Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 8829845, 12 pages
https://doi.org/10.1155/2020/8829845

Research Article

An Algorithm Based on Influence to Predict Invisible Relationship

Junfeng Tian, Lizheng Xue, and Hongyun Cai

School of Cyber Security and Computer, Hebei University, Baoding, Hebei Province, China

Correspondence should be addressed to Hongyun Cai; chy_hbu@126.com

Received 22 June 2020; Revised 15 November 2020; Accepted 24 November 2020; Published 7 December 2020

Academic Editor: Ding Wang

Research on social networks is at its peak in the current era of big data, especially in the field of computer research. Link prediction in social networks has attracted an increasing number of researchers. However, most of the current studies have focused on the prediction of the visible relationships between users, ignoring the existence of invisible relationships. The same as visible relationships, invisible relationships are also an indispensable part of social networks, and they can uncover more potential relationships between users. To better understand invisible relationship, definition, types, and characteristics of invisible relationship have been introduced in this paper. Also an influence algorithm is proposed to speculate on the existence of invisible edges between users. The algorithm is based on three indicators, namely, the *occasional contact degree*, *interest coincidence degree*, and the *popularity of users*, and it takes the *influence* as reference. By comparing with the threshold, Θ, defined in advance, users with relationships stronger than Θ are viewed as possessing invisible relationships. The feasibility and accuracy of the algorithm are proven by extensive numerical experiments compared with one well-known and widely used method, i.e., the *common neighbors* (CN).

1. Introduction

The progress of science and technology makes communication more convenient, especially the development of instant messaging software and mobile networks. People can share and publish their experience to communicate with others. The data actually realized the automatic generation and the big data era has arrived [1]. As one of the great applications in the big data era, social networks have attracted many loyal users by the powerful social function. By analyzing the data in social networks, much information of users can be gained which can help to provide better personal services for users, e.g., recommendation of web pages or goods or prediction of new links [2]. Among these applications, the problem of prediction potential links has attracted more and more attentions in recent years [3–5]. However, the existing studies on link prediction have focused on prediction of the visible relationships between users, ignoring the existence of invisible relationships. Invisible relationship is a kind of secret relationships which exists in social network and does not want to be discovered. The discovery of invisible relationship is a warning for special users but significant value for other users in the social network. For example, two people, who are engaged in a special job, always pass information through public (e.g., billboards) or participate in some common topics together, but they never interact with each other directly. If the invisible relationship between them is discovered, this may reveal their true identities and result in some horrendous consequences. However, if these users are malicious, the exposure of their invisible relationships can help to provide a safer network for genuine users. Therefore, the prediction of invisible relationships can help to build a safer environment and better protect genuine users in social networks.

To understand and predict invisible relationships in social networks, we first introduce the definition, types, and characteristics of invisible relationship. Then, we propose a novel influence algorithm to speculate on the existence of invisible relationship. Particularly, three indicators are presented for calculating the influence between two users, i.e., occasional contact degree, interest coincidence degree, and popularity of users.

The main contributions of this paper are summarized as follows:

(1) Different from existing link prediction methods that focus mainly on visible relationships in social networks, we first present the definition of invisible relationship between different users. Prediction of invisible relationship is an important complement of link prediction in social networks, which can help to enhance the security of social networks

(2) To predict the invisible relationships between users, we propose an influence algorithm based on three indicators, i.e., occasional contact degree, interest coincidence degree, and popularity of users.

The rest of the paper is organized as follows. "Related Work" introduces the related work about individual influence and link prediction. "Invisible Relationships" describes the definition, types, and characteristics of invisible relationship. "Influence Algorithm" proposes an influence algorithm for predicting invisible relationships between users. The experimental results are reported and analyzed in "Design and Analysis of Experiment." "Conclusion" concludes the proposed invisible relationship and prediction algorithm and discusses the future work.

2. Related Work

The research on individual influence is mainly focused on degree, closeness, and betweenness [6, 7]. Chintakunta et al. [8] proposed the SoCap method to find the influential nodes in a social network. In this method, the allocated value indicates the individual social capital. Subbian et al. [9] proposed a matrix factorization-based method to measure the nodes' influence. Liu et al. [10] introduced the trust-oriented social influence method to assess individual influence. Deng et al. [11] evaluated the influence of different nodes by combining the time-critical aspect with the characteristics of the nodes. Wang et al. [12] studied the influence of microblog opinion leaders by analyzing and modeling message propagation. Cao et al. [13] put forward a recognition algorithm named MFP (Multi-Feature PageRank), used to identify opinion leaders.

The research on link prediction can be classified as different categories. The main methods in the previous work were based on the Markov chain [14, 15] and machine learning [16]. Currently, research methods can be divided into three categories based on network structure, i.e., similarity, maximum likelihood estimation, and the probability algorithm. The Jaccard index [17, 18] is the earliest local link prediction algorithm proposed by Jaccard in 1901. In 2003, Adamic et al. [19] proposed the similarity method based on the inverse log frequency of the occurrence between users and predict relationships by similarity rank. Next, Liben-Nowell [20] proposed the well-known common neighbor index (CN). Zhou et al. [21] put forward the RA (Resource Allocation) similarity measure to predict missing links in networks. Recently, Xu et al. [22] proposed the CRA index algorithm to predict the hidden links based on the node attributes and local information. Wang et al. [23] proposed a novel index for link prediction based on the topology information and community information. This method calculates the likelihood of a link between two disconnected nodes by a similarity index. Sun et al. [24] presented a novel similarity index named the LAS method, which considers not only the common neighbors of nodes but also their community structure. Muniz et al. [25] combined the contextual, temporal, and topological information together for link prediction in social networks. Xu et al. [26] analyzed the dynamic properties of the interactions between different nodes and proposed a distributed temporal link prediction method, which uses the label propagation to update the similarity values of labels. Das et al. [27] proposed a Markov prediction model for link prediction. This method takes into account the effect of time scales and predicts the links based on the time-varying graph. Wang et al. [28] proposed a fusion probability matrix factorization framework to predict hidden links, which considers both symmetric metrics and asymmetric metrics. Shang et al. [29, 30] discussed the role of time and proposed the methods of link prediction in evolving networks. Rafiee et al. [31] proposed the CNDP method for link prediction based on common neighbor degree penalization, which determines the similarity score by combining the common neighbors of two nodes and the clustering coefficient of the network. Moreover, Zhang et al. [32] discussed the bipartite graph link prediction and proposed a novel method based on attribute extraction and similarity calculation of nodes.

This study benefits from the above research. Aiming at exploring the relationships between users in social networks, the invisible relationship is introduced and the related concepts are defined and explained in detail. Moreover, the occasional contact degree, interest coincidence degree, and user popularity are defined to measure invisible relationships. The randomness of links can be represented by the occasional contact degree, and interest coincidence degree is similar to the homogeneity implied in the proverb "Like attracts like, birds of a feather flock together" and user popularity can be gotten through common neighbors between users. The invisible relationships between users can be obtained by analyzing the occasional contact degree, interest coincidence degree, and user popularity. The proposed theory can be widely used in many fields of social networks including enrichment and perfection of the relationships between users.

