



Research Article

Measure User Intimacy by Mining Maximum Information Transmission Paths

Lin Guo¹ and Dongliang Zhang²

¹School of Economics and Management, Changchun University of Science and Technology, Changchun, Jilin 130022, China

²Institution of Technical Science, Fudan University, Shanghai 200000, China

Correspondence should be addressed to Lin Guo; guolin@cust.edu.cn

Received 17 June 2019; Accepted 29 November 2019; Published 21 March 2020

Guest Editor: Thiago Christiano Silva

Copyright © 2020 Lin Guo and Dongliang Zhang. 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.

The Internet has become an important carrier of information. Its data contain abundant information about hot events, user relations and attitudes, and so on. Many enterprises use high-impact Internet users to promote products, so it is very important to understand the mechanism of information transmission. Mining social network data can help people analyze the complex and changing relationships between users. The traditional method for doing this is to analyze information such as common interests and common friends, but this data cannot truly describe the degree of intimacy between users. What really connects different users on the Internet is the delivery of information. The algorithm proposed in this paper considers the dynamic characteristics of information transmission, finds maximum transmission paths from information transmission results, and finally calculates the intimacy degrees between users according to all the maximum information transmission paths within a certain period.

1. Introduction

Social network data contains a wealth of information about events, relationships, and attitudes. On the basis of fully understanding and analyzing the data, a series of technologies, such as text mining, statistical theory, association analysis, and visualization technologies, are adopted to realize emotional orientation analysis, information extraction, user influence analysis, and so on. Many current methods of computing user intimacy can be applied to static networks. However, users might unfollow certain friends and their interests might shift to new and different topics. In other words, the tie strengths between different users change over time. The algorithm proposed in this paper takes the dynamic nature of data into account to improve information transmission analysis in social networks. After that maximum transmission paths are identified in the information transmission results, and then the intimacy degrees between nodes can be computed according to multiple groups of maximum transmission paths.

The remainder of this paper is organized as follows: Section 2 introduces the related work of this paper. Section 3

proposes the concept of the information transmission matrix. Section 4 introduces the computational process of tie strength. Section 5 states the experimental results. Section 6 introduces the conclusions.

2. Related Works

Many enterprises use influential users to promote new products, but the mechanism of how information spreads through the network still needs to be further studied. It is very important to understand the communication mechanism of information, which can be applied in many fields, such as viral marketing, social behavior prediction, social recommendation, and community detection. These problems attract the attention of researchers from different fields, such as epidemiology, computer science, and sociology, who propose different information diffusion models to describe and simulate the process of information transmission, such as the independent cascade model, linear threshold model, and epidemic model. These models are mainly applied to influence evaluation, influence maximization, and information source detection. Most models recognize that

information is transmitted from a source node set and other nodes can only obtain information from the nodes that neighbor the source node set.

Social networking service providers, such as Twitter and Facebook, have grown rapidly in recent years, with increasing number of users sharing information with their friends. There are more than 2 million active users on Facebook every month from all over the world and about 5 billion new tweets on Twitter every day. Social network analysis can be divided into the following aspects [1, 2]: (a) studying the network structure and trends [3], (b) online learning of complex networks [4], (c) comparing different models, and (d) predicting node status [1, 5]. The focus of social influence study is to investigate neighbors and associations to predict the impact and influence of the occurrence of an action [2, 6].

Researchers have examined information transfers, including the analysis of relationships [7], social action tracking [1], and other types of relationship transfer [8]. The algorithm proposed in this paper constructs a matrix based on the information transmission between users to describe the complex correlation relations. By making certain changes to the matrix, information transmission paths can be identified and the tie strengths between nodes can be calculated. Due to the small computational difficulty involved in constructing a matrix, the algorithm proposed in this paper performs more efficiently than other algorithms.

3. Information Transmission Matrix

A piece of information is very valuable at one time, but after that it may be worthless. From the perspective of information transmission, the degree of interaction between users can be calculated. By analyzing information transmission paths, an information transmission tree can be generated to describe the information transmission rules and be used to analyze the dynamic changes of the correlations between users.

Definition 1. Let \mathbf{G} be a graph with n nodes and e edges. If

$$a_{ij} = \begin{cases} 1, & \text{if } e_j \text{ is associated with } v_i, \\ 0, & \text{if } e_j \text{ is not associated with } v_i, \end{cases} \quad (1)$$

then the $n \times e$ matrix composed of element a_{ij} ($1 \leq i \leq n$, $1 \leq j \leq e$) is constructed. $\mathbf{M} = (a_{ij})_{n \times e}$ is the complete incidence matrix of graph \mathbf{G} , namely, the information transmission matrix.

In this paper, information transmission data is used to construct and update \mathbf{M} . The construction process of \mathbf{M} is given below.

