IJICTR
International Journal of Information & Communication Technology Research

Volume 14- Number 3 – 2022 (19 -26)

DSMN: A New Approach for Link Prediction in Multilplex Networks

Samira Rafiee
Department of Computer Engineering
University of Kurdistan
Sanandaj, Iran
samira.rafiee@eng.uok.ac.ir

Alireza Abdollahpouri*
Department of Computer Engineering
University of Kurdistan
Sanandaj, Iran
abdollahpouri@uok.ac.ir

Received: 6 March 2022 – Revised: 15 May 2022 - Accepted: 3 September 2022

Abstract—**In a multiplex network, there exists different types of relationships between the same set of nodes such as people which have different accounts in online social networks. Previous researches have proved that in a multiplex network the structural features of different layers are interrelated. Therefore, effective use of information from other layers can improve link prediction accuracy in a specific layer. In this paper, we propose a new inter-layer similarity metric DSMN, for predicting missing links in multiplex networks. We then combine this metric with a strong intra-layer similarity metric to enhance the performance of link prediction. The efficiency of our proposed method has been evaluated on both real-world and synthetic networks and the experimental results indicate the outperformance of the proposed method in terms of prediction accuracy in comparison with similar methods.**

Keywords: Link prediction; Multilayer networks; Inter-layer similarity metric; structural information.

Article type: Research Article

[DOI: 10.52547/itrc.14.3.19]

I. INTRODUCTION

Recently, network analysis has become a novel branch of science, and many social and natural systems can be modeled as networks in which separate units are interconnected by links [1, 2]. Transportation networks and online social networks are some examples of these systems. One of the most important issues in network analysis is the problem of link prediction which aims to predict the missing links [3]. Link prediction is used in various fields such as biological networks [4, 5] and social networks [3]. For example, in biological networks, link prediction algorithms can be applied to discover missing interactions between proteins [6] and in social networks it can be used to suggest friendship requests to individuals [7].

Various methods have been suggested in the literature for link prediction that can be divided into two general categories: similarity-based methods and learning-based methods [7]. In similarity-based methods, a similarity score is assigned to each pair of unconnected nodes, and it is assumed that the higher the similarity score between the two nodes, the more likely the link between them will appear in the future. But, in learning-based methods, machine learning techniques are used. Usually the performance of link prediction in

* Corresponding Author

learning-based methods is better than similarity-based methods, but their computational complexity is also higher [8]. Most real-world systems are modeled as single-layer networks. But, there are some systems that can be best modeled in multiple layers [9, 10]. For instance, online social networks used by the same people are represented as a multiplex network. The links between these individuals in each social network can be considered as a layer of a larger network. As another example, cities may be connected by air, road, and rail transportations; the connections in each of the networks can be thought of as a layer of a larger transportation network [11].

Multilayer networks can be considered as a network of networks. In real multilayer networks, there is a strong correlation between the properties of nodes in different layers [12]. As a result, link prediction in one of the layers, in addition to the information of this layer, also depends on the structural information of other layers. The importance of this issue becomes clearer when it is proven that link prediction performance increases by considering the structural information of all layers [13]. This motivates us to propose a new link prediction method in multilayer networks with respect to the structural information of other layers.

Recently, several methods have been proposed to deal with the problem of link prediction in multiplex networks. Pujari et al. proposed an approach to predict links of co-authorship based on using information contained in bibliographical multiplex networks [14]. Their method is based on machine learning and they use a set of topological properties to describe positive and negative examples. Sharma et al. [15] have proposed an approach for link prediction in multiple networks. This method is also based on machine learning in which associations use multiple network layers for learning. Hajibagheri et al. [16] proposed a comprehensive framework called MLP in which the probability of having a link for a target layer is determined by the presence of links in other layers of the network. These probabilities are used to re-weight the output of a one-layer link prediction method that uses rank agglomeration to merge a set of topological criteria. The authors of [17], have proposed a systematic approach based on inter-layer similarity and features based on intra-layer proximity for link prediction in multiplex networks.

In [18] an approach based on the information related to the two-layer communication is proposed by Ruoqian Yang et al. to predict links for a user in a two-layer social network. By extracting the topological features from the network structure and using reliable paths between users, they developed a similarity measure for link prediction. Using these criteria, both the presence of links and their weight can be predicted. The authors of [19] proposed SimBins which is an automatic similarity-based multiplex link prediction method. Based on observed inter-layer correlations, it aims to quantify the amount of connection uncertainty in a multiplex network. Meanwhile, in SimBins, the prediction quality of the target layer is enhanced by combining the effect of link overlap in the layers.
Nasiri et al. proposed Multiplex Local Random Walk (MLRW) which is an extension of local random walk based on pure random walking for predicting links in multiplex networks [20]. In their proposed approach, they used inter-layer and intra-layer information extracted from multiplex network to define a biased random walk for finding the appearance likelihood of a new link in target layer.