3. Invisible Relationship

To discuss conveniently, social networks are represented as undirected acyclic graphs, denoted as $G = (V, E)$, where V and E represent the set of nodes and the set of edges in the networks, respectively.

3.1. Definition. Invisible relationship is different from visible relationship widely studied in previous research. To understand invisible relationship, the definitions of two kinds of relationships are given as follows:

Definition 1. Visible relationship. This refers to the real connection relationships that a user has in the social network diagraph. The edges between these users are called visible edges, noted as E_{visible} and represented as formula (1) as follows:

$$E_{\text{visible}} = \{(u, v) \mid u \in F(v), u, v \in V\}, \quad (1)$$

where $F(v)$ represents the friend collection of the user node v.

Visible relationships are pervasive links between social network users. As extension and reflection of interpersonal relationships in real life, social networks generally show visible relationships with relatives and friends, and to a certain extent, they can meet user demand for making friends.

However, the existence of invisible relationships will provide users with more potential connections.

Users in social networks may have had the experience that some of the other users' opinions and ideas coincide with their own ideas, or their needs can be satisfied by other unknown users. It is a seemingly nonexistent relationship that the invisible relationship represents. Moreover, social networks are becoming more and more open, and social network relationships show a strange social paradox. Take WeChat as an example. WeChat will be added to the activities, the same as the dinner party. The difference between familiarity and strangers is gradually disappearing. There are no friends in the "friends circle," and the "human relationship" is fading out.

The invisible relationship can be regarded as a weakened relationship in a sense.

Definition 2. Invisible relationship. This refers to the relationship that is not represented between users in the social network diagram, but that can be of help in one's social life.

The relationship is an inherent and weak link between users.

The edges between users with invisible relationships are called invisible edges, noted as $E_{\text{invisible}}$ and calculated as equation (2) as follows:

$$E_{\text{invisible}} = \{(u, v) \mid u \notin F(v), u, v \in V, \text{influece}(u, v) \geq \Theta\}, \quad (2)$$

where influence (u, v) represents the influence strength of users to establish invisible relationships; a specific definition will be presented in the next section. Θ is a predefined threshold.

Furthermore, in order to have a deep understanding of invisible relationships, from different points of view, two meanings of invisible edges are given by the following:

The view of graph theory. There is no direct representation in the social network topology diagraph, but it can affect the behavior of users, thereby making it possible for unlinked users to establish new links (not necessarily having a direct connection, similar to users in the same social group can communicate with each other, such as a QQ group). Edges such as this can be called invisible edges.

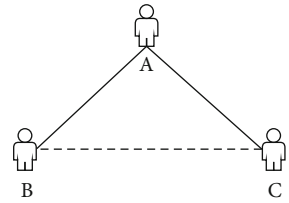

Figure 1: Triangle relationship type.

The view of the user's aim. There is no connection or direct connection between users, but there is a common goal or common interest, which can drive them to achieve the goal and realize a "win-win" situation in the end. The edges between these users are called invisible edges.

The concept of invisible relationships is introduced into social networks in this section; thus, the linked set of social network graphs can be extended and enriched to detailed classification. Then, the edge set E of social network graph G can be expressed as formula (3) as follows:

$$E = E_{\text{invisible}} \cup E_{\text{visible}}. \quad (3)$$

3.2. Types. Both the visible and invisible are objective relationships between users. The user invisible relationships possess more potential value for connections. They can reflect not only potential friend relationships but also the user's personal information (e.g., the types and interests of potential friends, and the targeted user may be affected by what kinds of friends that have). To study invisible relationships better, according to the different manifestations, we divide invisible relationships into three types, namely, triangle relationships, common interest, and demand-interest types.

3.2.1. Triangle Relationship Type. In social network link prediction research, a triangle structure is common in the topological graph. Similar to the triangle structure, a triangle relationship type (also named the common-friends type) is the simplest form of invisible relationship. That is, there might be invisible relationships between users with common friends.

As shown in Figure 1, nodes A, B, and C are three different users in the social network. User A is a common friend of users B and C (shown in the solid line in the figure, similarly hereafter). Then, there may be an invisible edge between users B and C (shown in the dotted line in the figure, similarly hereafter). The common-friends type of invisible relationship is easy to understand, but taking into account the similarity with the triangle structure in visible relationships, users with common friends such as users B and C are more likely to establish visible relationships. This is unable to highlight the existence of the invisible relationship. Therefore, more attention is paid to the following types.

3.2.2. Common Interest Type. People in social life generate social groups. As the extension of social life, the social networks are no exception. The homogeneity implied in the proverb "Like attracts like, birds of a feather flock together" also applies to social networks. It is the homogeneity that

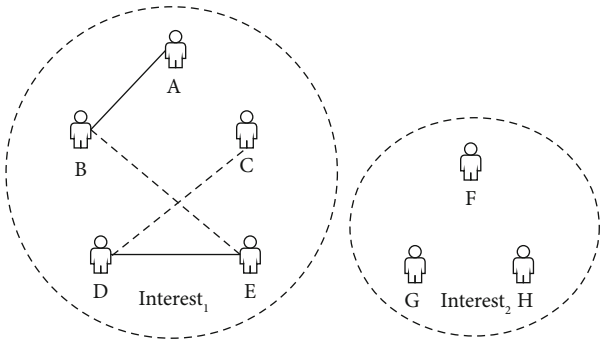

Figure 2: Common interest type.

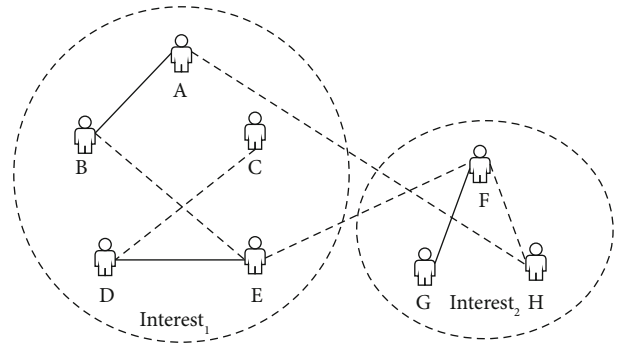

Figure 3: Demand-interest type.

makes users with the same interest have a tendency to establish invisible relationships. The common interest type of invisible relationships is similar to social interest, but the details are different. Users belonging to interest social will establish connection in certain ways, e.g., an online community. The invisible relationship between users is intrinsic, it is the manifest problem of invisible relationships that a connection is established or not.

In addition, even if a community is established, if there is no direct connection between users, it can still be considered an invisible relationship. That is, the concept of invisible relationship is larger and more specific than interest social.

As shown in Figure 2, Interest_1 and Interest_2 are two different interest groups (in order to simplify the representation, we only draw the invisible relationship in Interest_1). Users *A* and *B* and *D* and *E* are friends, respectively.

