

Research Article

Fariba Sarhangnia*, Shima Mahjoobi, and Samaneh Jamshidi

A novel similarity measure of link prediction in bipartite social networks based on neighborhood structure

<https://doi.org/10.1515/comp-2022-0233>

received December 31, 2020; accepted February 15, 2022

Abstract: Link prediction is one of the methods of social network analysis. Bipartite networks are a type of complex network that can be used to model many natural events. In this study, a novel similarity measure for link prediction in bipartite networks is presented. Due to the fact that classical social network link prediction methods are less efficient and effective for use in bipartite network, it is necessary to use bipartite network-specific methods to solve this problem. The purpose of this study is to provide a centralized and comprehensive method based on the neighborhood structure that performs better than the existing classical methods. The proposed method consists of a combination of criteria based on the neighborhood structure. Here, the classical criteria for link prediction by modifying the bipartite network are defined. These modified criteria constitute the main component of the proposed similarity measure. In addition to low simplicity and complexity, this method has high efficiency. The simulation results show that the proposed method with a superiority of 0.5% over MetaPath, 1.32% over FriendLink, and 1.8% over Katz in the f -measure criterion shows the best performance.

Keywords: social networks, bipartite, link prediction, similarity measure, neighborhood structure

1 Introduction

Social networks are a new generation of databases that are in the spotlight of Internet users these days [1]. Such databases operate on the basis of online organizations, and each has brought together a group of Internet users with a specific feature [2]. Today, social networks have become widespread and there is currently no convenient way to manage and categorize them [3–5]. Of course, some social networks have allowed users to categorize their friends into social circles (such as circles on Google Plus and friends list on Facebook and Twitter). However, these methods do not work well because with the addition of other people to the friends of these circles, they must be updated again by users [4]. Therefore, a mechanism is needed to learn and identify people and be able to automatically form and update social circles. In this case, we have the information of a person and his friends in a social network, and the aim is to find the social circles of the person in question, where each circle is a subset of the person's friends. As shown in Figure 1, the user is marked with u and the friends associated with u are marked with v , where the aim is to find some circles [5].

Social networks are constantly increasing the number of users and the communications between them, and unfortunately these communications may be lost for various reasons [6]. In relation with these links and the communications between them, the problem of link prediction, which is an important topic for social media analysis, has become important [4,7]. This means predicting the likelihood of a link between two users, knowing that there is currently no link between the two users. Predicting the occurrence of links is a fundamental problem in network analysis [8]. In the subject of link prediction, a view of a network is given and we want to know what transactions are likely to take place between current members of the network in the near future [8]. Although this problem has been extensively studied, however, the problem of how to optimally and

* **Corresponding author: Fariba Sarhangnia**, Department of Computer Engineering and Information Technology, Bushehr Branch, Islamic Azad University, Bushehr, Iran, e-mail: sarhangnia.fariba@gmail.com

Shima Mahjoobi: Department of Electronics Technology Engineering, Pishtazan of University, Shiraz, Iran, e-mail: shima.mahjoobi@gmail.com

Samaneh Jamshidi: Department of Computer Engineering and Information Technology, Bushehr Branch, Islamic Azad University, Bushehr, Iran, e-mail: sma.jamshidi@gmail.com

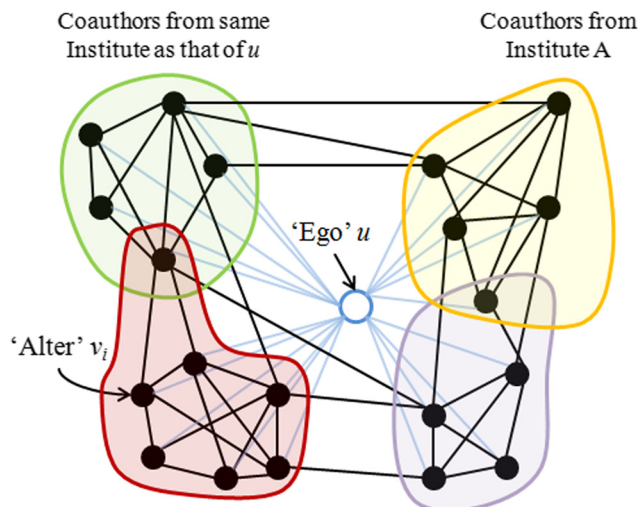


Figure 1: Social networks with circle labels [6].

effectively combine the information obtained from the network structure with the abundant descriptive data related to nodes and links, remains to a large extent. Real networks show a range of interesting features and patterns. One of the important topics in this field of research is the design of models that predict and reproduce the occurrence of such network structures [9]. Therefore, research processes seek to develop models that accurately predict the overall structure of the network.

Many types of networks are highly dynamic [9–11]. These networks are rapidly growing and changing by adding new nodes that represent the existence of new transactions between network nodes [9]. Therefore, the study of networks is considered at the level of creating separate links and is even more difficult in some respects than global modeling [10,11]. Identifying the mechanisms by which these networks grow at the level of individual ridges is not yet fully understood, and is in fact the impetus for research into link prediction. In general, we

consider the classic problem of link prediction. In this case, we have a view of the network at time t , and we seek to accurately link prediction that are added to the network in the time period t to $t + 2a$ (i.e., $t + a$) [12]. Figure 2 shows the concept of link prediction problem.

Bipartite networks are one of the most important types of complex networks in which nodes are divided into two parts [11]. In these networks, the links connect the nodes of different sections and there is no link between two nodes of the same section. Many real-world networks are essentially bipartite networks, such as people and items purchased, people and diseases, diseases and genes, papers and authors, words and texts, and investors and companies [12]. Link prediction is one of the most important issues in bipartite social media analysis. Figure 3 provides an overview of a bipartite network. In this figure, there are two types of nodes that connect users to items of interest. Here there is a weight as a similarity between each user and item, indicated by W_{uv} . The purpose of link prediction measurement is to predict the likelihood of users' future atoms [13].