Figure 1 depicts the information transmission relationships between nodes. If there is an edge between nodes, then it means that information has been successfully passed

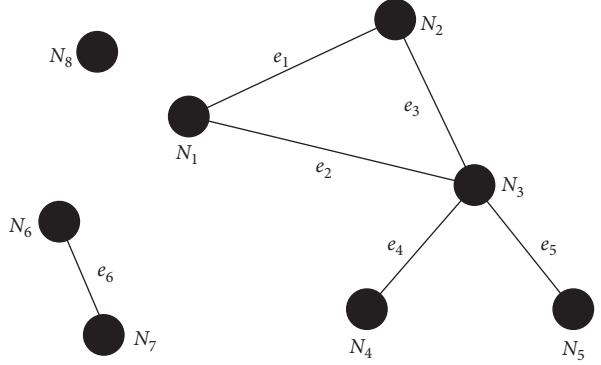


FIGURE 1: Information flow between nodes.

between them. Otherwise, no information has been passed. We construct matrix \mathbf{M} according to Figure 1, which describes the mapping relationship between nodes and edges. If there is an association between N_x and e_y , then $a_{xy} = 1$. Otherwise, $a_{xy} = 0$.

Because there is a large number of inactive nodes, most of the actions of the nodes on the Internet are from browsing information while actions such as commenting and forwarding are rare. Therefore, matrix \mathbf{M} is a sparse matrix. To reduce the negative impact of a large number of meaningless zeros in the matrix on subsequent calculations, further analysis of \mathbf{M} is required to delete redundant nodes. In Section 3.1, we describe a quick and effective way to remove redundant nodes.

3.1. Isolated Nodes

Definition 2. If the determinant of n th order matrix \mathbf{M} is not zero, that is, $|\mathbf{M}| \neq 0$, then \mathbf{M} is called a nonsingular matrix or full rank matrix. Otherwise, \mathbf{M} is called a singular matrix or reduced-rank matrix.

Definition 3. Nodes in graph \mathbf{G} are connected if and only if the rank of the complete incidence matrix is $n - 1$. The matrix whose order is $\min\{p, q\}$ is called a large submatrix of the $p \times q$ matrix.

By calculating whether $|\mathbf{M}|$ is 0, we can judge whether the nodes in \mathbf{G} are connected or not. A reduced matrix \mathbf{D} can be achieved by deleting redundant nodes in \mathbf{M} . \mathbf{D} is a full rank matrix, that is, $|\mathbf{D}| \neq 0$. At this time, \mathbf{D} is the maximum complete incidence matrix. That is, all the nodes in the new graph \mathbf{G} that are formed by \mathbf{D} are reachable, and there are no isolated nodes for information transmission.

Take matrix \mathbf{M} in Figure 2 as an example to illustrate the process of removing isolated nodes. The rank of \mathbf{M} is obtained by calculating the maximum number of linearly-independent crossings (that is, the maximum order of the nonzero submatrix):

$$\begin{aligned}
 \mathbf{M} = & \begin{array}{c} e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6 \\ \left(\begin{array}{cccccc} N_1 & 1 & 1 & 0 & 0 & 0 & 0 \\ N_2 & 1 & 0 & 1 & 0 & 0 & 0 \\ N_3 & 0 & 1 & 1 & 1 & 1 & 0 \\ N_4 & 0 & 0 & 0 & 1 & 0 & 0 \\ N_5 & 0 & 0 & 0 & 0 & 1 & 0 \\ N_6 & 0 & 0 & 0 & 0 & 0 & 1 \\ N_7 & 0 & 0 & 0 & 0 & 0 & 1 \\ N_8 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \end{array} \\
 & \text{addition and subtraction} \\
 & \text{of row vectors} \\
 & = \\
 & \begin{array}{c} e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6 \\ \left(\begin{array}{cccccc} N_1 & 1 & 1 & 0 & 0 & 0 & 0 \\ N_2 & 0 & 1 & 0 & 0 & 0 & 0 \\ N_3 & 0 & 0 & 1 & 0 & 0 & 0 \\ N_4 & 0 & 0 & 0 & 1 & 0 & 0 \\ N_5 & 0 & 0 & 0 & 0 & 1 & 0 \\ N_6 & 0 & 0 & 0 & 0 & 0 & 1 \\ N_7 & 0 & 0 & 0 & 0 & 0 & 0 \\ N_8 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \end{array} \\
 \xrightarrow{\text{remove redundant nodes}} & \begin{array}{c} e_1 \ e_2 \ e_3 \ e_4 \ e_5 \ e_6 \\ \left(\begin{array}{cccccc} N_1 & 1 & 1 & 0 & 0 & 0 & 0 \\ N_2 & 0 & 1 & 0 & 0 & 0 & 0 \\ N_3 & 0 & 0 & 1 & 0 & 0 & 0 \\ N_4 & 0 & 0 & 0 & 1 & 0 & 0 \\ N_5 & 0 & 0 & 0 & 0 & 1 & 0 \end{array} \right) \end{array} \\
 \xrightarrow{\text{remove redundant edges}} & \begin{array}{c} e_1 \ e_2 \ e_3 \ e_4 \ e_5 \\ \left(\begin{array}{ccccc} N_1 & 1 & 1 & 0 & 0 & 0 \\ N_2 & 0 & 1 & 0 & 0 & 0 \\ N_3 & 0 & 0 & 1 & 0 & 0 \\ N_4 & 0 & 0 & 0 & 1 & 0 \\ N_5 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{array} \\
 \end{aligned} \tag{2}$$

According to the abovementioned calculation results, $R(\mathbf{M}) = 6$. This indicates the existence of isolated nodes in \mathbf{M} . It can be seen that rows N_7 and N_8 are $\vec{0}$, so N_7 and N_8 are redundant, isolated nodes. Because the original data in line N_6 and N_7 are same and N_7 was determined to be an isolated node to be deleted, N_6 is also an isolated node. In conclusion, N_6 , N_7 , and N_8 are isolated nodes. After removing redundant

nodes, it is necessary to determine whether there are redundant edges in the matrix. Because column e_6 is $\vec{0}$ after the redundant nodes are deleted, e_6 is a redundant edge that needs to be deleted.