In this paper, we propose a new inter-layer similarity metric, DSMN (Degree Similarity in Multilayer Networks) to solve the link prediction problem, which pays attention to the structural information of other network layers as well, to predict the links in one layer. We then combine this inter-layer similarity metric with our strong previously proposed intra-layer similarity metric to increase the performance of link prediction in multilayer networks.

The rest of this article is organized as follows. Section II gives some background information and introduces inter and intra layer similarity metrics in multi-layer networks. Section III describes the proposed approach; and the results of evaluating the proposed approach are reported in Section IV. Finally, the paper is concluded in Section V.

II. BACKGROUND

As mentioned earlier, some real-world systems are modeled as multi-layered networks. Each of these layers is a network communicates through a number of links and forms a larger network. These links contain information that is useful to solve the link prediction problem. If this information, known as interlayer information, is ignored, the performance of the link prediction algorithm in multilayer networks is reduced [21].

To better understand the problem, assume a multilayer network G = (L1,…, LN) where Li = (V, E_i) is the i^{th} layer of the network G, N is the number of layers, V is the set of nodes and E_i is the set the edges of the layer *i*. The problem of link prediction in network G is the prediction of potential links between pairs of disconnected nodes in one of the layers (see Figure 1). Solving this problem requires paying attention to the inter-layer properties along with the intra-layer properties. According to the inter-layer properties, the inter-layer similarity metrics are defined and according to the intra-layer properties, the intra-layer similarity metrics are defined which will be discussed in the following subsection.

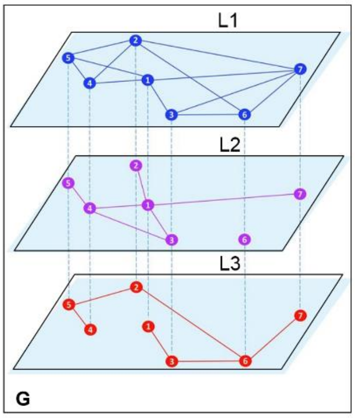

Figure 1. Multilayer network G with three layers L1, L2 and L3.

[DOI: 10.52547/itrc.14.3.19]

A. Intra-layer similarity metrics

Several similarity metrics have been suggested to solve the problem of link prediction in single-layer networks. From these criteria, it is possible to obtain the probability of existence a link between two nodes. These criteria, known as "inter-layer similarity metrics", are generally divided into two categories: neighborhood-based and path-based.

Neighborhood-based similarity metrics use only the properties of node neighbors, so they are based on local information. A wide variety of neighborhood-based link prediction methods have been proposed previously, including Common Neighbors (CN) [22], Jacquard Coefficient (JC) [22], and Adamic Adar (AA) [23], Preferred Supplement (PA) [24], Resource Allocation (RA) and Local Path (LP) [25].

Path-based similarity metrics are based on the global information of the nodes. Some examples of path-based similarity metrics are Katz [26], Page Rank (PR) [27], and Rooted Page Rank (RPR) [28].

We have recently introduced a new intra-layer similarity measure known as CNDP [29]. CNDP is a neighborhood-based similarity metric that, in addition to the degree of common neighbors of two nodes, the relationship between these neighbors is also considered. On the other hand, it also uses the average clustering coefficient of the network. As a result, it is suitable for predicting links in any type of network. In this paper, we combine the CNDP similarity metric with the proposed inter-layer similarity metric to improve the performance of link prediction in multilayer networks.

B. Inter-layer similarity metrics

So far, various methods have been proposed to calculate the inter-layer similarity in multilayer networks [30-33]. The following are the most commonly used:

- Degree-Degree Correlation (DDC): This metric calculates the inter-layer correlation of degrees of nodes in different layers. If DDC <0 the two layers are negatively correlated, if DDC> 0 the two layers are positively correlated and if DDC = 0 the two layers are not correlated.
- Based on betweenness (BW): To calculate this metric, at first the betweenness is calculated for a given node in each layer; then the difference of betweenness of this node is calculated for each two layers. This value is always between 0 and 1. Then, the betweenness similarity of this node in the two layers is obtained as 1 minus the betweenness distance. Finally, the betweenness similarity of all nodes is averaged for both layers to calculate the similarity of two layers.
- Average Similarity of Neighbors (ASN): Consider a two-layer network G = (L1, L2). There exist three types of links in G; intra-layer links in L1, intra-layer links in L2, and Inter-layer links between L1 and L2. For each node *i*, $k_{L_1}(i)$ is the degree of node *i* in layer L1, $k_{L_2}(i)$ is degree node *i* in layer L2, and $k_L(i)$ is degree of node *i* between layers L1 and L2. ASN, which is the similarity of the two layers L1 and L2, is defined as Equation (1) [33]:

$$ASN = \frac{\sum_i k_L(i)}{\sum_i (k_{L_1}(i) + k_{L_2}(i) - k_L(i))} \tag{1}$$

III. Proposed Method

In this section, at first the proposed inter-layer similarity metric is explained in detail. Thereafter, the way of its combination with our previously proposed intra-layer similarity metric (CNDP) is discussed. In the following, the proposed algorithm is introduced and its pseudo-code is given.

A. proposed interlayer similarity metric

Given a multilayer network G = (L1, …, LN) where N is the number of layers, the distance of node degree in the layer Lm and Lk which m, k∈ {1,…, N } is calculated as Equation (2):

$$D_{i_{(Lm,Lk)}} = \left| k_{i_{Lm}} - k_{i_{Lk}} \right| \tag{2}$$

Where $k_{i_{Lm}}$ is the degree of node *i* in the Lm layer and $k_{i_{Lk}}$ is the degree of node *i* in the Lk layer. Note that the value of $D_{i_{(Lm,Lk)}}$ is always between 0 and N-1; because the minimum degree of node *i* is 0 (when node *i* is not connected to any other node in the network) and the maximum degree of node *i* is N-1 (when node *i* is connected to all of the nodes in the network).

The similarity of degree of node *i* in two layers Lm and Lk is inversely related to the distance of degree of node *i* in these two layers; that is calculated as Equation (3):

$$S_{i_{(Lm,Lk)}} = \frac{1}{D_{i_{(Lm,Lk)}} + 1} \tag{3}$$

$S_{i_{(Lm,Lk)}}$ is the "similarity of node *i* in the Lm layer to itself in the Lk layer", which is symmetric according to (2). Note that $S_{i_{(Lm,Lk)}}$ is always greater than 0 and less than 1. If $D_{i_{(Lm,Lk)}} = 0$ then $S_{i_{(Lm,Lk)}} = 1$.

The similarity of node *i* to node *j* in two layers Lm and Lk is obtained from (4):

$$S_{(i,j)_{(Lm,Lk)}} = \frac{S_{i_{(Lm,Lk)}} + S_{j_{(Lm,Lk)}}}{I_{L_k}(i) + k_{j_L}} \tag{4}$$

Where, k_{i_L} is the number of connections between node *i* in two layers Lm and Lk.

The name of our proposed inter-layer similarity metric is degree similarity (abbreviated as DS), which is obtained from the degree distance. The equation for calculating the proposed inter-layer similarity metric is as follows:

[DOI: 10.52547/itrc.14.3.19]

$$S^{inter}_{(i,j)_{L_m}} = \begin{cases} \sum_{k=1,k\neq m}^{N} S_{(i,j)_{(L_m,L_k)}} & if\ I_{(i,j)_{Lk}} = 1 \\ \sum_{k=1,k\neq m}^{N} (1 - S_{(i,j)_{(L_m,L_k)}}) & if\ I_{(i,j)_{Lk}} = 0 \end{cases} \quad (5)$$

where I_{L_k} is the adjacency matrix of the layer Lk in network G, and N is the number of layers.

Suppose we want to calculate the inter-layer similarity between nodes *i* and *j* in the Lm layer. If there is a link between *i* and *j* in the Lk layer ($I_{(i,j)_{Lk}} = 1$), the greater the similarity between the two layers Lm and Lk, that is, the greater the similarity between nodes *i* and *j* due to their presence in two layers Lm and Lk ($S_{(i,j)_{(L_m,L_k)}}$), the interlayer similarity score of nodes *i* and *j* according to the presence in the Lm layer ($S^{inter}_{(i,j)_{L_m}}$) increases. So it is more likely that there exists a link between nodes *i* and *j* in layer Lm as well, and that makes perfect sense. Now, if there is no link between node *i* and node *j* in the Lk layer ($I_{(i,j)_{Lk}} = 0$), the similarity between two layers Lm and Lk, increases. It should be more likely that there is no link between *i* and *j* in the Lm layer. That is, the inter-layer similarity of *i* and *j* should be less due to their presence in layer Lm ($S^{inter}_{(i,j)_{L_m}}$). We subtract $S_{(i,j)_{(L_m,L_k)}}$, which is always less than 1, from 1 to reduce $S^{inter}_{(i,j)_{L_m}}$ and therefore, reduce the likelihood that there is a link between *i* and *j*.