The problem of link prediction arises mainly in the case of classical networks. Therefore, the appropriate method for link prediction in bipartite networks needs to be studied more carefully. Depending on the field of application, it is better to choose a link prediction method that the criteria used in it match the context of the problem. The purpose of this study is to develop a link prediction method for bipartite networks that examines the network from different perspectives based on neighborhood structure and has apt performance.

The main contribution of this study is as follows:

- Development of a novel similarity measure based on bipartite social network topology
- Measuring similarity between users based on neighborhood structure in bipartite social networks

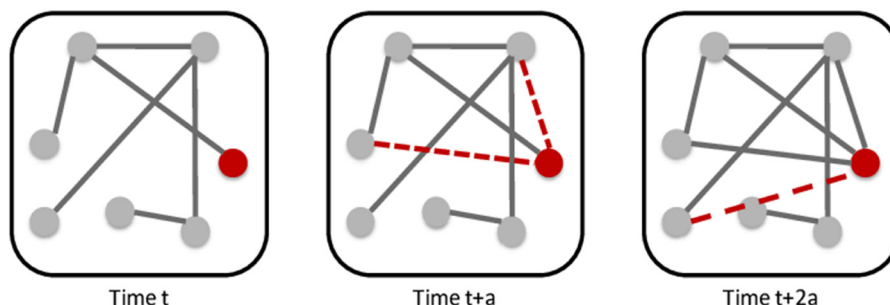


Figure 2: Concept of link prediction problem.

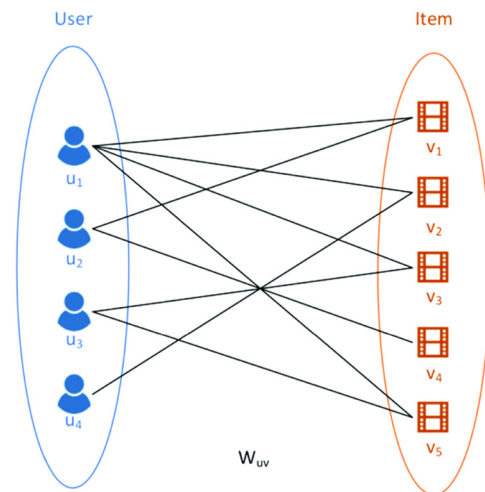


Figure 3: Overview of bipartite networks.

- Evaluation of the proposed algorithm with extensive simulation on real social networks using MATLAB.

The rest of the paper is organized as follows: Section 2 discusses the research literature on the problem of link prediction. Section 3 describes, in detail, the proposed method. Section 4 deals with simulations and experiments results. Finally, Section 5 concludes this study.

2 Literature review

Much research has been done on the problem of link prediction in social networks [14–16]. The first link prediction model, which was explicitly used in social networks, was proposed by Liben-Nowell and Kleinberg [14]. They defined the method of prediction by the similarity between the two nodes with the possibility of future friendship. They then ranked the nodes based on similarity scores and suggested the highest ranked nodes. Al Hasan and Zaki later developed this approach [15]. They showed that the use of external data can improve the performance of link prediction. The authors formulated the link prediction problem as a binary classification problem.

In ref. [16], a link prediction approach based on similarity in social networks was used using latent relationships between users. In this method, a new measurement is proposed to determine the similarity of each pair of nodes based on the number of common neighbors and the correlation between the neighboring vectors of the nodes. In ref. [17], a link prediction model for

complex networks is introduced. In this model, four similarity indices including CN, LHN-II, COS+, and MFI are combined to define a new index for link prediction in complex networks. The combination model through logistic regression introduces the Ensemble-Model-Based Link Prediction algorithm.

In ref. [18], the Common Neighbors Degree Penalization (CNDP) method is introduced to link prediction in social networks. CNDP offers a new criterion for link prediction by considering clustering coefficient as a structural feature of the network. In ref. [19], the detection of communities in complex networks with ambiguous structure is proposed to improve central node-based link prediction. In this study, a new link prediction strategy is designed that identifies communities in complex networks with ambiguous structures.

In ref. [20], the Stretch Shrink Distance Based Algorithm (SSDBA) is introduced to link prediction in social networks. SSDBA is a short-distance contraction-based algorithm that solves community prediction based on community identification. In this algorithm, first the associations of a social network are identified and then the active nodes are identified based on community average threshold and node average threshold in each community. Next, the Stretch Shrink Distance model is used to calculate the distance changes between active nodes and local neighbors. In ref. [21], a multilayer link prediction model for complex dynamic networks is proposed. The authors developed a method for modeling multilayer networks based on the evolution of each node membership at different layers. This evolution was formulated using the Infinite Hidden Markov Model through intralayer and interlayer bonds.

In ref. [7], a new approach to link prediction in multiplex networks is proposed as Multiple Stochastic Local Walking (MLRW). Local Random Walk is one of the most popular methods for link prediction for multiplex social networks, which records network structure through pure random walking to measure similarity between nodes. MLRW uses biasing functions to calculate the weight between different layers. In ref. [22], a link prediction accounting interlayer similarity framework and proximity-based features for multiplex social networks are proposed. The authors examine the effect of interlayer similarities on link prediction in artificial and real multiplex social networks.

In ref. [23], a supervised-learning approach is proposed to link prediction in single layer and multiplex social networks. The authors use improved structural features and similarity criteria. Here, community-based

features are used to develop this approach. In ref. [24], a supervised approach to solving the problem of link prediction in multiplex social networks is introduced. The authors derive a binary classification model from complex structural features of layers, where they consider the information of all layers at the same time. The Meta-Path algorithm is presented in ref. [25], which is a way to link prediction in multiplex social networks. MetaPath performs link prediction for Foursquare social network users based on node-based features as well as meta-path-based features on Twitter. The node-based features used are optimism and reputation, and the meta-path-based are derived from the path of multiplex networks.

