Learning Network-Invariant and Label-Discriminative Representations for Cross-Network Node Classification

Linyao Yang, Yancai Xu
The State Key Laboratory for Management and Control of Complex Systems
Institute of Automation, Chinese Academy of Sciences
University of Chinese Academy of Sciences
Beijing, China
yanglinyao2017@ia.ac.cn, yancai.xu@ia.ac.cn

Jiachen Hou
Qingdao Academy of Intelligent Industries
Qingdao, China
Institute of Systems Engineering
Macau University of Science and Technology
Macau, China
jiachen.hou@qaii.qc.cn

Yuxin Dai
School of Electrical Engineering and Automation
Wuhan University
Wuhan, China
YuxinDai@whu.edu.cn

Chen Lv
China Electric Power Research Institute
Beijing, China
lvchen@epri.sgcc.com.cn

2021 China Automation Congress (CAC) | 978-1-6654-2647-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/CAC53003.2021.9727541

***Abstract*—Networks are ubiquitous data structures in the real world. The accurate and efficient analysis of networks is critical to realizing many intelligent network-based services. However, most existing network analysis methods are developed for single networks and require a lot of labeled data, which is costly and time-consuming to acquire. Transfer learning has been widely accepted as an effective paradigm for tackling this problem by reusing the model trained on a supervised task. However, transfer learning on the non-euclidean network data has been investigated by no more than a few studies. To realize accurate node classification based on the knowledge learned from the labeled source network, this paper proposes to learn network-invariant and label-discriminative representations based on graph embedding and linear discriminant analysis. Specifically, we embed the source and target networks into adjacent vector spaces based on the graph attention network by minimizing the Sinkhorn distributional distances between their embeddings. To obtain label-discriminative features for learning better classification models, we then utilize a transferable linear discriminative analysis method to project the embeddings into label-discriminative subspaces. In the end, a support vector machine model trained on the labeled source network is utilized to classify the target nodes. Experiments on two pairs of networks illustrate that our method achieves the best performance and evaluates the effectiveness of the proposed modules.**

***Index Terms*—Transfer learning, node classification, Sinkhorn distance, linear discriminative analysis**

I. Introduction

Node classification is a fundamental task of network analysis for realizing many network-based intelligent services, such as function prediction of proteins in protein-protein interaction networks, user classification in social networks [1], [2], and rumor recognition on social media [3]. Node classification is to predict nodes' labels in a network. Traditional node classification methods apply classifiers trained on a set of labeled data with the input of node features. The representative classifiers include support vector machine (SVM), neural network, and logistic regression [4]. Recently, the development of graph embedding has greatly benefited the node classification task, which represents nodes with low-dimensional vectors while preserving their structural and attribute proximities [5]. Then, classifiers are trained on the embeddings to predict nodes' class labels. In addition, some methods utilize unified frameworks to learn class-specific embeddings for nodes by jointly optimizing the embedding module and the prediction module. However, most existing node classification methods learn and predict node classes on a single network and require a large percent of labeled data to provide supervision for the training, which is costly and hard to acquire.

Most recently, transfer learning has been deeply researched and used as an effective paradigm to deal with the lack of labeled data and has been successfully applied to solve various problems, such as medical image classification [6], cross-domain recommendation [7], and object instance segmentation [8]. The basic idea of transfer learning is to improve the classifier's performance on the target domain by transferring the knowledge contained in the source domain [9]. Despite their success in many domains, little research has been conducted on transfer learning for networked data structures. This is because that most existing transfer learning methods require identically and independent distributed (IID) data to make the knowledge

This work was supported by the National Key R&D Program of China under Grant 2018AAA0101502 and the Science and technology project of SGCC (State Grid Corporation of China): fundamental theory of human-in-the-loop hybrid-augmented intelligence for power grid dispatch and control.

learned from the source domain can be used to solve problems in the target domain, which cannot be satisfied by the non-euclidean graph-structured data. Most existing cross-network analysis studies transfer knowledge across networks based on plenty of anchor links [10], which is not the case in most situations.

