ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 1 - 2026

# A Study on Content Recommendation and Influence Analysis of Cross-cultural Social Network Communication Based on Matrix Decomposition Methods

*Hongwei Zhang*[1,*]

[1] School of Sports Management and Communication, Capital University of Physical Education and Sports, Beijing, 100191, China

**SUMMARY:** *This paper focuses on the method of content recommendation and influence analysis based on matrix decomposition in cross-cultural social network communication. Through in-depth study of the basic parameters of social network structure and methods such as matrix decomposition, it provides a solid theoretical foundation for the construction of influence calculation and content recommendation model. User rating information and social network structure information are combined to calculate the influence between users, and the influence between users and users' personal influence are combined to calculate the influence between asymmetric users. On the basis of the SoReg method, this paper proposes the SoInf recommendation model based on matrix segmentation and integrating community structure and social influence, so as to utilize the influence information between users and user rating information for recommendation. The results show that the index values of the influence calculation method in this paper show good correlation with the real spreading influence of individual nodes, and the Top-5 nodes calculated by this paper's method have stronger spreading ability compared with the nodes derived by other methods. In addition, the SoInf recommendation model constructed in this paper has higher accuracy and recall as well as lower Mean Absolute Error (MAE) than the benchmark model that does not consider the social network structure.*

***KEYWORDS:*** *social networks; matrix decomposition; SoInf recommendation model; influence calculation*

# 1 Introduction

With the continuous and deep development of globalization process, the economy, politics and culture between the countries in the world are more and more closely linked, the importance of cultural development is also becoming more and more prominent, and culture has become an important factor in the competition of the comprehensive national power of each country [1-3]. In order to enhance the cultural soft power of each country, more and more countries have elevated the construction and development of cultural industry to the height of national strategy in the new era [4, 5]. Accompanied by social networks and other electronic media and the continuous progress of information technology, the time and space distance between people has been further narrowed, and the whole earth seems to be linked together as a "village". The emergence of the global village has also made it possible for people from different countries and regions to spread culture across time and space [6]. Social networks such as Facebook,

*zhanghongwei198004@163.com
https://doi.org/10.65102/is2026032
Published: 30 April, 2026

Instagram, Tumblr, Twitter, TikTok, Google, WeChat, and so on, as a universal platform for social capital, have flourished in the field of culture and other areas beyond the initial expected function of online personal interaction. The initial expected function of social interaction, with the principles of participation, sharing, openness and collective wisdom, will be cultural sustainability, and the promotion and dissemination scope of social networks is more international, so social networks have become the mainstream path of cross-cultural communication [7-10]. However, due to the differences in cultural acceptance, cultural understanding, intercultural competence, semantic expression, etc. of social network users in different cultural backgrounds, resulting in differences in content preferences, dissemination mechanisms, influence, etc., which poses a challenge to the accuracy of cross-cultural communication content recommendation in social networks [11-14]. Based on this, in order to match the characteristics of cross-cultural communication, it is of great significance to study the cross-cultural social network recommendation system and its influence mechanism in depth.

In this paper, the basic structural parameters of social networks, matrix decomposition methods and global influence-related evaluation algorithms are analyzed in depth, and the relevant theoretical foundations of this paper for constructing a communication content recommendation and influence analysis model based on matrix decomposition methods in cross-cultural social networks are fully elaborated. The degree-corrected randomized block model DCBM is used for community discovery of user networks, and each user is divided into each community. The influence between users is calculated by user rating information and social network structure information, and the personal influence of users in social networks is analyzed by social network topology. Combine the influence between users and the personal influence of users to calculate the asymmetric influence between users. Based on the SoRge model, we propose this paper's SoInf recommendation model based on matrix decomposition and integrating community structure and social influence, which utilizes the influence information between users and the user rating information to recommend communicated content in cross-cultural social networks, and enhances this paper's model by using the users' indirect social relationships in the community. Experiments are conducted to test the effectiveness and superiority of the influence calculation method and SoInf recommendation model in the application of this paper, and to demonstrate the application value of this paper's method in improving the accuracy of communication content recommendation in cross-cultural social networks.

# 2 Relevant theoretical foundations

## 2.1 Basic structural parameters of social networks

With the increase of multiple information dissemination channels, information dissemination networks are becoming more and more complex, and the simple network structure cannot satisfy the research of multi-information dissemination, so multi-layer complex social networks are introduced with some characteristics of complex networks, such as small-worldness and scale-free. In order to portray the topology of social networks, the degree can be used to describe the importance of a user in the network, the average path portrays the speed of information transfer between users, and the clustering coefficient is used to describe the characteristics of the social network in which users are grouped together by their class and divided by their group.

Modern research has focused on portraying complex networks from a graph-theoretic perspective. A network $G=(V,E)$ consists of nodes $V=\{v_1,v_2,v_3,\cdots,v_N\}$ and connecting edges $E=\{e_1,e_2,e_2e_3,\cdots,e_M\}$ constitutes the network where the number of nodes is $N$ and the number of edges is $M$. The topology of the network can be represented by the adjacency matrix $A$, and the element $a_{ij}$ in the matrix indicates the connectivity between nodes $i$ and $j$, the number of elements is $N \times N$, and if there is a connecting edge between nodes $i$ and $j$, then $a_{ij} = 1$, otherwise $a_{ij} = 0$.

Figure 1 shows a simple undirected unweighted network graph where $N = 4$ and $M = 5$. Its adjacency matrix $A$ is:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix} \tag{1}$$

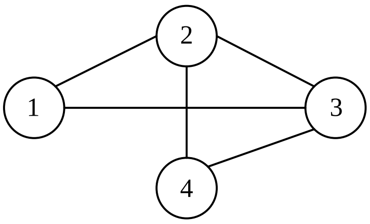



*Figure 1: Simple network diagram*

### 2.1.1 Degree and degree distribution

In a network, the degree of a node is the number of connected edges of the node. The degree of node $i$ in the adjacency matrix $A$ is denoted as:

$$k_i = \sum_{j=1}^{N} a_{ij} \tag{2}$$

The average degree of the network is defined as:

$$\langle k \rangle = \frac{1}{N} \sum_{j=1}^{N} k_i = \frac{2M}{N} \tag{3}$$

The degree distribution $P(k)$ of a network is the probability that the network randomly selects a node with degree $k$ that satisfies:

$$\sum_{k}^{\infty} P(k) = 1 \tag{4}$$

The average degree $\langle k \rangle$ of the network can also be expressed through the degree distribution $P(k)$ of the nodes as:

$$\langle k \rangle = \frac{1}{N} \sum_{j=1}^{N} kP(k) \tag{5}$$

As in the network in Figure 1, the degree of each node $\{k_1, k_2, k_3, k_4\} = \{2,3,2,2\}$, the average degree of the network $\langle k\rangle = 9/4$, the distribution of the degrees $P(1)=0$, $P(2)=2/4=1/2$, and $P(3)=2/4=1/2$.

### 2.1.2 Average path length

In a complex network, there may be multiple paths between nodes, with $d_{ij}$ denoting the shortest path, i.e., distance, from node $i$ to node $j$. Then the average path length of the network is:

$$L = \frac{2}{N(N-1)}\sum_{i<j} d_{ij} \tag{6}$$

The connectivity of a node is the ability of nodes to reach each other by connecting edges and the average network distance represents the average value of node connectivity.

Network diameter denotes the maximum value of the distance between any two nodes in the network, denoted as:

$$D = \max\left(d_{ij}\right) \tag{7}$$

### 2.1.3 Clustering coefficients

In a social relationship network, a person's friends are likely to be friends as well, referring to this property as the clustering property of the network. Let a node $v_i$ in the network have $k_i$ edges connected to other nodes, and these $k_i$ nodes are called neighbor nodes. There may be at most $C_{k_i}^2$ edges between that node. In fact the ratio of the number of edges $E_i$ present between these $k_i$ nodes to the total number of edges $C_{k_i}^2$ is defined as the clustering coefficient $C_i$ of node $v_i$, i.e.:

$$C_i = \frac{E_i}{C_{k_i}^2} \tag{8}$$

### 2.1.4 Feature vector centrality

The importance of a node depends on both the importance of neighboring nodes and the number of neighboring nodes. Given a relative score for each node in the network, the contribution of a connection to a node with a high score is greater than the contribution of a connection to a node with a low score. The eigenvector centrality is defined by the adjacency matrix $A$.