3 Proposed method

The idea of the proposed method is to use the well-known and classic criteria of link prediction that have been developed to adapt to the bipartite network. For better understanding, our focus is on criteria based on neighborhood structure. Criteria based on neighborhood structure are the most important set for link prediction. Therefore, in order to take advantage of different perspectives to solve the link prediction problem, we use a combination of different neighborhood-based criteria to define a new similarity measure. The main focus of the proposed method is on the importance of weight between users in calculating similarity. In this regard, among the criteria based on the neighborhood structure, we use the classical similarity criteria by weight, where they assign a higher score to the more dependent nodes. The classical similarity criteria used in the proposed method of Common Neighbors (CNs) [26], Jaccard Coefficient (JC) [27], Adamic-Adar (AA) [28], Preferential Attachment (PA) [29], Katz (KT) [30], and FriendLink (FL) [31]. All of these criteria calculate the similarity between two nodes based on the neighborhood structure. In these criteria, nodes with a higher degree are more important. The general process of this research is shown in Figure 4.

Due to the fact that some of the implicit information is lost in the conversion of the bipartite network to a one-part network, so the weighted version of the similarity criteria is used. Hence, the weight-less network graph is mapped to a weighted network. Users profile information is used to calculate the weight of links to express their common interests in communication. $G(V, E, W)$ is a weighted social network in which V is a set of nodes, E is a set of links, and W is a weight between links. $(u, v) \in E$ is a link between nodes u and v in graph G ,

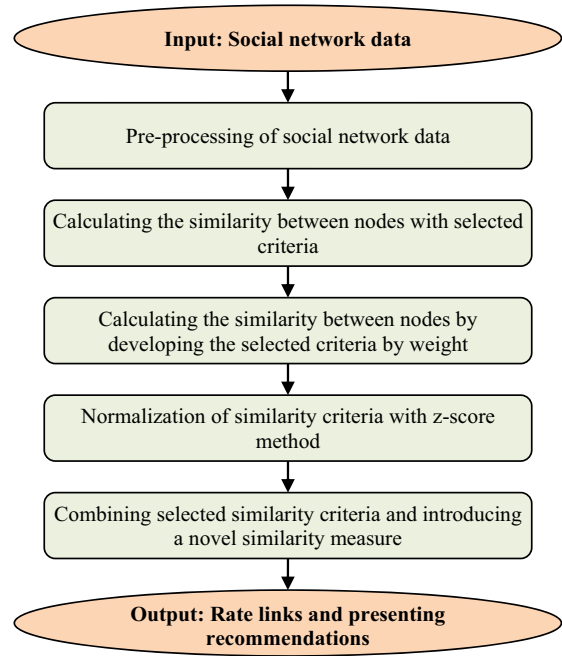


Figure 4: Conceptual diagram of the proposed method for link prediction.

and $w_{u,v} \in W$ represents the weight of the link between nodes u and v . Here, links between users are assumed to be directional, so $w_{u,v}$ and $w_{v,u}$ are different. In this study, $w_{u,v}$ is calculated based on the cosine coefficient, as shown in equation (1).

$$w_{u,v} = \frac{\sum_{k \in I} (r_u^k) \cdot (r_v^k)}{\sqrt{\sum_{k \in I} (r_u^k)^2 \cdot \sum_{k \in I} (r_v^k)^2}}, \quad (1)$$

where k is a member of set I and I refers to user profile features. r_u^k and r_v^k are the k th features of u and v users, respectively.

The following are the classical similarity criteria used in the proposed method, including CNs, JC, AA, PA, KT, and FL. All of these criteria are considered weight-less as well as weighted, where we use the weighted version in this paper because they conform to the structure of bipartite networks.

CN: This criterion in weight-less social networks refers to the number of common nodes that are directly connected to the two nodes under evaluation. The greater the number of common neighbors between the two nodes, the more likely it is that a direct link will be established between the two nodes in the future. Equation (2) shows the $CN(u, v)$ criterion for calculating the similarity between users u and v , and equation (3) defines this criterion as weighted [26].

$$CN(u, v) = |\Gamma(u) \cap \Gamma(v)|, \quad (2)$$

$$\text{WCN}(u, v) = \sum_{z \in |\Gamma(u) \cap \Gamma(v)|} w_{u,z} + w_{z,v}. \quad (3)$$

JC: This criterion refers to the highest value between a pair of nodes that has a number of common neighbors compared to the number of its neighbors. Equation (4) shows the JC(u, v) criterion for calculating the similarity between users u and v , and equation (5) defines this criterion as weighted [27].

$$\text{JC}(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}, \quad (4)$$

$$\text{WJC}(u, v) = \sum_{z \in |\Gamma(u) \cap \Gamma(v)|} \frac{w_{u,z} + w_{z,v}}{\sum_{x \in \Gamma(u)} w_{u,x} + \sum_{y \in \Gamma(v)} w_{y,v}}. \quad (5)$$

AA: This criterion is related to the JC. This criterion gives more importance to common neighbors who have a small number of neighbors. Thus, AA measures how strong the relationship between the common neighbors and the two nodes evaluated is. Equation (6) shows the AA(u, v) criterion for calculating the similarity between users u and v , and equation (7) defines this criterion as weighted [28].

$$\text{AA}(u, v) = \sum_{z \in \Gamma(u) \cap \Gamma(v)} \frac{1}{\log k_z}, \quad (6)$$

$$\text{WAA}(u, v) = \sum_{z \in \Gamma(u) \cap \Gamma(v)} \frac{w_{u,z} + w_{z,v}}{\log \left(1 + \sum_{c \in \Gamma(z)} w_{z,c} \right)}, \quad (7)$$

PA: This criterion assumes that the probability of creating a new link from node u depends on the degree of this node (i.e., $|\Gamma(u)|$). In other words, nodes that already have a large number of relationships will be more likely to have communications in the future. Equation (8) shows the PA(u, v) criterion for calculating the similarity between users u and v , and equation (9) defines this criterion as weighted [29].

