\pm0.55}$	66.3$_{\pm0.83}$	65.9$_{\pm0.85}$	62.4$_{\pm1.39}$	64.3$_{\pm0.88}$	61.8$_{\pm0.76}$	65.6$_{\pm0.97}$	54.7$_{\pm1.24}$	56.2$_{\pm0.78}$	53.4$_{\pm0.87}$	68.5$_{\pm0.44}$	67.4$_{\pm0.92}$
TFGDA-D	75.8$_{\pm0.38}$	70.7$_{\pm0.69}$	71.2$_{\pm0.61}$	67.5$_{\pm1.15}$	68.5$_{\pm0.57}$	63.7$_{\pm1.02}$	72.2$_{\pm0.88}$	68.1$_{\pm1.13}$	63.1$_{\pm0.74}$	58.5$_{\pm0.79}$	73.4$_{\pm0.51}$	71.1$_{\pm0.88}$
TFGDA-R	74.4$_{\pm0.46}$	70.1$_{\pm0.77}$	68.8$_{\pm0.54}$	64.7$_{\pm0.98}$	65.6$_{\pm0.49}$	62.4$_{\pm0.85}$	69.7$_{\pm0.94}$	63.3$_{\pm1.22}$	62.7$_{\pm0.60}$	56.8$_{\pm0.83}$	72.1$_{\pm0.43}$	69.6$_{\pm0.85}$
TFGDA-TD	78.9$_{\pm0.59}$	76.9$_{\pm0.62}$	72.9$_{\pm0.86}$	70.8$_{\pm1.34}$	70.1$_{\pm0.65}$	68.9$_{\pm0.89}$	73.7$_{\pm1.14}$	71.1$_{\pm1.39}$	64.8$_{\pm0.68}$	62.6$_{\pm0.76}$	75.2$_{\pm0.62}$	72.4$_{\pm0.93}$
TFGDA-TR	78.4$_{\pm0.64}$	75.8$_{\pm0.48}$	72.3$_{\pm0.60}$	68.3$_{\pm1.13}$	70.5$_{\pm0.58}$	69.6$_{\pm0.95}$	73.4$_{\pm0.93}$	70.8$_{\pm1.52}$	64.2$_{\pm0.71}$	61.8$_{\pm0.81}$	76.3$_{\pm0.54}$	72.6$_{\pm0.83}$
TFGDA-DR	79.2$_{\pm0.41}$	77.4$_{\pm0.50}$	73.2$_{\pm0.49}$	71.6$_{\pm0.94}$	72.0$_{\pm0.53}$	71.4$_{\pm0.92}$	74.5$_{\pm1.10}$	71.5$_{\pm1.47}$	65.3$_{\pm0.63}$	63.0$_{\pm0.70}$	77.1$_{\pm0.49}$	72.9$_{\pm0.87}$
TFGDA	**81.0**$_{\pm0.34}$	**78.9**$_{\pm0.46}$	**75.3**$_{\pm0.51}$	**73.2**$_{\pm0.89}$	**73.6**$_{\pm0.61}$	**72.3**$_{\pm0.94}$	**76.0**$_{\pm1.02}$	**72.6**$_{\pm1.35}$	**66.9**$_{\pm0.59}$	**64.3**$_{\pm0.72}$	**78.9**$_{\pm0.47}$	**74.4**$_{\pm0.91}$

and achieves outstanding results in some easy transfer scenarios, such as **A→C** and **D→C**, implying that TFGDA can successfully minimize domain discrepancy. More importantly, the results with a smaller fluctuation range indicate not only the stability of our framework, but also its robustness in the face of different scenarios. Furthermore, we find that most competitors, such as CDAN$_{GCN}$, UDA-GCN, and AdaGCN, exhibit poor performance due to their limited ability to align domains and ineffective utilization of unlabeled nodes. In contrast, TFGDA effectively solves these challenges.

5.3 Analysis and Ablation Study

Due to the limitation of page size, more experiments and analysis are placed on the **Appendix**.

1) Ablation Study: To investigate the contribution of each component in TFGDA, we compare TFGDA and its 7 variants on various tasks. The variants of TFGDA are shown in Table 2, and the detailed ablation study results are gathered in Table 1.

Contribution of Each Component: The results in Table 1 reflect the following observations: **(1)** TFGDA-S (baseline) performs poorly on all tasks because it encounters overfitting problem. **(2)** Compared to TFGDA-S, variants TFGDA-T, TFGDA-D and TFGDA-R achieve significant performance gains, indicating that preserving topological structure information, reducing domain discrepancy on the sphere, and guiding discriminative clustering of unlabeled nodes all facilitate the learning of domain-invariant node features.

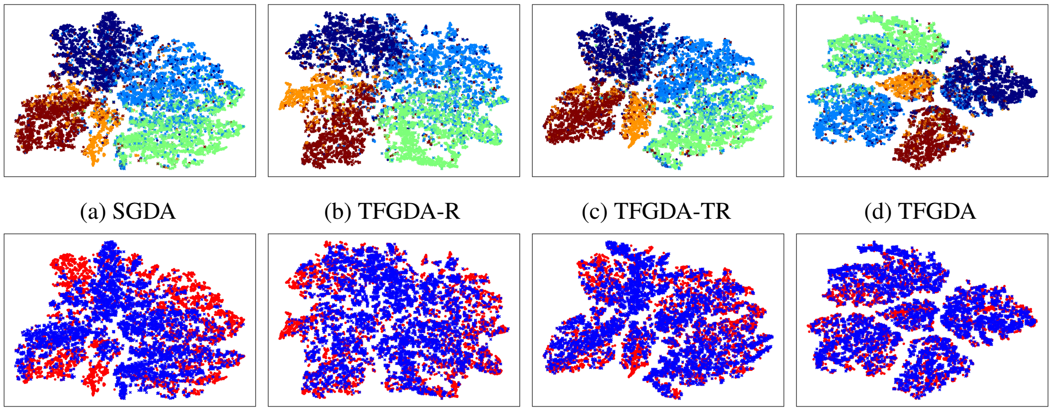
Table 2: Variants of TFGDA.