Let the centrality score $x$ of a node be proportional to the sum of the centrality scores of all nodes connected to it, then:

$$x_i = \frac{1}{\lambda}\sum_{j=1}^{N} a_{ij} x_j \tag{9}$$

where $N$ is the total number of nodes and $\lambda$ is a constant. Described in terms of vectors, Eq. (9) can be written as an eigenvector equation:

$$x = \frac{1}{\lambda} Ax \tag{10}$$

To wit:

$$Ax = \lambda x \tag{11}$$

### 2.1.5 Joint probability distribution

The joint probability $P(k_i, k_j)$ is defined as the probability that an edge is randomly selected from the network and the degrees of the two nodes connected to that edge are $k_i$ and $k_j$, respectively, i.e., as a proportion of the total number of edges in the network:

$$P(k_i, k_j) = \frac{M(k_i, k_j) N(k_i, k_j)}{2M} \tag{12}$$

where $M(k_i, k_j)$ is the total number of edges connected to nodes of degree $k_i$ and $k_j$, and $M$ is the total number of edges of the network; if $k_i = k_j$, then $N(k_i, k_j) = 2$, otherwise $N(k_i, k_j) = 1$.

The joint probability distribution has the following properties:
(1) Symmetry: $P(k_i, k_j) = P(k_j, k_i)$.
(2) Normalization: $\sum_{k_i, k_j = k_{\min}}^{k_{\max}} P(k_i, k_j) = 1$.

If the degree of two nodes connected by a randomly chosen edge in the network is completely random, i.e., whether or not an edge is connected between two nodes in the network is independent of the degree values of these two nodes, i.e., there is:

$$P(k_i, k_j) = P(k_i) P(k_j) \tag{13}$$

The network is said to be neutral and the network is not degree correlated, otherwise the network is said to be degree correlated.

## 2.2 Matrix decomposition

Socialization matrix decomposition algorithms need to consider the influence of the target user's friends as well as other users on the target user [15]. The matrix decomposition recommendation algorithm incorporating social information is called social matrix decomposition, which has a generalized optimization loss function as in Equation (14):

$$\begin{aligned} &\min_{U,V,\Omega} \left\| \left( R - U^T V \right) \right\|_F^2 + \alpha Social\left( T, S, \Omega \right) \\ &+ \lambda \left( \|U\|_F^2 + \|V\|_F^2 + \|\Omega\|_F^2 \right) \end{aligned} \tag{14}$$

where $R \in R_{i \times j}$ is the matrix of user ratings of items, $T \in R_{k \times l}$ is the matrix of user-user social relationships, $U \in R_{k \times d}$ is the matrix of potential user features, $V \in R_{k \times j}$ is the matrix of potential item features, and $Social\left( T, S, \Omega \right)$ is the socialization information obtained from the analysis of social networks, and $\Omega$ is the parameter learned from the socialization information. Where the coefficient $\alpha$ is used to control the effect of $Social(T,S,\Omega)$.

### 2.2.1 Co-decomposition methods

The principle of co-decomposition method is that social information and rating information can be connected by sharing the implicit feature space of users.SoRec algorithm is one of the most typical algorithms of this kind.

The SoRec algorithm uses the social relationship matrix $T$ to represent the user's social relationship information, and by decomposing the matrix $T$ we can get the user potential feature matrix $U$ and trust potential feature matrix $Z$. From the idea of matrix decomposition, we can get $R=U^TV$ and $T=U^TZ$ .The SoRec method combines the user's rating information with the user's social relationship by sharing the user's implicit feature space. It works as shown in Fig. 2.

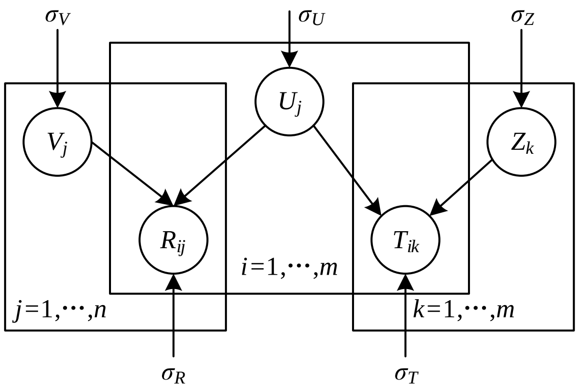



*Figure 2: The Principle of SoRec*

The optimized loss function of the SoRec algorithm is shown in Equation (15):

$$\begin{aligned}&\min_{U,V,Z}\left\|\left(R-U^TV\right)\right\|_F^2+\alpha\left\|T-U^TZ\right\|_F^2\\&+\lambda\left(\|U\|_F^2+\|V\|_F^2+\|Z\|_F^2\right)\end{aligned} \tag{15}$$

In this method, $Social(T,S,\Omega)$ is defined as in equation (16):

$$\left\|T-U^TZ\right\|_F^2 \tag{16}$$

The relation matrix $T$ is a 0-1 matrix and each element in the matrix represents a trust relationship between two users, so the matrix $T$ is asymmetric. The conditional probability distribution of $T$ is calculated as in Equation (17):

$$p\left(T\middle|U,Z,\sigma_T^2\right)=\prod_{i=1}^{m}\prod_{k=1}^{m}\left[N\left(T_{ik}\middle|g\left(U_i^TZ_k\right),\sigma_T^2\right)\right]^{I_{ik}^T} \tag{17}$$

where $I_{ik}^T$ is the indicator function, which takes the value of 1 when there is a trust relationship between users. Vice versa, it takes the value of 0. The function $g(x)=1/\left(1+e^{-x}\right)$ can constrain the value of $U_i^T Z_k$ to $[0,1]$. Assume that $U$, $Z$ obey a 0-expectation spherical Gaussian distribution as in Eq. (18):

$$
\begin{aligned}
p\left(U \middle| \sigma_U^2\right) &= \prod_{i=1}^{m} N\left(U_i \middle| 0, \sigma_U^2 I\right) \\
p\left(Z \middle| \sigma_Z^2\right) &= \prod_{i=1}^{m} N\left(Z_k \middle| 0, \sigma_Z^2 I\right)
\end{aligned}
\tag{18}
$$

The joint posterior probability of the potential feature matrices $U$, $V$, and $Z$ is shown in Equation (19):

$$
\begin{aligned}
& p\left(U,V,Z \middle| R,T,\sigma_R^2,\sigma_T^2,\sigma_U^2,\sigma_V^2,\sigma_Z^2\right) \\
& \propto P\left(R \middle| U,V,\sigma_R^2\right) P\left(T \middle| U,Z,\sigma_T^2\right) \\
& \times p\left(U \middle| \sigma_U^2\right) p\left(V \middle| \sigma_V^2\right) p\left(Z \middle| \sigma_Z^2\right) \\
& = \prod_{i=1}^{m}\prod_{j=1}^{n}\left[N\left(R_{ij} \middle| g\left(U_i^T V_j\right),\sigma_R^2\right)\right]^{I_{ij}^R} \\
& \times \prod_{i=1}^{m}\prod_{k=1}^{m}\left[N\left(T_{ik} \middle| g\left(U_i^T Z_k\right),\sigma_T^2\right)\right]^{I_{ik}^T} \\
& \times \prod_{i=1}^{m} N\left(U_i \middle| 0,\sigma_U^2 I\right) \times \prod_{j=1}^{n} N\left(V_j \middle| 0,\sigma_V^2 I\right) \\
& \times \prod_{k=1}^{m} N\left(Z_k \middle| 0,\sigma_Z^2 I\right)
\end{aligned}
\tag{19}
$$

### 2.2.2 Integration methods

Although the SoRec algorithm has received a lot of attention as a novel social recommendation algorithm, the way of combining social relationship and recommendation algorithm in this algorithm does not truly reflect the recommendation process in real scenarios, and the algorithm is poorly interpretable and lacks sufficient persuasive power.

Unlike the SoRec algorithm, the integrated approach assumes that the user's preferences are jointly determined by the user's personal preferences and the preferences of his/her friends. The final predicted score of the target user can be obtained by weighted average of the predicted score obtained from matrix decomposition of the target user's score matrix and the predicted score of his/her friends. The most typical algorithm in this category is the RSTE algorithm. Its working principle is shown in Figure 3.