$$\text{PA}(u, v) = |\Gamma(u)| \times |\Gamma(v)|, \quad (8)$$

$$\text{WPA}(u, v) = \sum_{x \in |\Gamma(u)|} w_{u,x} \times \sum_{y \in |\Gamma(v)|} w_{y,v}. \quad (9)$$

KT: This criterion is one of the most successful global metrics for calculating similarity between users and link prediction. KT calculates the similarity according to the number and length of paths between the two users. The characteristic of this criterion is the assignment of coefficients to the paths between two users, which decreases exponentially with respect to the path length. Thus, KT attaches less importance to paths with longer lengths in

calculating the final similarity. Equation (10) shows the KT(u, v) criterion for calculating the similarity between users u and v , and equation (11) defines this criterion as weighted [30].

$$\text{KT}(u, v) = \sum_{l=1}^{\infty} \beta^l \cdot |\text{paths}_{u,v}^l|, \quad (10)$$

$$\text{WKT}(u, v) = \sum_{l=1}^{\infty} \beta^l \cdot \left[\sum_{P \in \text{paths}_{u,v}^l} \prod_{(x,y) \in P} w_{x,y} \right], \quad (11)$$

where β is a constant coefficient to reduce the impact of long distances. $\text{paths}_{u,v}^l$ and $|\text{paths}_{u,v}^l|$ refer to the set of paths and the number of paths with length l between the two users u and v , respectively.

FL: Like KT, this criterion uses factors such as number and path length to calculate similarity. The only difference is considering the attenuation factor of $1/(l-1)$ for longer paths and also considering the total number of possible paths between two users by the $\sum_{j=2}^l (n-j)$ term. In addition, FL is a semi-local, similarity metric that considers finite length paths to calculate similarity. Equation (12) shows the FL(u, v) criterion for calculating the similarity between users u and v , and equation (13) defines this criterion as weighted [31].

$$\text{FL}(u, v) = \sum_{l=2}^L \frac{1}{l-1} \cdot \frac{|\text{paths}_{u,v}^l|}{\prod_{j=2}^l (n-j)}, \quad (12)$$

$$\text{WFL}(u, v) = \sum_{l=2}^L \frac{1}{l-1} \cdot \frac{\sum_{p \in \text{paths}_{u,v}^l} \prod_{(x,y) \in p} w_{x,y}}{\prod_{j=2}^l (N-j)}, \quad (13)$$

where L is the maximum path length and N is the total number of network nodes.

Due to the different amplitude and difference of the values of these criteria and in order to have the same effect in calculating the proposed similarity measure, the values of the introduced criteria are normalized using the z-score method [32], as shown in equation (14). This normalization is to map the amount of data from the current interval to another interval with the aim of increasing scalability.

$$x_i^{z\text{-score}} = \frac{x_i - \mu}{\sigma}, \quad (13)$$

where x_i is the actual value of the i th similarity criterion and $x_i^{z\text{-score}}$ is its normalized value. μ and σ are the mean and standard deviation of the i th similarity criterion, respectively. In this method, the mean and standard deviation for each criterion after normalization are 0

and 1, respectively. Therefore, the z -score places the data at center 0.

The proposed similarity measure for link prediction in a bipartite network is calculated based on different similarity criteria, where it can combine the information obtained from each criterion according to different concepts. Here, the average scores of these criteria are used to calculate the proposed criterion, as shown in equation (14).

$$\text{Sim}(u, v) = \frac{\text{WCN} + \text{WJC} + \text{WAA} + \text{WPA} + \text{WKT} + \text{WFL}}{M}, \quad (14)$$

where WCN, WJC, WAA, WPA, WPA, WKT, and WFL refer to Common Neighbors, JC, AA, PA, KT, and FL, respectively. Also, M is the number of criteria considered, which here is $M = 6$.

4 Experimental results

This section is related to the evaluation and comparison of the proposed method in solving the problem of link prediction in bipartite networks. Evaluation and comparison are based on various criteria such as precision, recall, f -measure and mean average precision (MAP). To compare the performance of the proposed method, the classical similarity criteria of KT [30] and FL [31] as well as the MetaPath algorithm [25] have been used. The simulation was performed by MATLAB R2019a on HP Pavilion 15 Laptop with 11th Gen Intel Core i7-1165G7 Processor at 4.2 GHz and 16 GB RAM. In addition, the simulation is based on the Twitter and Foursquare social network datasets.

All results are based on the 10-fold cross-validation method to ensure. In this validation, training users include 90% and testing users 10% of the total social network users. At each validation step, the same users are split between the two social networks Twitter and Foursquare into two sets, training (E^T) and probing (E^P) users, so that $E = E^T \cup E^P$ and $E^T \cap E^P = \emptyset$, where E refers to all links between users on the Foursquare social network. Here, the aim is to links prediction for Foursquare social network users based on data from both networks.

4.1 Evaluation criteria

In this study, various criteria such as precision, recall, f -measure and MAP have been used to evaluate the results of different algorithms in solving the problem of

link prediction [13,14]. These criteria are calculated based on two factors, including actual related users and recommended users. Let $\text{relevant_users}(u)$ be the set of users associated with user u . Also $\text{retrieved_users}(u)$ are the set of users recommended by the link prediction method for user u . The number of recommended users is determined based on the Top_K parameter. Accordingly, the precision criterion for user u shows the ratio of the number of recommended related users to the number of actual users, as shown in equation (15). The recall criterion for user u is the ratio of the recommended number of related users to the total number of related users of the E^P set with user u , as shown in equation (16). The f -measure criterion is the harmonic mean of the precision and recall criteria, as shown in equation (17). The MAP criterion represents the average precision with respect to the recommended user rating, as shown in equation (18).

$$\text{Precision}(u) = \frac{|\text{relevant users}(u) \cap \text{retrieved users}(u)|}{|\text{retrieved users}(u)|}, \quad (15)$$

$$\text{Recall}(u) = \frac{|\text{relevant users}(u) \cap \text{retrieved users}(u)|}{|\text{relevant users}(u)|}, \quad (16)$$

$$F\text{-Measure}(u) = \frac{2 \times \text{Precision}(u) \times \text{Recall}(u)}{\text{Precision}(u) + \text{Recall}(u)}, \quad (17)$$

$$\text{MAP}(u) = \frac{1}{|N|} \sum_{U=1}^{|N|} \frac{1}{r_u} \sum_{k=1}^{r_u} P[u@k] \quad (18)$$

where $|N|$ the number of users in the test set, r_u is the number of real related users for u , and $P[u@k]$ is the precision value for u in the k th rank.