Variant	$\mathcal{L}_{cls}$	$\mathcal{L}_{stsa}$	$\mathcal{L}_{sda}$	$\mathcal{L}_{rnc}$
TFGDA-S	✓			
TFGDA-T	✓	✓		
TFGDA-D	✓		✓	
TFGDA-R	✓			✓
TFGDA-TD	✓	✓	✓	
TFGDA-TR	✓	✓		✓
TFGDA-DR	✓		✓	✓
TFGDA	✓	✓	✓	✓

Correlation of Our Strategies: As shown in Table 1, the combination of different strategies improves the model's transfer performance, implying a clear complementary relationship among the STSA, SDA, and RNC strategies.

2) Visualization of Node Features: To show the superior transfer ability of our model, we employ t-SNE to visualize node features on task **A→C** under the same 5% label rate setting, as depicted in Figure 2. Although the SOTA method SGDA reduces domain discrepancy to a certain, there are some overlaps between different clusters, causing some hard-to-transfer nodes to be easily misclassfied. In comparison, TFGDA achieves exactly 5 clusters with clean decision boundaries, indicating that our model can capture more fine-grained transferable features as well as align more complex distributions.

3) Effect of STSA: To showcase the effectiveness of preserving topological structure information in assisting GTL, we conduct in-depth experiments from both quantitative and visual aspects: **(1)** As depicted in Table 1, TFGDA-TD and TFGDA-TR greatly outperform TFGDA-D and TFGDA-R respectively, indicating that aligning the topological structures of the input and latent spaces can effectively boost the model's generalization. **(2)** As shown in Figure 2, compared to TFGDA-R,



(a) SGDA (b) TFGDA-R (c) TFGDA-TR (d) TFGDA

(e) SGDA (f) TFGDA-R (g) TFGDA-TR (h) TFGDA

Figure 2: The t-SNE visualization of representations learned by SGDA, TFGDA and its two variants on **A**→**C** task with 5% label rate. In all subfigures, the marks • and × represent the source domain and target domain, respectively. **Fig 2(a-d)** depict category alignment (Different colors denotes different classes). **Fig 2(e-h)** depict domain alignment (Red: Source domain; Blue: Target domain).

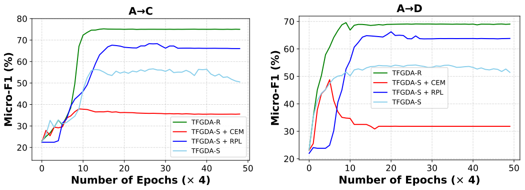


Figure 3: The trend of Micro-F1 during training.

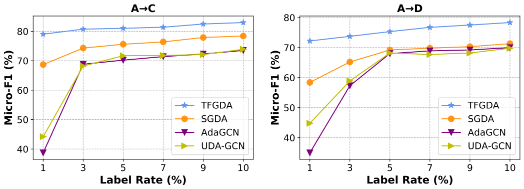


Figure 4: Performance with different label rates.

variant TFGDA-TR exhibits better intra-class compactness and inter-class separability in feature space, implying that the incorporation of topological structure information helps guide the discriminative clustering of unlabeled nodes, thereby promoting the learning of domain-invariant nodes features.

4) Effect of RNC: To show the effectiveness of our RNC strategy, we compare it with existing node clustering strategies, including conditional entropy minimization strategy (**CEM**) [75] and the recently proposed re-weighted pseudo-labeling strategy (**RPL**) [2]. We investigate the trend of Micro-F1 score during training on tasks **A**→**C** and **A**→**D**. The curves in Figure 3 reflect the following observations: **(1)** TFGDA-R converges more smoothly and quickly, achieving higher transfer performance, which suggests that RNC strategy can effectively accelerate the learning of domain-invariant features. **(2)** Compared to TFGDA-S (baseline), TFGDA-S + CEM suffers from severe performance degradation, as CEM may enforce over-confident probability on some misclassified unlabeled nodes. **(3)** RPL strategy fails to achieve satisfactory performance as it's sensitive to pseudo-label noise.

5) Effect of Label Rate: To verify the model's robustness under different label scarcity settings, we evaluate the performance of different methods on tasks **A**→**C** and **A**→**D**, using the following label rates for the source graph: 1%, 5%, 7%, 9%, and 10% respectively, as shown in Figure 4. It can be observed that our TFGDA significantly outperforms other competitors, even in the most challenging environment of 1% label rate, indicating the superiority of TFGDA in capturing transferable features.

6 Conclusion

In this paper, we develop a novel model named TFGDA for SGDA. Specially, we propose a STSA strategy to encode critical structure information into latent space, significantly improving model's transfer performance. Moreover, to stably reduce domain discrepancy, the SDA strategy is introduced to align features distributions on sphere. We also devise the RNC strategy to guide the clustering of unlabeled nodes to address the overfitting issue, greatly enhancing the model's robustness. Comprehensive experiments and analysis verify the superiority of our TFGDA.

7 Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grants (No.62302098, No.62401355), and the Start-up Program for New Young Teacher of Shanghai Jiao Tong University (KJ3-0221-22-6349).