B. Proposed algorithm

As mentioned earlier, our proposed algorithm is based on the use of both intra-layer properties and inter-layer similarity. In fact, in the proposed method for predicting the link between two nodes, first, the intra-layer similarity grade of these two nodes is calculated using the CNDP intra-layer similarity metric [26]. Then, using the DS inter-layer similarity metric, the inter-layer similarity of these two nodes is obtained. Finally, the average value of intra-layer and inter-layer similarities is calculated and considered as the final similarity score of two nodes. Thus, in general the similarity score between a pair of nodes (*i*, *j*) in the Lm layer of the network G is calculated as follows:

$$S_{(i,j)_{L_m}} = \frac{S_{L_m}{}^{intra}{}_{(i,j)} + S^{inter}{}_{(i,j)_{L_m}}}{2} \quad (6)$$

Here, $S_{(i,j)_{L_m}}$ is the total similarity between two node *i* and node *j* in the layer Lm, which is obtained from the average intra-layer similarity of *i* and *j* in the Lm layer ($S_{L_m}{}^{intra}{}_{(i,j)}$) and the inter layer similarity of *i* and *j* between the Lm layer and other network layers ($S^{inter}{}_{(i,j)_{L_m}}$). In addition, $S_{L_m}{}^{intra}{}_{(i,j)}$ is the same as the CNDP metric [26] and $S^{inter}{}_{(i,j)}$ is calculated from Equation (5). For the networks with more than two layers, the similarity between nodes *i* and *j* is obtained by getting similarities for all pairs of layers. Algorithm1 shows the pseudo-code of the proposed method.

Algorithm1: link Prediction in multilayer networks using intra-layer features and inter-layer similarity

---

Input: Lm= (V, E_m) layer of G=(L1,…,LN) Network where N is number of layers
Output: AUC , Precision
Begin algorithm:
Divide the Lm of into train set L_{train} and test set L_{test}.
For each non-observation edge (i, j) in L_{train} do

$$D_{i_{(Lm,Lk)}} = |k_{i_{Lm}} - k_{i_{Lk}}|$$ and $$D_{j_{(Lm,Lk)}} = |k_{j_{Lm}} - k_{j_{Lk}}|$$

$$S_{i_{(Lm,Lk)}} = \frac{1}{D_{i_{(Lm,Lk)}} + 1}$$ and $$S_{j_{(Lm,Lk)}} = \frac{1}{D_{j_{(Lm,Lk)}} + 1}$$

$$S_{(i,j)_{(Lm,Lk)}} = \frac{S_{i_{(Lm,Lk)}} + S_{j_{(Lm,Lk)}}}{k_{i_L} + k_{j_L}}$$

$$S^{inter}_{(i,j)_{L_m}} = \begin{cases} \sum_{k=1,k\neq m}^{N} S_{(i,j)_{(L_m,L_k)}} & if\ I_{(i,j)_{Lk}} = 1 \\ \sum_{k=1,k\neq m}^{N} (1 - S_{(i,j)_{(L_m,L_k)}}) & if\ I_{(i,j)_{Lk}} = 0 \end{cases}$$

$$S_{(i,j)_{L_m}} = \frac{S_{L_m}{}^{intra}{}_{(i,j)} + S^{inter}{}_{(i,j)_{L_m}}}{2}$$

End For
Sort the list of all Similarity scores in descending order.
Insert the edges of the sorted list to L_{train}.
Compute **AUC** and **Precision** from (7) and (8).
End algorithm

IV. Experimental Results

Our main goal in this paper is to combine intra-layer and inter-layer information to improve the performance of link prediction in multilayer networks. We evaluate the performance of our proposed method on both artificial and real data sets and report the results. To evaluate the efficiency of the proposed method, we compare it with the link prediction method [17]. Details of the experiments, such as selected datasets, evaluation criteria, and numerical results, are given below.