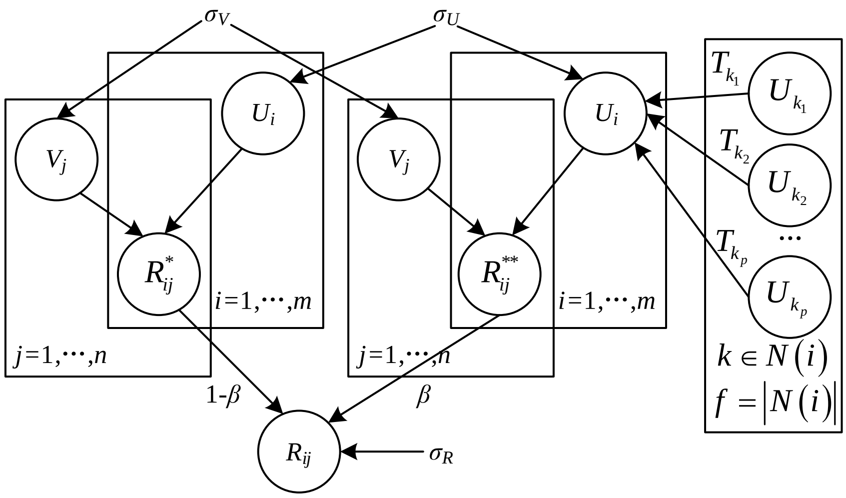



*Figure 3: The Principle of RSTE*

where $N(i)$ denotes the set of friends of user $u_i$ and $|N(i)|$ denotes the number of friends of user $u_i$. The linear expression $R_{ij}$ for predicting the scores in the RSTE algorithm is represented as in Eq. (20):

$$R_{ij} = (1-\beta)U_i^T V_j + \beta \sum_{k \in N(i)} S_{ik} U_k^T V_j \tag{20}$$

where $S_{i,k}$ is the regularization term of the sum of ratings of all friends of user $i$, and $\beta$ is used to control the proportion of ratings of the target user and his/her friends in the result, and its optimal loss function is shown in Equation (21):

$$\min_{U,V} \left\| \left(R - U^T V\right) - \beta S U^T V \right\|_F^2 + \lambda \left( \|U\|_F^2 + \|V\|_F^2 \right) \tag{21}$$

In this method, $Social(T, S, \Omega)$ is defined as in equation (22):

$$\left\| R - \beta S U^T V \right\|_F^2 - 2Tr\left( (R - U^T V) \beta S U^T V \right) \tag{22}$$

where $Tr(\ )$ denotes the trace of the matrix.

The algorithm produces predicted ratings that are influenced by the target user's friends, and the conditional distribution of its predicted ratings is shown in Equation (23):

$$p\left(R \middle| T, U, V, \sigma_T^2\right) = \prod_{i=1}^{m} \prod_{j=1}^{n} \left[ N\left( R_{ij} \middle| g\left( \sum_{k \in N(i)} T_{ik} U_k^T V_j \right), \sigma_R^2 \right) \right]^{I_{ij}^R} \tag{23}$$

Assuming that $T$ and $U$, $V$ are independent of each other, the posterior probabilities of the potential feature matrices $U$, $V$ of the unintegrated target user's friend ratings can be obtained according to the Bayesian formula as shown in Equation (24) below:

$$p\left(U,V\left|R,T,\sigma_R^2,\sigma_U^2,\sigma_V^2\right.\right)\propto p\left(R\left|T,U,V,\sigma_R^2\right.\right)p\left(U\left|\sigma_U^2\right.\right)p\left(V\left|\sigma_V^2\right.\right)$$
$$=\prod_{i=1}^{m}\prod_{j=1}^{n}\left[N\left(R_{ij}\left|g\left(\sum_{k\in N(i)}T_{ik}U_k^TV_j\right),\sigma_R^2\right.\right)\right]^{I_{ij}^R}$$
$$\times\prod_{i=1}^{m}N\left(U_i\left|0,\sigma_U^2I\right.\right)\times\prod_{j=1}^{n}N\left(V_j\left|0,\sigma_V^2I\right.\right) \tag{24}$$

The posterior probabilities of $U$ and $V$ after incorporating the target user's friend ratings are shown in Equation (25):

$$p\left(U,V\left|R,T,\sigma_R^2,\sigma_U^2,\sigma_V^2\right.\right)\propto p\left(R\left|T,U,V,\sigma_R^2\right.\right)p\left(U\left|\sigma_U^2\right.\right)p\left(V\left|\sigma_V^2\right.\right)$$
$$=\prod_{i=1}^{m}\prod_{j=1}^{n}\left[N\left(R_{ij}\left|g\left((1-\beta)U_i^TV_j+\beta\sum_{k\in N(i)}S_{ik}U_k^TV_j\right),\sigma_R^2\right.\right)\right]^{I_{ij}^R}$$
$$\times\prod_{i=1}^{m}N\left(U_i\left|0,\sigma_U^2I\right.\right)\times\prod_{j=1}^{n}N\left(V_j\left|0,\sigma_V^2I\right.\right) \tag{25}$$

### 2.2.3 Regularization methods

Regularization methods assume that a user's preferences will approximate those of his friends, and SocialMF is the most typical of such methods. The SocialMF algorithm incorporates the idea of trust propagation, since in real life trust spreads in social networks. The algorithm takes into account the influence of the target user's trusted users on his user's potential feature matrix in the calculation process. The principle of this algorithm is shown in Fig. 4.

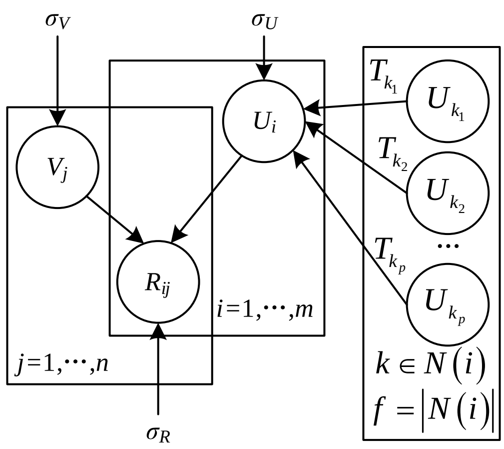



*Figure 4: The Principle of SocialMF*

There is a dependency of the user's latent feature matrix on the latent feature matrix of his/her friends. As shown in equation (26):

$$U_i=\frac{\sum_{k\in N(i)}S_{ik}U_k^T}{|N(i)|} \tag{26}$$

where $U_i$ is the potential feature matrix of user $u_i$ obtained from the potential feature matrices of all the friends of user $u_i$, and since the currently used social relationship matrix is large as a 0-1 matrix, the value of $S_{ik}$ can only be either 1 or 0. The $S_{ik}$ is normalized, and so Eq. (26) can be transformed into Eq. (27):

$$U_i = \sum_{k \in N(i)} S_{ik} U_k^T \tag{27}$$

The optimal loss function of the SocialMF algorithm is shown in Equation (28):

$$\begin{aligned} &\min_{U,V} \left\| R - U^T V \right\|_F^2 + \lambda \left( \|U\|_F^2 + \|V\|_F^2 \right) \\ &+\alpha \sum_{i=1}^{n} \left( U_i - \sum_{k \in N(i)} S_{ik} U_k^T \right)^2 \end{aligned} \tag{28}$$

In the SocialMF algorithm, $Social(T, S, \Omega)$ is defined as in Equation (29):

$$\sum_{i=1}^{n} \left( U_i - \sum_{k \in N(i)} S_{ik} U_k^T \right)^2 \tag{29}$$

In this algorithm, the posterior probabilities of the feature matrices $U$ and $V$ are shown in Equation (30):

$$\begin{aligned} &p\left(U,V \middle| R,T,\sigma_R^2,\sigma_T^2,\sigma_U^2,\sigma_V^2\right) \propto p\left(R \middle| U,V,\sigma_R^2\right) \\ &p\left(U \middle| T,\sigma_U^2,\sigma_T^2\right) p\left(V \middle| \sigma_V^2\right) \\ &= \prod_{i=1}^{m} \prod_{j=1}^{n} \left[ N\left(R_{ij} \middle| g\left(U_i^T V_j\right), \sigma_R^2\right) \right]^{I_{ij}^R} \\ &\times \prod_{i=1}^{m} N\left( U_i \middle| \sum_{k \in N(i)} T_{ik} U_k, \sigma_T^2 I \right) \times \prod_{i=1}^{m} N\left(U_i \middle| 0, \sigma_U^2 I\right) \\ &\times \prod_{j=1}^{n} N\left(V_j \middle| 0, \sigma_V^2 I\right) \end{aligned} \tag{30}$$

## 2.3 Global Influence Correlation Evaluation Algorithm

### 2.3.1 Degree centrality

For the graph $G := (V, E)$, the degree centrality of the nodes [16] can be expressed by equation (31):

$$C_D(v_i) = d_i / (n-1) \tag{31}$$

where $d_i$ denotes the degree of node $v_i$, $n$ denotes the total number of nodes in the network structure to which node $v_i$ belongs, and $n-1$ denotes the maximum number of possible connections of the node to other $n-1$ nodes.

### 2.3.2 Cartesian Centrality

For a graph $G := (V,E)$ of $n$ nodes, the median of node $v_i$ (edge) is computed as shown in equation (32):

$$C_B(v_i) = \sum_{v_s \neq v_i, v_t \in V, s<t} \frac{\sigma_{st}(v_i)}{\sigma_{st}} \tag{32}$$

where $\sigma_{st}$ denotes the number of shortest paths from $s \to t$, and $\sigma_{st}(v_i)$ denotes the number of shortest paths from $s \to t$ passing through the node $v_i$ (edge).

