Attribute Role in Graph Domain Adaption

Chen Wang
Western University
cwan372@uwo.ca

Ruiyi Fang
Western University
rfang32@uwo.ca

Liangjian Wen*
Southwestern University of Finance and Economics
wenlj@swufe.edu.cn

ABSTRACT

Graph Domain Adaptation (GDA) addresses a pressing challenge in cross-network learning, particularly pertinent due to the absence of labeled data in real-world graph datasets. The limited alignment information between the target and source graphs has been noted to hinder the adaptability of graph domain adaptation significantly. GDA predominantly relies on adaptation-based techniques, which struggle to capture the intricate components of graph data. However, despite recent advancements in theoretical and empirical studies on GDA capabilities, there remains a lack of formal and direct theoretical results to effectively capture the inherent characteristics of graphs, including their topology and attributes. This paper successfully incorporates the attribute graph adaptation method for GDA in our derived adaptation ability upper-bound, especially overcoming previous works ignorance of node attributes. In other words, we quantitatively show how the interplay between topology structure and node attributes influences the adaptation ability. This also partly explains the limitation of some existing methods for GDA. Based on this result, we further design a cross-channel module for graph data to utilize node attribute embeddings for alignment, in contrast to the structure-based method in previous GDA models. Experimental results on a variety of publicly available datasets reveal the effectiveness of our GFA model. The source code is available at https://drive.google.com/file/d/1mooWMdhJnT0bLohyt5fBYpnPZ017AkGE/view?usp=sharing

KEYWORDS

self-supervised learning, domain adaption, graph convolutional network

ACM Reference Format:
Chen Wang, Ruiyi Fang, and Liangjian Wen. 2025. Attribute Role in Graph Domain Adaption. In . ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In the area of widespread internet data, graph vertices are frequently associated with content information, referred to as node attributes within basic graph data. However, these data frequently suffer from label scarcity, as annotating complex structured data is both costly and challenging [27]. To address this challenge, transferring abundant labeling knowledge from task-related graphs has been proposed as an effective solution [1]. Specifically, using labeled graphs as sources to support the learning on unlabeled target graphs is known as graph domain adaptation (GDA). GDA aims to facilitate effective knowledge transfer across graphs by addressing distribution shifts [19].

Basically, GDA early works directly use deep domain adaptation (DA) techniques, ordinarily applied to unstructured data, to learn label-discriminative and domain-invariant embeddings[19]. Due to that influence, most early works[1, 15, 16, 25] in GDA have mainly relied on source data. However, they completely rely on the information of the source domain and ignore the target domain. Recently, some methods[12, 33] based on eliminating complex graph domain shift information have gained attention by using precisely designed graph feature extraction frameworks. While the above strategy may align overall embedding distributions, it fails to capture the nuances of differences in complex graph structures and task-specific components, leading to subpar outcomes. To address the above issue, researchers have proposed methods focusing on eliminating graph structural divergence[17, 18, 24, 28], which usually directly analyze the inherent properties of graph topology (e.g adjacency matrix). While these methodologies consider the graph's structural characteristics, they overlook a fundamental aspect of node classification tasks: the impact of node attributes.

In this paper, we focus on a study of graph data grouping effect, where the data features can be divided into two main things, graph structure data(topology structure) and node attributes (node features). This problem has received extensive attention in graph clustering and graph representation learning[9]. However, GDA's current research does not consider these issues, especially graph node attributes.

Distribution shift has a significant influence in the domain adaptation field [20]. In graph data view, most previously research in GDA focus on structural shift or topology information[4, 10]. Based on basic properties of GNN the label propagation on the graph is based on topological information(edges)[8]. For this reason, Alex[32] works realized used a low-rank adjacency matrix to overcome label-conditioned structural shift. Discovering from the aforementioned theoretical framework, the crucial idea is to characterize the cross-network knowledge transferability from the perspective of node attribute information.

