Research Article

Link Prediction Model for Weighted Networks Based on Evidence Theory and the Influence of Common Neighbours

Miaomiao Liu ^{1,2}, Yang Wang ¹, Jing Chen ³ and Yongsheng Zhang ¹

¹School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, Heilongjiang, China

²Key Laboratory of Petroleum Big Data and Intelligent Analysis of Heilongjiang Province, Daqing 163318, Heilongjiang, China

³College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, Hebei, China

Correspondence should be addressed to Yang Wang; 2296498823@qq.com

Received 22 November 2021; Revised 15 January 2022; Accepted 20 January 2022; Published 1 March 2022

Academic Editor: Siew Ann Cheong

Copyright © 2022 Miaomiao Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A link prediction model for weighted networks based on Dempster–Shafer (*DS*) evidence theory and the influence of common neighbours is proposed in this paper. First, three types of future common neighbours (*FCNs*) and their topological structures are proposed. Second, the concepts of endpoint weight influence, link weight influence, and high-strength node influence are introduced. Then, the similarity based on the impacts of current common neighbours (*CCNs*) and *FCNs* is defined, respectively. Finally, the two similarity indices are fused by the *DS* evidence theory. This model effectively integrates multisource information and completely exploits the influence of all *CCNs* and *FCNs* on similarity. Experiments are performed on 9 real and 40 simulation-weighted datasets, and these findings are compared with several classic algorithms. Results show that the proposed method has higher precision than other methods, which can achieve good performance in link prediction in weighted networks.

1. Introduction

The social network comprises many nodes that contribute to the social structure. Typically, nodes refer to individuals or organisations, and edges (also called links) between nodes represent all types of social relations, such as among friends, classmates, and business partners [1]. A weighted network is a social network with an edge weight, which reflects the degree of link compactness between nodes; that is, the higher the edge weight, the stronger the degree of a link between nodes [2]. Nowadays, link prediction has become a hot topic in social network analysis, which aims to analyse the network topology information to predict links that exist but are not detected and links that do not exist now but may occur in the future; that is, link prediction is used to detect missing links and predict future links [3]. In a weighted network, link prediction can not only help analyse a network with missing data but also provide a basis for the study of network evolution mechanism. It has considerable research and application values in many fields, such as recommendation systems [4] in informatics and protein-protein interactions in biology [5].

Scholars have proposed several link prediction models, such as a Markov chain-based probabilistic model [6, 7], a machine learning-based model [8], a matrix decomposition-based model [9], local similarity-based model [10], and global similarity-based model [11]. The Markov chain- and machine learning-based models can achieve high prediction accuracy, but their application in large-scale networks is limited because of the high complexity of algorithms and the difficulty of obtaining correct information and evaluating the models. The similarity-based model, which has become the mainstream link prediction method, can prevent this type of problem and can easily provide network information. For example, the classical common neighbour (*CN*) algorithm only calculates the number of *CNs* between the predicted node pair, and the resource allocation (*RA*) algorithm effectively improves the prediction accuracy by suppressing the contribution of *CNs* with a large degree. Zhu et al. [12] proposed the concept of an *H* index based on the *CN* index and effectively fused these two for link prediction in complex networks. In [13], a link prediction algorithm based on the node degree and the *H* index was proposed. Yi Can et al. [14] developed a link

prediction algorithm based on community relations and the *CN* index. Moreover, Li et al. [15] presented a method based on a topologically valid connected path, which quantified the local influence of nodes and realised link prediction in directed networks. To achieve high prediction accuracy and applicability, Wang et al. [16] constructed an algorithm based on the combined effect of the predicted nodes and their neighbours. By introducing parameters to adjust the link effect between neighbours and paths, Li et al. [17] proposed a prediction algorithm based on relative paths. In [18], an effective model was developed in performing link and sign prediction, which integrated algorithms comprising network embedding, network feature engineering, and integrated classifier. Experiments showed that the proposed model can offer a powerful methodology for multitask prediction in complex networks.

All the aforementioned methods are based on the link prediction of unweighted networks and are not suitable for weighted networks. This paper focuses on link prediction methods based on the similarity of weighted networks. Although few related studies have been conducted, some good methods have emerged. Tsuyoshi et al. [19] proposed weighted *CN* (*WCN*) and weighted Adamic-Adar (*WAA*) algorithms. Zhang et al. [20] introduced the concept of weight in the preferential attachment (*RA*) algorithm, and the results showed an improved prediction accuracy. Lü et al. [21] applied the *RA* algorithm to a weighted network and proposed the weighted preferential attachment (*WRA*) algorithm; however, the prediction results of these indexes in some weighted networks, such as *USAir* and *NetScience*, were unsatisfactory. Li et al. [22] proposed an algorithm based on a structure-weighted network by fusing the real weight and structure weight. In literature [23], the algorithm based on triangle structure and *RA* index (*TRA*) uses the number of triangles formed by nodes and their neighbours to realise link prediction, and the algorithm based on community membership model and *CN* (*CMS-CN*) employs the relationship between nodes and their communities to complete link prediction. Chen et al. [24] presented the node clustering coefficient plus (*NCCP*) algorithm, which used the degree of nodes and the clustering information of neighbours to predict the links of temporal networks. By introducing the concept of an asymmetric edge aggregation coefficient and using an adaptive function to punish *CN* nodes, the degree penalty asymmetric link clustering coefficient algorithm was proposed in [25], and good prediction results were obtained on a classical weighted network, *NetScience*. Jia et al. [26] studied the role of weak links and discussed the influence of weak links on the degree of nodes and *H* index. In [27, 28], the link prediction accuracy in weighted networks has been improved by adjusting the centrality of nodes and the weight of edges, respectively. Atiya et al. [29] analysed the influence of weights on community structure and used the fairness and goodness of fit of community structure to predict the weights of missing edges in networks. Guo et al. [30] developed a novel similarity algorithm based on transmission nodes of multipath (*STNMP*) and achieved good results in weighted network link prediction. However, with network scale expansion, its computational complexity increased. Naderi P T et al. [31]

constructed an algorithm to improve trust prediction in weighted signed networks by using local variables. However, the method focuses on the prediction of the sign of edges, and there is not much research on the role of the weight in link prediction in weighted networks. Most of the aforementioned link prediction algorithms based on node similarity only consider the number and weight of current *CNs* and do not consider the impact of potential future common neighbours (*FCNs*) on the link, for example, a node that is not a *CN* at present but can become a *CN* in the future. Such nodes raise several new questions that are worth exploring. First, do these *FCNs* help capture highly structural information in weighted networks to improve the link prediction accuracy? Second, what are the types of *FCNs*, and how can they be determined? Finally, how can the contribution of *FCNs* to the link be measured? To answer these questions, a link prediction model for weighted networks based on Dempster-Shafer (*DS*) evidence theory and node influence is proposed.