4.2 Dataset

This study uses the same users on Twitter and Foursquare to evaluate link prediction algorithms. Twitter is a directional microblogging social network, and Foursquare is a unidirectional social platform based on location. Foursquare social networking information is available at <https://sites.google.com/site/yangdingqi/home/foursquare-dataset> and Twitter social networking information is available at <https://snap.stanford.edu/data/egonets-Twitter.html>. Details of the dataset used by these networks are shown in Table 1.

4.3 Discussion and comparison

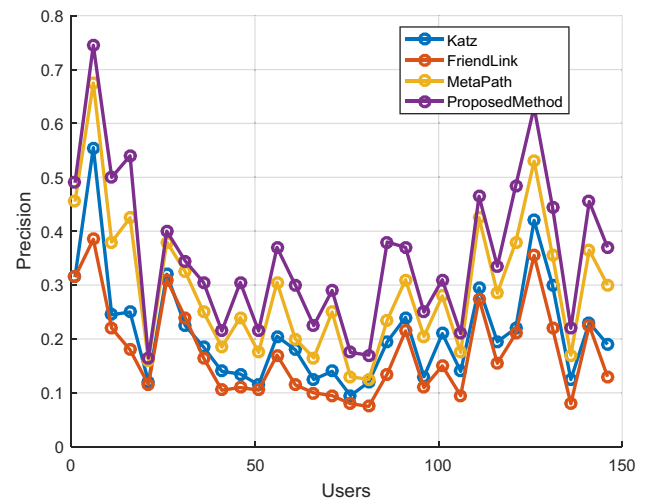
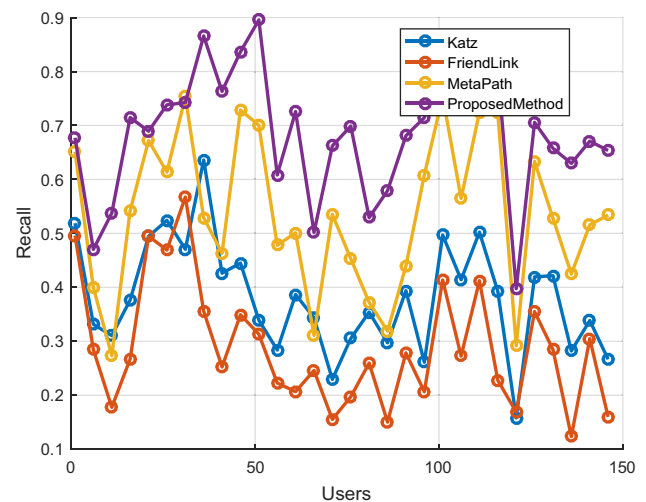
In this study, extensive experiments have been performed to evaluate the method in comparison with KT

Table 1: Details of the dataset used

Networks	#Links	#Nodes	#Common nodes	#Common links	Average degree	Average nodes
Twitter	81,306	1,768,149	1,508	6551	10.05	in = 10.05, out = 10
Foursquare	266,909	3,680,126			24.41	24.4

and FL similarity criteria as well as the MetaPath method. Considering the use of 10-fold cross-validation, there is 10% of the total users (i.e., 150 users) from the E^P set for which we have calculated the precision. Figure 5 shows a comparison of different methods based on each user in this criterion. The results of superior performance clearly show the proposed method for different users. In a similar experiment, recall was calculated for each user and reported for different methods. The results of this comparison are shown in Figure 6. In this experiment, the proposed method often shows superior performance, and in the next rank, the MetaPath method is the best. The f -measure criterion, considering the harmonic quantity of the two precision and recall criteria, has a suitable property for evaluating link prediction methods. Figure 7 shows a comparison based on the f -measure criterion based on each user, and here too the proposed method has superior performance. The MAP criterion represents the geometric combination of the two criteria precision and recall, and in fact shows the recall-precision curve. Figure 8 shows a comparison of different methods based on each user in this criterion. The results of this comparison also report better performance of the proposed method.

In the experiments, the evaluations were calculated and presented separately for each user, and the number of recommendations made to each user was considered 10 ($\text{Top}_K = 10$). Therefore, to show better performance and make a fairer comparison, different conditions of the recommendations are examined. Here, the number of recommendations for each user is analyzed on a scale of 1–20. In this case, the evaluation with different criteria is reported as an average for all users of the E^P set on a Top_K mode. Figure 9 shows a comparison of different methods based on the average precision criterion for all users. The results show that the proposed method in most Top_K could provide better precision results than other methods such as MetaPath. At best, it has reached 0.972 with $\text{Top}_K = 1$. In general, the amount of precision decreases as the number of recommendations increases. The reason for this condition is the nature of this criterion because at the denominator of this criterion, is the value of Top_K and makes the final value smaller.

**Figure 5:** Comparison of different methods based on each user in the precision criterion.**Figure 6:** Comparison of different methods based on each user in the recall criterion.