To better explore the graph node attribute, we propose a cross-channel GDA Feature alignment (GFA) method that aligns both the attribute and topology structure embedding. We first construct an attribute graph (feature graph) to provide an attribute view, which allows the feature information to propagate through the feature space[23], thus helping us get better attribute view information.

---

*Corresponding author

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

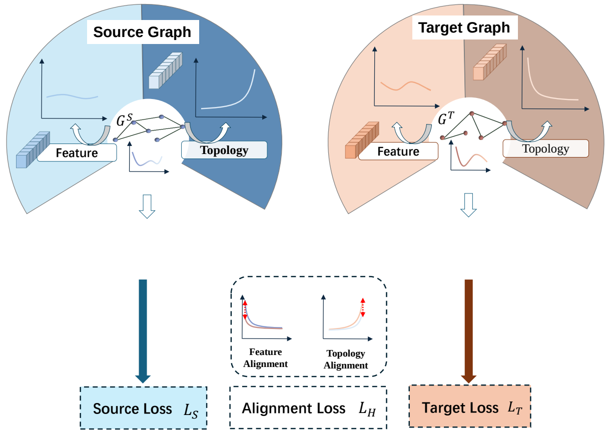


Figure 1: An overview of our method. GFA gives attribute and topology view graph representation, where minimizing source and target distribution shift through two views.

Then, we use inter-graphed contrastive methods to fuse local relations by combining the local relationship for a comprehensive node representation of each graph[2]. As shown in Fig.1, we obtain graph information in both the topology view and attribute view, and then we minimize the distribution shift in both views. Then, we using output embedding Z_t^T from A^T, Z_f^T from $\hat{A}^T$ and Z_t^S from A^S, Z_f^S from $\hat{A}^S$ are obtained through GCN. Ultimately, we enhance GDA by minimizing the domain shift across the graph's node attribute and topology perspectives.

Our main contributions are summarized as follows:

- We first emphasize the impact of node attributes in GDA. Our theory indicates that the GDA generalization error bounds of cross-network transfer learning can be derived based on node attributes and graph structure.
- We proposed a novel model, GFA, for cross-network node classification tasks. Based on the significant attribute shift we observed, we jointly minimize both attribute and topology distribution shifts based on the inherent graph property.
- Comprehensive experiments on benchmarks show the superior performance of our method compared to other state-of-the-art methods in real-world datasets of the cross-network node classification tasks.

2 RELATED WORKS

Transfer learning

Unsupervised domain adaptation (UDA) is a widely studied approach in transfer learning, aiming to minimize the distribution discrepancy between a labeled source domain and an unlabeled target domain [21]. To address cross-domain classification tasks, most UDA methods rely on deep feature representations [35], which map data from different domains into a shared feature space. A representative example is DANN [3], which employs a gradient reversal layer to encourage the learning of domain-invariant features by adversarially training the feature extractor against a domain classifier.

Recent work has sought to address challenges such as domain imbalance and class label distribution shifts to improve model transferability [5, 26]. Additionally, novel settings like source-free domain adaptation (SFDA)[29] and test-time domain adaptation (TTDA)[22] have gained increasing attention. For instance, Yi et al.[30] leverage early-stage training dynamics to address label noise, a phenomenon observed in both conventional noisy label scenarios and SFDA. CAN[6] learns transferable node embeddings for domain adaptation tasks by minimizing the maximum mean discrepancy (MMD) loss. However, these methods are primarily designed for grid-structured data and face significant challenges when applied directly to graph-structured data due to the unique properties and complexities of graphs.