Due to various reasons, such as geographical location, users *B*, *C*, and *E* and users *A*, *C*, and *D* have no direct contact but belong to the same group named Interest_1. Therefore, users *B* and *E* and *C* and *D* may have invisible relationships (also maybe users *A* and *C*; the relationship in the figure is just an example).

The common interest type is the most common type of invisible relationship. In scientific research, taking link prediction in social networks as an example, there is an invisible relationship between the researchers who are concerned and interested in this direction. A common research interest makes it possible for researchers to communicate and discuss with each other, such as the circulation of relevant papers and the convening of thematic meetings.

3.2.3. Demand-Interest Type. The enemy of my enemy is my friend. From the point of common purpose, there is the demand-interest type of invisible relationship. For instance, if user *A* is the enemy of user *B*, and user *C* is at odds with user *B* for some reason (such as interests and disputes), then there is an invisible relationship between user *C* and user *A*.

The demand-interest type is defined according to the view of the user's aim for the invisible relationship. One released his or her requirement, and it is completed by the other user who has the ability or is just interested in it. These users complete their goals and each takes what he needs eventually. The specific form is shown in Figure 3.

Due to lack of ability and interest, user *F* in interest group Interest_2 needs to ask other users for instruction and help in completing a mission. User *E* has the ability to satisfy the demand just right. Then, there is a demand-interest type of invisible relationship between user *E* and user *F*.

Compared with the common friend type, the demand-interest type is relatively few, but it is also a common form of invisible relationship. The relationship between teachers and users of online open courses is a typical demand-interest relationship. Teachers are interested in lectures; meanwhile, users who need this course can obtain relevant professional knowledge.

3.3. Characteristic

3.3.1. Universality. All things are related, the same as people in social life. According to the principle of a small world (also named six-degree segmentation theory) [33], any two strangers in the world can establish a connection within six people. That is, a person can know any stranger through five persons at most.

For example, for the whole world, there may be no direct connection between two specific individuals (such as a Chinese and a non-Chinese). However, living in the global village is a kind of invisible relationship between the two persons despite generalization. Therefore, the invisible relationship is ubiquitous, which is why it has the following characteristics.

Universality is the premise and foundation of the existence of invisible relationships.

3.3.2. Weak Connection. Users with invisible relationships are relatively independent from each other and seek common needs, and everyone takes what he or she needs. It is necessary to point out that the "weak" (similar to the weak tie in [34]) here is relative to the "strong" of visible relationships. The weak connection is the inherent attribute of invisible relationships.

It is the existence of the weak connection that makes users with common friends have a lower possibility to possess invisible relationships.

3.3.3. Randomness. Users who establish invisible relationships have strong subjective consciousness. They may establish contacts with others according to their needs or interests, or even just whims.

Randomness is a reflection of the subjective consciousness of the user. Randomness is an indispensable attribute of invisible relationships.

3.3.4. Transient. Invisible edges can exist "off and on." When they are used, they appear; otherwise, they are implicit. The edges in link prediction are real or going to exist. This notable feature of invisible edges is significantly different from visible edges.

Crowdfunding is a typical example. There are invisible relationships between the sponsors and participants of crowdfunding. Relationships appear during crowdfunding and disappear after crowdfunding. The relationship between many crowdfunding participants also has this characteristic.

The conjecture of invisible relationships in social networks is a problem between link inference and prediction. Both the invisible and visible relationships are inherent connections between users. Relative to the latter, the former are unstable, so it is necessary to infer their existence. Therefore, they have the attribute of link inference. Users with invisible relationships are more likely to establish direct links. Therefore, they also have the attribute of link prediction, and the problem of establishing links between users with invisible relationships can be viewed as the explicit representation of invisible edges.

Thus, the conjecture of invisible relationships can refer to or use methods and algorithms of link inference and link prediction.

This paper focuses on the common interest type of invisible relationships, and in a related concept, a method to conjecture its existence has been designed.

4. Influence Algorithm

To speculate on the existence of invisible relationships, based on previous link inference and prediction studies, considering the randomness and weak connection between users at the same time, three indices are proposed: the *occasional contact degree*, *interest coincidence degree*, and *user popularity*. Additionally, the influence algorithm is established to conjecture invisible relationships based on a comprehensive index of the three indicators, i.e., the influence factor.

4.1. Occasional Contact Degree. The establishment of invisible relationships has randomness. The *occasional contact degree* is used to measure this randomness.

Definition 3. Occasional contact degree $Connection\,(u, v)$. This refers to the possibility of establishing a random connection between users u and v in a social network. That is, the randomness of friendship between users is measured by $Co\ nnection\,(u, v)$.

Randomness is used to reflect the subjective consciousness of users, which is often represented by random numbers in the algorithm. Therefore, the *occasional contact degree* between users can be generated by random numbers; it can be expressed as formula (4) as follows:

$$\mathrm{Connection}(u, v) = \mathrm{unifrnd}(a, b, \mathrm{row}, \mathrm{col}), \tag{4}$$

where (a, b) is the interval in which values are generated, a is the lower bound of the interval and b is the upper bound, row and col represent the rows and columns of the matrix, respectively. $Connection\,(u, v)$ denotes the possibility of random connections between users, so a initializes 0, and b initializes 1. To ensure the equality of probability between user connections, let $Connection\,(u, v)$ be a matrix obeying a uniform random distribution.

The *occasional contact degree* is the first step to measure the connection between users. It is a reflection of user subjective consciousness. Because of the universality of invisible relationships, the *occasional contact degree* is set to a random uniform matrix in order to avoid big differences.

4.2. Interest Coincidence Degree. The degree to which the user is interested in something can be seen by his understanding of it. And the degree of understanding is reflected in users' descriptions.

As the name suggests, the *interest coincidence degree* measures the similarity between users from the view of interest. The degree user u is interested in something can be measured by $Hobby\,(u)$.

Definition 4. Interest degree $Hobby\,(u)$. This refers to the degree user u is interested in a specific thing. It can be measured by the ratio of the total views of all users describing the thing divided by the views of user u who described it. The specific expression is shown in equation (5) as follows:

$$\mathrm{Hobby(u)} = \frac{\mathrm{thing}_u}{\sum_{v \in T} \mathrm{thing}_v}. \tag{5}$$

In equation (5), thing_u represents the view describing the thing user u is interested in, and T is the set of views described by all users interested in the same. Similar to the blind men and the elephant, thing_u is the specific part of the elephant that one blind man touched, for example, ears, whereas $\sum_{v \in T} \mathrm{thing}_v \sum_{v \in T} \mathrm{thing}_v$ refers to the parts all blind men have touched, including the ears, tail, and nose.