This study focuses on the link prediction method based on the similarity for undirected weighted networks. The model proposed mainly uses the local and semiglobal structure information of nodes to define the similarity. And the *DS* evidence theory with multisource information fusion ability is used for synthesizing similarities based on current common neighbours (*CCNs*) and *FCNs* so as to improve the prediction accuracy on the premise of ensuring the execution efficiency of the algorithm. The main contributions and innovations of this study are as follows:

- (1) Three types of *FCN* nodes are proposed, and the corresponding topological structure definitions are provided.
- (2) Given the influence of the degree, strength, and edge weight of nodes on similarity, this paper proposes the node influence based on *CCNs*, called weighted strength-*CCN* (*WS-CCN*), which is used to measure the contribution of *CCNs* to the similarity of the node pair.
- (3) Three concepts of influence value based on *FCNs* are introduced, namely, endpoint weight influence (*EWI*), link weight influence (*LWI*), and high-strength node influence (*HSNI*), which can effectively explore the impact of *FCNs* on potential links.
- (4) Based on the definitions of *EWI*, *LWI*, and *HSNI*, this paper proposes the node influence based on *FCNs*, called *ELH-FCN*, which is used to measure the contribution of *FCNs* on the existence or establishment of links.
- (5) According to *DS* evidence theory with multisource knowledge and information fusion ability, the node influence index based on *CCN* and the node influence index based on *FCN* are effectively fused, and a new metric, called *CCN* influence and *FCN* influence based on *DS* (*CCNI-FCNI_DS*), is proposed to comprehensively measure the influence of various factors of common neighbours on the similarity.
- (6) Experiments are performed on nine real weighted networks and 40 artificial datasets, and the results are

compared with six benchmark-weighted similarity indexes and a related algorithm, namely, *WCN*, *WAA*, *WRA*, *WPA*, *WDijkstra*, *WJaccard*, and *STNMP*; the results showed that the proposed model has an overall high prediction accuracy. In addition, by changing the ratio of the training set to the test set and the corresponding parameters in the evaluation index, the experiment and analysis proved that the proposed method has better stability and robustness for link prediction in weighted networks.

2. Theoretical Basis

2.1. DS Evidence Theory. The evidence theory proposed by Dempster can deal with uncertain information [32]. It satisfies weaker conditions than Bayesian probability theory does and can directly express “uncertainty” and “ignorance.” In this theory, a set comprising a complete set of incompatible basic propositions is used as the recognition framework, and the basic probability distribution function, which can reflect the multisource fusion information, is calculated by combining rules.

For the whole domain $U = \{A_1, A_2, \dots, A_n\}$, the possible hypothesis $\{\emptyset, \{A_1\}, \{A_2\}, \dots, \{A_n\}, \{A_1, A_2\}, \dots, U\}$ is a basic recognition framework. The basic probability assignment function is the trust degree of each hypothesis, which is expressed using a basic probability assignment. Assuming that X is a recognition framework, the basic probability distribution function on X is a mapping function of $2^X \rightarrow [0,1]$, which is used to calculate the probability of each hypothesis. For an event $A(A \neq \emptyset)$ on an arbitrary recognition frame X , two main basic probability distribution functions on X are denoted as m_1 and m_2 , and their DS fusion rules can be expressed with formulas (1) and (2) and denoted as $m(A)$.

$$m(A) = m_1(A) \oplus m_2(A) = \frac{1}{K} \sum_{B \cap C = A} m_1(B) \cdot m_2(C), \quad (1)$$

$$K = \sum_{B \cap C \neq \emptyset} m_1(B) \cdot m_2(C) = 1 - \sum_{B \cap C = \emptyset} m_1(B) \cdot m_2(C). \quad (2)$$

2.2. Problem Description. To describe the method proposed more accurately, the variables involved and their symbolic representation are declared, as shown in Table 1. The meanings of symbols used in the following text are the same as those in Table 1.

Given a weighted network graph $G = (V, E, W)$, to find missing links in the network and possible links in the future, $\forall v_x, v_y \in V \wedge e(x, y) \notin E$, a similarity value $S_{x,y}$ is assigned to each node pair in the unknown link set (namely, $U-E$) according to a certain calculation method to quantify link possibility. The higher the similarity is, the higher the possibility of the edge between the two nodes is. All unconnected node pairs are arranged in descending order according to the similarity score, and the links in the front can be regarded as links with a high probability of existence.

Because the DS evidence theory can deal with uncertain information and multisource knowledge and has strong data fusion ability, it is completely integrated with support vector machines, neural networks, and other theories [33] and is widely used in reasoning models, decision systems, and other fields, playing an important role in medical diagnosis, target recognition, and many other aspects. Mao et al. [34] proposed a corn disease recognition algorithm based on the fusion of support vector machines and the DS evidence theory. Liu et al. [35] used the evidence theory to fuse the aggregation coefficient of nodes and realised link prediction in traditional unweighted social networks. In [36], a link prediction algorithm for weighted networks combining Dempster–Shafer evidence theory and node multifeatures is proposed, which made full use of the node’s degree, strength, edge weight, path information, triangular feature, and other pieces of information. The experimental results showed the good prediction performance of the algorithm. However, the algorithm did not take into account the impact of the characteristics of future common neighbours on node similarity. Therefore, this paper uses the evidence theory to fuse various factors that influence the similarity of nodes in weighted networks and then obtains a new weighted similarity index, which can be used to measure the probability of establishing or existing potential links.