In another similar experiment, the results of a comparison of the recall criteria for different Top_K states from 1 to 20 were reported. Figure 10 shows a comparison of different methods based on the average recall criterion

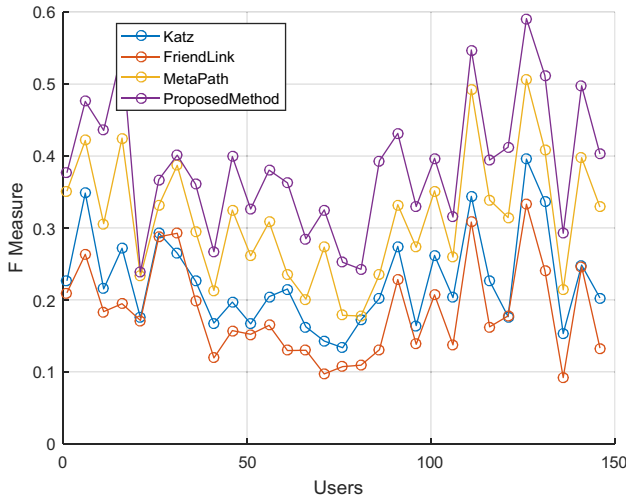


Figure 7: Comparison of different methods based on each user in the f -measure criterion.

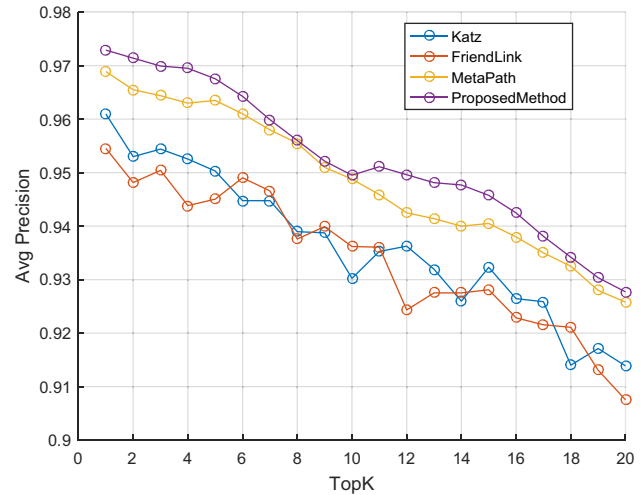


Figure 9: Comparison of different methods based on the average precision criterion for all users.

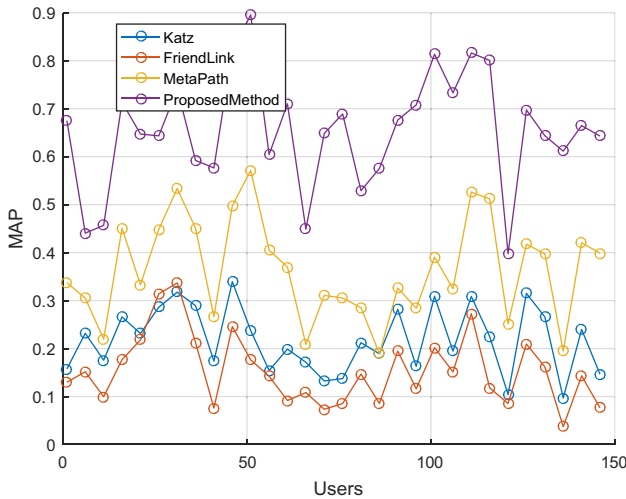


Figure 8: Comparison of different methods based on each user in the MAP criteria.

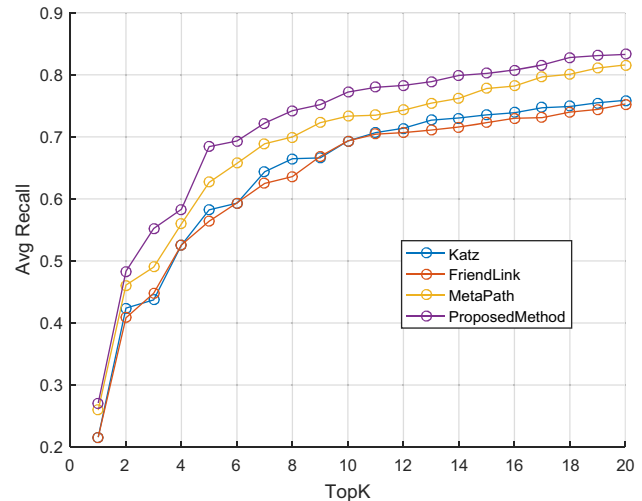


Figure 10: Comparison of different methods based on the average recall criterion for all users.

for all users. Based on the results, the proposed method has a better performance. The results of this experiment also show that the proposed method in most Top_K could provide better recall results than other methods such as MetaPath. At best, it has reached 0.832 with $Top_K = 20$. In general, the amount of recall increases as the number of recommendations increases. The reason for this condition is the specificity of this criterion, because in the denominator of this criterion, the number of related users is real, and the deduction will increase the number of recommended related users by increasing the number of recommendations. As a result, the final recall value is larger.

Figures 11 and 12 show the results for the f -measure and MAP criteria, respectively, with different Top_K states. In these experiments, the results of better performance of the proposed method are clearly visible. Both of these criteria are in fact a combination of precision and recall criteria that are widely used to express the quality of link prediction methods. In general, the MAP criterion shows the average precision according to the rank of the recommended links in the similarity matrix. To calculate the MAP for all E^P users, the precision value is calculated for all related neighbors in E^P , where the MAP value is calculated based on the user's rating. The results in f -measure criterion show that with $Top_K = 17$ the best result is

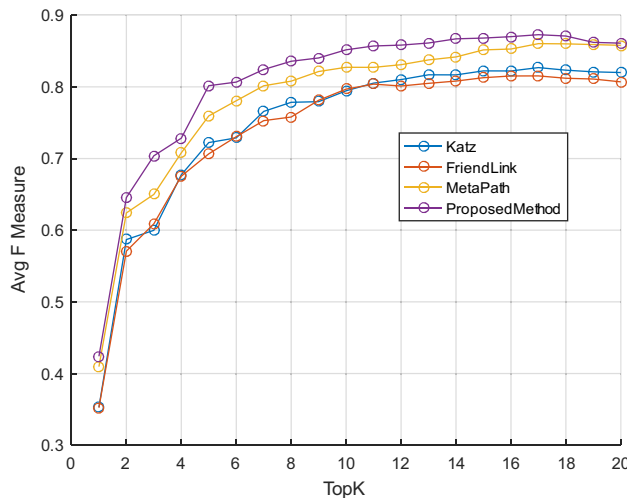


Figure 11: Comparison of different methods based on the average f -measure criterion for all users.