### 2.3.3 Close centrality

Tightness centrality considers nodes with shorter shortest path lengths to be more tightly packed. In network analysis, tightness usually denotes the average shortest distance from node $v_i$ to other reachable nodes, which is calculated as shown in Equation (33):

$$C_c(v_i) = \frac{\sum_{t \in V} d_G(v_i, t)}{n-1} \tag{33}$$

where $d_G(v_i, t)$ denotes the shortest distance from node $v_i$ to node $t$ and $n-1$ denotes the number of other nodes in the graph that can be reached from node $v_i$. The tightness calculation method can also be expressed as Equation (34):

$$C_c(v_i) = \frac{1}{\sum_{t \in V} d_G(v_i, t)} \tag{34}$$

### 2.3.4 Feature vector centrality

The process of computation of eigenvector centrality requires the use of adjacency matrix [17]. Firstly, the graph $G := (V,E)$ needs to be transformed into an adjacency matrix $A$, and the element $a_{v,t}$ in $A$ is the connectivity between node $v$ and node $t$, and the eigenvector centrality of node $v$ is computed as in Eq. (35):

$$C_e(v) = \frac{1}{\lambda} \sum_{t \in M(v)} C_e(t) = \frac{1}{\lambda} \sum_{t \in G} a_{v,t} C_e(t) \tag{35}$$

where $M(v)$ is the set of all neighboring nodes of node $v$, $\lambda$ is a constant, and Eq. (35) can be expressed as Eq. (36) by the eigenvector equation:

$$AC_c = \lambda C_c \tag{36}$$

When calculating the eigenvalues, all elements in the matrix must be non-negative, this constraint makes only one eigenvalue eligible for eigenvector centrality calculation. Power Iteration algorithm is one of the effective algorithms for eigenvector solution.

Before calculating the PageRank value, it is necessary to divide each web page into an initial value, which is usually set as $1/N$. Where $N$ is the total number of web pages. Then the Page Rank algorithm is iterated until a smooth distribution is reached. The formula of this algorithm is shown in Equation (37):

$$PR(p_i) = \alpha \sum_{p_j \in M_{p_i}} \frac{PR(p_j)}{L(p_j)} + \frac{(1-\alpha)}{N} \tag{37}$$

where $M_{p_i}$ is the set of all web pages that have outgoing links to $p_i$ web pages, $L(p_j)$ is the number of outgoing links to web page $p_j$, $N$ is the total number of web pages, and $\alpha$ is the damping coefficient, which has a value between 0 and 1, and is generally defined as 0.85.

# 3 Influence calculation and content recommendation methods

## 3.1 Interpretation of variables

$U = \{u_1, u_2, \cdots, u_m\}$ denotes the set of users in the recommender system, and $V = \{v_1, v_2, \cdots, v_n\}$ denotes the set of collection of items. The matrix of user ratings of items is denoted as $R = (R_{ij})^{m\times n}$, where $R_{ij} \in \{1, 2, 3, 4, 5\}$ denotes the ratings of the user $u_i$ on the item $v_j$. $T = (T_{ij})^{m\times m}$ denotes the user's social relationship matrix, $T_{ij} = 1$ means that there is a buddy relationship between user $u_i$ and user $v_j$, otherwise $T_{ij} = 0$, and the set of buddies directly connected to user $u_i$ is $N_i$. Suppose there are $x$ communities and $g = \{g_1, \cdots, g_x\}$ is the set of communities. For user $u_i$, let $u_i \in g_i$ and $C(i) \in (g_i \cup N_i)$ denote the set of the user's friends and other users in the community to which they belong. The matrix decomposition recommendation algorithm approximates the user rating matrix $R$ into a low-order user feature matrix $U \in R^{l\times m}$ and a low-order item feature matrix $V \in R^{l\times n}$, where $l$ is the dimensionality of the hidden eigenvector.

## 3.2 Community Discovery

Community discovery in this paper uses the degree-corrected randomized block model DCBM.DCBM uses the Kernighan-Lin algorithm, a heuristic based on the greedy principle. The network is first randomly partitioned into an initial set of $x$ communities. Then a vertex is repeatedly moved from one set to another, choosing at each move either the move that results in the largest increase in the objective function or the move that results in the smallest decrease in the objective function, and each vertex can only be moved once. When all vertices have been moved, examine the states passed from the beginning to the end of the move process, select the move state with the highest objective score, and use this state as the starting point for a new iteration of the same process. When there is no increase in the objective function, a complete iteration of the process ends and the final community segmentation of the network is obtained.

## 3.3 Impact calculations

### 3.3.1 Influence based on user rating information

Influence between users is the size of mutual influence between any two users, and both user history rating information and social network structure information are used in the calculation. Through the user history rating information, the similarity between users can be analyzed to get the similarity between users, this paper adopts Pearson similarity to calculate the similarity, the formula is as follows:

$$sim(i,k) = \frac{\sum_{j\in I}(R_{ij}-\bar{R}_i)(R_{kj}-\bar{R}_k)}{\sqrt{\sum_{j\in I}(R_{ij}-R_i)^2}\sqrt{\sum_{j\in I}(R_{kj}-R_k)^2}} \tag{38}$$

where $I$ denotes the set of items that have been rated by both user $i$ and user $k$. $R_{ij}$ is the rating of item $j$ by user $i$. $\bar{R}_i$ is the average rating of user $i$. In the set of jointly rated items, the closer the deviation of two users' ratings of the same item from the users' average ratings, the more similar the users' preferences are and the higher the similarity is. The function $f(x)=(x+1)/2$ is used to map the value domain of similarity to the interval [0, 1], with larger values representing more similarity between users.

### 3.3.2 Influence based on social network structure

However, the similarity calculated through user ratings only utilizes the rating information and does not fully utilize the information between users, so this paper also uses social network structure information to calculate the influence between users.

The influence between users is calculated by SimRank algorithm using social network structure data. Define the Sim Rank similarity $inf_s(i,k)$ between two users $i$ and $k$. Sim Rank calculates the similarity between users based on the idea that similar users have similar neighbors, and the calculation method is as follows:

(1) $inf_s(i,k)=0$ when $I(i)=\varnothing$ or $I(k)=\varnothing$.

(2) In other cases:

$$\mathrm{inf}_s(i,k) = \begin{cases} 1, & i=k \\ \dfrac{d}{|I(i)||I(k)|}\sum_{b\in I(k)}\sum_{a\in I(i)} inf_s(a,b), & i\neq k \end{cases} \tag{39}$$

where $I(i)$ denotes the set of all users pointing to user $a$ and $d\in(0,1)$ is a damping factor.

In social networks, users seeking advice are more likely to be influenced by other users with high personal influence. Moreover, when users choose other users to communicate with, they will be more inclined to choose users in the same community with them rather than users outside the community.

Define the local influence of each user in a community as $s_i$, and $S$ is a personalization vector consisting of the local influence of all users in the community. In each community, use Page Rank to find the local influence $s_i$ of each user in the community.

Define user's personalized influence $f_i$, $F$ is the vector of user influence scores in the network, $f_i \in F$ .

In this paper, the personalization vector $S$ calculated from the community information is combined to Personalized PageRank, and the user's personal influence is calculated, see Equation (40):

$$F = aP^T F + (1-a)S \tag{40}$$

$P$ is the transfer matrix of the network graph and $a$ is the jump factor. Using the Personalized Page Rank algorithm across the social network, the global personal influence $f_i$ of each user can be calculated. The global personal influence of users is then data normalized to map the influence onto the $[0,1]$ interval.

The calculated influence between users is basically symmetric, which will be biased when used directly to describe the influence between users. In practice, the influence of user $i$ on user $k$ may be different from the influence of user $k$ on user $i$. Users with high personal influence are more likely to influence other users. Combining the personal influence of users with the influence among users yields the asymmetric influence between user $i$ and user $k$ computed using the social network structure information:

$$inf(i,k) = \sqrt{inf_s(i,k) \times f_k} \tag{41}$$

### 3.3.3 Combined inter-user influence calculation

Combining the influence between users $sim(i,k)$ computed using information about users' historical ratings and the influence between users $inf(i,k)$ computed using information about the structure of the social network yields the combined influence of user $i$ over $k$ :

$$f(i,k) = \sqrt{sim(i,k) \times inf(i,k)} \tag{42}$$

## 3.4 Content recommendation algorithms for social network communication

In this paper, we propose a matrix decomposition recommendation algorithm SoInf that integrates community structure and social influence on the basis of SoReg recommendation algorithm, which utilizes the influence information between users and user rating information.