Graph Domain Adaption

For graph-structured data, several methods have been proposed to enable cross-graph knowledge transfer under the Graph Domain Adaptation (GDA) setting [1, 14, 19]. ACDNE [15] utilizes a k-hop PPMI matrix to capture high-order proximity, leveraging it as global consistency to preserve source information. CDNE [16] learns cross-network embeddings by directly minimizing the maximum mean discrepancy (MMD) between source and target domains. GraphAE [28] addresses node degree distribution shifts by aligning message-passing routes to reduce structural discrepancies. Similarly, DM-GNN [17] propagates label information by integrating both a node's own and its neighbors' edge structures. UDAGCN [25] introduces a dual graph convolutional network that captures both local and global knowledge through adversarial training. ASN [33] separates domain-specific and domain-invariant features via a private encoder, enhancing domain-invariant representation across networks. SOGA [10] is the first to apply the concept of structural consistency to improve discriminability in source-free graph domain adaptation, encouraging consistency among target nodes within the same class. Another variation of GraphAE [4] further explores the impact of degree distribution shifts on node embeddings by minimizing discrepancies between router embeddings to eliminate structural noise. SpecReg [31] leverages optimal transport theory to derive a GDA generalization bound and demonstrates that modifying the Lipschitz constant of GNNs via spectral smoothness and maximum frequency response improves adaptation performance. JHGDA [18] focuses on shifts in hierarchical graph structures, aggregating domain discrepancies at multiple levels to construct a comprehensive discrepancy measure. ALEX [32] introduces a label shift-enhanced augmented graph by constructing a low-rank adjacency matrix via singular value decomposition, guided by a contrastive loss. Finally, SGDA [12] enhances source graphs by introducing trainable perturbations (adaptive shift parameters) into embeddings. It jointly trains the graph encoder and perturbations through adversarial learning to minimize marginal shifts between domains.

3 THE PROPOSED METHODOLOGY

3.1 Problem Definition

Feature Graph Definition.

In our study, we derivate the corresponding feature graph $\hat{G} = \{\hat{\mathcal{V}}, \hat{\mathcal{E}}, \hat{A}, X, Y\}$, which shares the same X with G, but has a different adjacency matrix. Therefore, topology graph and feature graph refer to G and $\hat{G}$ respectively.

3.2 The Framework of GFA

As shown in Fig.1, we use topology graph and feature graph to capture the underlying information in topology space and feature space. Our model mainly contains two components: the feature abstraction module that uses GCN to extract features from graph, the self-supervision module that measures the consistency between the representations learned from the topology graph and feature graph.

Feature Extraction Module.
To extract meaningful features from graphs, we adopt GCN that is comprised of multiple graph convolutional layers. With the input graph G, the $(l+1)$-th layer's output $H^{(l+1)}$ can be represented as:

$$H^{(l+1)} = ReLU(D^{-\frac{1}{2}}AD^{-\frac{1}{2}}H^{(l)}W^{(l)}) \tag{1}$$

where $ReLU$ is the Relu activation function ($ReLU(\cdot) = max(0, \cdot)$), D is the degree matrix of A, $W^{(l)}$ is a layer-specific trainable weight matrix, $H^{(l)}$ is the activation matrix in the l-th layer and $H^{(0)} = X$. In our study we use two GCNs to exploit the information in topology and feature space. For source graph, output is donated by $Z^S = \left\{z_1^s, z_2^s, \cdots, z_N^s\right\}$ generated from G^S and $Z_f^S = \left\{z_{f1}^s, z_{f2}^s, \cdots, z_{fN}^s\right\}$ generated from $\hat{G}^S$, where $N = \left|V^S\right|$ is the number of nodes in G^S. Respectively, target graph is donated by $Z^T = \left\{z_1^t, z_2^t, \cdots, z_M^t\right\}$ and $Z_f^T = \left\{z_{f1}^t, z_{f2}^t, \cdots, z_{fM}^t\right\}$, where $M = \left|V^T\right|$ is the number of nodes in G^T.

Source Classifier Loss.
The source classifier loss $\mathcal{L}_S\left(f_s\left(Z^s\right), Y^s\right)$ is to minimize the cross-entropy for the labeled data node in the source domain:

$$\mathcal{L}_S\left(f_s\left(Z^s\right), Y^s\right) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log\left(\hat{y}_i\right), \tag{2}$$

where y_i denotes the label of the i-th node in the source domain and $\hat{y}_i$ are the classification prediction for the i-th source labeled node z_i^t.