To overcome these problems, we devise a transfer learning method for unsupervised node classification. Our motivation is to learn network-invariant embeddings to make the knowledge can be transferred between different networks and project the embeddings to label-discriminative subspaces to improve the classification performance. In specific, we utilize a pair of graph attention networks (GATs) [11] with parameter sharing to represent nodes from both networks, whose objectives are to minimize the classification loss on the labeled source nodes and the Sinkhorn distributional distance between two embedding spaces. Then, inspired by the transferable linear discriminant analysis (TLDA) [12], we learn label-discriminative projections by preserving the low-rank structure of the two domains while enlarging the distances between different classes. Based on the obtained representations, an SVM classifier with margin maximization is utilized to predict the labels of target nodes based on the parameters learned from the labeled source nodes. This paper's contributions are concluded as follows.

- A transfer learning model for the unsupervised node classification is proposed, which addresses this task without the need for anchor links.
- Our method learns network-invariant embeddings with supervision from the Sinkhorn distance.
- We adopt the TLDA method to project nodes into label-discriminative subspaces, which improves the separability of nodes' representations.
- Empirical evaluations are carried out, our model's superiority and the usefulness of each component is evaluated.

We organize the reset paper as follows. Section II reviews the related studies of graph embedding and cross-network transfer learning. Section III states the problem and describes the model's details. In Section IV, we perform empirical evaluations and analyze the results. In Section V, we summarize this paper.

II. Related Work

In the following, recent advances in graph embedding and cross-network transfer learning are briefly reviewed.

A. Graph Embedding

Over the past decade, significant progress has been made in graph embedding research. Graph embedding is to represent nodes with embedding vectors that preserve their proximities to simplify the graph calculations [13]. Most existing graph embedding techniques can be categorized into random walk, matrix factorization or deep learning methods.

The earliest graph embedding methods are mostly matrix factorization based models, which utilize matrices to represent the edges between nodes and factorize the matrices to obtain nodes' embeddings. The matrices utilized by these methods include the adjacency matrix, Laplacian matrix, and so on. Representatives of these methods such as graph factorization [14] that factorizes the adjacency matrix and Laplacian eigenmaps [15] that factorizes the Laplacian matrix. Matrix factorization based methods are usually used to embed homogeneous graphs and can hardly tackle large-scale graphs as their computational complexities increase with the matrix scale.

Random walk based methods learn embeddings from a few random walk paths. For example, DeepWalk [16] utilizes SkipGram to learn latent representations based on the sample paths from truncated random walks, which optimizes the embeddings to maximize the softmax probability between neghboring nodes in the paths. node2vec [17] generates more informative embeddings than DeepWalk based on the biased random walks. Random walks can effectively tackle the graph embeddings of large-scale or incomplete graphs. However, this kind of methods can hardly preserve high order proximities between nodes.

Deep learning based methods utilize deep autoencoders to map graphs into embedding matrices while preserving the proximity contained in the data. Typical methods of this kind such as SDNE [18] and LINE [4]. In addition, the increasing research on graph neural networks has led to the development of graph convolutional networks (GCNs) [19] based graph embedding methods. GCNs iteratively aggregate the neighborhoods' embeddings of a node and update its embedding based on a convolution operator, which can preserve the global proximities between high-order neighborhoods.

B. Cross-Network Transfer Learning

Transfer learning aims to tackle problems in the target domain based on the knowledge studied from the source domain [20]. The main objective is to achieve effective knowledge transfer by minimizing the distributional divergence between the two domains [9]. Most commonly used distance measurements such as the KL divergence and maximum mean discrepancy (MMD). Recently, a few studies have implemented transfer learning on the research of multiple network analysis. For instance, Lee et al. [21] designed a graph transfer learning framework and investigated the conditions to achieve successful knowledge transfers between graphs. Salem et al. [22] devised a pipeline model which utilizes GCNs to tackle the virtual screening task and transfers the learned knowledge between multiple molecular datasets, which effectively alleviates the over-training and generalization problems caused by unbalanced and small datasets. Ye et al. [23] utilized instance weighting to employ the knowledge trained on the most valuable source instances to solve the edge classification problem in the target network. Fang et al. [24] mapped the structural features of different networks into the same latent space based on label propagation. Existing methods do not jointly consider the domain invariance and label discrimination while representing nodes.