2.3. Classic Weighted Similarity Indices. The classic weighted similarity indices include *WCN*, *WAA*, *WRA*, *WJaccard*, and *WPA*, as shown in formulas (3)–(7). However, these methods only consider the influence of CCNs.

$$S_{x,y}^{WCN} = \sum_{z \in \Gamma_1(x) \cap \Gamma_1(y)} w(x, z) + w(y, z), \quad (3)$$

$$S_{x,y}^{WAA} = \sum_{z \in \Gamma_1(x) \cap \Gamma_1(y)} \frac{w(x, z) + w(y, z)}{\log(1 + s_z)}, \quad (4)$$

$$S_{x,y}^{WRA} = \sum_{z \in \Gamma_1(x) \cap \Gamma_1(y)} \frac{w(x, z) + w(y, z)}{s_z}, \quad (5)$$

$$S_{x,y}^{WJaccard} = \frac{\sum_{z \in \Gamma_1(x) \cap \Gamma_1(y)} w(x, z) + w(y, z)}{\sum_{a \in \Gamma_1(x)} w(x, a) + \sum_{b \in \Gamma_1(y)} w(y, b)}, \quad (6)$$

$$S_{x,y}^{WPA} = \sum_{z \in \Gamma_1(x) \cap \Gamma_1(y)} w(x, z) \times w(y, z). \quad (7)$$

2.4. Evaluation Indicators. AUC and precision are the commonly used evaluation indicators of link prediction. AUC [37] is defined as formula (8), and the computational procedure is as follows. Conduct independent experiments for n times; randomly select one link in the test set to compare with the nonexistent link in $U-E$ each time; and when the similarity score of the link in the test set is greater than that of the nonexistent link, increase n' by 1. If the two scores are equal, increase n'' by 1; that is, the randomly

TABLE 1: Symbols used and their meanings.

Symbolic representation	Description
$G = (V, E, W)$	Graph of an undirected weighted network
V	Node set. $V = \{v_1, v_2, \dots, v_n\}$, $ V = n$
E	Edge set. $E = \{e(i, j) \mid v_i, v_j \in V, i \neq j\}$, $ E = m$
W	Weight set. $W = \{w(i, j) \mid v_i, v_j \in V, i \neq j\}$
U	Collection of links in the complete graph of G
$w(x, y)$	Edge weight value between nodes v_x and v_y
k_x	Degree of node v_x
s_x	Sum of the weights of all edges connected to node v_x
$\Gamma_1(x)$	The first-order neighbour set of node v_x
$\Gamma_2(x)$	The second-order neighbour set of node v_x
$S_{x,y}$	Similarity between nodes v_x and v_y
CCN	Current common neighbour
FCN	Future common neighbour
EWI	Endpoint weight influence
LWI	Link weight influence
$HSNI$	High-strength node influence
$S_{x,y}^{WS-CCN}$	Node influence based on CCNs
$S_{x,y}^{ELH-FCN}$	Node influence based on FCNs
$S_{x,y}^{CCN-FCN}$	Total similarity based on the influence of CCNs and FCNs
$S_{x,y}^{CCNI-FCNI-DS}$	Similarity based on DS and the influence of all common neighbours

selected link in the test set has a higher probability than the nonexistent link, and the larger the AUC value is, the higher the prediction accuracy is. In this study, n is set to 10000.

$$AUC = \frac{n' + 0.5n''}{n}. \quad (8)$$

In the link prediction experiment, the set comprising all links in the network except the training set is called the unknown edge set, that is, E^{uk} . It contains the edges in the test set and the links that do not exist. The indicator of precision [38] is used to calculate the existence probability of all the links in set E^{uk} , and these links are arranged in descending order. In the first L links in the descending order, if there are m links belonging to the test set, the prediction precision is evaluated using m/L , as shown in formula (9). It can be seen that the value of *precision* depends on the value of L , and in the initial experiment in our study, L is set to 10.

$$Precision = \frac{m}{L}. \quad (9)$$

3. Proposed Method

Most existing weighted similarity indices are simply weighted based on the CN algorithm and only consider the influence of CCN nodes. To relatively better mine the influence of node information on similarity, three types of FCN nodes are proposed, and three concepts of EWI, LWI, and HSNI are introduced to capture the influence of node information on similarity from different angles. On this basis, node influences based on CCNs and FCNs are defined. Finally, the DS evidence theory is used to reasonably and effectively combine them, and a new index, CCNI-FCNI-DS, which can measure the comprehensive influence of different factors on node similarity in weighted networks, is obtained.

3.1. Question Posed

3.1.1. Problem Definition. The proposed link prediction model for weighted networks can detect all the FCNs of the predicted node pair and effectively measure the contribution of this type of node to similarity. As shown in Figure 1, suppose $\langle a, b \rangle$ is the seed node pair to be predicted, node c is the CCN of $\langle a, b \rangle$, and nodes d, e , and f are the three FCNs of the seed node pair. If node $d, d \in \Gamma_1(a) \cap \Gamma_2(b)$, can be directly linked to node b in the future, then d can be considered as an FCN of $\langle a, b \rangle$. In this paper, an FCN is a node that is not a first-level CN of a node pair at present but can be its first-level CN in the future. To measure the contribution of the information of FCNs to the similarity, this paper presents a detailed definition of the type of FCN and its topology; on this basis, the node influence based on the FCN is proposed.

3.1.2. Topological Structure of FCN Nodes. In the weighted graph shown in Figure 1, there are three types of FCN nodes. As shown in Figure 2, $\langle a, b \rangle$ is the seed node pair to be predicted, and our goal is to predict whether a link will be established between the node a and node b in the future. In terms of the node pair $\langle a, b \rangle$, its first type of FCN means that there is a node c which is directly connected to node a , but node c is not connected to node b at present. Then, we can say that node c belongs to the first type of FCN of $\langle a, b \rangle$; the current node c is directly connected with a and is not connected with b , that is, $c \in \Gamma_1(a) \wedge c \notin \Gamma_1(b)$, as shown in the T1 structure. By calculating the similarity between current nodes c and b , it is found that the greater the similarity is, the higher the probability of a link forming between the two nodes is, and when it is easier for current node c to link with node b , node c becomes the CCN of node pair $\langle a, b \rangle$. The algorithm calculates the contribution of CCNs and FCNs to the seed node pair and measures the influence of all neighbour nodes on the similarity for achieving

FIGURE 1: Example of a weighted network topology.