Matrix \mathbf{D} is obtained after deleting the redundant nodes in \mathbf{M} . Next, whether the nodes in \mathbf{D} are connected must be calculated as follows:

$$\begin{aligned}
 \mathbf{D} = & \begin{array}{c} e_1 \ e_2 \ e_3 \ e_4 \ e_5 \\ \left(\begin{array}{ccccc} N_1 & 1 & 1 & 0 & 0 & 0 \\ N_2 & 1 & 0 & 1 & 0 & 0 \\ N_3 & 0 & 1 & 1 & 1 & 1 \\ N_4 & 0 & 0 & 0 & 1 & 0 \\ N_5 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{array} \\
 & \text{addition and subtraction} \\
 & \text{of row vectors} \\
 & = \\
 & \begin{array}{c} e_1 \ e_2 \ e_3 \ e_4 \ e_5 \\ \left(\begin{array}{ccccc} N_1 & 1 & 1 & 0 & 0 & 0 \\ N_2 & 0 & -1 & 1 & 0 & 0 \\ N_3 & 0 & 0 & 2 & 0 & 0 \\ N_4 & 0 & 0 & 0 & 1 & 0 \\ N_5 & 0 & 0 & 0 & 0 & 1 \end{array} \right) \end{array} \\
 \end{aligned} \tag{3}$$

The result is $R(\mathbf{D}) = 5$. That is, $|\mathbf{D}| \neq 0$, so \mathbf{D} is a full rank matrix. The conclusion is that all nodes in \mathbf{D} are connected. In other words, there are no isolated nodes of information transmission.

To discover all information transmission paths in \mathbf{M} , it is necessary to further determine which nodes can be tentatively considered to be redundant. The deleted redundant nodes are reconstituted into a new matrix \mathbf{M} and the abovementioned operations are repeated to obtain a matrix \mathbf{D} . Finally, multiple matrix \mathbf{Ds} are obtained.

3.2. Information Transmission Path. To study the information transmission mechanism, it is necessary to identify all the information transmission paths from the information matrix. Therefore, further processing of the set of \mathbf{Ds} is required.

Definition 4. Submatrix \mathbf{A} is obtained by removing one row from the complete incidence matrix \mathbf{D} . For \mathbf{A} to be nonsingular, the edges that correspond to the columns of \mathbf{A} must form a spanning tree of G .

Definition 4 provides a method for calculating all spanning trees in the connected graph G . By removing one row from matrix \mathbf{D} and then calculating all the maximized nonsingular submatrices of the newly-generated matrix \mathbf{D} , the edges that correspond to the columns of each nonsingular submatrix form a spanning tree of G .

The matrix \mathbf{D} obtained in the previous section is taken as an example to illustrate the process of identifying information transmission paths according to Definition 4. Remove one row from \mathbf{D} (delete row 5 here) to get a matrix \mathbf{A} :

	e_1	e_2	e_3	e_4	e_5	e_6
N_1	1	1	0	0	0	0
N_2	1	0	1	0	0	0
N_3	0	1	1	1	1	0
N_4	0	0	0	1	0	0
N_5	0	0	0	0	1	0
N_6	0	0	0	0	0	1
N_7	0	0	0	0	0	1
N_8	0	0	0	0	0	0

FIGURE 2: Matrix \mathbf{M} .

$$A = \begin{array}{l} \begin{matrix} & e_1 & e_2 & e_3 & e_4 & e_5 \\ N_1 & \left(\begin{matrix} 1 & 1 & 0 & 0 & 0 \end{matrix} \right) \\ N_2 & \left(\begin{matrix} 1 & 0 & 1 & 0 & 0 \end{matrix} \right) \\ N_3 & \left(\begin{matrix} 0 & 1 & 1 & 1 & 1 \end{matrix} \right) \\ N_4 & \left(\begin{matrix} 0 & 0 & 0 & 1 & 0 \end{matrix} \right) \end{matrix} \end{array} \quad (4)$$

By calculating the rank of A , we can get $R(A) = 4$. This value indicates that the nodes in A are connected. Although all nodes are connected to other nodes, there may be redundant edges. For example, the nodes N_1 , N_2 , and N_3 in Figure 1 have three edges, and these three nodes can be completely connected to each other by two of the edges. To remove redundant edges, we apply the following rules to the matrix:

Rule 1: the i th row of the matrix can be added and subtracted to the j th row

Rule 2: repeat the operation of Rule 1 until there is no operable item

Rule 3: the row vectors of the matrix are not interchangeable

Rule 4: the column vectors of the matrix are not interchangeable

Through the transformation of the matrix, the number of 1 in the matrix is reduced and the most concise matrix is finally obtained. At this time, the nodes in the matrix are connected by the minimum number of edges. Take matrix \mathbf{M} in Figure 2 as an example to illustrate the process of removing redundant edges. Matrix \mathbf{D} is obtained by deleting isolated nodes in \mathbf{M} , and matrix \mathbf{A} with 4 rows and 5 columns is obtained after deleting one row from \mathbf{D} . Matrix \mathbf{A} is a full rank matrix. To delete the redundant information transmission path, it is necessary to delete one column from \mathbf{A} to form multiple sets of different column combinations. The different combinations of columns are $\{(e_1, e_2, e_3, e_4); (e_1, e_2, e_3, e_5); (e_1, e_2, e_4, e_5); (e_1, e_3, e_4, e_5); (e_2, e_3, e_4, e_5)\}$. Then, perform row operations on each of the above-mentioned matrices according to the rules.

The first two rows in Table 1 describe the cases in which the constructed matrix does not meet the judgment condition for generating a maximum information transmission path. In the

first combination, edges $(e_1, e_2, e_3, \text{ and } e_4)$ are selected. It is found that in the matrix $M_{N_5, e_1}, M_{N_5, e_2}, M_{N_5, e_3}$, and M_{N_5, e_4} are all 0, so this path does not contain N_5 . That is, it is not a maximum information transmission path, so the combination of $(e_1, e_2, e_3, \text{ and } e_4)$ is deleted and the calculation is stopped. Similarly, in the second combination, $M_{N_4, e_1}, M_{N_4, e_2}, M_{N_4, e_3}$, and M_{N_4, e_4} is 0, so the calculation result obtained by this structure does not include N_4 , that is, it is not a maximum information transmission path. The ranks of the third, fourth, and fifth matrixes are all 4, so they are full rank matrices that satisfy the condition of generating maximum information transmission paths. The fourth column in rows 3, 4, and 5 in Table 1 show the row transformation process. Number 1 is the lowest in the transformed matrix, so the matrix does not have redundant edges. Column 5 shows the graph structure of the matrix obtained after eliminating the redundant edges. It can be seen from the graphs that the method proposed in this paper can be used to identify all maximum information transmission paths.

4. Tie Strength between Nodes

According to the characteristics of information transmission, it is reasonable to assume that there must be some association between the nodes in the same transmission path. Here, it is assumed that if information is transmitted frequently between two nodes, then the degree of intimacy between these two nodes is high. After a period of data accumulation, data about maximum information transmission paths is added to the correlation strength matrix (denoted as T). Because the construction of T is executed according to information transmission flows, matrix T also keeps changing with the change of information transmission state. In matrix T , $T_{i, I}$ represents the occurrence number of node i in the process of information transmission and $T_{i, j}$ represents the information transfer times between nodes i and j .

The following is the formula for calculating the weight of node i :

$$\text{node}_i = \frac{T_{i,i}}{\sum_{x=1}^n T_{x,x}}. \quad (5)$$

The following is the formula for calculating the ties between node a and b :

$$\text{edge}_{a,b} = \frac{T_{a,b}}{\sum_{\substack{x=1 \\ x \neq a}}^n T_{a,x} - \{T_{a,a}\}} \quad (a > 1). \quad (6)$$

According to formulas (5) and (6), the degree of intimacy between different users is calculated. The specific algorithm is shown in Algorithm 1.

5. Experiments

Five datasets are used in this paper. For detailed information about the datasets, please refer to our paper [2] published earlier.

TABLE 1: The process of mining maximum information transmission paths.

	Matrix	Result	Reasons for failure to meet the criteria/matrix transformation process/the structure corresponding to the calculation result
1	(e_1, e_2, e_3, e_4)	Not satisfied	Reason: $M_{N_5, e_1}, M_{N_5, e_2}, M_{N_5, e_3}$, and M_{N_5, e_4} in M are 0; therefore, N_5 is excluded, and it is not a maximum information transmission path
2	(e_1, e_2, e_3, e_5)	Not satisfied	Reason: $M_{N_4, e_1}, M_{N_4, e_2}, M_{N_4, e_3}$, and M_{N_4, e_5} in M are 0; therefore, N_4 is excluded, and it is not a maximum information transmission path
3	(e_1, e_2, e_4, e_5) $R = 4;$ full rank matrix		$\begin{array}{l} \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 \\ N_1 & 1 & 1 & 0 & 0 \\ N_2 & 1 & 0 & 0 & 0 \\ N_3 & 0 & 1 & 1 & 0 \\ N_4 & 0 & 0 & 1 & 0 \end{matrix} & \Rightarrow & \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 \\ N_1 & 0 & 1 & 0 & 0 \\ N_2 & 1 & 0 & 0 & 0 \\ N_3 & 0 & 0 & 0 & 0 \\ N_4 & 0 & 0 & 1 & 0 \end{matrix} \end{array}$
4	(e_1, e_3, e_4, e_5) $R = 4;$ full rank matrix		$\begin{array}{l} \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 \\ N_1 & 1 & 0 & 0 & 0 \\ N_2 & 1 & 1 & 0 & 0 \\ N_3 & 0 & 1 & 1 & 1 \\ N_4 & 0 & 0 & 1 & 0 \end{matrix} & \Rightarrow & \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 \\ N_1 & 1 & 0 & 0 & 0 \\ N_2 & 0 & 1 & 0 & 0 \\ N_3 & 0 & 0 & 0 & 1 \\ N_4 & 0 & 0 & 1 & 0 \end{matrix} \end{array}$
5	(e_2, e_3, e_4, e_5) $R = 4;$ full rank matrix		$\begin{array}{l} \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 \\ N_1 & 1 & 0 & 0 & 0 \\ N_2 & 0 & 1 & 0 & 0 \\ N_3 & 1 & 1 & 1 & 1 \\ N_4 & 0 & 0 & 1 & 0 \end{matrix} & \Rightarrow & \begin{matrix} e_1 & e_2 & e_3 & e_4 & e_5 \\ N_1 & 1 & 0 & 0 & 0 \\ N_2 & 0 & 1 & 0 & 0 \\ N_3 & 0 & 0 & 0 & 1 \\ N_4 & 0 & 0 & 1 & 0 \end{matrix} \end{array}$