In general, link prediction is to predict links that do not exist in the train set and may appear in the future on the network. To identify these links, a similarity score must be given to each of them to determine the possibility of its presence in the future. In a single-layer network, this similarity score is determined by the

[DOI: 10.52547/itrc.14.3.19]

metric of similarity within the layer. However, in a multilayer network, in addition to an intra-layer similarity metric, we must also have an inter-layer similarity metric; and combine it with the intra-layer similarity metric so that we can suggest a suitable method for predicting links in multilayer networks. The effectiveness of the proposed method is ultimately measured by evaluation criteria. The evaluation metric used in this work is the AUC metric. This metric, which is explained below, is an outstanding evaluation metric in link prediction research.

A. *Datasets*

To evaluate the performance of the proposed method, both real-world and synthetic datasets have been used. At first, the performance of the method on synthetic datasets is evaluated. For the sake of simplicity, a Two-layer synthetic network is created and our proposed method is evaluated on it. To create the first layer, we use Barabasi-Albert (BA) model. Using a preferential attachment method, this model generates random scale-free networks. In this model, the network starts with an initial component of m nodes. Then, new nodes are attached to the network one by one. Each new node is likely to be connected to one or more existing nodes according to the degree of the existing nodes.

Thus, we created layer L1 of the two-layer network G using BA model with total number of 1000 nodes and 5 initial nodes (m = 5). We then copied all the nodes of this layer and some of its edges to layer L2. We assumed that each edge of L1 layer is copied to L2 with the probability of PCopy. Our goal is to predict missing links in L2 with respect to layer L1. In this way, we can find out if the proposed method works properly for predicting missing links. The results of the implementation of the algorithm on the BA artificial network can be seen in Section Evaluation results.

In the following, we run the proposed method on 4 real-world multilayer networks. These networks include the two-layer network Twitter-Foursquare, the two-layer network Twitter-Instagram, the five-layer network online and offline relationships, and the 37-layer European air transportation network. The specifications of these networks are given in Table 1.

Twitter-Foursquare [17] is a two-layer network that includes Twitter and Foursquare as a micro-blogging location and service-based social network, respectively. This network contains 1565 users. User's communication is directional on Twitter and undirected on Foursquare.

Twitter-Instagram [17] is a two-layer network including Twitter and Instagram. This network has 3298 users. Instagram is a directed network for uploading photos and videos which allows users to upload their contents to other social networks such as Twitter.

The 5-layer network of offline and online relationships [17] include, work, leisure, lunch, co-authorship, and Facebook. This network actually shows the relationships between the computer science employees of Aarhus University. This dataset was obtained through questionnaires distributed among 62 employees.

The EU air transportation network includes 37 layers, each of which belongs to an airline [17]. Each of these layers contains 450 nodes, indicating the airports, and the connections in each layer are an airway between two airports. The results of the proposed algorithm implementation on the four real networks used are shown in Section IV.C (Evaluation results).

TABLE I. SPECIFICATIONS OF REAL DATASETS USED IN THIS ARTICLE

Multilayer network	Layers	No. of nodes	No. of edges	Ratio of Common links
Twitter-Foursquare	Twitter	1,565	2,663	0.55
	Foursquare	1,565	36,056	0.098
Twitter-Instagram	Twitter	13298	52668	0.45
	Instagram	13298	227794	0.06
Online and Offline Relationships	Facebook	62	193	0.031
	Leisure	62	124	0.030
	Work	62	21	0.065
	CO-authorship	62	88	0.044
	Launch	62	194	0.031
EU-Air Transportation	Lufthansa	450	244	0.015
	Ryan air	450	601	0.004
	Easy jet	450	307	0.011
	British airways	450	66	0.016
	Turkish airlines	450	118	0.006
	Air Berlin	450	184	0.013
	Air France	450	69	0.018
	Scandinavian airlines	450	110	0.019
	KLM	450	62	0.019
	Alitalia	450	93	0.015

B. *Evaluation Metrics*

In this paper, two metrics *AUC* (area under the curve), and *precision* have been used to measure the performance of the proposed method and compare it with other methods. AUC is the probability that the score of similarity of a selected edge from the test set is larger than the similarity score of a selected edge from a set of non-existent edges in the network. The calculation of AUC is as follows:

$$AUC = \frac{n_1 + 0.5 n_2}{n} \tag{7}$$

To calculate the AUC, we must select one edge from the test set and one edge from the set of non-existent edges in the network and then compare the scores of these two edges. We have to do this several times to be able to calculate the AUC correctly. In (7), n is the number of comparisons, n_1 is the number of times the edge score from the test set is higher, and n_2 is the number of times the score of both edges is equal.