References

[1] Quanyu Dai, Xiao-Ming Wu, Jiaren Xiao, Xiao Shen, and Dan Wang. Graph transfer learning via adversarial domain adaptation with graph convolution. *IEEE Transactions on Knowledge and Data Engineering*, 35(5):4908–4922, 2023.

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

[3] Xiao Shen, Quanyu Dai, Sitong Mao, Fu-Lai Chung, and Kup-Sze Choi. Network together: Node classification via cross-network deep network embedding. *IEEE Transactions on Neural Networks and Learning Systems*, 32(5):1935–1948, 2021.

[4] Xiaowen Zhang, Yuntao Du, Rongbiao Xie, and Chongjun Wang. Adversarial separation network for cross-network node classification. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 2618–2626, 2021.

[5] Leonardo Tadeu Lopes and Daniel Carlos Guimarães Pedronette. Self-supervised clustering based on manifold learning and graph convolutional networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 5634–5643, January 2023.

[6] Michael Moor, Max Horn, Bastian Rieck, and Karsten Borgwardt. Topological autoencoders. In *International conference on machine learning*, pages 7045–7054. PMLR, 2020.

[7] Wenjie Zhu and Bo Peng. Manifold-based aggregation clustering for unsupervised vehicle re-identification. *Knowledge-Based Systems*, 235:107624, 2022.

[8] Shunjie Dong, Zixuan Pan, Yu Fu, Dongwei Xu, Kuangyu Shi, Qianqian Yang, Yiyu Shi, and Cheng Zhuo. Partial unbalanced feature transport for cross-modality cardiac image segmentation. *IEEE Transactions on Medical Imaging*, 42(6):1758–1773, 2023.

[9] Mushui Liu, Yuhang Ma, Yang Zhen, Jun Dan, Yunlong Yu, Zeng Zhao, Zhipeng Hu, Bai Liu, and Changjie Fan. Llm4gen: Leveraging semantic representation of llms for text-to-image generation. In *AAAI*, 2025.

[10] Herbert Edelsbrunner, John Harer, et al. Persistent homology-a survey. *Contemporary mathematics*, 453(26):257–282, 2008.

[11] Xiao Shen, Quanyu Dai, Fu-lai Chung, Wei Lu, and Kup-Sze Choi. Adversarial deep network embedding for cross-network node classification. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 2991–2999, 2020.

[12] Mingsheng Long, ZHANGJIE CAO, Jianmin Wang, and Michael I Jordan. Conditional adversarial domain adaptation. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[13] P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi. Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE transactions on medical imaging*, 23(8):995–1005, 2004.

[14] Clément Bonet, Paul Berg, Nicolas Courty, François Septier, Lucas Drumetz, and Minh-Tan Pham. Spherical sliced-wasserstein. *arXiv preprint arXiv:2206.08780*, 2022.

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

[16] Jun Dan, Weiming Liu, Mushui Liu, Chunfeng Xie, Shunjie Dong, Guofang Ma, Yanchao Tan, and Jiazheng Xing. Hogda: Boosting semi-supervised graph domain adaptation via high-order structure-guided adaptive feature alignment. In *Proceedings of the 32nd ACM International Conference on Multimedia*, MM '24, page 11109–11118, New York, NY, USA, 2024. Association for Computing Machinery.

[17] Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, and Jie Tang. Gcc: Graph contrastive coding for graph neural network pre-training. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 1150–1160, 2020.

[18] Ziyue Qiao, Yanjie Fu, Pengyang Wang, Meng Xiao, Zhiyuan Ning, Denghui Zhang, Yi Du, and Yuanchun Zhou. Rpt: toward transferable model on heterogeneous researcher data via pre-training. *IEEE Transactions on Big Data*, 9(1):186–199, 2022.

[19] Ziniu Hu, Yuxiao Dong, Kuansan Wang, Kai-Wei Chang, and Yizhou Sun. Gpt-gnn: Generative pre-training of graph neural networks. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1857–1867, 2020.

[20] Weiming Liu, Chaochao Chen, Xinting Liao, Mengling Hu, Jianwei Yin, Yanchao Tan, and Longfei Zheng. Federated probabilistic preference distribution modelling with compactness co-clustering for privacy-preserving multi-domain recommendation. In *IJCAI*, pages 2206–2214, 2023.

[21] Weiming Liu, Xiaolin Zheng, Mengling Hu, and Chaochao Chen. Collaborative filtering with attribution alignment for review-based non-overlapped cross domain recommendation. In *Proceedings of the ACM web conference 2022*, pages 1181–1190, 2022.

[22] Shikun Liu, Tianchun Li, Yongbin Feng, Nhan Tran, Han Zhao, Qiang Qiu, and Pan Li. Structural re-weighting improves graph domain adaptation. In *International Conference on Machine Learning*, pages 21778–21793. PMLR, 2023.

[23] Ziqian Lu, Fengli Shen, Mushui Liu, Yunlong Yu, and Xi Li. Improving zero-shot generalization for clip with variational adapter. ECCV, 2024.

[24] Jun Dan, Tao Jin, Hao Chi, Shunjie Dong, Haoran Xie, Keying Cao, and Xinjing Yang. Trust-aware conditional adversarial domain adaptation with feature norm alignment. *Neural Networks*, 168:518–530, 2023.

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

[26] Gaoyang Guo, Chaokun Wang, Bencheng Yan, Yunkai Lou, Hao Feng, Junchao Zhu, Jun Chen, Fei He, and Philip Yu. Learning adaptive node embeddings across graphs. *IEEE Transactions on Knowledge and Data Engineering*, 2022.