INPUT: matrix M
 OUTPUT: intimacy correlation graph
 (1) generate matrix D according to matrix M
 (2) construct matrix A from D
 (3) FOREACH full rank matrix X IN matrix A
 (4) update matrix T according to X
 (5) ENDFOREACH
 (6) calculate *weights* and *ties* in T
 (7) construct a graph describing the degree of intimacy between nodes

ALGORITHM 1: TieCP.

- (1) Coauthor (<https://www.aminer.cn/data>): a dynamic coauthor network from ArnetMiner (<http://www.aminer.cn/>). We collected publications published from 2010 to 2016 by 100,000 authors.
- (2) DBLP (<http://www.vldb.org/dblp/>): the dataset is derived from a snapshot of the bibliography for 10 years, where each vertex represents a scientist and two vertices are connected if they work together on an article.
- (3) Twitter (<https://twitter.com>): we crawled the following links between 19,000,000 users from Twitter at 10 different time stamps from October to December 2017.
- (4) Weibo (<http://code.google.com/p/weibo4j/>): the most popular Chinese microblogging site. The data are crawled from March 8, 2014 when the crash of MH370 happened to April 8, 2014.

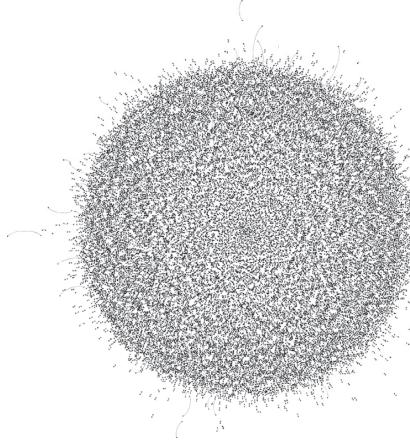


FIGURE 3: Unprocessed graph of Coauthor dataset.