Attribute-Driven Domain Adaptive Loss The proposed framework follows the transition learning paradigm, where the model maximizes the agreement of the two views. In detail, GFA jointly optimizes two views of GDA alignment. To be specific, $\mathcal{L}_{\dashv}$ is the Mean Squared Error (MSE) loss between the source graph Z_t^S and Z_f^S and the target graph Z_t^T and Z_f^T, which can be formulated as:

$$\mathcal{L}_a = -\left(||Z_t^S - Z_t^T||_2^2 + ||Z_f^S - Z_f^T||_2^2\right) \tag{3}$$

We adapt the domain in two views, domain classifier loss in the topology view is $||Z_t^S - Z_t^T||_2^2$ enforces that the topology graph node representation after the node feature extraction from source domain network G^S and target domain network G^t are similar.

Target Node Classification.
We use the Gradient Reversal Layer (GRL) [3] for adversarial training. Mathematically, we define the GRL as $Q_\lambda(x) = x$ with a reversal gradient $\frac{\partial Q_\lambda(x)}{\partial x} = -\lambda I$. Learning a GRL is adversarial in such a way that: on the one side, the reversal gradient enforces $f_s(Z^S)$ to be maximized; on the other side, θ_D is optimized by minimizing the cross-entropy domain classifier loss:

$$\mathcal{L}_D = -\frac{1}{N+M}\sum_{i=1}^{N+M} m_i \log\left(\hat{m}_i\right) + \left(1 - m_i\right)\log\left(1 - \hat{m}_i\right) \tag{4}$$

To utilize the data in the target domain, we use entropy loss for the target classifier f_t :

$$\mathcal{L}_T\left(f_t\left(Z^T\right)\right) = -\frac{1}{M}\sum_{i=1}^{M} \hat{y}_i \log\left(\hat{y}_i\right) \tag{5}$$

where $\hat{y}_i$ are the classification prediction for the i-th node Z_i^s in the target domain. Finally, by combining L_c, L_s ,L_D and L_T, the overall loss function of our model can be represented as:

$$\mathcal{L} = \mathcal{L}_A + \alpha\mathcal{L}_s + \beta\mathcal{L}_D + \gamma\mathcal{L}_T, \tag{6}$$

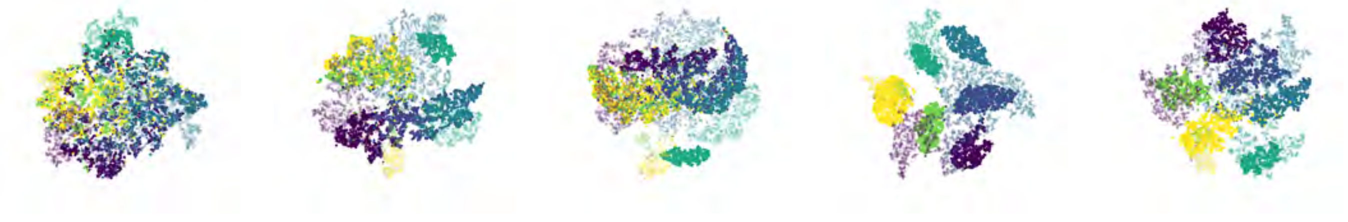
where α, β and γ are trade-off hyper-parameters. The parameters of the whole framework are updated via backpropagation. A detailed description of our algorithm is provided in Algorithm 1.