Recently, with the development of optimal transport, the Sinkhorn distance has been widely utilized as an effective

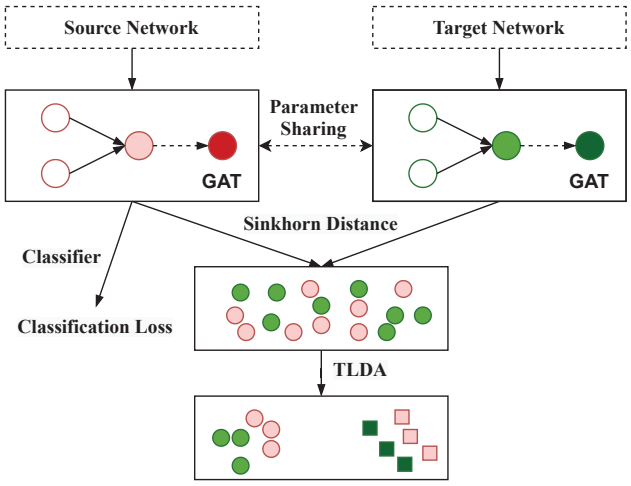

Fig. 1: Framework of the proposed method.

distance measure between different embeddings. For example, Xu et al. [25] proposed matching embeddings of two languages by minimizing their Sinkhorn distance for unsupervised cross-lingual translation and demonstrated the superiority of Sinkhorn distance over KL divergence. To learn label-discriminative representations for knowledge transfer tasks, Han et al. [12] devised the TLDA method to project embeddings into label discriminative subspaces. Inspired by these methods, we propose to learn network-invariant and label discriminative embeddings for for both source and target networks by minimizing their Sinkhorn distance and maximizing the separability for data from different classes.

III. Proposed Method

In the following, we first propose the mathematical definition of the cross-network node classification task. Then, we propose the technical details of our method.

A. Problem Definition and Framework Overview

Suppose that $G = (V, E, X, Y)$ is a network, in which V represents the nodes, E denotes the edges, X and Y are the attribute and label matrices of the nodes. The edge set can be described in the adjacency matrix. Given a source network $G_s = (V_s, E_s, X_s, Y_s)$ whose labels are known and a target network $G_t = (V_t, E_t, X_t, Y_t)$ whose labels are unknown, the cross-network node classification problem is to predict the target nodes' labels based on the knowledge learned from the source network. Note that the node categories of the target network must be the same as that of the source network. In addition, there are no equivalent nodes between G_s and G_t, and G_s and G_t are two disconnected networks, which is distinct from those anchor links based multi-network analysis studies.

To achieve this goal, we propose a transfer learning framework based on the representation learning, as is shown in Fig. 1. In this framework, we first utilize two GATs with parameter sharing to learn network-invariant embeddings by minimizing the Sinkhorn distance. Then, we adopt TLDA to project the embeddings into label-discriminative subspaces.

B. Graph Attention Network

Given G_s and G_t, we devise a graph embedding method based on the GAT model. Typically, GATs update nodes' representations by attending over their neighborhoods' features, which can be computed by

$$\vec{h}_i^{(l)} = \sigma\left(\sum_{j \in N_i} \alpha_{ij} \mathbf{W}^{(l)} \vec{h}_j^{(l-1)}\right), \tag{1}$$

where $\vec{h}_i^{(l)}$ represents the i-th node's hidden representation at the l-th layer, $\sigma(\cdot)$ denotes the activation function $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$, $\mathbf{W}^{(l)}$ is a linear transformation parameter matrix shared by all nodes, N_i is single-hop neighbors of the i-th node containing itself, α_{ij} represents the attention weight between the i-th node and its j-th neighbor, which is computed by