FIGURE 2: Three topologies of FCN nodes.

a highly accurate prediction. Similarly, as shown in Figure 2, the second FCN node is $c \in \Gamma_1(b) \wedge c \notin \Gamma_1(a)$, which is denoted as the T2 structure. The third FCN node is $c \notin \Gamma_1(a) \wedge c \notin \Gamma_1(b)$, which is denoted as the T3 structure.

The degree, strength, and edge weight between the current node and its neighbours all influence the similarity between the seed node pair. The role of FCN nodes in link prediction is further illustrated in Figure 3, where $\langle a, b \rangle$ is the predicted node pair, and $c \in \Gamma_1(a) \cap \Gamma_1(b)$ is the CCN node of $\langle a, b \rangle$. According to the three topological structures described in Figure 2, nodes d , e , and f shown in Figure 3 are the first, second, and third type FCN nodes of $\langle a, b \rangle$, respectively. By calculating the similarity between the FCN node and seed node, the local and semiglobal influence of FCNs on the seed node pair to be predicted can be determined.

3.2. Node Influence Based on CCNs. The classic weighted similarity index is often simply a weighted superposition based on CN, without the consideration of the influence between CCNs and other nodes on the similarity. Based on this, this paper comprehensively uses the degree, strength, and edge weight to calculate the node influence based on the CCN, called WS-CCN. $\forall x, y \in V$, the similarity contribution of the CCNs of $\langle x, y \rangle$ to this node pair is denoted as $S_{x,y}^{WS-CCN}$, as shown in the following formula:

$$S_{x,y}^{WS-CCN} = \sum_{z \in \Gamma_1(x) \cap \Gamma_1(y)} \frac{w(x,z) + w(y,z)}{s_z \times k_z}. \quad (10)$$

3.3. Node Influence Based on FCNs. In the discussion of the contribution of FCNs to links, this paper uses multisource information of FCNs to define three influence values from different angles.

3.3.1. EWI. EWI is defined as the ratio of the weight value between the predicted node pair and the total strength of their CNs, which is defined as shown in formula (11), for calculating the influence of the local topology formed by the CN nodes on the potential link.

$$EWI(x, y) = \sum_{z \in \Gamma_1(x) \cap \Gamma_1(y)} \frac{w(x, z) + w(y, z)}{s_z}. \quad (11)$$

3.3.2. LWI. LWI is defined as the ratio of the edge weight between nodes to the strength sum of the two nodes, as shown in formula (12). It considers the influence of the link weight between CNs and the current node on the strength sum of the two nodes, which can also be understood as the influence of the local aggregation of links on the similarity.

$$LWI(x, y) = \sum_{z \in \Gamma_1(x) \cap \Gamma_1(y)} \left(\frac{w(x, z)}{(s_x + s_z)} + \frac{w(y, z)}{(s_y + s_z)} \right). \quad (12)$$

3.3.3. HSNI. HSNI is defined as the ratio of the number of CCNs to the maximum strength of the two nodes, as shown in formula (13), which is used to further measure the similarity influence of the CCNs on the high-strength node pair.

$$HSNI(x, y) = \frac{|\Gamma_1(x) \cap \Gamma_1(y)|}{\max(s_x, s_y)}. \quad (13)$$

Based on the definitions of these three values, the node influence based on FCN is obtained, which is called ELH-FCN, and its similarity contribution to the target node pair is denoted as $S_{x,y}^{ELH-FCN}$, as shown in the following formula:

FIGURE 3: Example of three types of FCN node.

$$S_{x,y}^{\text{ELH-FCN}} = \text{EWI}(x, y) \times \text{LWI}(x, y) \times \text{HSNI}(x, y). \quad (14)$$

3.4. Proposed Model. Considering that the contribution of CCN to the similarity is higher than that of FCN, influence factors 9 and 1 are given to the two neighbour nodes, respectively. Then the total similarity based on the influence of CCNs and FCNs is obtained, which is denoted as $S_{x,y}^{\text{CCN-FCN}}$, as shown in the following formula:

$$S_{x,y}^{\text{CCN-FCN}} = 9 \times S_{x,y}^{\text{WS-CCN}} + S_{x,y}^{\text{ELH-FCN}}. \quad (15)$$

Based on this definition, the DS evidence theory is used to fuse the similarity contribution based on CCNs and FCNs. In this paper, the identification framework of seed node pair $\langle x, y \rangle$ based on the evidence theory is denoted as $\{m_{x,y}, \overline{m}_{x,y}\}$, where $m_{x,y}$ represents the probability of a link existing between nodes x and y , and its definition is shown in formula (16); $\overline{m}_{x,y}$ represents the probability that there is no link between nodes x and y , and its definition is shown in formula (17). The new fusion weighted similarity index is $S_{x,y}^{\text{CCNI-FCNI-DS}}$, as shown in formulas (18) and (19).

$$m_{x,y} = (9 \times S_{x,y}^{\text{WS-CCN}}) \oplus S_{x,y}^{\text{ELH-FCN}}, \quad (16)$$

$$\overline{m}_{x,y} = (1 - (9 \times S_{x,y}^{\text{WS-CCN}})) \oplus (1 - S_{x,y}^{\text{ELH-FCN}}), \quad (17)$$

$$M = m_{x,y} \otimes \overline{m}_{x,y} = \frac{m_{x,y}}{m_{x,y} + \overline{m}_{x,y}}, \quad (18)$$

$$S_{x,y}^{\text{CCNI-FCNI-DS}} = 1 - \prod (1 - M). \quad (19)$$

Based on these definitions, the corresponding contribution of the neighbours in Figure 3 is calculated. In terms of the seed node pair $\langle a, b \rangle$, node c is its CCN, node d is the first FCN, node e is the second FCN, and node f is the third FCN. The corresponding results are as follows:

$$\begin{aligned}
 S_{a,b}^{\text{WS-CCN}}(c) &= 0.5; \\
 S_{a,b}^{\text{ELH-FCN}}(d) &= 0.2418; \\
 S_{a,b}^{\text{ELH-FCN}}(e) &= 0.1753; \\
 S_{a,b}^{\text{ELH-FCN}}(f) &= 0.1273.
 \end{aligned} \quad (20)$$

3.5. Algorithm Description

Input: adjacency matrix of an undirected graph $G = (V, E, W)$.