In SoReg, the mutual influence relationship between users and their friends is influenced by the similarity calculated through the user history rating information. In contrast, in SoInf, both user rating information and social network structure information are used to compute the influence $f(i,k)$ between users, and to fuse the influence between users and the user's personal influence computed by combining community information. The implicit feature vector $U_i$ of user $u_i$ and the implicit feature $U_k$ of user $u_k$ are constrained by the influence $f(i,k)$ between users:

$$\frac{\alpha}{2}\sum_{i=1}^{m}\sum_{k\in C(i)} f(i,k)\left\|U_i - U_k\right\|_F^2 \tag{43}$$

where $C(i) \in (g_i \cup N_i)$ denotes the set of the user's friends and other users in the community to which they belong, and $\|\cdot\|_F^2$ is the Frobenius paradigm.

Assuming that $U_i$, $V_j$ prior probabilities follow a Gaussian distribution and are independent of each other, there are:

$$p\left(U \mid \sigma_U^2\right) = \prod_{i=1}^{m} N\left(U_i \mid 0, \sigma_U^2 I\right) \tag{44}$$

$$p\left(V \mid \sigma_V^2\right) = \prod_{j=1}^{n} N\left(V_j \mid 0, \sigma_V^2 I\right) \tag{45}$$

Given the implicit features of other users in the same community of the user and the user's direct neighbors, an optimization task is performed in the space of the user's implicit features to obtain the conditional distribution of the user's implicit features:

$$\begin{aligned} &p\left(U \mid U_k, T, \sigma_U^2, \sigma_T^2\right) \propto p\left(U \mid \sigma_U^2\right) \\ &\times \prod_{i=1}^{m} \prod_{k \in C(i)} p\left(U \mid U_k, \frac{\sigma_T^2}{f(i,k)}\right) \\ &= \prod_{i=1}^{m} N\left(U_i \mid 0, \sigma_U^2 I\right) \times \prod_{i=1}^{m} \prod_{k \in C(i)} p\left(U \mid U_k, \frac{\sigma_T^2}{f(i,k)}\right) \end{aligned} \tag{46}$$

The conditional distribution of predicted rating values is as follows:

$$p\left(R \mid U, V, \sigma_R^2\right) = \prod_{i=1}^{m} \prod_{j=1}^{n} \left[N\left(R_{ij} \mid U_i^T V_j, \sigma_R^2\right)\right]^{I_{ij}^R} \tag{47}$$

where $I_{ij}^R$ is the indicator function, user $u_i$ has generated ratings for item $v_j$, then $I_{ij}^R = 1$, otherwise $I_{ij}^R = 0$.

Using Bayes' rule, the posterior distributions of the feature matrices $U$ and $V$ can be computed by the following method:

$$\begin{aligned} &p\left(U, V \mid R, T, \sigma_R^2, \sigma_U^2, \sigma_V^2, \sigma_T^2\right) \times p\left(R \mid U, V, T, \sigma_R^2\right) \\ &\times p\left(U \mid U_k, T, \sigma_U^2, \sigma_T^2\right) \times p\left(V \mid \sigma_V^2\right) \\ &= p\left(R \mid U, V, \sigma_R^2\right) \times p\left(U \mid \sigma_U^2\right) \times \prod_{i=1}^{m} \prod_{k \in C(i)} p\left(U \mid U_k, \frac{\sigma_T^2}{f(i,k)}\right) \\ &\times p\left(V \mid \sigma_V^2\right) = \prod_{i=1}^{m} \prod_{j=1}^{n} \left[N(R_{ij} \mid U_i^T V_j, \sigma_R^2)\right]^{I_{ij}^R} \\ &\times \prod_{i=1}^{m} N\left(U_i \mid 0, \sigma_U^2 I\right) \times \prod_{i=1}^{m} \prod_{k \in C(i)} p\left(U \mid U_k, \frac{\sigma_T^2}{f(i,k)}\right) \\ &\times \prod_{j=1}^{n} N\left(V_j \mid 0, \sigma_V^2 I\right) \end{aligned} \tag{48}$$

Taking the logarithm of the posterior distribution gives:

$$\begin{aligned}
&\ln p\left(U,V \mid R,T,\sigma_R^2,\sigma_U^2,\sigma_V^2,\sigma_T^2\right) = -\frac{1}{2\sigma_R^2}\sum_{i=1}^{m}\sum_{j=0}^{n} I_{ij}^R\left(R_{ij}-U_i^T V_j\right)^2 \\
&-\frac{1}{2\sigma_T^2}\sum_{i=1}^{m}\sum_{k\in C(i)} f(i,k)(U_i-U_k)^2 \\
&-\frac{1}{2\sigma_U^2}\sum_{i=1}^{m}U_i^T U_i-\frac{1}{2\sigma_V^2}\sum_{j=1}^{n}V_j^T V_j-\frac{1}{2}\sum_{i=1}^{m}\sum_{k\in C(i)}\ln\sigma_T^2 \\
&-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=0}^{n} I_{ij}^R \ln\sigma_R^2-\frac{1}{2}\left(ml\ln\sigma_U^2+nl\ln\sigma_V^2\right)+Con
\end{aligned} \tag{49}$$

where $Con$ is a constant. Maximizing the above log posterior probability is equivalent to minimizing the following objective function:

$$\begin{aligned}
E\left(U,V,R,T\right) = &\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n} I_{ij}\left(R_{ij}-U_i^T V_j\right)^2 \\
&+\frac{\alpha}{2}\sum_{i=1}^{m}\sum_{k\in C(i)} f(i,k)\left\|U_i-U_k\right\|_F^2 \\
&+\frac{\lambda_U}{2}\left\|U\right\|_F^2+\frac{\lambda_V}{2}\left\|V\right\|_F^2
\end{aligned} \tag{50}$$

where $\lambda_U=\sigma_R^2/\sigma_U^2$, $\lambda_V=\sigma_R^2/\sigma_V^2$, $\alpha=\sigma_R^2/\sigma_T^2$. By stochastic gradient descent method, the local optimal solutions of user feature matrix $U$ and item feature matrix $V$ are obtained. The corresponding partial derivatives are shown in Eq. (51) and Eq. (52):

$$\begin{aligned}
\frac{\partial E}{\partial U_i} = &\sum_{j=1}^{n} I_{ij}\left(U_i^T V_j-R_{ij}\right)V_j \\
&+\lambda_U U_i+\alpha\sum_{k\in C(i)} f(i,k)(U_i-U_k)
\end{aligned} \tag{51}$$

$$\frac{\partial E}{\partial V_j} = \sum_{i=1}^{m} I_{ij}\left(U_i^T V_j-R_{ij}\right)U_i+\lambda_V V_j \tag{52}$$

# 4 Experimental results and analysis

## 4.1 Comparative Experiments and Analysis of Impact Calculations

In order to verify the accuracy of the influence calculation method proposed in this paper in calculating influential users in social networks, the following experiments are designed and compared with other classical calculation methods in this section.

### 4.1.1 Experimental setup

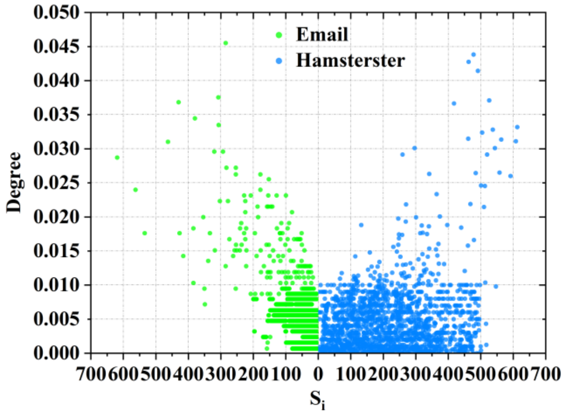
In conducting the analysis of the propagation performance of nodes, most of the studies use the classical disease propagation model SI for simulation studies, which can well simulate the propagation process of information and viruses. In the SI disease propagation model, the nodes in the network have two possible states, susceptible state (S) and infected state (I), at any given moment. A node in the infected state spreads the virus to its neighboring nodes with a propagation probability of $\beta_c \in (0,1)$ at each time period, and a node in the susceptible state S is infected and transforms to the infected state I and cannot recover. Let the initial propagation source be any node $i$ in the network, and define the propagation impact of node $i$ as the total number of infected nodes $S_i$ after a specified propagation time $t$. In order to make the results more reliable, the average propagation influence obtained after repeating the experiment for M times for node $i$ will be taken as the actual propagation influence, as shown in equation (53):

$$\bar{S}_i = \frac{1}{M}\sum_{m=1}^{M} S_i \tag{53}$$

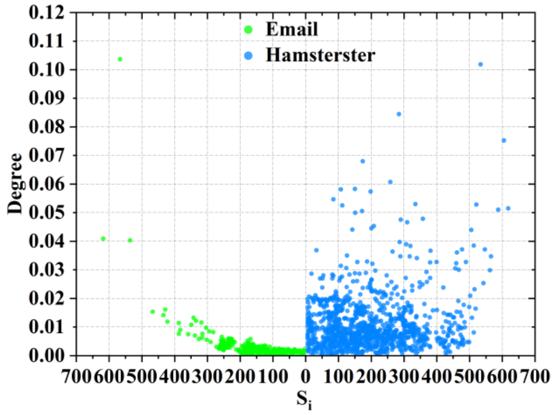
In this paper, different types of real social network data are selected as datasets for experiments, which are Email, a university members' email communication relationship network, and the friendship network of Hamster website. Among them, the number of nodes, the number of edges, the average degree, the clustering coefficient, and the infection rate of the Email dataset are 1130, 5450, 9.627, 0.23, and 0.054, respectively.The relevant data of the Hamster dataset, on the other hand, are 2425, 16635, 13.387, 0.535, and 0.046, respectively.At the same time, in order to validate the influence of the influence proposed in this paper calculation method is effective, the following classical metrics are selected for comparison, including DC, PageRank and the traditional structural hole algorithm SH, respectively.