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

[28] Jun Dan, Tao Jin, Hao Chi, Yixuan Shen, Jiawang Yu, and Jinhai Zhou. Homda: High-order moment-based domain alignment for unsupervised domain adaptation. *Knowledge-Based Systems*, 261:110205, 2023.

[29] Jun Dan, Tao Jin, Hao Chi, Shunjie Dong, and Yixuan Shen. Uncertainty-guided joint unbalanced optimal transport for unsupervised domain adaptation. *Neural Computing and Applications*, 35(7):5351–5367, 2023.

[30] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2016.

[31] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

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

[33] Jiarong Xu, Yang Yang, Junru Chen, Xin Jiang, Chunping Wang, Jiangang Lu, and Yizhou Sun. Unsupervised adversarially robust representation learning on graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 4290–4298, 2022.

[34] Wei Jin, Yaxing Li, Han Xu, Yiqi Wang, Shuiwang Ji, Charu Aggarwal, and Jiliang Tang. Adversarial attacks and defenses on graphs. *ACM SIGKDD Explorations Newsletter*, 22(2):19–34, 2021.

[35] Yiwei Wang, Wei Wang, Yuxuan Liang, Yujun Cai, Juncheng Liu, and Bryan Hooi. Nodeaug: Semi-supervised node classification with data augmentation. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 207–217, 2020.

[36] Yongduo Sui, Qitian Wu, Jiancan Wu, Qing Cui, Longfei Li, Jun Zhou, Xiang Wang, and Xiangnan He. Unleashing the power of graph data augmentation on covariate distribution shift. *Advances in Neural Information Processing Systems*, 36, 2024.

[37] Louis-Pascal Xhonneux, Meng Qu, and Jian Tang. Continuous graph neural networks. In *International Conference on Machine Learning*, pages 10432–10441. PMLR, 2020.

[38] Ziyue Qiao, Pengyang Wang, Pengfei Wang, Zhiyuan Ning, Yanjie Fu, Yi Du, Yuanchun Zhou, Jianqiang Huang, Xian-Sheng Hua, and Hui Xiong. A dual-channel semi-supervised learning framework on graphs via knowledge transfer and meta-learning. *ACM Transactions on the Web*, 2023.

[39] Ziyue Qiao, Pengyang Wang, Pengfei Wang, Zhiyuan Ning, Yanjie Fu, Yi Du, Yuanchun Zhou, Jianqiang Huang, Xian-Sheng Hua, and Hui Xiong. A dual-channel semi-supervised learning framework on graphs via knowledge transfer and meta-learning. *ACM Transactions on the Web*, 18(2):1–26, 2024.

[40] Afra Zomorodian and Gunnar Carlsson. Computing persistent homology. In *Proceedings of the twentieth annual symposium on Computational geometry*, pages 347–356, 2004.

[41] Chi Seng Pun, Si Xian Lee, and Kelin Xia. Persistent-homology-based machine learning: a survey and a comparative study. *Artificial Intelligence Review*, 55(7):5169–5213, 2022.

[42] Jose A Perea and John Harer. Sliding windows and persistence: An application of topological methods to signal analysis. *Foundations of Computational Mathematics*, 15:799–838, 2015.

[43] Adrien Poulenard, Primoz Skraba, and Maks Ovsjanikov. Topological function optimization for continuous shape matching. In *Computer Graphics Forum*, volume 37, pages 13–25. Wiley Online Library, 2018.

[44] Gunnar Carlsson and Rickard Brüel Gabrielsson. Topological approaches to deep learning. In *Topological Data Analysis: The Abel Symposium 2018*, pages 119–146. Springer, 2020.

[45] Rickard Brüel Gabrielsson, Bradley J Nelson, Anjan Dwaraknath, and Primoz Skraba. A topology layer for machine learning. In *International Conference on Artificial Intelligence and Statistics*, pages 1553–1563. PMLR, 2020.

[46] Christoph Hofer, Florian Graf, Bastian Rieck, Marc Niethammer, and Roland Kwitt. Graph filtration learning. In *International Conference on Machine Learning*, pages 4314–4323. PMLR, 2020.

[47] James R Clough, Nicholas Byrne, Ilkay Oksuz, Veronika A Zimmer, Julia A Schnabel, and Andrew P King. A topological loss function for deep-learning based image segmentation using persistent homology. *IEEE transactions on pattern analysis and machine intelligence*, 44(12):8766–8778, 2020.

[48] Mushui Liu, Jun Dan, Ziqian Lu, Yunlong Yu, Yingming Li, and Xi Li. Cm-unet: Hybrid cnn-mamba unet for remote sensing image semantic segmentation. *arXiv preprint arXiv:2405.10530*, 2024.

[49] Ekaterina Khramtsova, Guido Zuccon, Xi Wang, and Mahsa Baktashmotlagh. Rethinking persistent homology for visual recognition. In *Topological, Algebraic and Geometric Learning Workshops 2022*, pages 206–215. PMLR, 2022.

[50] Vinay Venkataraman, Karthikeyan Natesan Ramamurthy, and Pavan Turaga. Persistent homology of attractors for action recognition. In *2016 IEEE international conference on image processing (ICIP)*, pages 4150–4154. IEEE, 2016.

[51] Jun Dan, Yang Liu, Jiankang Deng, Haoyu Xie, Siyuan Li, Baigui Sun, and Shan Luo. Topofr: A closer look at topology alignment on face recognition. *arXiv preprint arXiv:2410.10587*, 2024.

[52] Mushui Liu, Fangtai Wu, Bozheng Li, Ziqian Lu, Yunlong Yu, and Xi Li. Envisioning class entity reasoning by large language models for few-shot learning. In *AAAI*, 2025.