$$\alpha_{ij}^{(l)} = \mathrm{softmax}\left(a_{ij}^{(l)}\right) = \frac{\exp\left(a_{ij}^{(l)}\right)}{\sum_{k \in N_i} \exp\left(a_{ij}^{(l)}\right)}, \tag{2}$$

where $a_{ij}^{(l)}$ is the attention coefficient between the i-th node and its j-th neighbor, which is computed by the following equation:

$$a_{ij}^{(l)} = \eta\left(\vec{q}^T \left[\mathbf{W}^{(l)} \vec{h}_i^{(l)} \oplus \mathbf{W}^{(l)} \vec{h}_j^{(l)}\right]\right), \tag{3}$$

where $\eta(\cdot)$ denotes the LeakyReLu activation function, $\vec{q}^T$ is a learnable parameter vector, $\cdot^T$ represents the transposition operation, and $\oplus$ represents vector concatenation. We restrict the transformation matrix $\mathbf{W}^{(l)}$ to be a diagonal matrix to avoid overfitting caused by too much parameters and increase the generalization.

The GATs are trained with two aspects of supervision, the first of which is to make the embeddings of the source network discriminative according to their classes, while the second is to decrease the divergence between the target embeddings and the source embeddings to make the knowledge can be effectively transferred. To achieve this, we add a classifier above the embedding layers to utilize the source labels to optimize the embeddings. Specifically, the classification output is computed by the following equation:

$$\hat{y}_i = \phi\left(\vec{h}_i \mathbf{W}_c + \vec{b}_c\right), \tag{4}$$

where $\hat{y}_i \in R^c$ represents the predicted probabilities of the i-th node over all categories, $\phi(\cdot)$ is the ReLu function, and $\mathbf{W}_c$ and $\vec{b}_c$ are trainable parameters of the classifier. The classification loss to optimize the model with the source nodes, which is computed by the following cross-entropy loss:

$$\mathcal{L}_c = -\frac{1}{|V_s|} \sum_{i \in V_s} \sum_{k=1}^{c} y_{ik} \log(\hat{y}_{ik}) + (1 - y_{ik}) \log(1 - \hat{y}_{ik}), \tag{5}$$

in which $|V_s|$ is number of node in the source network. y_{ik} is the true class of the i-th node, $y_{ik} = 1$ represents that the i-th node is the k-th class; otherwise, $y_{ik} = 0$. $\hat{y}_{ik}$ denotes the predicted probability of the i-th node to be the k-th class. This loss function is also backpropagated to optimize the embeddings.

Sinkhorn distance is an optimal transportation distance that measures the divergence between two probability distributions. Besides, Sinkhorn distance is an approximation of the Wasserstein distance [26]. Therefore, we employ the Sinkhorn distance to calculate the closeness between the embeddings learned for the two domains. Our embedding module optimizes the GATs for lower Sinkhorn distance to make target embedding close to the distributions of the source embeddings. In particular, outputs of the GATs are normalized embeddings. Therefore, the target embeddings that are similar to the source embeddings in distribution are also adjacent to the source embeddings in the vector space.

The Sinkhorn distance between embeddings H_1 and H_2 is defined as:

$$d(H_1, H_2) = \min_{T \in U_\alpha(r,c)} \langle T, M \rangle, \tag{6}$$

where $\langle \cdot, \cdot \rangle$ denotes the Forbenius dot product. M is the transportation cost matrix, which we adopt the square root cosine distance proposed by Xu et al. [25]. $U_\alpha(r, c)$ is the transport polytope with the entropy constraint, defined as:

$$\begin{aligned} U_\alpha(r, c) = \{T \in \mathbb{R}_+^{n \times n} | T\vec{1}_n = r, T^T\vec{1}_n = c, \\ \varepsilon(T) \leq \varepsilon(r) + \varepsilon(c) - \alpha\}, \end{aligned} \tag{7}$$

in which $\varepsilon(\cdot)$ is the entropy function. r and c are the weights for the source nodes and the target nodes, which are set to be proportional to their degrees.