Output: the similarity matrix of G and the corresponding prediction results.

Step 1: read the dataset file, and store the data as $n \times n$ adjacency matrix

Step 2: calculate the similarity contribution of CCNs according to formula (10)

Step 3: calculate the corresponding influence according to formulas (11)–(13) and the similarity contribution of FCNs according to formula (14)

Step 4: calculate the total similarity based on CCNs and FCNs according to formula (15)

Step 5: calculate the total similarity after the fusion of the DS evidence theory according to formulas (16)–(19), and store it in the similarity matrix

Step 6: traverse the similarity matrix, sort the elements in descending order, and output the corresponding prediction results

4. Experiment and Analysis

To verify the correctness and effectiveness of the proposed algorithm, experiments were performed on nine real weighted network datasets and 40 simulation datasets, with AUC and precision as evaluation indicators. The experimental results of the proposed algorithm are compared with those of six classic weighted similarity indexes and several

related weighted network link prediction algorithms, such as STNMP [30]. Given that a large number of studies have shown that the 10-fold cross-validation method [39, 40] can achieve the best tradeoff between the computational complexity and performance, we use it to divide the dataset into a training set and a test set. Furthermore, the robustness of the algorithm is verified by adjusting the training set ratio and parameters in evaluation indicators.

4.1. Description of Experimental Process. The implementation of the proposed algorithm is based on Windows 10 operating system and the MyEclipse10 development tool through Java and Python language coding, and the Gephi software is used to complete the topology analysis of datasets. The experiment process is as follows:

- (i) Preprocess the obtained datasets. For example, ignore the direction of the edge and convert the dataset into an undirected graph; remove duplicate edges and isolated nodes with a degree of 0. Then the dataset is converted into *.csv format for storage, and we use three pieces of data to represent the topology information of each link, namely, the number label of the two nodes and the weight between them. Subsequently, every dataset in *.csv format is analysed through Gephi, and the corresponding topology information is obtained, such as the average degree of nodes and the clustering coefficient.
- (i) Every dataset is split into a training set and a test set by the 10-fold cross-validation method, and the ratio of the number of links in the two sets is 9:1. Namely, for each dataset, 10% of links are randomly selected as the test set, and the remaining 90% of links are the training set. Moreover, the division is repeated 10 times to ensure that all data are both trained and tested.
- (ii) Using the links in the training set as known information, randomly select links from the test set and the nonexisting edge set and calculate the similarity of the two node pairs corresponding to the two selected links.
- (iii) AUC and precision are used as evaluation indicators, and the average value is obtained after 10 independent experiments to evaluate the prediction accuracy and verify the correctness and effectiveness of the algorithm.

4.2. Experiments on Real Weighted Networks

4.2.1. Real Datasets. Nine real weighted networks were obtained, and their topology information is shown in Table 2, where $|V|$ is the number of nodes, $|E|$ is the number of edges, $\bar{k}$ is the average degree of nodes, WAD is the weighted average degree, N_d is the graph density, C is the network clustering coefficient, and APL is the average path length.

4.2.2. Results and Comparative Analysis. Ten experiments were conducted independently, and the average values of AUC and *precision* were calculated, as shown in Figures 4 and 5. From Figures 4 and 5, we can find that, in the USAir network with the smallest average path length, NetScience network with the largest average path length and the smallest graph density, Reco-Net network with the smallest aggregation coefficient, TrainBomb network with a larger aggregation coefficient, Animal-Social network with the largest graph density, and Sandi-auths network with a smaller graph density, the proposed algorithm always shows high performance, and its prediction accuracy is better than that of other algorithms which only consider a single factor. This result further verifies the correctness and effectiveness of using the DS evidence theory to fuse the influence of CCNs and FCNs to define the similarity.

4.3. Experiments on Artificial Weighted Networks

4.3.1. Artificial Datasets. To further verify the accuracy of the proposed method, artificial weighted networks were used. The research in [41] showed that the degree (recorded as k), the strength (recorded as s), and the edge weight (recorded as w) always satisfy power-law distribution, namely, $p(k) \propto k^{-\gamma}$, $p(s) \propto s^{-\gamma}$, $p(w) \propto w^{-\gamma}$, where $\gamma \in [2, 3]$ in most real weighted networks. According to this, four networks with a power-law distribution of nodes are generated using Python complex network analysis library, NetworkX. Each type of network includes 10 simulation datasets, and their corresponding number of nodes is 100, 200, ..., 1000, respectively. Thereafter, the edges in the four networks are weighted, and then 40 artificial weighted networks are formed. The probability distributions of weights of edges in the four networks are uniform distributions with a random integer between 1 and 10, namely, $\forall x, y \in V, w(x, y) \in [1, 10]$; the power-law distribution with γ equals 2, 2.5, and 3.

4.3.2. Comparative Analysis of Results on Artificial Datasets. Experiments were performed on 40 artificial datasets. The results of AUC and *precision* are shown in Figure 6, Figure 7, and Figure 8. Results showed that the proposed algorithm achieved a good effect on datasets with uniform or power-law distribution. With network scale expansion, the prediction accuracy of related algorithms on each network with 100–1000 nodes gradually showed a downward trend, but the precision of the proposed algorithm on the same dataset was always the highest, showing its sufficient robustness.

4.4. Parameter Sensitivity Analysis

4.4.1. Comparative Analysis under Different Training Set Proportion on Real Datasets. To further verify the robustness of the proposed algorithm, the proportion of the training set was adjusted from 90% to 80% and 70% in turn, recorded as $|E_{tr}|/|E|$. For example, for each dataset, 20% of links in the graph are randomly selected as the test set, and the remaining 80% of links are the training set. Seven datasets were selected,

TABLE 2: Basic topological structure of real weighted datasets used in the experiment.