### 4.1.2 Relevance of metrics and communication capacity

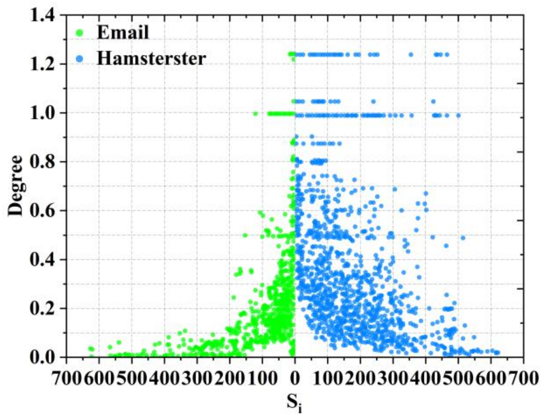
Firstly, the simulation experiment on the propagation ability of each node in a real complex network is carried out using the SI model. Let the initial propagation source be any node in the network, the total number of infected nodes after the propagation time $t = 30$ is the actual propagation influence of the node, repeat the experiment for 200 times, and take the average value of the results of 200 times $\bar{S}_i$ as the final number of nodes in the infected state. Figure 5 shows the relationship between different influence metrics and the actual influence $\bar{S}_i$, and (a)-(d) are the calculation results of DC, PageRank, SH and the algorithm of this paper, respectively. It can be found that the DC method shows a roughly positive correlation trend with influence, i.e., the spreading influence of a node increases with the degree. However, multiple nodes with high degree but average propagation influence can be found in both Email and Hamster, illustrating the limitations of relying on node degree to calculate influence. Similarly, the PageRank method is positively correlated with influence, which is particularly evident in the Email network. However, in the Hamster network, it can be observed that nodes with similar values of metrics have vastly different propagation capabilities, and thus the correlation curve shows a divergent trend. Unlike DC and PageRank, SH and this paper's method show a negative correlation with the spreading influence, i.e., the spreading influence decreases with the increase of metrics. Compared with SH, this paper's method has a more obvious progress in the problem of dispersion of correlation curve, and at the same time, this paper's method has a good performance in calculating the extreme influence nodes (nodes with the highest and lowest influence). This is because it not only takes into account the degree of connection between the node and its neighbors, but also the information of the node itself and the information of the social network structure. All in all, the metric values of this paper's method show a good correlation with the real spreading influence of individual nodes, which is better than other methods.



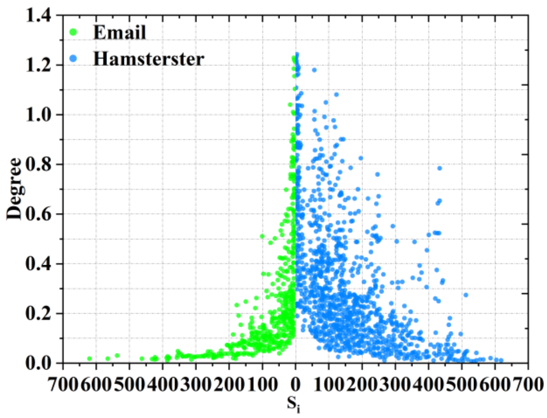

(a) DC




(b) PageRank




(c) SH




(d) Ours

*Figure 5: Correlation analysis of algorithms and actual communication ability*

### 4.1.3 Propagation Performance Analysis of Top-K Nodes

In order to further analyze the difference between the method of this paper and the metric effect of each algorithm, this paper conducts propagation performance experiments on the influential Top-K nodes calculated by each algorithm. The same SI model is used for the simulation study, and unlike the previous experiment which used any node as the propagation source, in this experiment the initial propagation source is made to be the Top-5 node of each algorithm, and the change of the number of nodes in the infected state with time $t$ is analyzed. Tables 1 and 2 show the Top-5 nodes of each algorithm in Email and Hamster networks, respectively. Fig. 6 shows the propagation capability curves of these nodes obtained after simulation by the SI propagation model, (a) and (b) denote the Email network and Hamsterster network, respectively.

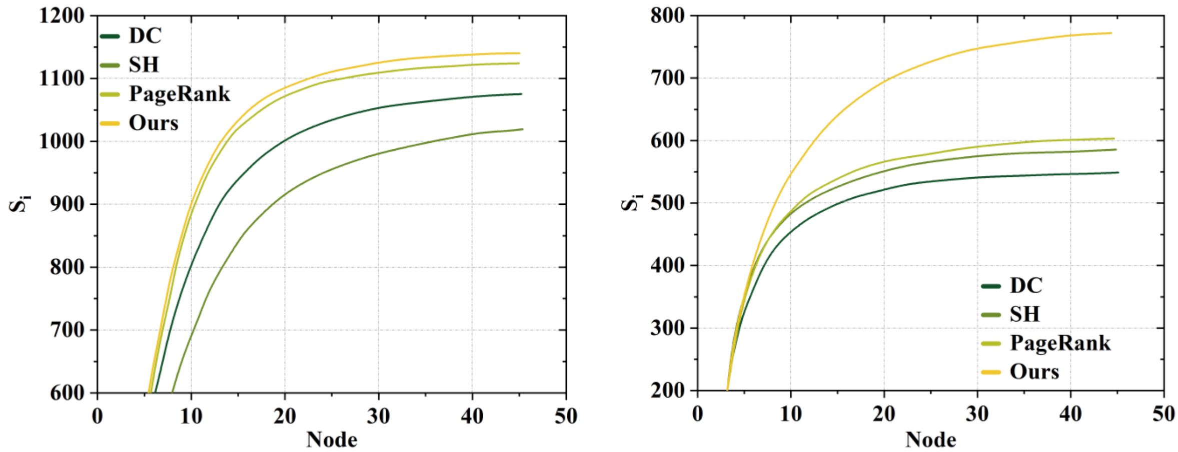
By analyzing the actual situation of Top-5 nodes of different algorithms, the top-5 nodes of this paper's algorithm are different from the top-5 nodes of other algorithms. The results in the figure show that both in Email network and Hamsterster network, the Top-5 nodes of this paper's algorithm have a stronger propagation ability compared to the nodes derived from other algorithms. It is these nodes found only by this paper's algorithm that have stronger propagation ability as well as faster propagation rate than the difference nodes found by DC, PageRank, and SH methods.The less desirable results obtained by the DC and SH methods, which are weaker than PangRank, aptly illustrate that when measuring the importance of a node in a directed graph, if we only take into account the degree of the node or ignore the influence of the node itself, we cannot calculate high influence well.

*Table 1: Top-5 ranking nodes for each algorithm on Email dataset*

| Rank | DC | PageRank | SH | Ours |
|---|---|---|---|---|
| 1 | 103 | 1 | 305 | 25 |
| 2 | 41 | 5 | 107 | 4 |
| 3 | 22 | 3 | 239 | 1 |
| 4 | 40 | 6 | 574 | 20 |
| 5 | 14 | 18 | 40 | 15 |

*Table 2: Top-5 ranking nodes for each algorithm on Hamsterster dataset*

| Rank | DC | PageRank | SH | Ours |
|---|---|---|---|---|
| 1 | 628 | 813 | 628 | 889 |
| 2 | 619 | 628 | 813 | 2017 |
| 3 | 623 | 889 | 623 | 874 |
| 4 | 819 | 623 | 619 | 628 |
| 5 | 611 | 619 | 889 | 2218 |




(a) Email (b) Hamsterster

*Figure 6: The propagation ability of the Top-5 nodes ranked by each algorithm*

## 4.2 Social network communication content recommendation

### 4.2.1 Experimental data

The corpus used in the experiment comes from Sina Weibo, and this paper chooses a telling IT enterprise as the seed for data collection, and collects the data through Sina Weibo's open API, using the “snowball” method to collect the data. As of November 1, 2024, the company has followed 890 users with 1567840 followers, and obtained its user data through Sina Weibo's open API. After the pre-processing process, 850 users remain after excluding inactive users (less than or equal to 15 followers and less than 50 tweets posted). For each user, the last 150 tweets are extracted, and finally 126348 tweets are obtained (some users post less than 150 tweets).