Algorithm 1: Calculation of Sinkhorn distance

Input: M, r, c, μ, T
Output: $d(H_1, H_2)$

```
K = e^{-μM};
v = 1⃗/n;
for t from 1 to T do
    u = r./Kv;
    v = c./K^T u;
    d(H_1, H_2) = u^T ((K ⊗ M), v);
```

For a batch of samples with n source nodes and n target nodes, the Sinkhorn distance between their embeddings is calculated by Algorithm 1, in which μ is the Lagrange multiplier for the entropy constraint and $\otimes$ is the inner-product calculation. For each batch, the calculation of the Sinkhorn distance is repeated T times.

The distributional divergence loss is then defined as the Sinkhorn loss, which is denoted as:

$$\mathcal{L}_d = d(H_s, H_t). \tag{8}$$

The total loss for the optimization of the embedding module is $\mathcal{L}_c + \lambda_1 \mathcal{L}_d$, where λ_1 is the relative weight of the distributional divergence loss.

C. Transferable Linear Discriminant Analysis

The core idea of TLDA [12] is to reduce the variants of distributions within the same subspace and enlarge the vector distances between different classes while preserving the low-rank structure of both domains during linear discriminant analysis. Therefore, inspired by TLDA, the loss function of projecting the embeddings into different subspaces for cross-network node classification is defined as:

$$\begin{aligned} \mathcal{L}(P) = \min_P \operatorname{tr}\left(P\left(S_w - \xi S_b\right)P^T\right) \\ + \lambda_2 \left\{ \sum_{c=1}^{C} \|PH_s^c\|_* - \|PH\|_* \right\} \\ \text{s.t. } \|P\|_2 = 1 \end{aligned} \tag{9}$$

where P represents the linear projection matrix, S_w represents the intra-class scatter matrix while S_b denotes the inter-class scatter matrix, H is the concatenation of the source embedding and the target embedding, $\|\cdot\|_*$ denotes the nuclear norm, and λ_2 is a weight hyperparameter.

The projection matrix is optimized with the projected sub-gradient optimization algorithm proposed by Han et al. [12], i.e., $P_{t+1} = P_t - \rho \triangle P_t$, where P_t is P at the t-th iteration and ρ is the step size. P is renormalized after each iteration. The subgradient is calculated by

$$\begin{aligned} \triangle P =& 2P\left(S_w - \xi S_b\right) + \\ & \lambda_2 \left(\sum_{c=1}^{C} \triangle \|PH_s^c\|_* \left(H_s^c\right)^T - \triangle \|PH\|_* H^T \right), \end{aligned} \tag{10}$$

in which $\triangle \|\cdot\|_*$ represents the nuclear norm's subdifferential. With the optimized projection matrix, the node representations with better label-discriminant can be obtained. Then, we train an SVM classifier on the projected representations of the source network with their labels and utilize this model to predict the categories of target nodes with their projected representations.

TABLE I: Description of the network datasets.

Network	Nodes	Edges	Common Attributes	Categories
Blog1	2300	33471	8189	6
Blog2	2896	53836		
Citationv1	8935	15113	6775	5
DBLPv7	5484	8130		

IV. Experimental Evaluations

In this section, we carry out cross-network node classification experiments on four networks to evaluate the proposed method. We conduct comparisons with both network transfer learning methods and graph embedding based methods and analyze the experimental results in detail. An ablation study is also conducted to evaluate each module's effect.

TABLE II: Overall node classification results of different models.

G_s	G_t	F1	MMD	DANN	NetTr	ANRL	LANE	SEANO	GCN	Ours
Blog1	Blog2	Micro-F1	0.4384	0.4492	0.5011	0.4772	0.4701	0.4985	0.5111	**0.6350**
		Macro-F1	0.4368	0.4481	0.4914	0.4587	0.4572	0.4953	0.4785	**0.6239**
Blog2	Blog1	Micro-F1	0.4592	0.4652	0.5242	0.4415	0.4955	0.5022	0.4981	**0.5591**
		Macro-F1	0.4579	0.4643	0.5150	0.4223	0.4942	0.4981	0.4631	**0.5468**
Citationv1	DBLPv7	Micro-F1	0.5702	0.5782	0.5985	0.6601	0.5855	0.6929	0.7122	**0.7201**
		Macro-F1	0.5355	0.5514	0.5516	0.6275	0.5506	0.6691	0.6810	**0.6868**
DBLPv7	Citationv1	Micro-F1	0.5341	0.5625	0.5910	0.6661	0.5694	0.7145	**0.7163**	0.7082
		Macro-F1	0.4960	0.5411	0.5552	0.6342	0.5381	0.6951	0.6717	**0.6968**