Algorithm 1: The proposed algorithm GFA

```
Input: Source node feature matrix $X^S$; source original
    graph adjacency matrix $A^S$; Target node feature
    matrix $X^T$; Target original graph adjacency matrix
    $A^T$ source node label matrix $Y^S$; maximum number
    of iterations $\eta$
for $it = 1$ to $\eta$ do
    $Z^S = GCN(A^S, X^S)$
    $Z_f^S = GCN(A_f^S, X^S)$// embedding of source graph
    $Z^T = GCN(A^T, X^T)$
    $Z_f^T = GCN(A_f^T, X^T)$// embedding of target graph
    $Z^S$ and $Z_f^S$ interact with local feature fusion.
    $Z^T$ and $Z_f^T$ interact with local feature fusion.
    Domain Adaptive Learning between $Z^S$ and $Z^T$
    Domain Adaptive Learning between $Z_f^S$ and $Z_f^T$
        // adaptive in two views
    $y\hat{i}^s$constrained by$y_i^s$ and $y\hat{i}^t$constrained by$y\hat{i}^t$
    Calculate the overall loss with Eq.(6)
    Update all parameters of the framework according to
      the overall loss
end for
Predict the labels of target graph nodes based on the trained
  framework.
Output: Classification result $\hat{Y}^T$
```

4 EXPERIMENT

In this section, we conduct extensive experiments to evaluate The effectiveness of the proposed method in two extensively used datasets.



(a) UDAGCN (b) JHGDA (c) GRADE (d) SpecReg (e) our method

Figure 2: Visualization of learnt representations of different methods on D-A task of dataset.

4.1 Datasets

To prove the superiority of our work on domain adaptation node classification tasks, we evaluate it on two types of datasets, including the Airport dataset[13] and the Citation dataset[25]. The airport dataset involves three countries' airport traffic networks: USA, Brazil, and Europe, in which the node indicates the airport and the edge indicates the routes between two airports. The citation dataset includes three different citation networks: DBLPv8, ACMv9, and Citationv2, in which the node indicates the article and the edge indicates the citation relation between two articles. Because these two groups of dataset ingredients are generated from different data sources, their distributions are naturally diverse. Thus, we give two groups of datasets: for Citation dataset(source → target): A→D, D→A, A→C, C→A, D→C, C→D, and for Airport dataset(source → target): USA→Brazil, Brazil→USA, USA→ Europe, Europe→USA, Brazil→Europe, Europe→Brazil.

4.2 Baselines

We choose some representative methods to compare.

- **GCN** [8] further solves the efficiency problem by introducing first-order approximation of ChebNet.
- **DANN** [3] uses a 2-layer perceptron to provide features and a gradient reverse layer (GRL) to learn node embeddings for domain classification
- **DANE** [34]shared distributions embedded space on different networks and further aligned them through adversarial learning regularization.
- **UDAGCN** [25] is a dual graph convolutional network component learning framework for unsupervised GDA, which captures knowledge from local and global levels to adapt it by adversarial training.
- **ASN** [33] uses the domain-specific features in the network to extract the domain-invariant shared features across networks.
- **EGI** [36] through Ego-Graph Information maximization to analyze structure-relevant transferability regarding the difference between the source-target graph.
- **GRADE-N** [24]propose a graph subtree discrepancy to measure the graph distribution shift between source and target graphs.
- **JHGDA** [18] explores information from different levels of network hierarchy by a hierarchical pooling model.
- **SpecReg** [31]achieve improving performance regularization inspired by cross-pollinating between the optimal transport DA and graph filter theories

4.3 Experimental Setting

The experiments are implemented in the PyTorch platform using an Intel(R) Xeon(R) Silver 4210R CPU @ 2.40GHz, and GeForce RTX A5000 24G GPU. Technically, a two-layer GCN is built, and we train our model by utilizing the Adam [7] optimizer with a learning rate ranging from 0.0001 to 0.0005. We repeatedly train and test our model five times with the same partition of the dataset and then report the average of ACC. For fairness, for all the cross-domain node classification methods in our experiment, we use the same parameter settings, except for some special cases. For GCN, AdaGCN, UDA-GCN, and JHGDA, the GCNs of both the source and target networks contain two hidden layers ($L = 2$) with structure as $128 - 16$. The dropout rate for each GCN layer is set to 0.3. We repeatedly train and test our model for five times with the same partition of dataset and then report the average of ACC.