Edge selection can be done in two approaches. In the first case, the choices can be random and the number of times the edge is selected (i.e., n in Formula 7) is arbitrary. In the second case, all pairs of edges from the two sets of test and non-existent should be compared. In this case, the number of selected edges is equal to the number of members of the test set multiplied by the number of members of the set of non-existent edges. The second case is more reliable, and we do the second case. Note that in the random case, AUC value is about 0.5. Therefore, if the AUC value is more than 0.5, it means that the performance of the algorithm is higher

[DOI: 10.52547/itrc.14.3.19]

than the random condition. The higher the AUC, the better the performance of the algorithm. Note that the maximum value of AUC is 1.

Precision can be formulated as follows:

$$\text{Precision}= TP/(TP+FP) \quad (8)$$

where, FP and TP represent the number of false positives and true positives respectively. Here, the parameter denotes the ratio of the number of links predicted correctly by the algorithm to all the prediction results of the algorithm.

C. Evaluation results

The proposed method is compared with three common inter-layer similarity metrics including ASN (Average Similarity of the Neighbors), BW (Betweenness), and DDC (Degree-Degree Correlation) on the BA artificial network. The AUC values are shown in Fig. 2 for predicting links in layer L2. As mentioned previously, Pcopy is the probability of copying a link from the first layer to the second layer in generation process of two-layer BA synthetic network. As can be evidently seen, the higher the Pcopy value is, the higher the inter-layer similarity between L1 and L2. The accuracy of the proposed algorithm increases regardless of the inter-layer similarity metric used (i.e., the AUC increases). On the other hand, according to the AUC value, our algorithm works better than other similarity criteria in all cases with the help of proposed inter-layer similarity metric.

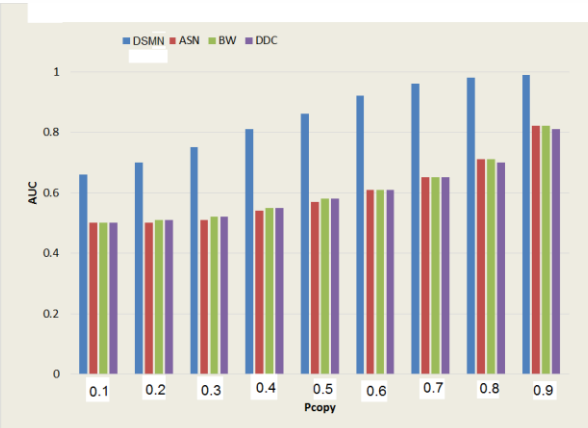

Figure 2. Comparison of the AUC values for the proposed method (DS: Distance Similarity) on the BA artificial network along with three common metrics.

Similarly, for real-world networks, the performance of the proposed method in terms of AUC is evaluated and the results are shown in Table2. The DSMN inter-layer similarity and CNDP intra-layer similarity metrics are used to predict missing links. The performance of the proposed method is compared with LPIS [17], which is recently proposed by Najari et al. for link prediction in multilayer networks.

TABLE II. AUC VALUES FOR REAL-WORLD MULTILAYER NETWORKS ARTICLE

Multilayer network	Layers	LPIS	DSMN
Twitter-Foursquare	Twitter	0.884	**0.904**
	Foursquare	0.938	**0.952**
Twitter-Instagram	Twitter	**0.995**	0.990
	Instagram	0.964	**0.971**
Online and Offline Relationships	Facebook	0.898	**0.957**
	Leisure	0.898	**0.929**
	Work	0.968	**0.973**
	Co-authorship	0.941	**0.959**
	Launch	0.789	**0.861**
EU-Air Transportation	Lufthansa	0.982	**0.991**
	Ryan air	0.966	**0.978**
	Easy jet	0.976	**0.984**
	British airways	**0.995**	0.991
	Turkish airlines	0.987	**0.994**
	Air Berlin	0.988	**0.995**
	Air France	0.991	**0.993**
	Scandinavian airlines	0.992	**0.996**
	KLM	0.996	**0.997**
	Alitalia	0.994	**0.996**

The authors of [17] provide an inter-layer similarity measure called AASN and a framework called LPIS for predicting missing links in multilayer networks. This framework is based on inter-layer similarity and neighborhood-based intra-layer features. LPIS is a systematic approach in which any inter-layer similarity metric and any neighborhood-based attribute can be used. We compare our proposed method with LPIS which uses the AASN similarity metric and the Adamic-Adar similarity measure. Table 2 shows the results of the proposed method and method [17] in each of the layers of TF, TI, OOR, and EAT network. As can be seen, except in two cases including the Twitter layer of the Twitter-Instagram network and the British airways layer of the EU-Air Transportation network, the DSMN method performs better than LPIS. It should be noted that even in these two cases, the AUC value in the LPIS method is comparable with the value in the DSMN method. These results indicate the superiority of the proposed method for predicting links in multi-layer networks.