[53] Bozheng Li, Mushui Liu, Gaoang Wang, and Yunlong Yu. Frame order matters: A temporal sequence-aware model for few-shot action recognition. In *AAAI*, 2025.

[54] Danijela Horak, Simiao Yu, and Gholamreza Salimi-Khorshidi. Topology distance: A topology-based approach for evaluating generative adversarial networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 7721–7728, 2021.

[55] Afra Zomorodian. Fast construction of the vietoris-rips complex. *Computers & Graphics*, 34(3):263–271, 2010.

[56] Chi-Chong Wong and Chi-Man Vong. Persistent homology based graph convolution network for fine-grained 3d shape segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 7098–7107, 2021.

[57] Yuning You, Tianlong Chen, Yongduo Sui, Ting Chen, Zhangyang Wang, and Yang Shen. Graph contrastive learning with augmentations. *Advances in neural information processing systems*, 33:5812–5823, 2020.

[58] Yanbei Liu, Yu Zhao, Xiao Wang, Lei Geng, and Zhitao Xiao. Multi-scale subgraph contrastive learning. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence*, pages 2215–2223, 2023.

[59] Yiding Yang, Jiayan Qiu, Mingli Song, Dacheng Tao, and Xinchao Wang. Distilling knowledge from graph convolutional networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7074–7083, 2020.

[60] Sen Jia, Shuguo Jiang, Shuyu Zhang, Meng Xu, and Xiuping Jia. Graph-in-graph convolutional network for hyperspectral image classification. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):1157–1171, 2022.

[61] Jiangzhang Gan, Rongyao Hu, Yujie Mo, Zhao Kang, Liang Peng, Yonghua Zhu, and Xiaofeng Zhu. Multigraph fusion for dynamic graph convolutional network. *IEEE Transactions on Neural Networks and Learning Systems*, 35(1):196–207, 2022.

[62] Yimeng Min, Frederik Wenkel, and Guy Wolf. Scattering gcn: Overcoming oversmoothness in graph convolutional networks. *Advances in neural information processing systems*, 33:14498–14508, 2020.

[63] Luca Cosmo, Giorgia Minello, Alessandro Bicciato, Michael M Bronstein, Emanuele Rodolà, Luca Rossi, and Andrea Torsello. Graph kernel neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

[64] Qingqing Long, Yilun Jin, Yi Wu, and Guojie Song. Theoretically improving graph neural networks via anonymous walk graph kernels. In *Proceedings of the Web Conference 2021*, pages 1204–1214, 2021.

[65] Ruijia Xu, Guanbin Li, Jihan Yang, and Liang Lin. Larger norm more transferable: An adaptive feature norm approach for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1426–1435, 2019.

[66] Xiang Gu, Jian Sun, and Zongben Xu. Unsupervised and semi-supervised robust spherical space domain adaptation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[67] Weiming Liu, Xiaolin Zheng, Chaochao Chen, Jiahe Xu, Xinting Liao, Fan Wang, Yanchao Tan, and Yew-Soon Ong. Reducing item discrepancy via differentially private robust embedding alignment for privacy-preserving cross domain recommendation. In *Forty-first International Conference on Machine Learning*.

[68] Weiming Liu, Chaochao Chen, Xinting Liao, Mengling Hu, Yanchao Tan, Fan Wang, Xiaolin Zheng, and Yew Soon Ong. Learning accurate and bidirectional transformation via dynamic embedding transportation for cross-domain recommendation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 8815–8823, 2024.

[69] Weiming Liu, Xiaolin Zheng, Chaochao Chen, Jiajie Su, Xinting Liao, Mengling Hu, and Yanchao Tan. Joint internal multi-interest exploration and external domain alignment for cross domain sequential recommendation. In *Proceedings of the ACM Web Conference 2023*, pages 383–394, 2023.

[70] Xinting Liao, Weiming Liu, Chaochao Chen, Pengyang Zhou, Fengyuan Yu, Huabin Zhu, Binhui Yao, Tao Wang, Xiaolin Zheng, and Yanchao Tan. Rethinking the representation in federated unsupervised learning with non-iid data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22841–22850, 2024.

[71] Julien Rabin, Julie Delon, and Yann Gousseau. Transportation distances on the circle. *Journal of Mathematical Imaging and Vision*, 41(1-2):147–167, 2011.

[72] Thomas Bendokat, Ralf Zimmermann, and P-A Absil. A grassmann manifold handbook: Basic geometry and computational aspects. *arXiv preprint arXiv:2011.13699*, 2020.

[73] Erik G Learned-Miller. Entropy and mutual information. *Department of Computer Science, University of Massachusetts, Amherst*, 4, 2013.

[74] Xu Ji, Joao F Henriques, and Andrea Vedaldi. Invariant information clustering for unsupervised image classification and segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9865–9874, 2019.

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

[76] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79:151–175, 2010.

[77] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. *Advances in neural information processing systems*, 19, 2006.

[78] Bharath Bhushan Damodaran, Benjamin Kellenberger, Rémi Flamary, Devis Tuia, and Nicolas Courty. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 447–463, 2018.

[79] Weiming Liu, Xiaolin Zheng, Chaochao Chen, Mengling Hu, Xinting Liao, Fan Wang, Yanchao Tan, Dan Meng, and Jun Wang. Differentially private sparse mapping for privacy-preserving cross domain recommendation. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 6243–6252, 2023.