In fact, thing_u can be seen as the point the user is interested in. Thus the *interest coincidence degree* is defined as the following:

Definition 5. Interest coincidence degree $Interest\,(u, v)$. This refers to the product of the interest points overlap between different users for the same thing and the attention they paid to it. If there is more than one interesting thing, they must be added up. The *interest coincidence degree* between any two users can be expressed by formula (6) as follows:

$$\mathrm{Interest(u, v)} = \sum_{i}^{C} w_{u_{c_i}} w_{v_{c_i}} \frac{\left|\mathrm{Hobby}_{C_i}(u) \cap \mathrm{Hobby}_{C_i}(v)\right|}{|C_i|}, \tag{6}$$

where $|\text{Hobby}_{C_i}(u) \cap \text{Hobby}_{C_i}(v)|/|C_i|$ represents the proportion that the overlap number of interest points between user u and user v takes in all users' interest points and denotes the attention user u paid to thing C_i. It will be defined in the following part. Obviously, $Interest\,(u, v) \in (0, 1)$.

According to formula (6), it is easy to see that when users u and v are entirely in different interest groups, the *interest coincidence degree* between them is 0, in line with the actual situation.

The *interest coincidence degree* is the second step to conjecture invisible relationships between users. It can be used to measure the possibility of establishing a dialogue between different users. The smaller the degree, the less possibility of common interests. The *interest coincidence degree* is an important embodiment of the common interest type invisible relationship.

4.3. User Popularity. Whether a person is popular or not can be reflected in the comments of the people around him. *User popularity* is used to analyze the influence between users. That is, the degree of other users' acquaintance and comments of the target user.

Definition 6. User popularity *Popular*. This refers to the degree user v has impact on user u or the degree user v is familiar with user u. In addition, the *Popular* of user u is shown in formula (7) as follows:

$$\text{Popular}_u = \deg_u, \tag{7}$$

where Popular_u is the *Popular* of user u, $\deg_u$ means the number of u's degree. *User popularity* can be obtained through the number of common friends between user u and user v divided by the *Popular* of user u. Thus, the *Popularity* that user u is popular to user v can be defined as formula (8) as follows:

$$\text{Popularity}(u, v) = \frac{|F(u) \cap F(v)|}{\text{Popular}_u}. \tag{8}$$

In formula (8), $|F(u) \cap F(v)|$ means the number of common friends between users u and v.

User popularity is an important way to understand target users by measuring the impact of common friends on them.

Even in the same group, it has different influence on different users, so it is necessary to determine the specific target user when calculating *user popularity*.

4.4. Influence Algorithm. There is a certain logical progressiveness between the occasional contact degree, interest coincidence degree, and user popularity.

For users A and B, the occasional contact degree is used when user A needs to know the existence of user B and want to know B. Then, the interest coincidence degree is used if user A needs to know some information about user B to judge whether they have something in common. Next, user popularity is used to know the comments about the target user, which are made by the common friends of users A and B.

As known to all, link prediction is a complex process, it needs the acknowledgement of users, and there must be something in common between the users. Therefore, we set an index denoted by influence as a balance of the three indices to measure invisible relationships.

Definition 7. Influence factor $Influence\,(u, v)$. This refers to the influential factor for establishing a connection between users, considering the randomness, homogeneity, and popularity of establishing links. It is a comprehensive index used to measure the invisible relationship between users, and it is the compromise of the occasional contact degree, interest coincidence degree, and user popularity, which is presented as formula (9) as follows:

$$\text{Influence}(u, v) = \frac{\text{Connection}(u, v) + \text{Interest}(u, v) + \text{Popularity}(u, v)}{\alpha}. \tag{9}$$

In formula (9), the parameter α is set to 3 based on the study of Dunbar [35] and the six-degree segmentation theory. Dunbar found that a person's core circle may have three or five people; they are the person's closest friends. Next, there are 12 to 15 people, whose death could bring heavy hurt to the person. Then, there are 50 people. The number increases by a multiplier of approximately 3, and the number of friends for one person is no more than 150. The six-degree segmentation theory points out that a stranger can be recognized by five people at most. Therefore, if the number 6 is regarded as the average number of core friends belonging to one person, the cube of 6 is just more than 150. Therefore, it is reasonable to set α to 3.

As the invisible relationship is uncertain, its existence can be speculated by the possibility, noted as P and calculated by formula (10):

$$P(\text{u}, \text{v}) = \text{Influence}(u, v). \tag{10}$$

The threshold Θ is set for determining the existence of invisible relationship. For convenience, Θ is defined referring to the value of α. The value of Θ is close to $1/\alpha$ to make difference, and it is fixed so that it can be convenient to observe the changes of other indices. When $P(u, v) > \Theta$, the existence of invisible relationship can be considered. On the other hand, the existence of invisible relationship is regarded as a small probability event, which can be ignored according to the feature of small probability event.

To conjecture the invisible relationship, the influence algorithm has been established based on the influence factor.

However, before using the influence algorithm, it needs user information preprocessing. The user information can be divided into user nodes and interest attributes. Then, the two parts need processing, respectively.

The process for user nodes is as follows:

(1) Generate the matrix of occasional contact degree according to the number of user nodes and keep the correspondence between the rows and columns of the matrix and the user nodes

(2) Generate an adjacency matrix based on user node pairs, and determine common friends between users.

The process of user interest attributes is as follows:

(1) Users are divided into different interest groups according to interest attributes that are digitalized

(2) Calculate the interest coincidence degree between users.

Therefore, the main steps in the influence algorithm are described as follows:

(1) Generate the matrix connection on base of occasional contact degree

(2) Deal with the attribute of users and then produce the matrix of interest coincidence degree, *Interest*

(3) Analyze the number of common friends and get the user popularity

(4) Generate the matrix of Influence based on the influence factor arise from ①②③

(5) Compare the element in Influence with Θ and then generate the matrix of invisible relationship.

The specific process is shown in Figure 4.

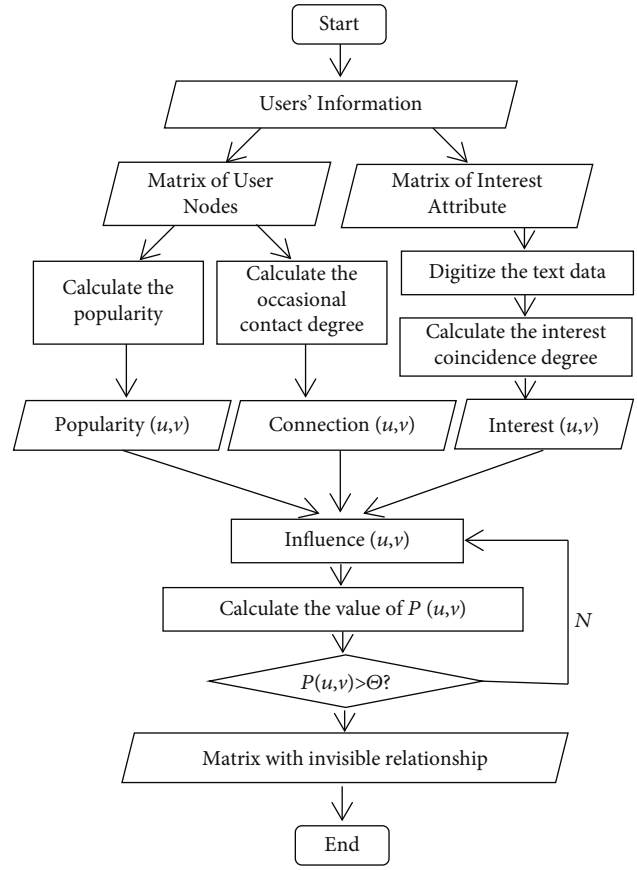

Figure 4: Influence algorithm.