As another evaluation, the precision values of the proposed method along with the compared LPIS method are shown in Fig. 3. The highest precision is obtained for the Ryan air layer of the EU-Air Transportation network. For all cases, our proposed method performs slightly better and predict missing links more precisely.

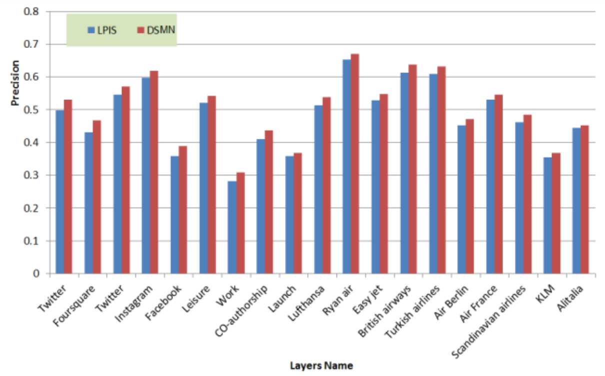

Figure 3. Comparison of the precision values in a real multilayer networks for our DSMN method and LPIS

[DOI: 10.52547/itrc.14.3.19]

V. Conclusion

Link prediction in multilayer networks becomes increasingly important in recent years. In this article, we presented a new method for this purpose with the aim of our previously proposed CNDP method for single-layer networks. Considering the fact that in real multilayer networks there is usually a significant inter-layer similarity between their layers, we predict missing links in one layer by considering the structural information of the other layers. To do that, inter-layer similarity is combined with intra-layer similarity to use the information of other layers in the process of link prediction and improve link prediction results. We evaluated our proposed method on both synthetic and real multilayer networks. Experimental results confirm the outperformance of the proposed method in comparison with other state-of-the-art methods in terms of AUC and precision.

As the future work, we are planning to use graph embedding methods to transform the graph structure into *d*-dimensional vectors of real numbers. Tools such as multi-node2vec can be used for this purpose; and the link prediction then can be performed in the embedding space.

References

[1] Barabási, A.-L., Network science. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2013. 371(1987): p. 20120375.

[2] Salehi, M., Sharma, R., Marzolla, M., Magnani, M., Siyari, P. and Montesi, D., Spreading processes in multilayer networks. IEEE Transactions on Network Science and Engineering, 2015. 2(2): p. 65-83.

[3] Liben Nowell, D. and J. Kleinberg, The link prediction problem for social networks. Journal of the American society for information science and technology, 2007. 58(7): p. 1019-1031.

[4] Barabási, A.-L., N. Gulbahce, and J. Loscalzo, Network medicine: a network-based approach to human disease. Nature reviews genetics, 2011. 12(1): p. 56-68.

[5] Barzel, B. and A.-L. Barabási, Network link prediction by global silencing of indirect correlations. Nature biotechnology, 2013. 31(8): p. 720-725.

[6] Guimerà, R. and M. Sales-Pardo, Missing and spurious interactions and the reconstruction of complex networks. Proceedings of the National Academy of Sciences, 2009. 106(52): p. 22073-22078.

[7] Lü, L. and T. Zhou, Link prediction in complex networks: A survey. Physica A: statistical mechanics and its applications, 2011. 390(6): p. 1150-1170.

[8] Al Hasan, M., Chaoji, V., Salem, S. and Zaki, M., Link prediction using supervised learning. in SDM06: workshop on link analysis, counter-terrorism and security. 2006.

[9] Ranjbar, V., Salehi, M., Jandaghi, P. and Jalili, M., QANet: Tensor Decomposition Approach for Query-based Anomaly Detection in Heterogeneous Information Networks. IEEE Transactions on Knowledge and Data Engineering, 2018. 31(11): p. 2178-2189.

[10] Wang, Z., L. Wang, and M. Perc, Degree mixing in multilayer networks impedes the evolution of cooperation. Physical Review E, 2014. 89(5): p. 052813.

[11] Gu, C.G., Zou, S.R., Xu, X.L., Qu, Y.Q., Jiang, Y.M., Liu, H.K. and Zhou, T., Onset of cooperation between layered networks. Physical Review E, 2011. 84(2): p. 026101.

[12] Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y. and Porter, M.A., Multilayer networks. Journal of complex networks, 2014. 2(3): p. 203-271.

[13] Hristova, D., et al., A multilayer approach to multiplexity and link prediction in online geo-social networks. EPJ Data Science, 2016. 5(1): p. 24.