No.	Dataset	$ V $	$ E $	$\bar{k}$	WAD	N_d	C	APL
1	Karate	34	78	4.588	13.588	0.139	0.588	2.408
2	Office	40	238	11.9	29.9	0.305	0.43	1.764
3	Sandi-auths	86	124	2.884	4.186	0.034	0.575	4.777
4	TrainBomb	64	243	7.594	8.812	0.121	0.711	2.691
5	NetScience	379	914	4.823	2.583	0.013	0.798	6.042
6	Reco-net	100	200	4	4.662	0.04	0.185	3.768
7	GD01-C	33	135	8.182	8.848	0.256	0.578	2.17
8	Animal-social	17	54	6.235	10.353	0.39	0.609	1.743
9	USAir	332	2126	12.807	0.924	0.039	0.625	0.738

FIGURE 4: AUC of different algorithms in real datasets.

FIGURE 5: Precision of different algorithms in real datasets.

and experiments were performed again in the same environment. The AUC values of eight algorithms on these datasets were obtained, as shown in Figure 9.

4.4.2. Comparative Analysis under Different Training Set Proportion on Simulation Datasets. Similar experiments are performed on simulation datasets, and results are shown in

Figures 10 and 11, from which we know that each algorithm has the optimum prediction effect when $|E_{tr}|/|E| = 0.9$, and the accuracy of all algorithms gradually decreases with an increase in the proportion of test sets. Moreover, the empirical research on the proportion of training set to test set in machine learning shows that 9:1 can achieve good results. So, the proportion 9:1 is set in subsequent experiments.

FIGURE 6: Results of AUC on uniformly distributed simulation datasets.

FIGURE 7: Results of precision on uniformly distributed simulation datasets.

4.4.3. Precision under Different L Values on Real Datasets.

When the precision indicator is used to evaluate the accuracy of the algorithm, the result depends on the value of L . With the increase in L , the precision tends to gradually decrease. To further verify the accuracy and robustness of the proposed algorithm, the precision of the algorithm under different L is statistically analysed on nine real datasets, and the results are shown in Figure 12.

Results in Figure 12 show that the proposed algorithm performs obvious advantages in almost all real datasets. With the increase in L from 10 to 100, the precision of the proposed algorithm always keeps almost the highest, it is not

greatly affected by the value of L , and the prediction accuracy slightly fluctuates as a whole, showing its high stability and robustness.

4.4.4. Precision under Different L Values on Simulation Datasets.

Experiments were also performed on 25 simulation datasets. The precision results on ten networks with uniform weight distribution and fifteen networks with power-law distribution are shown in Figure 13 and Figure 14, respectively. From them, we know that, with the increase in L , the precision of the eight algorithms on the same network shows a downward trend irrespective of

FIGURE 8: AUC of different algorithms under power-law distributed simulation datasets.

FIGURE 9: Comparative results of AUC under different training set proportions.

the type of datasets. With the expansion of the dataset scale, the performance of all algorithms decreases slightly, but the precision of the proposed method *CCNI-FCNI_DS* always remains the optimum on all types of artificial datasets.

4.5. Algorithm Robustness Verification. An unweighted network is a special weighted network with the weights of all edges of 1. When a weighted network is transformed into an unweighted network by ignoring the weight of the edge, then the node strength degenerates into the node degree. To further verify the robustness of the proposed algorithm, a large-scale network NetScience was selected, and the weights of edges were ignored to transform it into an unweighted network. Simultaneously, the real unweighted Karate

network was selected, and these two classic datasets were obtained as examples. Their basic attributes are shown in Table 3.

The proposed algorithm is compared with the recent algorithms *CMS-CN* [23], *TRA*[23], *NCCP*[24], and *IMP-CN* [35], as well as four other algorithms. They are the link prediction algorithm based on high-order path similarity by punishing the long path (*HPS-LP*) between the predicted node pairs [42], the link prediction method based on local and global structure information by measuring the relative entropy (*RE*) under the joint action of first-order and second-order neighbour information [43], the algorithm called *HD* that was proposed based on a new definition of global and quasiloal extensions of some commonly used local similarity indices [44], and the link prediction algorithm called *MSLPA* based on community preference information

FIGURE 10: AUC of uniformly distributed networks under different training set proportions.

FIGURE 11: AUC of power-law distributed networks under different training set proportions.

by considering the network structure attributes and interest preferences of users as the dominant factors in a Twitter dataset [45]. For the two unweighted networks, comparison results of these algorithms based on the AUC evaluation index are shown in Figure 15.

The prediction accuracy of the proposed method is better than that of other algorithms, whether it is a small-scale unweighted Karate network or a large-scale unweighted NetScience network. It can also achieve relatively higher

performance, robustness, and universality for link prediction in unweighted networks.

4.6. Algorithm Complexity Analysis. The model *CCNI-FCNI_DS* proposed in this paper combines local and semiglobal structure information to define node similarity. This algorithm uses the adjacency matrix to store the undirected weighted graph $G=(V,E,W)$, and the

FIGURE 12: Precision of various algorithms under different (L) values on real datasets.

space complexity is $O(n^2 + n * m)$, where n is the number of nodes and m is the number of edges in the graph G . When initializing the adjacency matrix, the corresponding time complexity is $O(n^2)$. When calculating the contribution of common neighbours, the corresponding

time complexity is $O(n^2)$. When calculating the contribution of three types of future common neighbours, the corresponding time complexity is $O(n^2m)$. Therefore, the total time complexity of the proposed algorithm is $O(n^2m + n^2)$. Compared with some classic algorithms

FIGURE 13: Precision on simulation datasets with uniform weights under different L . FIGURE 14: Precision on simulation datasets with power-law distribution under different L .

TABLE 3: Basic topological properties of unweighted network datasets used in experiments.

Dataset	$ V $	$ E $	$\bar{k}$	C	APL
Unweighted Karate	34	77	4.529	0.574	2.426
Unweighted NetScience	2149	5390	3.091	0.027	4.377

based on local similarity, the time complexity of the WCN algorithm is $O(n^2)$, and the time complexity of WAA, WRA, and WJaccard algorithm is $O(2n^2)$, while the algorithms based on global similarity, such as Katz and Random Walk, have a time complexity of $O(n^3)$. It can be seen that although the time complexity of the algorithm

proposed in this paper is slightly higher than that of the similarity algorithm that fuses local and global structural features, the proposed model CCNI-FCNI_DS can still guarantee the execution efficiency on the premise of improving the prediction accuracy, which shows good performance in link prediction in weighted networks.