4.4 GDA Node Classification Results

The results of experiments are summarized in Table 1 and 3, where the best performance is highlighted in boldface. Some results are directly taken from [11, 18]. We have the following findings: It can be seen that our proposed method boosts the performance of SOTA methods across most evaluation metrics on two group datasets with 12 tasks, which proves its effectiveness. Particularly, compared with other optimal performances, GFA achieves a maximum average improvement of 1.12% for ACC. This illustrates that our proposed model can effectively utilize node attribute information. Our GFA achieves much better performances than SpecReg and JHGDA on all of the metrics in a dataset of Airport and most of the metrics in a dataset of citation. This can be explained by our method's use of attribute and topology structure. In most cases, GFA produces better performance than SpecReg [31], JHGDA [18], which were published in 2023. This verifies the advantage of our approach.On most occasions, the feature graph produces a better result than the original graph. Our finding affirm that the observed discrepancy in node attributes surpasses that of the topological misalignment, thus suggesting that the alignment of node attributes holds potential for yielding more substantial enhancements.

For more intuitive understanding and comparison, we use t-SNE method to visualize the progression of the representation learned

Methods	USA → Brazil	USA → Europe	Brazil → USA	Brazil → Europe	Europe → USA	Europe → Brazil	Avg.
GCN [8]	0.366	0.371	0.491	0.452	0.439	0.298	0.403
DANN [3]	0.501	0.386	0.402	0.350	0.436	0.538	0.436
DANE [34]	0.531	0.472	0.491	0.489	0.461	0.520	0.494
UDAGCN [25]	0.607	0.488	0.497	0.510	0.434	0.477	0.502
ASN [33]	0.519	0.469	0.498	0.494	0.466	0.595	0.507
EGI [36]	0.523	0.451	0.417	0.454	0.452	0.588	0.481
GRADE-N [24]	0.550	0.457	0.497	0.506	0.463	0.588	0.510
JHGDA [18]	0.695	0.519	0.511	0.569	0.522	**0.740**	0.593
SpecReg [31]	0.481	0.487	0.513	0.546	0.436	0.527	0.498
GFA	**0.694**	**0.573**	**0.552**	**0.533**	**0.536**	0.641	**0.575**

Table 1: Cross-network node classification on the Airport network

Methods	A → D	D → A	A → C	C → A	C → D	D → C	Avg.
GCN [8]	0.632	0.578	0.675	0.635	0.666	0.654	0.638
DANN [3]	0.488	0.436	0.520	0.518	0.511	0.465	0.491
DANE [34]	0.661	0.619	0.642	0.653	0.661	0.709	0.641
UDAGCN [25]	0.684	0.623	0.726	0.661	0.712	0.645	0.684
ASN [33]	0.729	0.723	0.752	0.678	0.752	0.754	0.731
EGI [36]	0.641	0.553	0.676	0.598	0.662	0.652	0.32
GRADE-N [24]	0.701	0.660	0.736	0.687	0.722	0.687	0.699
JHGDA [18]	0.755	0.737	0.814	0.757	0.762	0.794	0.770
SpecReg [31]	0.762	0.654	0.753	0.680	0.768	0.727	0.724
GFA	**0.779**	**0.747**	**0.821**	**0.776**	**0.762**	**0.793**	**0.781**

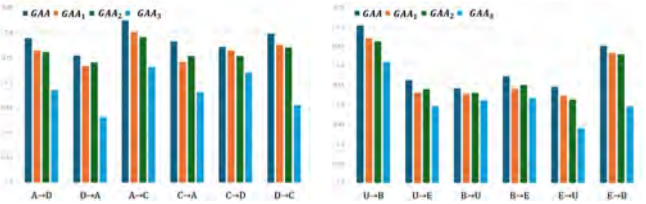
Table 2: Cross-network node classification on the Citation network