[80] Weiming Liu, Xiaolin Zheng, Jiajie Su, Mengling Hu, Yanchao Tan, and Chaochao Chen. Exploiting variational domain-invariant user embedding for partially overlapped cross domain recommendation. In *Proceedings of the 45th International ACM SIGIR conference on research and development in information retrieval*, pages 312–321, 2022.

[81] Fan Wang, Chaochao Chen, Weiming Liu, Tianhao Fan, Xinting Liao, Yanchao Tan, Lianyong Qi, and Xiaolin Zheng. Ce-rcfr: Robust counterfactual regression for consensus-enabled treatment effect estimation. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3013–3023, 2024.

[82] Chaoqi Chen, Weiping Xie, Wenbing Huang, Yu Rong, Xinghao Ding, Yue Huang, Tingyang Xu, and Junzhou Huang. Progressive feature alignment for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 627–636, 2019.

[83] Jun Dan, Mushui Liu, Chunfeng Xie, Jiawang Yu, Haoran Xie, Ruokun Li, and Shunjie Dong. Similar norm more transferable: Rethinking feature norms discrepancy in adversarial domain adaptation. *Knowledge-Based Systems*, 296:111908, 2024.

[84] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. Arnetminer: extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 990–998, 2008.

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

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

[87] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10285–10295, 2019.

A Appendix

A.1 Datasets

We run experiments on three real-world graphs: *ACMv9* (**A**), *Citationv1* (**C**), and *DBLPv7* (**D**). In these graphs, every node refers to a paper, and the attribute of each paper is represented as a sparse bag-of-words vector derived from its title. The edges in these graphs depict citation relationships among the papers. Each node is assigned a five-class label, determined by its relevant research areas, including *Artificial Intelligence*, *Computer Vision*, *Database*, *Information Security*, and *Networking*.

Table 3 displays various statistical information of three graphs, including graph scale, attributes, average degree, and label proportion. We can observe substantial intrinsic discrepancy among these graphs. In this paper, we adopt an alternating approach where we select one of these graphs as the source domain, while considering the remaining two as the target domains.

Table 3: The statistics of three real-world graphs. Note that '#' means 'the number of'. 'Attr.' refer to 'Attributes'. 'Avg.' represents 'Average'.

Graph	#Nodes	#Edges	#Attr.	Avg. Degree	Label Proportion (%)
ACMv9 (**A**)	9,360	15,602	5,571	1.667	20.5/29.6/22.5/8.6/18.8
Citationv1 (**C**)	8,935	15,113	5,379	1.691	25.3/26.0/22.5/7.7/18.5
DBLPv7 (**D**)	5,484	8,130	4,412	1.482	21.7/33.0/23.8/6.0/15.5

A.2 Implementation Details

Our experiments are implemented using Pytorch library. Following previous work [2], we choose a two-layer GCN as the feature extractor $\mathcal{F}$ of TFGDA model. We randomly initialize the shift parameter ξ^s and ξ^t using uniform distributions $\mathcal{U}_s(-\epsilon, \epsilon)$ and $\mathcal{U}_t(-\epsilon, \epsilon)$, respectively. In all experiments, we set the value of ϵ to 0.5. For all transfer tasks, we perform each random experiment 5 times and record the average accuracy with standard deviation. And we sample different label sets for each experiment to mitigate the randomness.

To optimize the network, we employ Adam optimizer with a weight decay of 0.001 for better convergence. The base learning rate is set to 0.002 for all tasks. In terms of the balanced coefficients ε and η, we choose $\varepsilon = 0.3$ and $\eta = 1$ for all transfer tasks. Furthermore, instead of fixing the trade-off parameter λ of RNC strategy, we adopt a progressive schedule as [2] to dynamically adjust λ from 0 to 1 by multiplying by $\frac{\theta}{\Theta}$ to more stably guide the discriminative clustering of unlabeled nodes, where θ is the current epoch and Θ is the maximum epoch. Parameter τ in Eq.(8) is set to 0.999, and the spherical radius r is set to the lower bound.

In our STSA strategy, we choose to retain 0-dimensional topological information in PDs. This is because some preliminary experiments have shown that using higher-dimension topological information does not lead to clear accuracy improvements but noticeably increases the model's training time. For the subgraph sampling operation, we set the number of subgraphs to 10 and the number of nodes in each subgraph to 800 to strike a balance between model performance and training efficiency. During the inference process, we disable the shift parameters branch in our RNC strategy. Notably, for all methods (including the compared methods), the dimension of node features is set to 512.

A.3 More Experiments and Analysis

A.3.1 More Ablation Study

6) Effectiveness of SDA: To demonstrate the effectiveness of SDA strategy, we compare it with existing domain alignment strategies, including adversarial training alignment strategy (**AT**) [1], sliced Wasserstein distance-based alignment strategy (**SWD**)[87], class-conditional MMD strategy (**CMMD**) [3], and the recently proposed shifting-guided adversarial training alignment strategy (**SAT**) [2]. We employ variant TFGDA-S as the baseline and evaluate the performance gains brought by these strategies on two typical transfer tasks: **A**→**C** and **A**→**D**.

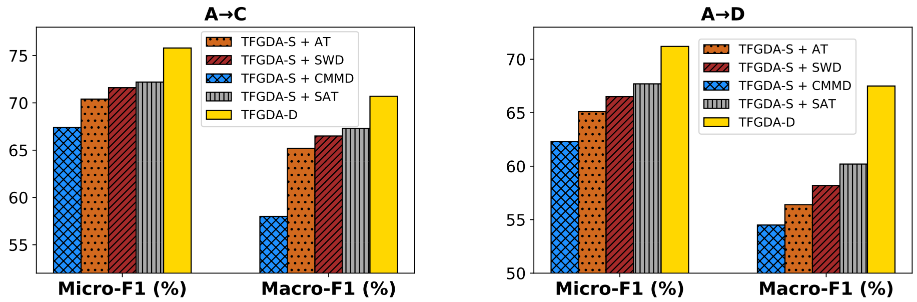


Figure 5: Transfer performance with different domain alignment strategies **A**→**C** and **A**→**D**.