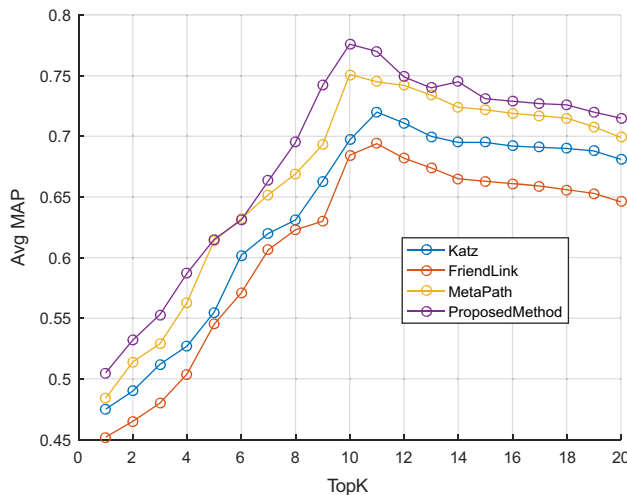


Figure 12: Comparison of different methods based on the average MAP criterion for all users.

0.872. Meanwhile, the MAP criterion with $\text{Top}_K = 10$ has reached its best result of 0.771. Due to the greater use of the f -measure criterion, in this study, $\text{Top}_K = 17$ is considered as the most appropriate number for the recommended users.

In order to better express the results of different methods, Table 2 is presented. In this table, the results of all different methods for precision, recall, f -measure, and MAP criteria are reported. According to the evaluation of the proposed method in the best case, here the results are reported based on $\text{Top}_K = 17$. In all experiments, the results are based on the 10-fold cross-validation criterion and the superior performance of the proposed method is clearly known.

Table 2: Comparison of the proposed method with similar methods

Methods	Precision	Recall	F -Measure	MAP
KT	92.15	66.15	81.52	80.11
FL	92.58	69.21	82.68	81.32
MetaPath	93.76	72.74	86.47	86.43
Proposed method	93.81	75.50	87.28	87.75

Table 3: Percentage of superiority of the proposed method over similar methods

Methods	Precision	Recall	F -Measure	MAP
KT	9.53	14.13	1.80	11.63
FL	7.90	9.09	1.32	9.23
MetaPath	1.52	3.79	0.50	1.60

The proposed method has reached the f -measure criterion of 87.28% according to all experiments performed. This advantage is achieved with 17 recommended users (i.e., $\text{Top}_K = 17$). MetaPath ranks next with an f -measure of 86.47%. After that, KT and FL similarity criteria are in the next ranks, respectively. Finally, the superiority of the proposed method in different criteria compared to KT, FL, and MetaPath methods is presented in Table 3. The proposed method shows the best performance with a superiority of 0.5% over MetaPath, 1.32% over FL and 1.8% over KT in f -measure criteria.

5 Conclusion

Social network analysis is an approach to the study of social structures. Link prediction is one of the important fields in social networks analyses. Link prediction tries to reach an appropriate answer to this question: what kinds of links among members of a network would possibly form in future, given a snapshot of the network in current time. Similarity based methods, due to simplicity and suitable performance, are among the most popular methods of link prediction. In this study, a neighborhood structure-based method for link prediction in bipartite networks is presented. In this method, the classical similarity criteria based on neighborhood structure were first defined by applying modifications to bipartite networks. These criteria have been developed from the mapping of weightless to weighted networks. Here, we used CNs, JC, AA, PA, KT, and FL criteria. The proposed similarity measure is a combination of these criteria that can have the

conceptual information of all of them. The evaluation results show that the proposed method has better performance than the basic methods such as KT and FL and also has a promising performance compared to the new MetaPath method. Therefore, the aim of the research is to achieve a criterion based on neighboring structure and optimal performance in bipartite networks. However, it is suggested that this method be analyzed for other networks as well, such as ego-centric and multiplex.

Funding information: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions: All authors contributed to the design and implementation of the research, analysis of the results and writing of the manuscript.

Conflict of interest: We certify that there is no actual or potential conflict of interest in relation to this manuscript.

Competing interests: There is no free code for this study.

Ethics approval: This material is the authors' own original work, which has not been previously published elsewhere.

Data availability statement: Data sharing is not applicable to this manuscript as no datasets were generated or analyzed during the current study.