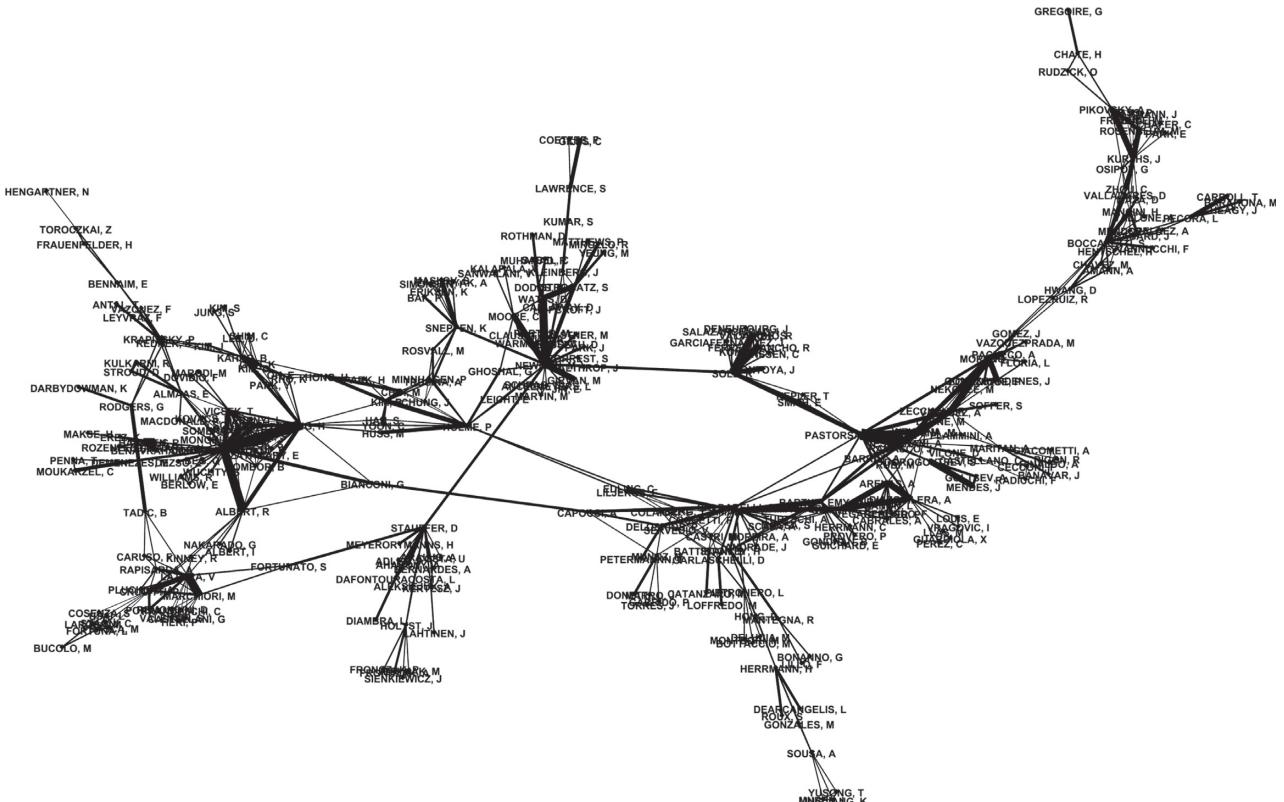


FIGURE 4: The maximum information transmission path in Coauthor dataset.

(5) Dolphin's Associations (<http://www-personal.umich.edu/~mejn/netdata/>): this dataset is an undirected social network of frequent associations between 62 dolphins, which has 62 nodes and 159 edges.

Three sets of baseline approaches are chosen for the experiments:

- (1) PTPMF [9]: this method uses neighborhood overlap to approximate tie strength and extend the popular Bayesian Personalized Ranking (BPR) model to incorporate the distinction of strong and weak ties

- (2) TrustMF [10]: this is a model-based method that adopts matrix factorization technique that maps users into low-dimensional latent feature spaces in terms of their trust relationship and aims to more accurately reflect the users' reciprocal influence on the formation of their own opinions and to learn better preferential patterns of users for high-quality recommendations.
 - (3) SBPR: this method presents a generic optimization criterion BPR-Opt for personalized ranking, that is,

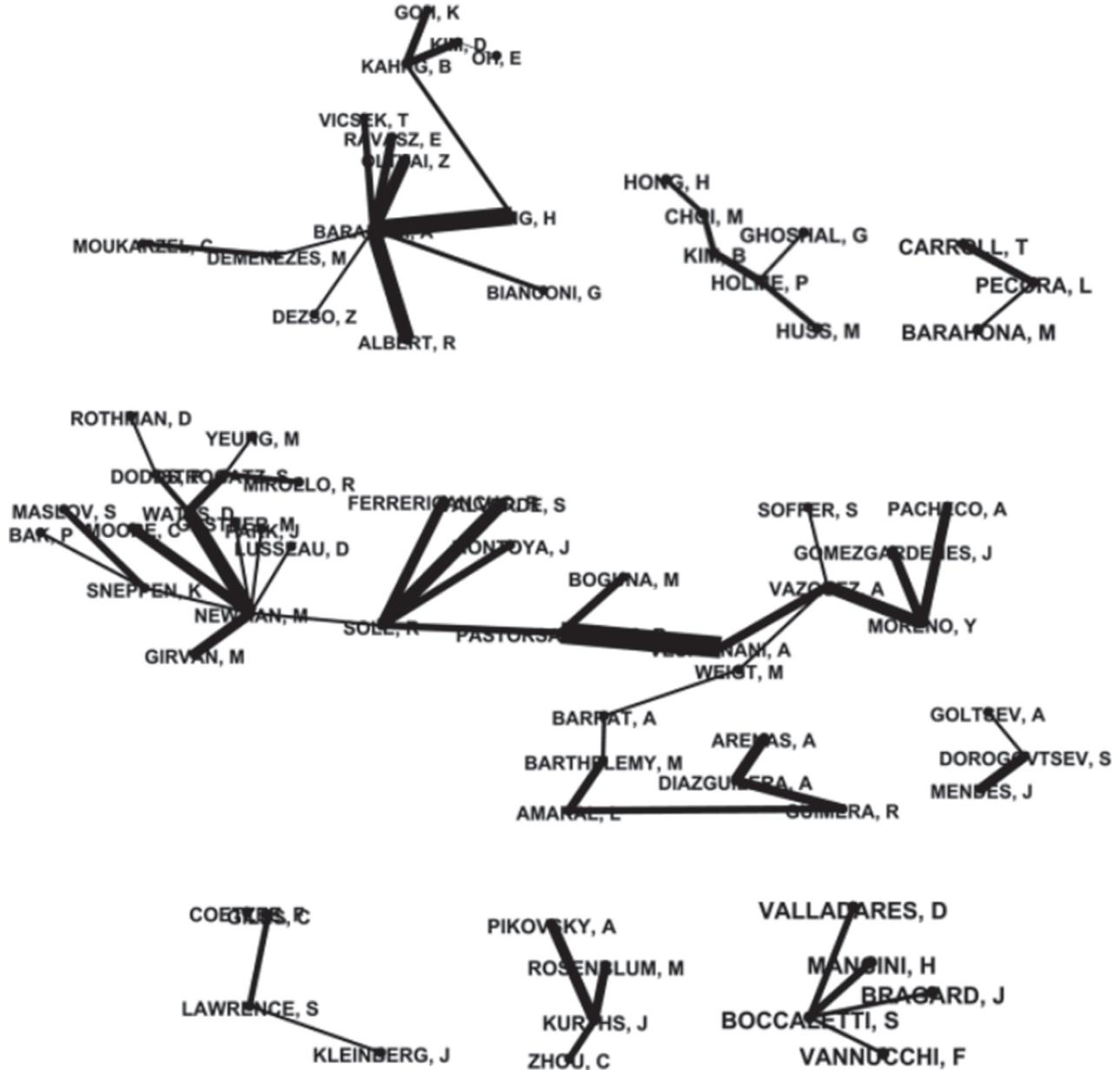


FIGURE 5: The maximum information transmission paths.