Methods	B1 → B2	B2 → B1	DE → EN	EN → DE
GCN [8]	0.632	0.578	0.675	0.635
DANN [3]	0.488	0.436	0.520	0.518
DANE [34]	0.664	0.619	0.642	0.653
UDAGCN [25]	0.684	0.623	0.728	0.663
ASN [33]	0.729	0.723	0.752	0.678
EGI [36]	0.647	0.557	0.676	0.598
GRADE-N [24]	0.701	0.660	0.736	0.687
JHGDA [18]	0.755	0.737	0.814	0.757
SpecReg [31]	0.762	0.654	0.753	0.680
GFA	**0.783**	**0.755**	**0.821**	**0.775**

Table 3: Cross-network node classification on the Citation network

by our GFA. Every color denotes a distinct class, with darker and lighter hues of the same color denoting nodes belonging to corresponding classes within the source and target domains, respectively. To demonstrate the advantage of our proposed method, we also visualize the embedding results on D->A generated by competitive methods UDAGCN, JHGDA, GRADE, and SpecReg, which are shown in Fig.2. While SpecReg appears to enhance clustering boundary definition, its overlap with data from different domains is inadequate, thereby elucidating the method's limitations. Our GFA method produces a compact cluster structure. In other words, our method has the highest intra-class similarity and the most distinct boundaries between different classes.

According to Fig.3, we can draw the following conclusions: (1) The results of GFA are consistently better than all variants, indicating the rationality of our model. (2) Both topology and feature information are crucial to domain adaptation. (3) The smaller gap in feature homophily rate between the target graph and the source graph effectively improves the migration effect.



(a) Citation **(b) Airport**

Figure 3: The classification accuracy of GFA and its variants on citation datasets and airport dataset.

5 CONCLUSION

In this paper, we propose a framework to solve the GDA problem in node classification tasks. The key idea is to utilize the intrinsic graph node attributes and structures of graphs to minimize domain discrepancy. In addition, we also theoretically confirmed that the generalization error bound of GDA is related to the distance between the topology and the attribute. Comprehensive experiments verify the superiority of our approach. In the future, we may strive to design new frameworks for other cross-network learning tasks, including link-level and graph-level. We will also delve into the graph domain adaptation theory for developing more powerful models.

REFERENCES

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

[2] Ruiyi Fang, Liangjian Wen, Zhao Kang, and Jianzhuang Liu. 2022. Structure-preserving graph representation learning. In *2022 IEEE International Conference on Data Mining (ICDM)*. IEEE, 927–932.

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

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

[5] Taotao Jing, Bingrong Xu, and Zhengming Ding. 2021. Towards fair knowledge transfer for imbalanced domain adaptation. *IEEE Transactions on Image Processing* 30 (2021), 8200–8211.

[6] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. 2019. Contrastive adaptation network for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 4893–4902.

[7] Diederick P Kingma and Jimmy Ba. 2015. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*.

[8] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[9] Xiang Li, Ben Kao, Caihua Shan, Dawei Yin, and Martin Ester. 2020. CAST: a correlation-based adaptive spectral clustering algorithm on multi-scale data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 439–449.

[10] Haitao Mao, Lun Du, Yujia Zheng, Qiang Fu, Zelin Li, Xu Chen, Shi Han, and Dongmei Zhang. 2024. Source free unsupervised graph domain adaptation. *WSDM2024* (2024).

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

[12] Ziyue Qiao, Xiao Luo, Meng Xiao, Hao Dong, Yuanchun Zhou, and Hui Xiong. 2023. Semi-supervised domain adaptation in graph transfer learning. *IJCAI* (2023).

[13] Leonardo FR Ribeiro, Pedro HP Saverese, and Daniel R Figueiredo. 2017. struc2vec: Learning node representations from structural identity. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*. 385–394.

[14] Xiao Shen and Fu Lai Chung. 2019. Network embedding for cross-network node classification. *arXiv preprint arXiv:1901.07264* (2019).

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

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

[17] Xiao Shen, Shirui Pan, Kup-Sze Choi, and Xi Zhou. 2023. Domain-adaptive message passing graph neural network. *Neural Networks* 164 (2023), 439–454.