A. Experimental Setup

Datasets: Our method is evaluated on two pairs of cross-network classification datasets [20], whose statistics are concluded in Table I. Blog1 and Blog2 are two disconnected datasets collected from the BlogCatalog network, while Citationv1 and DBLPv7 are two networks extracted from the ArnetMiner citation dataset. The experiments are conducted between Blog1 and Blog2 and between Citationv1 and DBLPv7.

Baselines: The following baseline are utilized to evaluate our method [27]: MMD [28], which utilizes the MMD measure to match the distributions of two domains; DANN [29], which incorporates the gradient reversal layer into the representation learning module to obtain features that are discriminative for the classification task while indiscriminative between different domains; NetTr [24], which constructs label propagation matrices in both source and target networks to map their latent features into a shared space; ANRL [30], which incorporates the structural and attribute information of nodes to learn network embeddings; LANE [31], which incorporates label information and preserves nodes' correlations while learning their embeddings; SEANO [32], which learns network representations incorporating the structural, attribute, and label information; GCN [19], which utilizes a GCN model to learn node representations. MMD, DANN, and NetTr are transfer learning models, while the other baselines are graph embedding models.

Parameter settings: We implement our method and the baseline models on a work station with an Intel Xeon(R) CPU E5-2630 v4 @ 2.2GHZ and GPU TITAN RTX (32G). The configurations of the embedding module are as follows: embedding dimension 64, learning rate 0.0001, batch size 100, training epochs 20, $\lambda_1 = 15$, and the calculation of the Sinkhorn distance is iterated for 20 times. For the TLDA module, ξ, ρ, and λ_2 are all set to 0.0001, and it is trained for 10 epochs.

B. Results and Analysis

In our proposed method, we train a SVM classifier on the source domain and then employ it to predict the categories of the target nodes. The Micro-F1 and Macro-F1 are utilized to compare different methods. We report the evaluation results in Table II.

The results indicate that the proposed method performs the best on these network knowledge transfer tasks. We can also see that NetTr achieves the highest F1 scores than other transfer learning based baselines. This is because MMD and DANN consider each sample independently while learning domain-invariant representations. However, networks are not IID data, for which the complex relationships between nodes should be considered. Among the graph-embedding based methods, GCN achieves the best cross-network classification performance in most experimental scenarios as it effectively utilizes the higher-order structural proximities, the attributes and labels of nodes to learn node embeddings. However, due to that GCN does not address the discrepancy between different domains, the proposed method still performs better than it.

C. Ablation Study

In the following, ablation studies are carried out to evaluate the usefulness of different components in the proposed method. To achieve this, we present two variants, -Sinkhorn and -TLDA, in which we remove the Sinkhorn distance-based distributional divergence loss in the embedding module and the TLDA module, respectively. Besides, we also propose a variant, +LDA, which replaces the TLDA with traditional LDA to demonstrate the difference of learning label-discriminative representations in the network transfer learning situations.

TABLE III: Experiment results of ablation studies of the proposed method.