Table 4: Notations.

Notation	Description
a	number of sampled subgraphs $\left\{\hat{\mathcal{G}}_1, \hat{\mathcal{G}}_2, \cdots, \hat{\mathcal{G}}_a\right\}$
b	number of great circles projections
ϖ	a coefficient for controling the scale of feature perturbation
η , ε , λ	balance parameters of loss terms $\mathcal{L}_{sda}, \mathcal{L}_{stsa}$ and $\mathcal{L}_{rnc}$
$\lvert\hat{\mathcal{V}}_i\rvert$	number of nodes in each subgraph

The results shown in Figure 5 reflect the following observations: **(1)** Compared to adversarial training-based strategies such as AT and SAT, our SDA strategy can bring greater performance improvements to the model, especially in the hard transfer task **A**→**D**. This is because SDA implements a more stable distributions alignment process, which promotes the model to capture more transferable features. **(2)** The proposed SDA strategy works better than the SWD strategy, as SDA takes into account the manifold structure of the data in the spherical space, while SWD primarily focuses on the distribution discrepancy in the Euclidean space. **(3)** TFGDA-D significantly outperforms all compared strategies, which indicates the superiority of our SDA strategy in mitigating domain discrepancy and extracting domain-invariant features.

A.3.2 Parameter Sensitivity

The key parameters and notations in our TFGDA model are summarized in Table 4.

7) Effect of Subgraphs: As depicted in Figure 6, we investigate the effect of different numbers of sampled subgraphs a and different subgraph sizes $|\hat{\mathcal{V}}_i|$ on the model's transfer performance, respectively.

We can obtain the following observations: **(1)** As the number of sampled subgraphs a increases, the model's transfer performance gradually improves until it converges. However, a large number of subgraphs can lead to slow model training. Therefore, in our model, we set the number of subgraphs to 10 to balance model's accuracy and training efficiency; **(2)** Similarly, as the size of the subgraphs $|\hat{\mathcal{V}}_i|$ increases, the model's accuracy gradually improves until it converges. However, excessively large subgraphs also significantly increase training time. Hence, to strike a balance between performance and training efficiency, we set the number of nodes in each subgraph to 800. **(3)** When the number a or size $|\hat{\mathcal{V}}_i|$ of the subgraphs is set too small, the model's performance becomes less robust, exhibiting a higher standard deviation. This is because, in such cases, the subgraphs struggle to capture the structure information present in the original graph sufficiently.

8) Effect of Feature Perturbation Scale: The constraint coefficient ϖ of shift parameters ξ_s and ξ_t is responsible for controlling the scale of node feature perturbation. To investigate the effect of coefficient ϖ in our framework, we train our TFGDA models with different ϖ and evaluate their performance on two tasks: **A**→**C** and **A**→**D**.

The results are illustrated in Figure 7 As ϖ increase, the model's accuracy first rises and then falls, which implies that properly perturbing the node feature can effectively enhance the model's robustness.

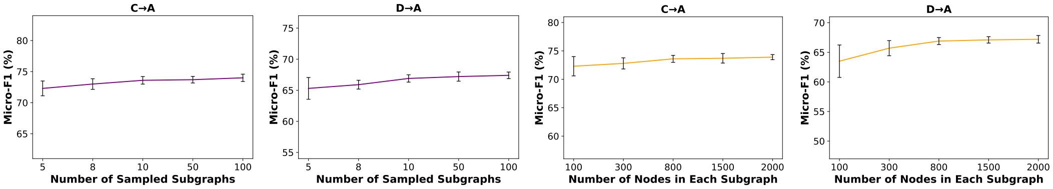


Figure 6: Transfer performance with subgraphs at different scales on **C**→**A** and **D**→**A** tasks.

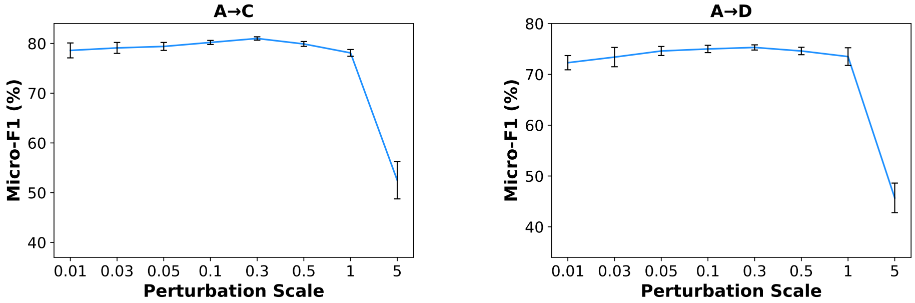


Figure 7: Transfer performance with different feature perturbation scale ϖ on **A**→**C** and **A**→**D** tasks.