Figure 7 shows the statistical results of the frequency of keyword usage in the corpus. The frequency of keyword usage obtained after processing the corpus by natural language processing techniques has a power rate distribution. The user-keyword preference matrix

$KS = U \times K$ based on user rating information is obtained through user preference idempotentization, where $U$ contains 850 users and $K$ contains 13468 keywords, and $KS$ is $850 \times 13468$ matrix, which is relatively sparse (sparsity is 97.38%). The user association matrix $R = U \times U$ is obtained by extracting the social network relations, $R$ is $850 \times 850$ matrix, which is sparse (sparsity is 93.21%). In order to test the effectiveness of the proposed method, the corpus is divided into a training set and a test set, where the training set accounts for 75% and the remaining part constitutes the test set.

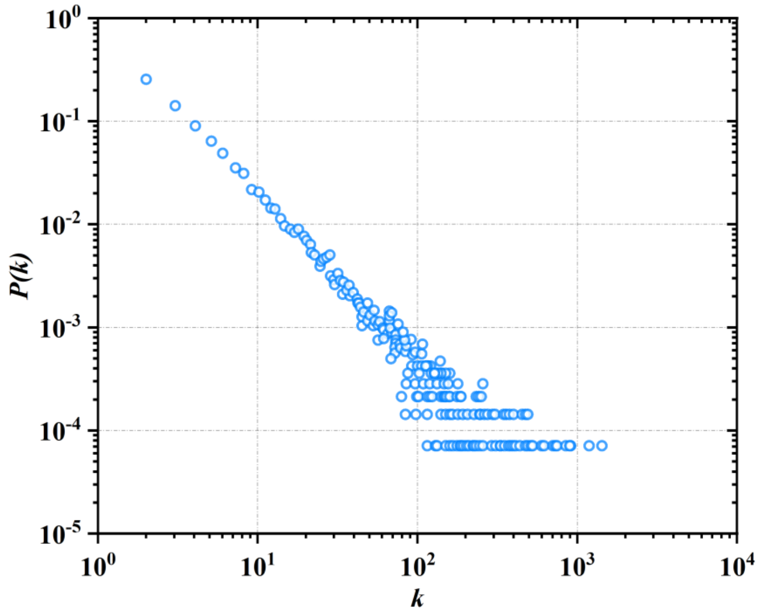



*Figure 7: Use frequency of keywords*

### 4.2.2 Benchmarking models and evaluation indicators

In order to verify the validity, this paper takes the standard collaborative filtering model as the benchmark model and compares the recommendation effect of the recommendation model (SoInf) proposed in this paper, which is based on matrix decomposition and incorporates user information and social network structure information, with it. Moreover, the benchmark model is a special case of the model designed in this study, i.e., when $\alpha = 1$, the recommendation model does not consider the social network relationship and degenerates into the benchmark model. When $\alpha = 0$, the recommendation model is the result of collaborative filtering of the potential user-keyword preference matrix obtained from the social network information dissemination model.

The recommendation process is the prediction of user preferences and actively retrieves and pushes the information or products needed by the user to the maximum extent. In order to evaluate the recommendation effect, this paper draws on the prediction theory and the commonly used evaluation indexes in the field of information retrieval, using the mean absolute error (MAE), the precision rate (P) and the recall rate (R), which measure the prediction ability of the recommendation model, the relevance of the recommendation results, and the coverage ability of the potential preferences, respectively.

### 4.2.3 Analysis of experimental results

Figure 8 shows the relationship between MAE and $\alpha$. The figure shows that the SoInf model has a high recommendation efficiency due to the effective combination of users' personal information and social network structure information. The MAE is used to evaluate the recommendation efficiency of the recommendation model and observe the change of MAE with the nearest neighbor set $L$ when the adjustment coefficient $\alpha$ is taken to different values.The MAE value decreases continuously with the increase of $\alpha$. Increasing the weight of the user content dimension in the recommendation model will effectively improve the recommendation efficiency. However, when $\alpha$ increases to 1, the recommendation model degrades to the baseline model, at which time the MAE value increases abruptly, indicating that the recommendation efficiency will be significantly reduced if the social network information is not considered. When $\alpha = 0.9$ is taken in the experiment, the MAE is the smallest, and its recommendation efficiency is about 10% higher than that of the benchmark model. Secondly, the figure shows that there are various functional relationships between the value of MAE and the size of $L$, e.g., when $\alpha = 0.9$, MAE increases with $L$, and when $\alpha = 1$, MAE decreases with $L$.

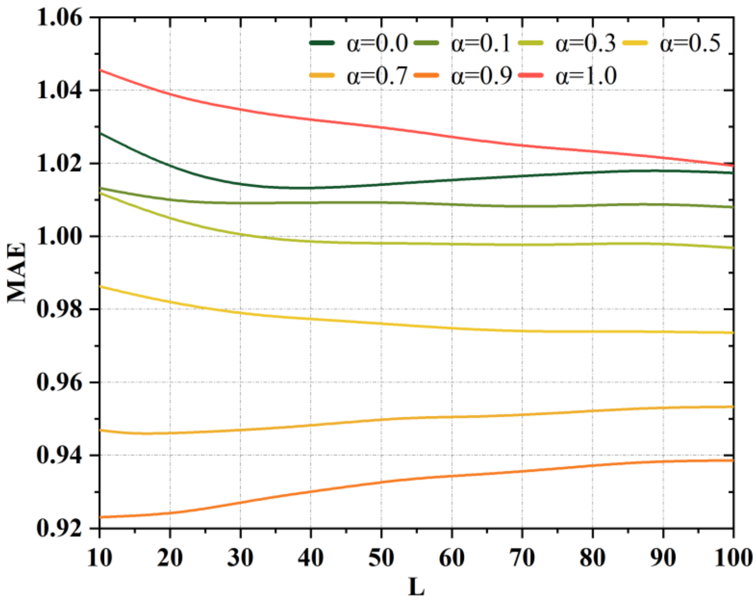
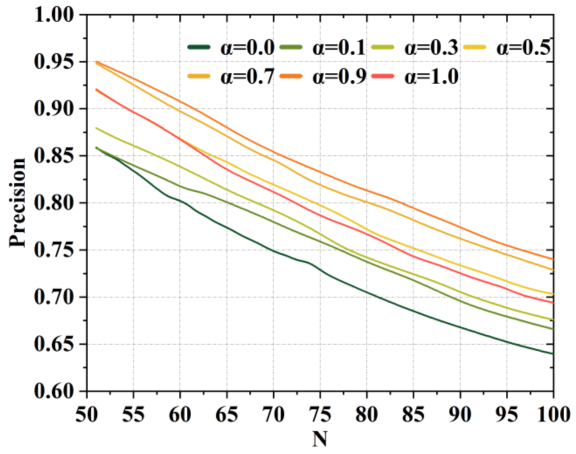
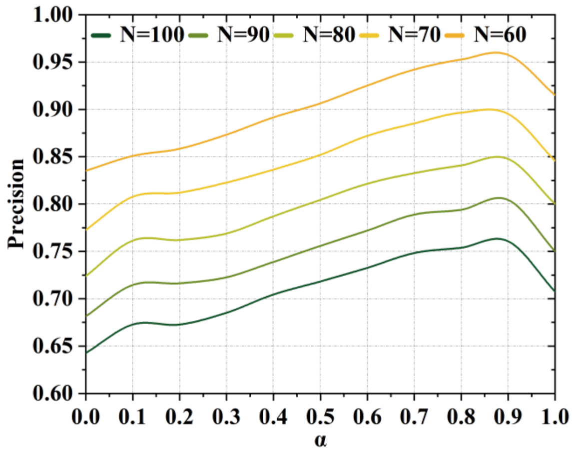



*Figure 8: The relationship between MAE and* $\alpha$

Fig. 9 shows the evaluation of the recommendation efficiency of the recommendation model using the accuracy rate P. Observe the change of the accuracy rate P with the number of recommended keywords $N$ when $\alpha$ takes different values, (a) and (b) indicate the relationship between the accuracy rate and $N$ and $\alpha$, respectively. Figure (a) shows that the accuracy rate P decreases as the number of recommended keywords $N$ rises, because the number of keywords recommended to the target user rises, the denominator in the formula rises and the accuracy rate value decreases. Secondly, the accuracy rate P rises as the adjustment factor $\alpha$ increases and the keywords recommended by the recommendation model include more keywords that are relevant to the target user. However, when $\alpha$ increases to 1 (i.e., degrades to the baseline model), the accuracy rate P decreases abruptly, which indicates that the accuracy rate recommended by the SoInf model is higher than that of the baseline model. The recommendation efficiency of the recommendation model without considering the social network structure information decreases. When $N = 100, \alpha = 0.9$, the accuracy of the designed model is improved by about 5% compared to the benchmark model.