G_s	G_t	F1	Ours	-Sinkhorn	-TLDA	+LDA
Blog1	Blog2	Micro-F1	**0.6350**	0.3053	0.6277	0.6336
		Macro-F1	**0.6239**	0.2450	0.6128	0.6219
Blog2	Blog1	Micro-F1	**0.5591**	0.3309	0.5483	0.5496
		Macro-F1	**0.5468**	0.3000	0.5203	0.5389
Citationv1	DBLPv7	Micro-F1	**0.7201**	0.6750	0.7125	0.7199
		Macro-F1	**0.6868**	0.6450	0.6858	0.6953
DBLPv7	Citationv1	Micro-F1	**0.7082**	0.6384	0.7045	0.7051
		Macro-F1	**0.6968**	0.5655	0.6883	0.6922

We report the ablation results in Table III. The results indicate that the distributional divergence loss is important for the proposed method. Without the supervision from the Sinkhorn distance, the GATs cannot learn network-invariant embeddings, thus the classifier learned from the source nodes cannot be effectively applied on the target network. We can also see that the TLDA model is beneficial to achieve better performance. Besides, compared with the traditional LDA

method, projections with TLDA are more label discriminative, which results in higher F1 scores.

V. Conclusion

In this paper, we propose a novel method achieve cross-network node classification with the network-invariant and label-discriminative representations. In specific, our proposed method utilizes GATs with the supervision from the Sinkhorn distance between the source embeddings and the target embeddings to mitigate the domain shift and employs TLDA to project the embeddings into label-discriminative subspaces. Based on these representations, the classifier learned from the source nodes can be effectively utilized to the target network and achieve better classification performance. Evaluation results indicate the superiority of the proposed method.

References

[1] X. Wang, X. Zheng, X. Zhang, K. Zeng, and F.-Y. Wang, "Analysis of cyber interactive behaviors using artificial community and computational experiments," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 6, pp. 995–1006, 2017.

[2] T. Wang, X. Wang, S. Tang, Y. Lin, W. Liu, Z. Liu, B. Xiu, D. Shen, X. Zhao, and Y. Gao, "Collaborations patterns and productivity analysis for ieee t-its between 2010 and 2013," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 6, pp. 2360–2367, 2014.

[3] L. Li, Q. Zhang, X. Wang, J. Zhang, T. Wang, T.-L. Gao, W. Duan, K. K.-f. Tsoi, and F.-Y. Wang, "Characterizing the propagation of situational information in social media during covid-19 epidemic: A case study on weibo," *IEEE Transactions on Computational Social Systems*, vol. 7, no. 2, pp. 556–562, 2020.

[4] J. Tang, M. Qu, M. Wang, M. Zhang, J. Yan, and Q. Mei, "Line: Large-scale information network embedding," in *Proceedings of the 24th International Conference on World Wide Web*, 2015, pp. 1067–1077.

[5] W. Xia, S. Wang, M. Yang, Q. Gao, J. Han, and X. Gao, "Multi-view graph embedding clustering network: Joint self-supervision and block diagonal representation," *Neural Networks*, vol. 145, pp. 1–9, 2022.

[6] T. Shen, C. Gou, F.-Y. Wang, Z. He, and W. Chen, "Learning from adversarial medical images for x-ray breast mass segmentation," *Computer Methods and Programs in Biomedicine*, vol. 180, pp. 1–13, 2019.

[7] X. Wang, Z. Peng, S. Wang, P. S. Yu, W. Fu, and X. Hong, "Cross-domain recommendation for cold-start users via neighborhood based feature mapping," in *Database Systems for Advanced Applications*, J. Pei, Y. Manolopoulos, S. Sadiq, and J. Li, Eds., 2018, pp. 158–165.

[8] S. Mulay, K. Ram, B. Murugesan, and M. Sivaprakasam, "Style transfer based coronary artery segmentation in x-ray angiogram," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, October 2021, pp. 3393–3401.

[9] L. Yang, X. Wang, Y. Dai, K. Xin, X. Zheng, W. Ding, J. Zhang, and F.-Y. Wang, "Hackrl: Reinforcement learning with hierarchical attention for cross-graph knowledge fusion and collaborative reasoning," *Knowledge-Based Systems*, vol. 233, p. 107498, 2021.

[10] X. Du, J. Yan, and H. Zha, "Joint link prediction and network alignment via cross-graph embedding," in *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, 2019, pp. 2251–2257.