[14] Pujari, M. and R. Kanawati, Link prediction in multiplex networks. Networks & Heterogeneous Media, 2015. 10(1): p. 17.

[15] Sharma, S. and A. Singh. An efficient method for link prediction in complex multiplex networks. in 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). 2015.

[16] Hajibagheri, A., G. Sukthankar, and K. Lakkaraju. A holistic approach for link prediction in multiplex networks. in International Conference on Social Informatics. 2016. Springer.

[17] Najari, S., Salehi, M., Ranjbar, V. and Jalili, M, Link prediction in multiplex networks based on inter-layer similarity. Physica A: Statistical Mechanics and its Applications, 2019. 536: p. 120978.

[18] Yang, R., Yang, C., Peng, X. and Rezaeipanah, A., A novel similarity measure of link prediction in multi-layer social networks based on reliable paths. Concurrency and Computation: Practice and Experience, p.e6829. 2021

[19] Jafari, S.H., Abdolhosseini-Qomi, A.M., Asadpour, M., Rahgozar, M. and Yazdani, N., 2021. An information theoretic approach to link prediction in multiplex networks. Scientific Reports, 11(1), pp.1-21. 2021

[20] Nasiri, E., Berahmand, K. and Li, Y., 2021. A new link prediction in multiplex networks using topologically biased random walks. Chaos, Solitons & Fractals, 151, p.111230. 2021

[21] Ramezanian, R., Magnani, M., Salehi, M. and Montesi, D., Diffusion of innovations over multiplex social networks. in 2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP). 2015. IEEE.

[22] Al Hasan, M. and M.J. Zaki, A survey of link prediction in social networks, in Social network data analytics. 2011, Springer. p. 243-275.

[23] Adamic, L.A. and E. Adar, Friends and neighbors on the web. Social networks, 2003. 25(3): p. 211-230.

[24] Newman, M.E., Clustering and preferential attachment in growing networks. Physical review E, 2001. 64(2): p. 025102.

[25] Zhou, T., L. Lü, and Y.-C. Zhang, Predicting missing links via local information. The European Physical Journal B, 2009. 71(4): p. 623-630.

[26] Katz, L., A new status index derived from sociometric analysis. Psychometrika, 1953. 18(1): p. 39-43.

[27] Brin, S. and L. Page, The anatomy of a large-scale hypertextual web search engine. 1998.

[28] Brin, S. and L. Page, Reprint of: The anatomy of a large-scale hypertextual web search engine. Computer networks, 2012. 56(18): p. 3825-3833.

[29] Rafiee, S., C. Salavati, and A. Abdollahpouri, CNDP: Link prediction based on common neighbors degree penalization. Physica A: Statistical Mechanics and its Applications, 2020. 539: p. 122950.

[30] Aliakbary, S., et al., Distance metric learning for complex networks: Towards size-independent comparison of network structures. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2015. 25(2): p. 023111.

[31] Battiston, F., V. Nicosia, and V. Latora, Structural measures for multiplex networks. Physical Review E, 2014. 89(3): p. 032804.

[32] Sun, Y., et al., Pathsim: Meta path-based top-k similarity search in heterogeneous information networks. Proceedings of the VLDB Endowment, 2011. 4(11): p. 992-1003.

[33] Zhao, D., Li, L., Peng, H., Luo, Q. and Yang, Y., Multiple routes transmitted epidemics on multiplex networks. Physics Letters A, 2014. 378(10): p. 770-776.

[DOI: 10.52547/itrc.14.3.19]

Samira Rafiee received her M.Sc. in Artificial Intelligence from University of Kurdistan, Sanandaj, Iran. She received the B.Sc. degree in Computer Engineering from University of Kurdistan, Sanandaj, Iran. Her research interests include Link Prediction in Complex Networks, Data Mining, and Machine Learning. She has been teaching computer related courses for ten years.

Alireza Abdollahpouri is an Associate Professor of Computer Networks at the Department of Computer Engineering, University of Kurdistan, Sanandaj, Iran. He has obtained Ph.D. (Computer Networks) in 2012 from University of Hamburg, Germany. He received the B.Sc. and M.Sc. degrees both in Computer Engineering from Isfahan University of Technology and Amirkabir University of Technology, respectively. He was working at Telecommunication Company of Tehran from 2001 until 2005 as a System Programmer. He joined the University of Kurdistan as a faculty member in 2005. His main research interests are in the field of Complex Networks and Wireless Communications.

[DOI: 10.52547/itrc.14.3.19]