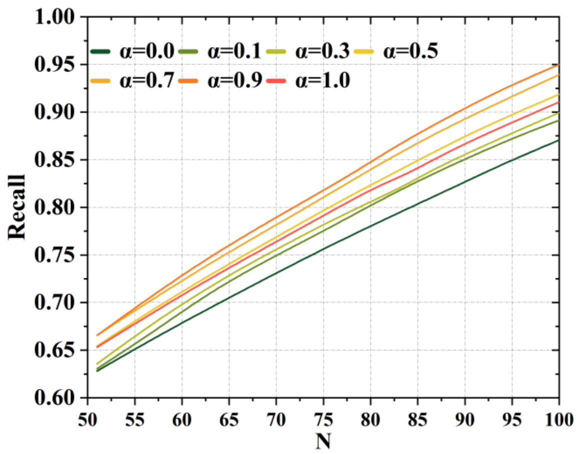

(a) Precision and N



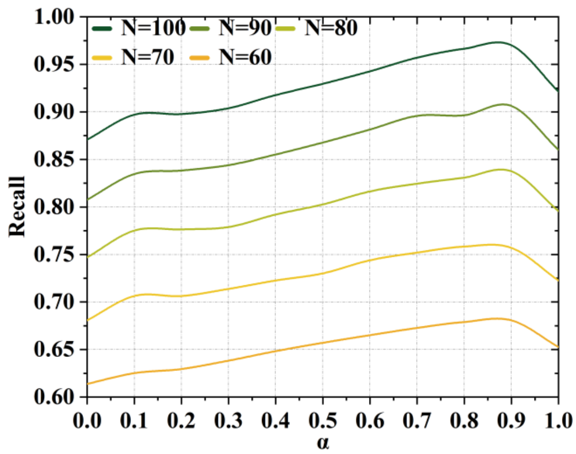

(b) Precision and $\alpha$

*Figure 9: The relationship among precision and N or* $\alpha$

Figure 10 uses the recall rate R to evaluate the recommendation efficiency of the recommendation model, and observes the change of the recall rate R with the number of recommended keywords N when $\alpha$ takes different values, (a) and (b) represent the relationship between the recall rate and $N$ and $\alpha$, respectively. Recall R increases with the increase of the number of recommended keywords N, because the more recommended keywords, the more keywords containing users' preferences, and the numerator in the formula becomes larger, which increases the recall R. Secondly, the increase of $\alpha$ increases the number of recommended keywords N, and the increase of $\alpha$ increases the number of recommended keywords. Secondly, increasing $\alpha$ increases the recall R of the recommendation model, i.e., the keywords covered by the recommendation model can contain more key wests that are relevant to the user. However, when $\alpha$ is increased to 1 (i.e., degraded to the baseline model), the recall R suddenly decreases, which indicates that the recall R of the SoInf model is higher than that of the baseline model, and the recommendation efficiency of the recommendation model that does not take into account the structural information of the social network decreases. When at $\alpha = 0.9$ and $N = 100$, the efficiency of the recommendation model increases by about 10%.




(a) Recall and N




(b) Recall and $\alpha$

*Figure10: The relationship among recall and N or* $\alpha$

The experimental results show that the proposed SoInf recommendation model outperforms the benchmark model in three evaluation indexes: MAE, accuracy P and recall R. This indicates that the integration of both user and social network information can effectively improve the recommendation efficiency and solve the information overload problem.

# 5 Conclusion

Based on the matrix decomposition method, this paper proposes a SoInf recommendation model that integrates community structure and social influence in cross-cultural social networks to realize accurate recommendation of social network communication content.

Experimentally, this paper's influence calculation method based on user rating information and social network structure information has significant improvement compared with DC and SH methods in the correlation curve dispersion problem. Moreover, in calculating the extreme influence nodes with the highest or lowest influence, this paper's method has a better performance. The index values of this method show good correlation with the real propagation influence of individual nodes, which is more advanced than other aspects. The Top-5 nodes calculated by this paper's method have stronger propagation ability than those of other algorithms due to the comprehensive consideration of the degree of the nodes and the influence of the nodes themselves.

In social network propagation content recommendation, this paper's method has more significant advantages in terms of mean absolute error (MAE), accuracy and recall than the benchmark collaborative filtering model that does not consider social network relationships. It is able to accurately predict user preferences in social networks with high coverage, and thus recommend personalized and high-quality communication content to users in cross-cultural social networks. It provides solid technical support for cultural communication and cultural integration in cross-cultural social networks.

# Funding

This work was supported by National Social Science Fund Project.

# About the Author

Hongwei Zhang was born in Inner Mongolia, China, in 1980. I obtained a doctor’s degree from Suchoow University in China. I am currently teaching at School of Sports Management and Communicationl of Capital University of Physical Education and Sports. My main research direction is sports event communication and new media.

# References

[1] Chen, S., & Lin, N. (2021). Culture, productivity and competitiveness: disentangling the concepts. Cross Cultural & Strategic Management, 28(1), 52-75.

[2] Yúdice, G. (2018). The globalization of culture and the new civil society. In Cultures of Politics/Politics of Cultures (pp. 353-379). Routledge.

[3] Collier, P. (2017). Culture, politics, and economic development. Annual Review of

Political Science, 20(1), 111-125.

[4] Sun, S. C. (2023). Confucius institutes: China’s cultural soft power strategy. Journal of Culture and Values in Education, 6(1), 52-68.

[5] Porta, A., Segre, G., & Arauzo-Carod, J. M. (2024). The Italian National Strategy for Inner Areas: first insights from regions’ specialization in cultural and creative industries. In Clusters in Times of Uncertainty (pp. 240-261). Edward Elgar Publishing.

[6] Khan, A. (2024). The Global Village: Intercultural Communication and Understanding in a Connected World. Liberal Journal of Language & Literature Review, 2(01), 13-22.

[7] Han, X., Goxe, F., & Freeman, S. (2024). Internationalization through social networks: A systematic review and future research agenda. International Business Review, 33(4), 102292.

[8] Chen, R., & Chen, X. (2025). Cultural sustainability through social networks: a moderated mediation model exploring the psychological dimensions of cultural dissemination. Frontiers in Psychology, 15, 1514693.

[9] Krishen, A. S., Leenders, M. A., Muthaly, S., Ziółkowska, M., & LaTour, M. S. (2019). Social networking from a social capital perspective: a cross-cultural analysis. European Journal of Marketing, 53(6), 1234-1253.

[10] Li, R. (2022). New Path of Stranger Interaction Platform and Network Cross-Cultural Communication in the Omnimedia Era. Wireless Communications and Mobile Computing, 2022(1), 6796160.

[11] Alsaleh, D. A., Elliott, M. T., Fu, F. Q., & Thakur, R. (2019). Cross-cultural differences in the adoption of social media. Journal of Research in Interactive Marketing, 13(1), 119-140.

[12] Sheldon, P., Rauschnabel, P. A., Antony, M. G., & Car, S. (2017). A cross-cultural comparison of Croatian and American social network sites: Exploring cultural differences in motives for Instagram use. Computers in human behavior, 75, 643-651.

[13] Kassymova, A., Tussupbekova, G., Sabyrbayeva, R., Zhumagulova, A., & Saganayeva, G. (2024). Structural-semantic and linguocultural aspects of vocabulary in social networks. Atlantic Journal of Communication, 1-16.

[14] Hong, M., An, S., Akerkar, R., Camacho, D., & Jung, J. J. (2024). Cross-cultural contextualisation for recommender systems. Journal of Ambient Intelligence and Humanized Computing, 1-12.

[15] Shengwei Sang, Mingyang Ma & Huanli Pang. (2024). Weighted Matrix Factorization Recommendation Model Incorporating Social Trust. Applied Sciences,14(2).

[16] ChaoWu & FengWu. (2025). Design and Implementation of Node Degree Centrality Computing of Network Security Database Based on Knowledge Graph. Security and Privacy,8(1),e473-e473.

[17] Mariam Elnour, Rachad Atat, Abdulrahman Takiddin, Muhammad Ismail & Erchin Serpedin. (2025). Eigenvector centrality-enhanced graph network for attack detection in power distribution systems. Electric Power Systems Research,241,111339-111339.