Grad-Align: Gradual Network Alignment via Graph Neural Networks (Student Abstract)

Jin-Duk Park[1], **Cong Tran**[2], **Won-Yong Shin**[1*], **Xin Cao**[3]

[1]Mathematics and Computing (Computational Science and Engineering), Yonsei University, Seoul 03722, Republic of Korea
[2]Faculty of Information Technology, Posts and Telecommunications Institute of Technology, Hanoi 100000, Vietnam
[3]Computer Science and Engineering, The University of New South Wales, Sydney 2052, Australia
jindeok6@yonsei.ac.kr, congtt@ptit.edu.vn, wy.shin@yonsei.ac.kr, xin.cao@unsw.edu.au

Abstract

Network alignment (NA) is the task of finding the correspondence of nodes between two networks. Since most existing NA methods have attempted to discover every node pair *at once*, they may fail to utilize node pairs that have strong consistency across different networks in the NA task. To tackle this challenge, we propose Grad-Align, a new NA method that *gradually* discovers node pairs by making full use of either node pairs exhibiting strong consistency or prior matching information. Specifically, the proposed method gradually aligns nodes based on both the similarity of embeddings generated using graph neural networks (GNNs) and the Tversky similarity, which is an asymmetric set similarity using the *Tversky index* applicable to networks with different scales. Experimental evaluation demonstrates that Grad-Align consistently outperforms state-of-the-art NA methods in terms of the alignment accuracy. Our source code is available at https://github.com/jindeok/Grad-Align.

Introduction

Considerable attention has been paid to conducting NA, which is often the very first step to solve many machine learning tasks on multiple networks. However, since most existing NA methods including (Zhang and Tong 2016; Trung et al. 2020) have attempted to discover all node pairs *at once* based on their own similarity measures, they do not distinctly utilize node pairs having *strong consistency*, indicating pairs of nodes that have highly consistent attributes and topological structures across different networks. To solve this problem, we propose Grad-Align, a new NA method taking advantage of strongly consistent node pairs by discovering node pairs *gradually* based on two types of similarity measures.

Methodology

Given a source network G_s and a target network G_t whose numbers of vertices are denoted by n_s and n_t, respectively, Grad-Align gradually finds node pairs by calculating two key similarities, including 1) the embedding similarity via graph convolutional network (GCN) that captures structural and attribute semantic relations between nodes and 2) the so-called Tversky similarity as an asymmetric set similarity based on the Tversky index (Tversky 1977) that enables us to overcome the problem raised by networks with different scales.

Embedding Similarity

The feed-forward process of GCN with shared weights can be formulated as follows:

$$H_*^{(l+1)} = \sigma(\hat{D}_*^{-\frac{1}{2}} \hat{A}_*^{-\frac{1}{2}} \hat{D}_*^{-\frac{1}{2}} H_*^{(l)} W^{(l)}), \qquad (1)$$

where the subscript * represents s and t for source and target networks, respectively; $H^{(l)}$ is a hidden representation matrix at layer l; $\hat{A} = A + I$ is the adjacency matrix with self-connections in which I is the identity matrix; $\hat{D}$ is a diagonal matrix with $\hat{D}_{i,i} = \sum_j \hat{A}_{i,j}$; $\sigma(.)$ is the activation function; and the $W^{(l)}$ is a trainable weight matrix at layer l and is shared for both source and target networks. Based on extracted representations at each layer, we compute the multi-layer embedding similarity matrix $\mathbf{S}_{emb} \in \mathbb{R}^{n_s \times n_t}$ whose entry represents the similarity of structural and attribute relations between nodes in the two networks:

$$\mathbf{S}_{emb} = \sum_l H_s^{(l)} {H_t^{(l)}}^\top. \qquad (2)$$

Tversky Similarity

In realistic scenarios, the number of nodes in source and target networks is often far different. To overcome this problem, we present an iterative asymmetric set similarity using the Tversky index. Given two sets of already aligned nodes at the i-th iteration in G_s and G_t, denoted as $\mathcal{A}_s^{(i)}$ and $\mathcal{A}_t^{(i)}$, respectively,[1] the Tversky similarity $\mathbf{S}_{Tve}^{(i)}(u, v)$ between two nodes u and v at the i-th iteration in source and target networks, respectively, measures how many aligned cross-network neighbor-pairs that they have upon alignment mapping $\pi^{(i-1)}$. Here, $\pi^{(i-1)}$ indicates one-to-one mapping $\pi^{(i-1)} : \mathcal{A}_s^{(i-1)} \rightarrow \mathcal{A}_t^{(i-1)}$ at the $(i-1)$-th iteration step.

*Corresponding author

[1]If $\mathcal{A}_s^{(0)}$ and $\mathcal{A}_t^{(0)}$ are $\emptyset$, meaning that no prior information is given, then we use only $\mathbf{S}_{emb}$ at the first matching step.