FIGURE 15: Experimental results on unweighted networks.

5. Conclusions

Through the in-depth study of the shortcomings of the existing link prediction algorithms based on node similarity, this paper proposes a link prediction model for weighted networks, which integrates the *CCN* influence and *FCN* influence by using the *DS* evidence theory. The algorithm comprehensively uses the degree, strength, and edge weight to define the influence of *CCNs*; based on the three types of *FCNs*, the influence of *FCNs* is defined by introducing *EWI*, *LWI*, and *HSNI*. Finally, the *DS* theory is used to effectively fuse multiple factors that affect the similarity of nodes, fully mining the local and global structural characteristics of the network and realising link prediction in weighted networks. The accuracy and effectiveness of the proposed algorithm are verified through experimental comparison on several real and artificial datasets. However, the prediction performance of the proposed method on a certain dataset is slightly lower than the benchmark-weighted similarity index; thus, exploring the reasons and improving the algorithm are the next steps. Moreover, for large-scale datasets, how to apply the network representation method based on deep learning to weighted network link prediction and how to improve the prediction accuracy and efficiency by optimising the representation of feature information are also major research directions that can be addressed in the future.

Data Availability

The data used to support this study are available at <http://snap.stanford.edu/data/>, <http://netwiki.amath.unc.edu/SharedData/SharedData>, and <http://www-personal.umich.edu/~mejn/netdata/>.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (42002138 and 61871465), Natural

Science Foundation of Heilongjiang Province (LH2019F042), Postdoctoral Scientific Research Development Fund of Heilongjiang Province (no. LBH-Q20073), and Excellent Young and Middle-Aged Innovative Team Cultivation Foundation of the Northeast Petroleum University (KYCXTDQ202101).

References

- [1] M. Liu, Q. Hu, J. Guo, and J. Chen, "Link prediction algorithm for signed social networks based on local and global tightness," *Journal of Information Processing Systems*, vol. 17, no. 2, pp. 213–226, 2021.
- [2] M. Liu, J. Guo, and J. Chen, "Community discovery in weighted networks based on the similarity of common neighbors," *Journal of Information Processing System*, vol. 15, no. 5, pp. 1055–1067, 2019.
- [3] H. Wang and Z. Le, "Seven-layer model in complex networks link prediction: A survey," *Sensors*, vol. 20, no. 22, p. 6560, 2020.
- [4] S. Forouzandeh, M. Rostami, and K. Berahmand, "Presentation a trust walker for rating prediction in recommender system with biased random walk: Effects of H-index centrality, similarity in items and friends," *Engineering Applications of Artificial Intelligence*, vol. 104, 2021.
- [5] E. Nasiri, K. Berahmand, M. Rostami, and M. Dabiri, "A novel link prediction algorithm for protein-protein interaction networks by attributed graph embedding," *Computers in Biology and Medicine*, vol. 137, Article ID 104772, 2021.
- [6] E. Nasiri, K. Berahmand, and Y. Li, "A new link prediction in multiplex networks using topologically biased random walks," *Chaos, Solitons & Fractals*, vol. 151, Article ID 111230, 2021.
- [7] K. Berahmand, E. Nasiri, S. Forouzandeh, and Y. Li, "A preference random walk algorithm for link prediction through mutual influence nodes in complex networks," *Journal of King Saud University - Computer and Information Sciences*, vol. 3, 2021.
- [8] W. Liu and J. Chen, "Link prediction in complex networks," *Journal of Information and Control*, vol. 49, no. 1, pp. 1–23, 2020.
- [9] S. Li, J. Huang, Z. Zhang, J. Liu, T. Huang, and H. Chen, "Similarity-based future common neighbors model for link prediction in complex networks," *Scientific Reports*, vol. 8, no. 1, Article ID 17014, 2018.