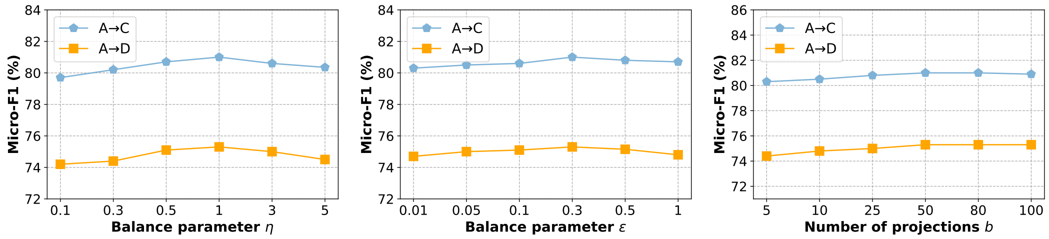


Figure 8: Parameter sensitivity analyses of parameters η, ε, and b on **A**→**C** and **A**→**D** tasks.

Specially, when $\varpi \in [0.05, 0.5]$, our model can achieve stable transfer performance with a very small standard deviation. However, when the coefficient ϖ is set to a large value, unlabeled nodes struggle to cluster in the correct direction, leading to a significant degradation in model performance.

9) Effect of Hyper-parameter: Figure 8 depicts our evaluation of the sensitivity of several hyper-parameters on transfer tasks **A**→**C** and **A**→**D**. The evaluated hyper-parameters include balance parameters η and ε, and the number of projections b. For the parameter η, we find that selecting an appropriate value to adjust the SDA loss $\mathcal{L}_{sda}$ can effectively decrease domain discrepancy. Moreover, we can observe that our model is robust to changes in parameter ε, indicating the stability of our STSA strategy. The results also implies that our SDA strategy maintains stability when aligning feature distributions across domains, regardless of variations in the specific value of the parameter b. When b is extremely small, accurately approximating the SSW distance (*i.e.*, domain discrepancy Eq.(12)) in the SDA strategy becomes difficult, leading to a slight decrease in model's transfer performance.

A.3.3 More Analysis

10) Why our RNC strategy can avoid degenerate clustering solutions ?: The mutual information (MI) objective function $\mathcal{L}_{mi}(\mathbb{Q})$ in Eq. 15 can be expanded to: $\mathcal{L}_{mi}(\mathbb{Q}) = \mathcal{L}_{mi}(\phi, \phi^{\xi}) = E(\phi) - E(\phi|\phi^{\xi})$ [73, 74]. Therefore, maximizing this objective $\mathcal{L}_{mi}$ involves a trade-off between minimizing the conditional cluster assignment entropy $E(\phi|\phi^{\xi})$ and maximizing the entropy of individual cluster assignments $E(\phi)$. Specially, the minimum value of $E(\phi|\phi^{\xi})$ is 0, which is achieved when the cluster assignments can be precisely predicted from each other. The maximum value of $E(\phi)$ is $\ln K$, achieved when all clusters have an equal probability of being selected, where K is the number of classes. This situation arises when the data is evenly distributed among the clusters, resulting in an equal distribution of their masses. Hence, the loss function cannot minimized when all samples are assigned to the same cluster. In such case, maximizing MI naturally achieves a balance between reinforcing predictions and equalizing the cluster masses, thereby avoiding the occurrence of degenerate clustering solutions. The t-SNE node features visualization results in the main text also validate this viewpoint.

A.3.4 Limitation

Our model might has the following two limitations:

(i) As depicted in Figure 6 of the Appendix, the effectiveness of our STSA strategy depends on the quality of the subgraphs (*i.e.*, subgraph size $|\hat{\mathcal{V}}_i|$ and the number of subgraphs a). As a result, the model's transfer performance may decrease if the subgraph size is small or the number of subgraphs is small, as these local subgraphs are difficult to sufficiently capture the properties of the original graph.

(ii) As depicted in Figure 7 of the Appendix, the transfer ability of the model is affected by the feature perturbation scale ϖ in the RNC strategy. Specially, when the feature perturbation scale ϖ is set to a large value, it is difficult for the RNC strategy to accurately capture the intrinsic invariant features of nodes, which affects the discriminative clustering of unlabeled nodes and leads to a decline in model's generalization performance.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The abstract and introduction section include the main claims made in the paper.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: We have discussed the limitations of this work. Please refer to the Section A.3.4 in the Appendix.

   Guidelines:

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

3. **Theory Assumptions and Proofs**

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

Answer: [NA]

Justification: This work does not involve any novel theoretical findings.

Guidelines:

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

4. **Experimental Result Reproducibility**

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

   Answer: [Yes]

   Justification: We have provided detailed implementation details in Section A.2 of the Appendix.

   Guidelines:

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

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

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

Answer: [No]

Justification: We have provided publicly available dataset information in Section A.1 of the Appendix.

Guidelines:

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

6. **Experimental Setting/Details**

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

   Answer: [Yes]

   Justification: We have provided detailed implementation details and training settings in Section A.2 of the Appendix.

   Guidelines:

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

7. **Experiment Statistical Significance**

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

   Answer: [Yes]

   Justification: Our experimental results already include error bars.

   Guidelines:

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

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

8. **Experiments Compute Resources**

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

   Answer: [Yes]

   Justification: We have provided detailed experiment settings in Section A.2 of the Appendix.

   Guidelines:

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

9. **Code Of Ethics**

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

   Answer: [Yes]

   Justification: The research conducted in the paper fully adheres to the NeurIPS Code of Ethics.

   Guidelines:

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

10. **Broader Impacts**

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

    Answer: [NA]

    Justification: All the datasets involved in this paper are publicly available.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

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

    Answer: [NA]

    Justification: The paper poses no such risks.

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: The data and code used in this paper have obtained legal permissions.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.

- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

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

    Answer: [NA]

    Justification: The paper does not release new assets.

    Guidelines:

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

14. **Crowdsourcing and Research with Human Subjects**

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

    Answer: [NA]

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

    Guidelines:

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

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

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

    Answer: [NA]

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

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.