5. Design and Analysis of Experiment

To verify the feasibility and accuracy of the influence algorithm, four indicators are used. They are *the number of user friends*, *the rate of determination*, *false positives*, and *false negatives* (refer to Definitions 10 and 11). In addition, because CN has higher robustness and stability than other algorithms, taking CN as the comparison method, the experiments were designed and implemented. The experimental data are a collection referred to in [36] named the Hamsterster friendships. They are undirected, acyclic and unweighted, and they denote friendships between users on the web named http://hamsterster.com. The average degree is nearly 13.492. The experiments were implemented on this data set. The experimental environment included an Intel (R) Core (TM) i5-4570CPU@3.20 GHz (3.20 GHz); 8.00 GB (1600 MHz) memory; Lenovo Lenovo-SSD-ST600-240G (TOSHIBA DT01ACA100 1 T); and Microsoft Windows 10 version, 64-bit operating system. The algorithms were implemented with MATLAB R2015a. Meanwhile, Excel 2013 was used to select and digitize text. Then, the feasibility and accuracy of the influence algorithm were verified from *the number of user friends*, *the rate of determinatio*n, *false positive*s, and *false negatives*. For measuring the rate of determination, since the experimental data are static, subjective logic is introduced to simulate the dynamic changes of relationships in the network. The data were marked and counted to get the data with identification. Finally, the results were calculated using the formula of subjective logic.

5.1. The Particularity of the Experimental Method. In previous experiments, data sets were usually divided into training set and test set. The training set was to find rules and the test set carried out experiments obeying the rules. This experiment was different from before. Subjective logic was introduced to mark data for simulating the dynamic changes of relationships. Then, the influence algorithm was used to predict invisible relationships, and the performance of the influence algorithm was analyzed from *the number of user friends*, *the rate of determination*, *false positives*, and *false negatives*. The user interest attributes are one important component in the influence algorithm. Interest attributes are usually text data, and they are hard to divide into two parts as before (the previous data set are just digitized data). In addition, it is hard to represent the rules obtained from text data in digital form. Therefore, the design of the experiment is reasonable.

5.2. Subjective Logic of Jøsang. Referring to the book [37], relevant knowledge about subjective logic was introduced. The subjective logic put forward by Jøsang [38] is used for expressing subjective uncertainty, and it has achieved fruitful results.

Subjective logic is based on the distribution of Beta describing the posterior probability of binomial events. A positive event number r and a negative event number s of the observed events are given to calculate the probability deterministic density function. On the basis of the density function, the credibility of each event produced by entities is calculated. Subjective logic can be more practical for modeling and analyzing the real world than traditional probability calculus and probability logic. When subjective logic is used in decision support, it enables decision makers to better assess the impact of uncertainty on future outcomes and make improvements in a timely manner. Since the data collected in experiment are static, they cannot analyze the dynamic changes of the relationship.

Therefore, subjective logic was used to simulate the dynamic changes of network relationships. To simplify the calculation, the simplest subjective logic is used and calculated by equation (11):

$$\text{Exp} = \text{lm} + \mu = \frac{N_1 + 1}{N_1 + N_2 + 2}, \tag{11}$$

where l is a priori probability and set to 0.5, μ represents the value of probability believed, m represents the value of probability of unbelief, N_1 denotes the number of users that marked with relationship, and N_2 denotes the number of users that marked without relationship. Here, $m = N_1/N_1 + N_2 + 2$ and $\mu = N_2/N_1 + N_2 + 2$.

In the experiment, the size of sample was 1800, and the limit of Exp can be determined by formula (11). That is, $\text{limit} = (N_1 + 1)/1800$. To make the experimental results more convincing, the number of friends predicted was approximately equal to the number N_1 on the condition that the number of Exp is approaching the limit, calculated as formula (12):

$$\text{limit} = \frac{N_1 + 1}{1800} = \frac{1}{N}, \tag{12}$$

where N denotes the number of users that can be conjectured by the influence algorithm. Calculated by formula (12), the values of N and N_1 are approximately 44. It was known that the average degree of the node was approximately 13.49, and N and N_1 could be reduced to 13 in the same proportion. However, due to the existence of randomness, the numerical value cannot be guaranteed to be 13 exactly. Therefore, the numerical value of N and N_1 was controlled between 10 and 20.

5.3. Data Processing. It is valuable to collect the user's interest attribute data since it can help conjecture invisible relationships. Since the user interest attribute data are text, they need to be preprocessed—text data can be well processed with the help of Natural Language Processing (NLP). In NLP and text analysis problems, Bag of Words (BOW) and Word Embedding are two commonly used algorithms. Word vector can only represent single words. It needs to do some extra processing to deal with text. The BOW is used to process text based on word frequency, ignoring word order and syntax, which are necessary in the experiments. Kim-Kwang and Raymond Choo et al. [39] proposed unsupervised and supervised approaches on various datasets and conducted experiments on tweets using their methods and achieved higher accuracy. Every attribute view has its own importance for one person, and the other calculation is related to the digitation of text. To ensure the accuracy and reliability of the following experiments, the text was manually processed using Excel 2013 to digitize text.

Table 1: AUC of different methods on four datasets.

	CN	Jaccard	AA	RA
USAir	0.9496	0.9104	0.9645	0.9749
Router	0.667	0.6676	0.6604	0.6691
Yeast	0.9348	0.9295	0.9313	0.9233
Hamsterster friendships	0.8214	0.8103	0.8228	0.8236

According to user interest attributes, on the basis of the principle that an able man is always busy and energy is limited, the specific interest of the user with multiple interests was assigned as follows:

Suppose the total category of user interests is C, then the value of loc assigned to these interests followed by C, $C - 1 \cdots 1$, along the positive X axis direction. Therefore, the attention w of the user for the specific interest C_i is calculated by formula (13) as follows:

$$w_{C_i} = \frac{\text{loc}_{C_i}}{\sum_j^C \text{loc}_{C_i}}. \tag{13}$$

After the text digitation, the concrete experiment operation is carried out according to the method given in the fourth part of this article. 2000 nodes were selected as samples for experiment; however, the results visualization was so intensive that it was difficult to observe the effect of the experiments. Therefore, the visualization of data was divided into groups with 50 nodes in a group. In addition, comparing with CN can highlight the reliability of the influence algorithm.

5.4. Preexperiment. Four methods were tested on network datasets, and besides the Hamsterster friendships, datasets of USAir, Router, and Yeast (http://www.linkprediction.org/index.php/link/resource/data) were included. We calculate the AUC accuracy on four datasets. The average results for ten experiments are listed in Table 1.