References

- [1] W. Yuan, K. He, D. Guan, L. Zhou, and C. Li, "Graph kernel based link prediction for signed social networks," *Inf. Fusion.*, vol. 46, pp. 1–10, 2019.
- [2] Z. Samei and M. Jalili, "Application of hyperbolic geometry in link prediction of multiplex networks," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, 2019.
- [3] P. Pei, B. Liu, and L. Jiao, "Link prediction in complex networks based on an information allocation index," *Phys. A: Stat. Mech. its Appl.*, vol. 470, pp. 1–11, 2017.
- [4] M. S. Aslanpour, S. E. Dashti, M. Ghobaei-Arani, and A. A. Rahmani, "Resource provisioning for cloud applications: a 3-D, provident and flexible approach," *J. Supercomput.*, vol. 74, no. 12, pp. 6470–6501, 2018.
- [5] M. Etemadi, M. Ghobaei-Arani, and A. Shahidinejad, "Resource provisioning for IoT services in the fog computing environment: An autonomic approach," *Comput. Commun.*, vol. 161, pp. 109–131, 2020.
- [6] T. M. Tuan, P. M. Chuan, M. Ali, T. T. Ngan, and M. Mittal, "Fuzzy and neutrosophic modeling for link prediction in social networks," *Evol. Syst.*, vol. 10, no. 4, pp. 629–634, 2019.
- [7] E. Nasiri, K. Berahmand, and Y. Li, "A new link prediction in multiplex networks using topologically biased random walks," *Chaos, Solitons Fractals*, vol. 151, p. 111230, 2021.
- [8] K. Berahmand and A. Bouyer, "LP-LPA: a link influence-based label propagation algorithm for discovering community structures in networks," *Int. J. Mod. Phys. B*, vol. 32, no. 06, p. 1850062, 2018.
- [9] R. Yang, C. Yang, X. Peng, and A. Rezaeipana, "A novel similarity measure of link prediction in multi-layer social networks based on reliable paths," *Concurrency Computation: Pract. Experience*, p. e6829, 2022. DOI: 10.1002/cpe.6829.
- [10] K. Berahmand, E. Nasiri, M. Rostami, and S. Forouzandeh, "A modified DeepWalk method for link prediction in attributed social network," *Computing*, vol. 103, no. 10, pp. 2227–2249, 2021.
- [11] S. Mallek, I. Boukhris, Z. Elouedi, and E. Lefèvre, "Evidential link prediction in social networks based on structural and social information," *J. Comput. Sci.*, vol. 30, pp. 98–107, 2019.
- [12] E. Nasiri, K. Berahmand, M. Rostami, and M. Dabiri, "A novel link prediction algorithm for protein-protein interaction networks by attributed graph embedding," *Comput. Biol. Med.*, vol. 137, p. 104772, 2021.
- [13] A. Rezaeipana, G. Ahmadi, and S. Sechin Matoori, "A classification approach to link prediction in multiplex online ego-social networks," *Soc. Netw. Anal. Min.*, vol. 10, no. 1, pp. 1–16, 2020.
- [14] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *J. Am. Soc. Inf. Sci. Technol.*, vol. 58, no. 7, pp. 1019–1031, 2007.
- [15] M. Al Hasan, M. J. Zaki, A survey of link prediction in social networks, *Social network data analytics*, Boston, MA, Springer, 2011, pp. 243–275.
- [16] A. Zareie and R. Sakellariou, "Similarity-based link prediction in social networks using latent relationships between the users," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020.
- [17] K. Li, L. Tu, and L. Chai, "Ensemble-model-based link prediction of complex networks," *Computer Netw.*, vol. 166, p. 106978, 2020.
- [18] S. Rafiee, C. Salavati, and A. Abdollahpour, "CNDP: link prediction based on common neighbors degree penalization," *Phys. A: Stat. Mech. its Appl.*, vol. 539, p. 122950, 2020.
- [19] H. Jiang, Z. Liu, C. Liu, Y. Su, and X. Zhang, "Community detection in complex networks with an ambiguous structure using central node based link prediction," *Knowl. Syst.*, vol. 195, p. 105626, 2020.
- [20] R. Yan, Y. Li, D. Li, W. Wu, and Y. Wang, "SSDBA: the stretch shrink distance based algorithm for link prediction in social networks," *Front. Comput. Sci.*, vol. 15, no. 1, pp. 1–8, 2021.
- [21] M. K. Manshad, M. R. Meybodi, and A. Salajegheh, "A new irregular cellular learning automata-based evolutionary computation for time series link prediction in social networks," *Appl. Intell.*, vol. 51, no. 1, pp. 71–84, 2021.
- [22] S. Najari, M. Salehi, V. Ranjbar, and M. Jalili, "Link prediction in multiplex networks based on interlayer similarity," *Phys. A: Stat. Mech. Appl.*, vol. 536, p. 120978, 2019.

- [23] D. Malhotra and R. Goyal, "Supervised-learning link prediction in single layer and multiplex networks," *Mach. Learn. Appl.*, vol. 6, p. 100086, 2021.
- [24] N. Shan, L. Li, Y. Zhang, S. Bai, and X. Chen, "Supervised link prediction in multiplex networks," *Knowl. Syst.*, vol. 203, p. 106168, 2020.
- [25] M. Jalili, Y. Orouskhani, M. Asgari, N. Alipourfard, and M. Perc, "Link prediction in multiplex online social networks," *R. Soc. open. Sci.*, vol. 4, no. 2, p. 160863, 2017.
- [26] F. Lorrain and H. C. White, "Structural equivalence of individuals in social networks," *J. Math. Sociol.*, vol. 1, no. 1, pp. 49–80, 1971.
- [27] S. Niwattanakul, J. Singthongchai, E. Naenudorn, and S. Wanapu, "Using of Jaccard coefficient for keywords similarity," *Proc. Int. Multiconference Eng. Comput. Sci.*, vol. 1, no. 6, pp. 380–384, 2013, March.
- [28] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Soc. Netw.*, vol. 25, no. 3, pp. 211–230, 2003.
- [29] H. Chen, X. Li, Z. Huang, Link prediction approach to collaborative filtering, *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'05) IEEE*, 2005, June, pp. 141–142.
- [30] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.
- [31] A. Papadimitriou, P. Symeonidis, and Y. Manolopoulos, "Fast and accurate link prediction in social networking systems," *J. Syst. Softw.*, vol. 85, no. 9, pp. 2119–2132, 2012.
- [32] C. Cheadle, M. P. Vawter, W. J. Freed, and K. G. Becker, "Analysis of microarray data using Z score transformation," *J. Mol. diagnostics*, vol. 5, no. 2, pp. 73–81, 2003.