[18] Boshen Shi, Yongqing Wang, Fangda Guo, Jiangli Shao, Huawei Shen, and Xueqi Cheng. 2023. Improving graph domain adaptation with network hierarchy. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*. 2249–2258.

[19] Boshen Shi, Yongqing Wang, Fangda Guo, Bingbing Xu, Huawei Shen, and Xueqi Cheng. 2024. Graph Domain Adaptation: Challenges, Progress and Prospects. *arXiv preprint arXiv:2402.00904* (2024).

[20] Remi Tachet des Combes, Han Zhao, Yu-Xiang Wang, and Geoffrey J Gordon. 2020. Domain adaptation with conditional distribution matching and generalized label shift. *Advances in Neural Information Processing Systems* 33 (2020), 19276–19289.

[21] Boyu Wang, Jorge Mendez, Mingbo Cai, and Eric Eaton. 2019. Transfer learning via minimizing the performance gap between domains. *Advances in neural information processing systems* 32 (2019).

[22] Qin Wang, Olga Fink, Luc Van Gool, and Dengxin Dai. 2022. Continual test-time domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 7201–7211.

[23] Xiao Wang, Meiqi Zhu, Deyu Bo, Peng Cui, Chuan Shi, and Jian Pei. 2020. Am-gcn: Adaptive multi-channel graph convolutional networks. In *Proceedings of the 26th ACM SIGKDD International conference on knowledge discovery & data mining*. 1243–1253.

[24] Jun Wu, Jingrui He, and Elizabeth Ainsworth. 2023. Non-iid transfer learning on graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 10342–10350.

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

[26] Pengcheng Xu, Boyu Wang, and Charles Ling. 2023. Class Overwhelms: Mutual Conditional Blended-Target Domain Adaptation. *Proceedings of the AAAI Conference on Artificial Intelligence* (2023).

[27] Zihao Xu, Hao He, Guang-He Lee, Yuyang Wang, and Hao Wang. 2022. Graph-relational domain adaptation. *arXiv preprint arXiv:2202.03628* (2022).

[28] Bencheng Yan and Chaokun Wang. 2020. Graphae: adaptive embedding across graphs. In *2020 IEEE 36th International Conference on Data Engineering (ICDE)*. IEEE, 1958–1961.

[29] Shiqi Yang, Yaxing Wang, Joost Van De Weijer, Luis Herranz, and Shangling Jui. 2021. Generalized source-free domain adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 8978–8987.

[30] Li Yi, Gezheng Xu, Pengcheng Xu, Jiaqi Li, Ruizhi Pu, Charles Ling, A Ian McLeod, and Boyu Wang. 2023. When source-free domain adaptation meets learning with noisy labels. *arXiv preprint arXiv:2301.13381* (2023).

[31] Yuning You, Tianlong Chen, Zhangyang Wang, and Yang Shen. 2022. Graph domain adaptation via theory-grounded spectral regularization. In *The Eleventh International Conference on Learning Representations*.

[32] Jingyang Yuan, Xiao Luo, Yifang Qin, Zhengyang Mao, Wei Ju, and Ming Zhang. 2023. Alex: Towards effective graph transfer learning with noisy labels. In *Proceedings of the 31st ACM International Conference on Multimedia*. 3647–3656.

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

[34] Yizhou Zhang, Guojie Song, Lun Du, Shuwen Yang, and Yilun Jin. 2019. Dane: Domain adaptive network embedding. *arXiv preprint arXiv:1906.00684* (2019).

[35] Lei Zhu, Qi She, Qian Chen, Yunfei You, Boyu Wang, and Yanye Lu. 2022. Weakly supervised object localization as domain adaption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14637–14646.

[36] Qi Zhu, Carl Yang, Yidan Xu, Haonan Wang, Chao Zhang, and Jiawei Han. 2021. Transfer learning of graph neural networks with ego-graph information maximization. *Advances in Neural Information Processing Systems* 34 (2021), 1766–1779.