As shown in Table 1, on the USAir and Hamsterster datasets, the AUC of CN is larger than that of Jaccard. On the Router dataset, the AUC of CN is better than that of AA. On the Yeast dataset, the AUC of CN is the best among four methods. Therefore, we can conclude that CN can work well on four datasets.

5.5. Analysis of Experimental Results. This section is mainly to analyze the statistical results of the experiments. By comparing with CN, the feasibility and accuracy of the influence algorithm were analyzed from *the number of user friends*, *the rate of determination*, *false positives*, and *false negatives*.

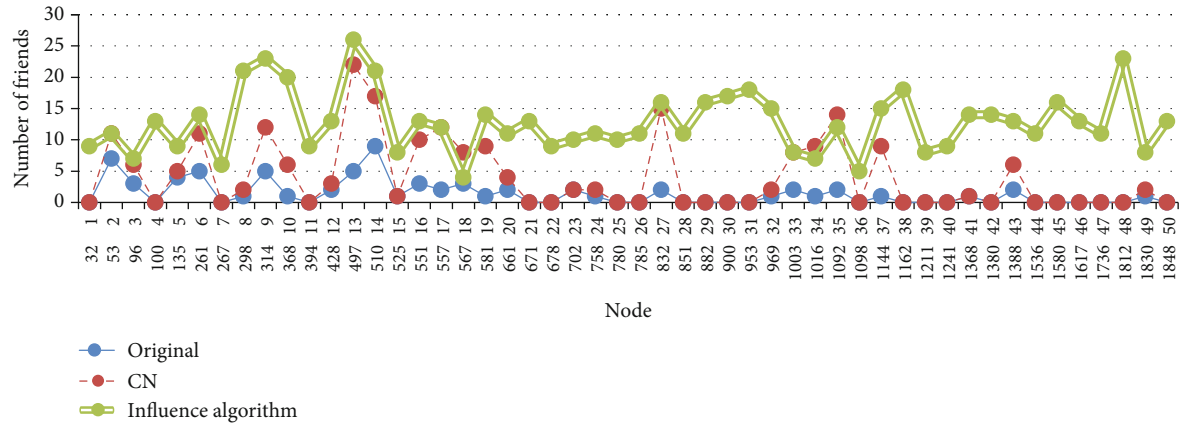

Figure 5: Comparison of user common friends between original, CN, and influence algorithm. Note: there are two rows on the horizontal axis. The first row is the ordinal number of nodes, and the second row is the user's identifier corresponding to node.

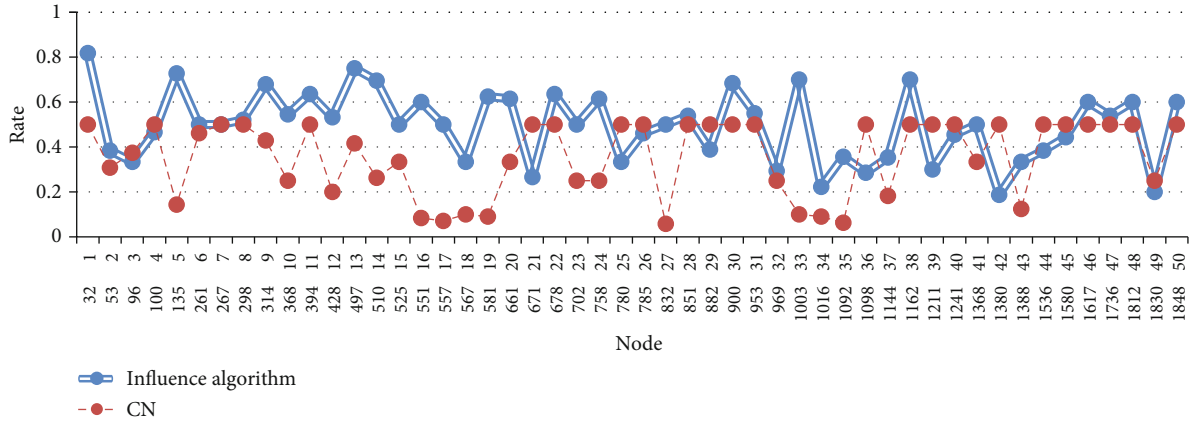

Figure 6: Comparison of determination rate between CN and influence algorithm.

5.5.1. The Number of User Friends. We analyze the number of friends discovered by three methods, i.e., original, CN, and the influence algorithm proposed in this paper. It can be seen from Figure 5 that the tendency predicted by the algorithms for the number of user friends is similar. CN predicts links according to the rule that friends have friends in common, so the tendency of growth is consistent with the tendency of the initial number. The consistency of the influence algorithm verified its feasibility. Moreover, the number of friends predicted by the influence algorithm is higher than CN in most cases, and the stability of the influence algorithm is better. The reason for different stability is that the result of CN depends more on the initial number because it has to use original users to find other friends. The number of original users is the base of the number that CN can predict.

5.5.2. The Rate of Determination

Definition 8. Determination number *Dm*. This refers to the total number of users with determined relationships. That is, the users were marked friends or nonfriends.

Definition 9. The rate of determination. This refers to the determined degree of the relationship predicted by the algorithms, and it can be obtained by the number of friends predicted with determined relationships divided by the determination number.

The experiments were performed using the sample, and 50 nodes were randomly selected to visualization. The results are shown in Figure 6. Marking nodes according to subjective logic, in influence matrix, the nodes with values greater than threshold Θ are marked with determined relationship, and the nodes with values lower than the threshold Δ (small probability event) are marked with determined nonrelationship. When calculating the determination rate of influence algorithm according to formula (11), N_1 represents the proportion that the number of users with determined friendship takes in the number of users that can be predicted by the algorithm. And the "determined" means the node marked by subjective logic. Similarly, N_2 denotes the proportion that the number of users with determined nonfriendship takes in the number of users that can be conjectured by the algorithm.

We can know from "Subjective Logic of Jøsang" that the determination rate can be considered equal in extreme case. Under the same limitation, the determination rate of the influence algorithm is higher than CN; here are three cases:

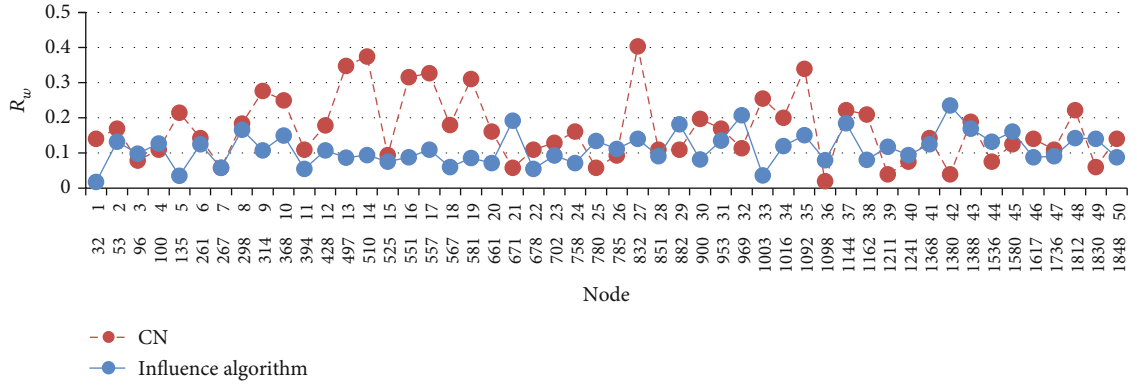

Figure 7: Comparison of false positive between CN and influence algorithm.