the maximum posterior estimator derived from a Bayesian analysis of the problem

Figure 3 shows the information transmission graph without data processing. It contains 38,501 nodes and 20,354 edges. If all nodes in the Coauthor dataset were displayed in Figure 3, then the picture would be black and the structure would not be visible. Therefore, only some of the nodes in the Coauthor dataset are shown in this figure. As can be seen in Figure 3, it is very difficult to process network data.

In the Coauthor dataset, the lengths of most information transmission paths are 2 or 3. Figure 4 shows the path with the maximum length in the Coauthor dataset.

By constructing a matrix according to the structure in Figure 4 and executing the algorithm proposed in this paper

on this matrix, it can be found that several groups of the largest and nonsegmented information transmission paths can be found, as shown in Figure 5. As can be seen from Figure 5, all the paths are loop free and achieve the maximum coverage of all nodes. Therefore, Figure 5 verifies the accuracy of the algorithm from the perspective of visualization.

Figure 6 depicts the degree of all nodes in the maximum information propagation path. It is found that the degree of most nodes is 1, the degree of a few nodes is greater than or equal to 2, and the highest degree value is 13. Figure 6 illustrates that the algorithm achieves the maximum removal of redundant edges.

The tie coefficients between different nodes are calculated according to information transmission paths. Figure 7

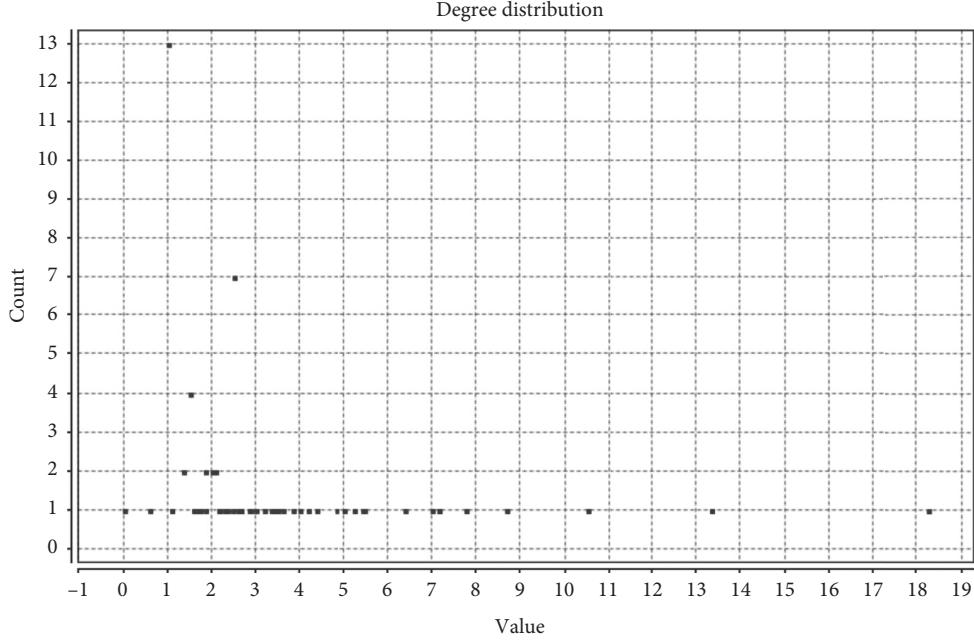


FIGURE 6: The degree of nodes in the maximum information transmission path.

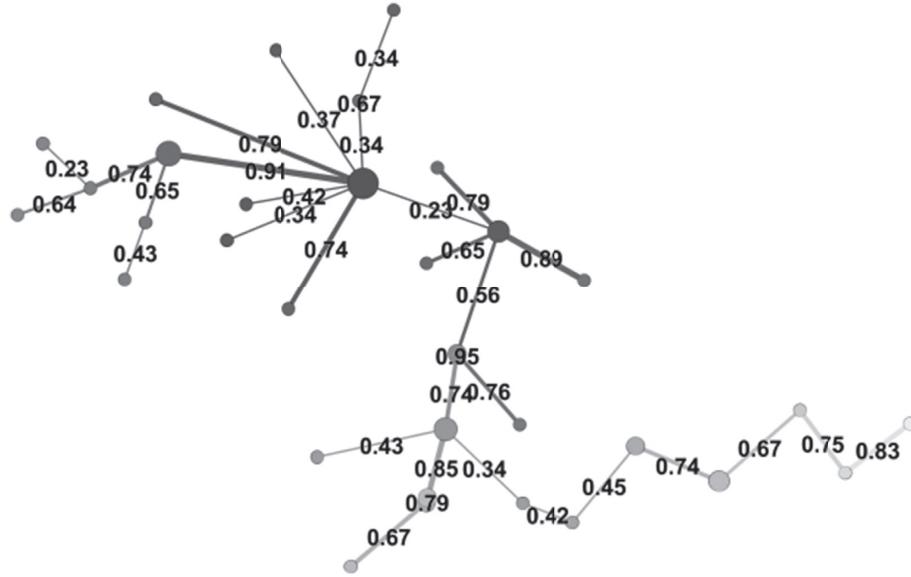


FIGURE 7: Tie coefficient.