- [10] Y. Li and T. Zhou, "Local similarity indices in link prediction," *Journal of University of Electronic Science and Technology of China*, vol. 50, no. 3, pp. 422–427, 2021.
- [11] Z. Ahmad and S. Rizos, "Similarity-based link prediction in social networks using latent relationships between the users," *Scientific Reports*, vol. 10, no. 1, Article ID 20137, 2020.
- [12] S. Zhu, W. Li, N. Chen, and X. Zu, "Weighted synthetical influence of degree and H-index in link prediction of complex networks," *International Journal of Modern Physics B*, vol. 34, no. 31, 2020.
- [13] M. Wang, X. Lou, and B. Cui, "A degree-related and link clustering coefficient approach for link prediction in complex networks," *The European Physical Journal B*, vol. 94, no. 1, pp. 1–12, 2021.
- [14] C. Yi, M. He, B. Wu, and L. Lv, "Link Prediction algorithm combining with community relations and community information of common neighbors," *Journal of Electronic and Instrumentation*, vol. 35, no. 5, pp. 174–181, 2021.
- [15] Z. Li, L. Ji, and S. Liu, "A method of link prediction in directed network based on effective connectivity path," *Journal of University of Electronic Science and Technology of China*, vol. 50, no. 1, pp. 127–137, 2021.
- [16] Y. Wang and J. Wang, "Design of link prediction algorithm for complex network based on the comprehensive influence of predicting nodes and neighbor nodes," *Journal of Forecasting*, vol. 40, no. 5, pp. 911–920, 2021.
- [17] S. Li, J. Huang, J. Liu, T. Huang, and H. Chen, "Relative-path-based algorithm for link prediction on complex networks using a basic similarity factor," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 1, Article ID 013104, 2020.
- [18] C. Liu, S. Yu, Y. Huang, and Z.-K. Zhang, "Effective model integration algorithm for improving link and sign prediction in complex networks," *IEEE Transactions on Network Science and Engineering*, vol. 8, pp. 2613–2624, 2021.
- [19] T. Murata and S. Moriyasu, "Link prediction of social networks based on weighted proximity measures," in *Proceedings of the IEEE International Conference on Web Intelligence*, pp. 85–88, IEEE, Fremont, CA, USA, November 2007.
- [20] S. Zhang and Y. Zhou, "Time-weighted link prediction algorithm for social network based on random walk," *Computer applications and software*, vol. 31, no. 7, pp. 28–30, 2014.
- [21] L. Lü and T. Zhou, "Link prediction in weighted networks: The role of weak ties," *EPL (Europhysics Letters)*, vol. 89, no. 1, Article ID 18001, 2010.
- [22] T. Li and H. Zhang, "Link prediction based on structure weighted network," *Journal of Northwestern Polytechnical University*, vol. 34, no. 3, pp. 544–547, 2016.
- [23] S. Bai, L. Li, J. Cheng, S. Xu, and X. Chen, "Predicting missing links based on a new triangle structure," *Complexity*, vol. 2018, Article ID 7312603, 2018.
- [24] D. Chen, Z. Yuan, and X. Huang, "Temporal network node similarity measure and link prediction Algorithm," *Journal of Northeastern University*, vol. 41, no. 1, pp. 29–344, 2020.
- [25] X. Yang, *Research on Link Prediction Algorithm Based on Path and Asymmetric Clustering Coefficient*, Zhejiang University of Technology, Zhejiang, China, 2019.
- [26] J. Jia, Y. Chen, Y. Li, T. Li, N. Chen, and X. Zhu, "Effect of weak ties on degree and H-index in link prediction of complex network," *Modern Physics Letters B*, vol. 35, no. 18, Article ID 2150301, 2021.
- [27] G. Zhao, P. Jia, and A. Zhou, "Improved degree centrality for directed-weighted network," *Journal of Computer Applications*, vol. 40, no. S1, pp. 141–145, 2020.
- [28] R. Yuan, Y. Song, and F. Meng, "Link prediction method based on weighted network topology weight," *Journal of Computer Science*, vol. 47, no. 5, pp. 265–270, 2020.
- [29] H. R. Atiya and H. N. Nawaf, "Community structure-aware fairness and goodness algorithm for link weight prediction," *Journal of Physics: Conference Series*, vol. 1804, no. 1, Article ID 012080, 2021.
- [30] J. Guo, M. Liu, and X. Luo, "Link prediction based on multipath node similarity in weighted networks," *Journal of Zhejiang University*, vol. 50, no. 7, pp. 1347–1352, 2016.
- [31] P. T. Naderi and F. Taghiyareh, "Strup: Stress-based trust prediction in weighted sign networks," *SN Computer Science*, vol. 2, no. 1, 2021.
- [32] B. Kang, G. Chhipi-Shrestha, Y. Deng, J. Mori, K. Hewage, and R. Sadiq, "Development of a predictive model for *Clostridium difficile* infection incidence in hospitals using Gaussian mixture model and Dempster-Shafer theory," *Stochastic Environmental Research and Risk Assessment*, vol. 32, no. 6, pp. 1743–1758, 2018.
- [33] S. Peñafiel, N. Baloian, H. Sanson, and J. A. Pino, "Applying Dempster-Shafer theory for developing a flexible, accurate and interpretable classifier," *Expert Systems with Applications*, vol. 148, Article ID 113262, 2020.
- [34] Y. Mao and H. Gong, "Identification of maize disease based on Svm and DS evidence theory," *Chinese Journal of Agricultural Machinery Chemistry*, vol. 41, no. 4, pp. 152–157, 2020.
- [35] Y. Liu, L. Li, and N. Dan, "A link prediction method based on aggregation coefficient fusion," *Journal of Computer Applications*, vol. 40, no. 1, pp. 28–35, 2020.
- [36] M. Liu, Y. Wang, J. Guo, J. Chen, J. Yang, and Z. Liu, "A link prediction algorithm for weighted networks based on dempster-shafer evidence theory and node multi-features," in *Proceedings of the 2021 IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET)*, pp. 302–307, Beijing, China, August 2021.
- [37] B. Liu, S. Xu, T. Li, J. Xiao, and X. K. Xu, "Quantifying the effects of topology and weight for link prediction in weighted complex networks," *Entropy (Basel, Switzerland)*, vol. 20, no. 5, p. 363, 2018.
- [38] Z. Samei and M. Jalili, "Application of hyperbolic geometry in link prediction of multiplex networks," *Scientific Reports*, vol. 9, no. 1, Article ID 12604, 2019.
- [39] J. D. Rodriguez, A. Perez, and J. A. Lozano, "Sensitivity analysis of k-fold cross validation in prediction error estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 569–575, 2010.
- [40] A. Kosir, O. Ante, and T. Marko, "How to improve the statistical power of the 10-fold cross validation scheme in recommender systems," in *Proceedings of the International Workshop on Reproducibility and Replication in Recommender Sytems Evaluation (RepSys'2013 ACM)*, pp. 3–6, ACM, Hong Kong, China, October 2013.
- [41] M. Liu, J. Guo, and J. Chen, "Partitioning weighted social networks based on the link strength of nodes and communities," *Journal of Information Hiding and Multimedia Signal Processing*, vol. 9, no. 1, pp. 21–32, 2018.
- [42] Q. Gu, B. Wu, and R. Chi, "Link Prediction method based on the similarity of high path," *Journal on Communications*, vol. 42, no. 7, pp. 61–69, 2021.
- [43] Y. Meng and J. Guo, "Link prediction algorithm based on node structure similarity measured by relative entropy," *Journal of Physics: Conference Series*, vol. 1955, no. 1, Article ID 012078, 2021.

- [44] F. Aziz, H. Gul, I. Uddin, and G. V. Gkoutos, "Path-based extensions of local link prediction methods for complex networks," *Scientific Reports*, vol. 10, no. 1, Article ID 19848, 2020.
- [45] J. Ge, L. L. Shi, L. Liu, H. Shi, and J. Panneerselvam, "Intelligent link prediction management based on community discovery and user behavior preference in online social networks," *Wireless Communications and Mobile Computing*, vol. 2021, Article ID 3860083, 13 pages, 2021.