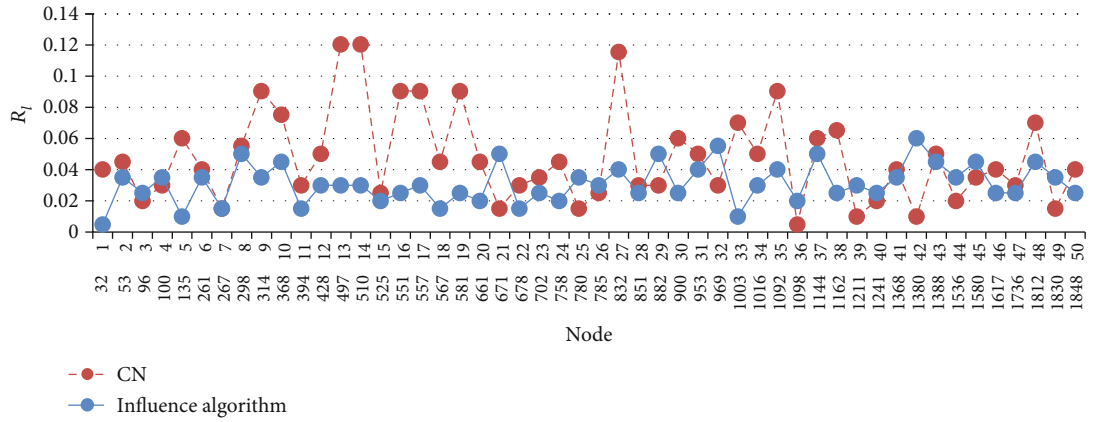

Figure 8: Comparison of false negative between CN and influence algorithm.

(1) When the denominators are equal, the numerator N_1 of the influence algorithm is greater than that of CN

(2) When the numerators are same, the denominator N_2 of the influence algorithm is lower than that of CN

(3) When the denominators and numerators are not equal, the numerator of the influence algorithm is relatively greater and the denominator is relatively lower.

We can conjecture from formula (11) that the above three cases can make the determination rate of the influence algorithm greater than that of CN. In other words, the influence algorithm possesses a higher accuracy rate and lower error rate. N_1 represents the number of users with determined friendship, the greater N_1 is, and the more accurate the result is. N_2 denotes the number of users with determined nonfriendships. With a certain maximum, the greater N_2 is, the lower N_1 is, and so the determination rate is relatively lower. Therefore, we can conclude that the influence algorithm is more reliable than CN.

5.5.3. False Positive and False Negative

Definition 10. False positive *Rw*. This refers to the proportion of users with relationships but judged nonrelationships in *Dm*; it is denoted by *Rw*. It is calculated by equation (14) as follows:

$$R_w = \frac{tf + \mathrm{ft}}{Dm}, \tag{14}$$

where tf represents the number of users predicted without relationships but marked with relationships, and ft denotes the number of users predicted with relationships but marked nonrelationships.

Definition 11. False negative *Rl*. This refers to the proportion of users with determined relationships but not predicted in *Dm*; it is denoted by *Rl*. It can be formulated as (15) as follows:

$$R_l = \frac{tm + ff}{Dm}, \tag{15}$$

where tm represents the unpredicted number of users with relationships and marked with relationships, and ff denotes the unpredicted number of users without relationships and marked nonrelationships.

False positives and false negatives are commonly used indicators to measure method accuracy. From Definitions 10

and 11, it is clear that there are two parts of the numerators. Whether it is the false positive or false negative, the denominators in formulas are both determination number *Dm*. Then, the larger the numerator is, the greater the result is. The experimental results of CN and influence algorithm are shown in Figures 7 and 8. As shown in Figures 7 and 8, in most cases, the results of CN are higher than the influence algorithm results. This illustrates the sum of unpredicted users measured by CN exceeding the results of the influence algorithm (the sum is obtained by adding the number of users with relationships marked with relationship to the number of users without relationships whereas marked nonrelationships). In addition, the number of user errors judged by CN is also higher than for the influence algorithm. Therefore, the accuracy of CN is lower than the influence algorithm. That is, the influence algorithm is more reliable than CN.

6. Conclusion

Link prediction is an important research field in social networks. The invisible relationships proposed in this paper will make the links in social networks more detailed and enriched. At the same time, the proposal of invisible relationships puts forward a new possibility for research of interpersonal relationships, i.e., there may be relationships between people seemingly without connection. To analyze invisible relationships between users in social networks, the definition, types, and characteristics of invisible relationship are introduced. In addition, an influence algorithm is proposed to predict the invisible relationship between users, which is based on occasional contact degree, interest coincidence degree, and the popularity of users. The experimental results on the Hamsterster friendships dataset show that the proposed influence algorithm is effective in predicting invisible relationship and outperforms the CN baseline.

As invisible relationship is a very important relationship which cannot be ignored in social networks, the prediction of invisible relationship is worthy of further researching and extending. In our future work, we will consider node attributes and topology of social networks and propose approaches to predict invisible relationship across multiple social networks.

Data Availability

http://konect.uni-koblenz.de/networks/petster-friendships-hamster

Additional Points

Statement. This paper is the extended version of the manuscript published in 2018 13th Asia Joint Conference on Information Security (AsiaJCIS). There are some differences between them: firstly, this paper added some related work and references; secondly, this paper specified the process of the algorithm; and third, this paper added preexperiment.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is partially supported by the Natural Science Foundation of Hebei Province, China (Nos. F2016201244 and F2020201023), and the Social Science Foundation of Hebei Province, China (HB18SH002).

References

[1] A. I. Naimi and D. J. Westreich, "Big data: a revolution that will transform how we live, work, and think," *Mathematics & Computer Education*, vol. 47, no. 17, pp. 181–183, 2014.

[2] N. N. Daud, S. H. A. Hamid, M. Saadoon, F. Sahran, and N. B. Anuar, "Applications of link prediction in social networks: a review," *Journal of Network and Computer Applications*, vol. 166, article 102716, 2020.

[3] V. Martínez, F. Berzal, and J.-C. Cubero, "A survey of link prediction in complex networks," *ACM Computing Surveys*, vol. 49, no. 4, pp. 1–33, 2017.

[4] E. Bütün, M. Kaya, and R. Alhajj, "Extension of neighbor-based link prediction methods for directed, weighted and temporal social networks," *Information Sciences*, vol. 463-464, pp. 152–165, 2018.