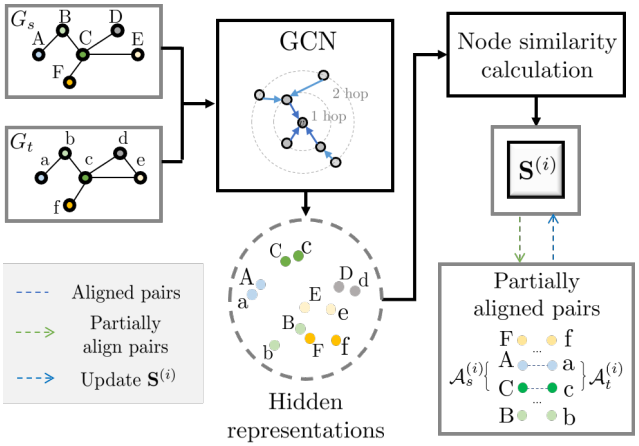

Figure 1: The schematic overview of our Grad-Align method

Gradual Alignment

Finally, our similarity matrix at the i-th iteration, $\mathbf{S}^{(i)} \in \mathbb{R}^{n_s \times n_t}$, is calculated by using the above two similarity matrices as follows:

$$\mathbf{S}^{(i)} = \mathbf{S}_{emb} \odot \mathbf{S}^{(i)}_{Tve}, \tag{3}$$

where $\odot$ indicates the element-wise matrix multiplication operator. We design a gradual assignment strategy on $\mathbf{S}^{(i)}$ according to the following iterative process. At the i-th step, we aim at finding $N(i) > 0$ node pairs corresponding to the top-$N(i)$ highest score entries in $\mathbf{S}^{(i)}$.[2] Then, we add newly aligned nodes to the sets $\mathcal{A}_s^{(i)}$ and $\mathcal{A}_t^{(i)}$, and update $\pi^{(i)}$, $\mathbf{S}^{(i+1)}_{Tve}$, and $\mathbf{S}^{(i+1)}$ accordingly. This process is repeated until all correspondences are found. The schematic overview of the proposed method is presented in Figure 1, where the two most consistent pairs (A, a) and (C, c) are aligned at the first step.

Experimental Evaluation

Datasets

1. **Facebook vs. Twitter** Subnetworks of Facebook and Twitter.
2. **Douban online vs. Douban offline** Subnetworks of an online and offline Chinese social network.
3. **Econ vs. Econ** The economic model network of Victoria state. Here, a perturbed version is made by dropping 10% of edges randomly for each network.

Experimental Results

We compare our method with five state-of-the-art NA methods including DeepLink (Zhou et al. 2018), PALE (Man et al. 2016), FINAL (Zhang and Tong 2016), CENALP (Du, Yan, and Zha 2019), and GAlign (Trung et al. 2020). As for experimental settings, 10% of ground truth node correspondences are randomly selected and used as the training set over all the NA methods as in (Trung et al. 2020). Table 1 shows the evaluation performance in terms of the *alignment accuracy*, which indicates the proportion of correct node correspondences out of the total correspondences. Grad-Align is shown to significantly outperform all state-of-the-art methods. Such a tremendous gain is possible due to the fact that Grad-Align leverages node pairs having strong consistency in order to find pairs revealing relatively weak consistency, resulting in a perfect matching score in a perturbed network setting (i.e., Econ vs. Econ). Moreover, since Grad-Align relies on the structural and/or attributed consistency for matching, it works successfully for both scenarios where no attributes are given (e.g., Facebook vs. Twitter) and the network structure is inconsistent between source and target networks (e.g., Douban online vs. Douban offline).

[2] Already aligned node pairs are excluded from the top-$N(i)$ calculation in $\mathbf{S}^{(i)}$.

Method	Facebook Twitter	Douban online Douban offline	Econ Econ
DeepLink	0.36	0.08	0.42
PALE	0.82	0.12	0.65
FINAL	0.87	0.37	0.50
CENALP	0.91	0.19	0.89
GAlign	0.02	0.28	0.85
Grad-Align	**0.95**	**0.50**	**1.00**

Table 1: Performance comparison among Grad-Align and five state-of-the-art NA methods

Acknowledgements

The work of W.-Y. Shin was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021R1A2C3004345), by Institute of Information & communications Technology Planning Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2021-0-00347, 6G Post-MAC (POsitioning- & Spectrum-aware intelligenT MAC for Computing & Communication Convergence)), and by the Yonsei University Research Fund of 2021 (2021-22-0083). The work of X. Cao was supported by ARC DE190100663.

References

Du, X.; Yan, J.; and Zha, H. 2019. Joint link prediction and network alignment via cross-graph embedding. In *IJCAI*, 2251–2257.

Man, T.; Shen, H.; Liu, S.; Jin, X.; and Cheng, X. 2016. Predict anchor links across social networks via an embedding approach. In *IJCAI*, 1823–1829.

Trung, H. T.; Van Vinh, T.; Tam, N. T.; Yin, H.; Weidlich, M.; and Hung, N. Q. V. 2020. Adaptive network alignment with unsupervised and multi-order convolutional networks. In *ICDE*, 85–96.

Tversky, A. 1977. Features of similarity. *Psychological review*, 84(4): 327.

Zhang, S.; and Tong, H. 2016. FINAL: Fast attributed network alignment. In *SIGKDD*, 1345–1354.

Zhou, F.; Liu, L.; Zhang, K.; Trajcevski, G.; Wu, J.; and Zhong, T. 2018. DeepLink: A deep learning approach for user identity linkage. In *INFOCOM*, 1313–1321.