[11] P. Velikovi, G. Cucurull, A. Casanova, A. Romero, P. Li, and Y. Bengio, "Graph attention networks," in *6th International Conference on Learning Representations*, 2017, pp. 1–12.

[12] N. Han, J. Wu, X. Fang, J. Wen, S. Zhan, S. Xie, and X. Li, "Transferable linear discriminant analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5630–5638, 2020.

[13] B. Jiang, L. Wang, J. Cheng, J. Tang, and B. Luo, "Gpens: Graph data learning with graph propagation-embedding networks," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–14, 2021.

[14] K. Zhang, X. Zhao, and S. Peng, "Multiple graph regularized semi-supervised nonnegative matrix factorization with adaptive weights for clustering," *Engineering Applications of Artificial Intelligence*, vol. 106, p. 104499, 2021.

[15] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic*, 2001, pp. 585–591.

[16] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2014, pp. 701–710.

[17] A. Grover and J. Leskovec, "Node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 855–864.

[18] D. Wang, P. Cui, and W. Zhu, "Structural deep network embedding," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2016, pp. 1225–1234.

[19] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, 2017, pp. 1–17.

[20] X. Shen, Q. Dai, S. Mao, F.-L. Chung, and K.-S. Choi, "Network together: Node classification via cross-network deep network embedding," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 5, pp. 1935–1948, 2021.

[21] J. Lee, H. Kim, J. Lee, and S. Yoon, "Transfer learning for deep learning on graph-structured data," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, pp. 2154–2160, 2017.

[22] M. Salem, A. Khormali, A. K. Arshadi, J. Webb, and J.-S. Yuan, "Transcreen: Transfer learning on graph-based anti-cancer virtual screening model," *Big Data and Cognitive Computing*, vol. 4, no. 3, pp. 1–20, 2020.

[23] J. Ye, H. Cheng, Z. Zhu, and M. Chen, "Predicting positive and negative links in signed social networks by transfer learning," in *Proceedings of the 22nd International Conference on World Wide Web*, 2013, pp. 1477–1488.

[24] M. Fang, J. Yin, and X. Zhu, "Transfer learning across networks for collective classification," in *2013 IEEE 13th International Conference on Data Mining*, 2013, pp. 161–170.

[25] R. Xu, Y. Yang, N. Otani, and Y. Wu, "Unsupervised cross-lingual transfer of word embedding spaces," in *Conference on Empirical Methods on Natural Language Processing*, 2018, pp. 1–10.

[26] G. Luise, A. Rudi, M. Pontil, and C. Ciliberto, "Differential properties of sinkhorn approximation for learning with wasserstein distance," in *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, 2018, pp. 5864–5874.

[27] X. Shen, Q. Dai, F.-l. Chung, W. Lu, and K.-S. Choi, "Adversarial deep network embedding for cross-network node classification," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 03, 2020, pp. 2991–2999.

[28] A. Gretton, K. Borgwardt, M. Rasch, B. Schölkopf, and A. Smola, "A kernel method for the two-sample-problem," in *Advances in Neural Information Processing Systems*, vol. 19, 2007, pp. 1–8.

[29] Y. Ganin, E. Ustinova, H. Ajakan, P. Germain, H. Larochelle, F. Laviolette, M. Marchand, and V. Lempitsky, "Domain-adversarial training of neural networks," *The Journal of Machine Learning Research*, vol. 17, no. 1, pp. 2096–2030, Jan. 2016.

[30] Z. Zhang, H. Yang, J. Bu, S. Zhou, P. Yu, J. Zhang, M. Ester, and C. Wang, "Anrl: Attributed network representation learning via deep neural networks," in *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018, pp. 3155–3161.

[31] X. Huang, J. Li, and X. Hu, "Label informed attributed network embedding," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 2017, pp. 731–739.

[32] J. Liang, P. Jacobs, J. Sun, and S. Parthasarathy, "Semi-supervised embedding in attributed networks with outliers," in *Proceedings of the 2018 SIAM International Conference on Data Mining*, 2018, pp. 1–10.