[5] K. Chi, G. Yin, Y. Dong, and H. Dong, "Link prediction in dynamic networks based on the attraction force between nodes," *Knowledge-Based Systems*, vol. 181, article 104792, 2019.

[6] K. Li, L. Zhang, and H. Huang, "Social influence analysis: models, methods, and evaluation," *Engineering*, vol. 4, no. 1, pp. 40–46, 2018.

[7] Y. Yang, Z. Wang, and T. Jin, "Tracking top-k influential users with relative errors," in *CIKM '19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 1783–1792, New York, NY, USA, November 2019.

[8] H. Chintakunta and A. Gentimis, "Influence of topology in information flow in social networks," in *2016 50th Asilomar Conference on Signals, Systems and Computers*, pp. 67–71, Pacific Grove, CA, USA, November 2016.

[9] K. Subbian, D. Sharma, Z. Wen, and J. Srivastava, "Finding influencers in networks using social capital," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13*, pp. 592–599, New York, NY, USA, August 2013.

[10] G. Liu, F. Zhu, K. Zheng et al., "TOSI: a trust-oriented social influence evaluation method in contextual social networks," *Neurocomputing*, vol. 210, pp. 130–140, 2016.

[11] X. Deng, Y. Pan, Y. Wu, and J. Gui, "Credit distribution and influence maximization in online social networks using node features," in *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp. 93–100, Zhangjiajie, China, August 2015.

[12] C. Wang, X. Guan, T. Qin, and Y. Zhou, "Algorithming on opinion leader's influence in microblog message propagation and its application," *Journal of Software*, vol. 26, no. 6, pp. 1473–1485, 2015.

[13] J. X. Cao, G. J. Chen, J. L. Wu, B. Liu, T. Zhou, and S. Xu, "Multi-feature based opinion leader mining in social networks," *Acta Electronica Sinica*, vol. 44, no. 4, pp. 898–905, 2016.

[14] R. R. Sarukkai, "Ramesh, Link prediction and path analysis using Markov chains," *Computer Networks*, vol. 33, no. 1-6, pp. 377–386, 2000.

[15] J. Zhu, J. Hong, and J. G. Hughes, *Using Markov Chains for Link Prediction in Adaptive Web Sites*, Springer, Berlin Heidelberg, 2002.

[16] A. Popescul and L. Ungar, "Statistical relational learning for link prediction," in *Proceeding of the Workshop on Learning Statistical Algorithms from Relational Data*, p. 81, New York, NY, USA, 2003.

[17] P. Jaccard, "Etude comparative de la distribution florale dans une portion des alpes et des Jura," *Bulletin de la Societe Vaudoise des Sciences Naturelles*, vol. 37, no. 142, pp. 547–579, 1901.

[18] K.-k. Shang, T.-c. Li, M. Small, D. Burton, and Y. Wang, "Link prediction for tree-like networks," *Chaos*, vol. 29, article 061103, pp. 1–10, 2019.

[19] L. A. Adamic and E. Adar, "Friends and neighbors on the web," *Social Networks*, vol. 25, no. 3, pp. 211–230, 2003.

[20] D. Liben-Nowell and J. Kleinberg, "The link-prediction problem for social networks," *Journal of the American Society for Information Science and Technology*, vol. 58, no. 7, pp. 1019–1031, 2007.

[21] T. Zhou, L. Lü, and Y. C. Zhang, "Predicting missing links via local information," *European Physical Journal B*, vol. 71, no. 4, pp. 623–630, 2009.

[22] M. Xu and Y. Yin, "A similarity index algorithm for link prediction," in *2017 12th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pp. 1–6, Nanjing, China, November 2017.

[23] J. Wang, Y. Ma, M. Liu, H. Yuan, W. Shen, and L. Li, "A vertex similarity index using community information to improve link prediction accuracy," in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 158–163, Banff, AB, Canada, October 2017.

[24] Q. Sun, R. Hu, Z. Yang, Y. Yao, and F. Yang, "An improved link prediction algorithm based on degrees and similarities of nodes," in *2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pp. 13–18, Wuhan, China, May 2017.

[25] C. P. Muniz, R. Goldschmidt, and R. Choren, "Combining contextual, temporal and topological information for unsupervised link prediction in social networks," *Knowledge-Based Systems*, vol. 156, pp. 129–137, 2018.

[26] X. Xu, N. Hu, T. Li et al., "Distributed temporal link prediction algorithm based on label propagation," *Future Generation Computer Systems*, vol. 93, pp. 627–636, 2019.

[27] S. Das and S. K. Das, "A probabilistic link prediction model in time-varying social networks," in *2017 IEEE International Conference on Communications (ICC)*, pp. 1–6, Paris, France, May 2017.

[28] Z. Wang, J. Liang, and R. Li, "A fusion probability matrix factorization framework for link prediction," *Knowledge-Based Systems*, vol. 159, pp. 72–85, 2018.

[29] K.-k. Shang, W.-s. Yan, and M. Small, "Evolving networks-using past structure to predict the future," *Physica A*, vol. 455, pp. 120–135, 2016.

[30] K.-k. Shang, M. Small, X.-k. Xu, and W.-s. Yan, "The role of direct links for link prediction in evolving networks," *EPL*, vol. 117, no. 1-8, article 28002, 2017.

[31] S. Rafiee, C. Salavati, and A. Abdollahpouri, "CNDP: Link prediction based on common neighbors degree penalization," *Physica A: Statistical Mechanics and its Applications*, vol. 539, article 122950, pp. 1–12, 2020.

[32] L. Zhang, M. Zhao, and D. Zhao, "Bipartite graph link prediction method with homogeneous nodes similarity for music recommendation," *Multimedia Tools and Applications*, vol. 79, no. 19-20, pp. 13197–13215, 2020.

[33] M. E. J. Newman, "Models of the small world," *Journal of Statistical Physics*, vol. 101, no. 3/4, pp. 819–841, 2000.

[34] M. E. Brashears and E. Quintane, "The weakness of tie strength," *Social Networks*, vol. 55, pp. 104–115, 2018.

[35] R. Dunbar, *How Many Friends does one Person Need?*, CITIC Publishing House, 1st edition, 2011.

[36] J. Kunegis, "KONECT – the Koblenz network collection," in *WWW '13 Companion: Proceedings of the 22nd International Conference on World Wide Web*, pp. 1343–1350, New York, NY, USA, May 2013.

[37] J. Tian, H. Jiao, and R. Du, *Subjective logic and its application*, vol. 9, Science Press, Beijing, 2015.

[38] A. Jøsang, "A logic for uncertain probabilities," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, no. 3, pp. 1–31, 2001.

[39] J. K. Rout, K. K. R. Choo, A. K. Dash, S. Bakshi, S. K. Jena, and K. L. Williams, "A model for sentiment and emotion analysis of unstructured social media text," *Electronic Commerce Research*, vol. 18, no. 2, pp. 181–199, 2018.