shows the tie coefficient of nodes. In it, the darkness of the edges represents the correlation strengths between the node and the ego node. The darker the color is, the stronger the correlation is and vice versa. The number in the edge represents the tie strength between two connected nodes, which is the final result obtained by fusing multiple sets of maximum information transmission paths.

In order to analyze the experimental results, we use the following measurement parameters [10]: Precision calculated by $P = tp/(tp + fp)$, Recall by $R = tp/(tp + fn)$, and F1-score by $F = P \times R \times 2/(P + R)$. tp is the number of

correctly identified examples, tn is the number of correctly identified nonrelated examples, fn is the number of not correctly identified the related examples, and fp is the number of not correctly identified nonrelated examples. Table 2 shows a comparison of the performances of different clustering algorithms on different datasets. It displays performance comparisons of SBPR, TrustMF, PTPMF, and TieCP using different datasets. According to Table 2, we can conclude that TieCP has the most stable execution effect and the best result regarding F-Score.

TABLE 2: The performance comparisons of different algorithms.

Dataset	SBPR			TrustMF			PTPMF			TieCP		
	R	P	F	R	P	F	R	P	F	R	P	F
DBLP	0.73	0.91	0.81	0.84	0.84	0.84	0.88	0.94	0.91	0.89	0.90	0.89
Twitter	0.63	0.98	0.77	0.76	0.87	0.81	0.80	0.94	0.86	0.83	0.92	0.87
Weibo	0.84	0.80	0.82	0.76	0.86	0.81	0.81	0.86	0.83	0.87	0.89	0.88
Coauthor	0.65	0.77	0.70	0.74	0.75	0.74	0.78	0.86	0.79	0.86	0.89	0.87

6. Conclusion

The algorithm proposed in this paper calculates the intimacy degrees between users according to the information transmission matrix. Compared to some mainstream methods, our method is simple and able to identify all the maximum information transmission paths. Beyond that our algorithm is relatively more stable when dealing with different kinds of data. Due to the small computational difficulty of constructing a matrix, the algorithm proposed in this paper performs more efficiently than other algorithms.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by Youth Program of the National Social Science Fund of China (Project name: Research on Online Behavior Pattern of Customers and Multidimensional Customer Insight Method under Big Data; Grant no. 19CGL024).

References

- [1] C. Tan, J. Tang, and J. Sun, “Social action tracking via noise tolerant time-varying factor graphs,” in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD’10*, pp. 1049–1058, ACM, Washington, DC, USA, July 2010.
- [2] L. Guo and B. Zhang, “Mining structural influence to analyze relationships in social network,” *Physica A: Statistical Mechanics and Its Applications*, vol. 523, pp. 301–309, 2019.
- [3] M. Welling and Y. W. Teh, “Bayesian learning via stochastic gradient langevin dynamics,” in *Proceedings of the 28th International Conference on Machine Learning*, pp. 681–688, Bellevue, WA, USA, 2011.
- [4] T. Shi and J. Zhu, “Online Bayesian passive-aggressive learning,” *Journal of Machine Learning Research*, vol. 1, pp. 1–48, 2014.
- [5] L. Guo, W. Zuo, and T. Peng, “Inference network building and movements prediction based on analysis of induced dependencies,” *IET Software*, vol. 11, no. 1, pp. 12–17, 2017.
- [6] Y. Li, Z. Liu, and H. Yu, “Advisor-advisee relationship identification based on maximum entropy model,” *Acta Physica Sinica*, vol. 62, no. 16, 2013.
- [7] J. Chen, Y. Liu, G. Yang, and M. Zou, “Inferring tag co-occurrence relationship across heterogeneous social networks,” *Applied Soft Computing*, vol. 66, pp. 512–524, 2018.
- [8] K. Tago and Q. Jin, “Influence analysis of emotional behaviors and user relationships based on twitter data,” *Tsinghua Science and Technology*, vol. 23, no. 1, pp. 104–113, 2018.
- [9] X. Wang, S. Hoi, M. Ester, J. Bu, and C. Chen, “Learning personalized preference of strong and weak ties for social recommendation,” in *Proceedings of the 26th International Conference on World Wide Web-WWW’17*, ACM, Rio de Janeiro, Brazil, May 2017.
- [10] E. Amigó, J. Gonzalo, J. Artiles, and F. Verdejo, “A comparison of extrinsic clustering evaluation metrics based on formal constraints,” *Information Retrieval*, vol. 12, no. 4, pp. 461–486, 2009.