Research on the Influence Maximization Problem in Social Networks Based on the Multi-Functional Complex Networks Model

Sheng Bin, Qingdao University, China*

Gengxin Sun, Qingdao University, China

ABSTRACT

Most of the existing influence maximization problems in social networks only focus on single relationship social networks, that is, there is only one relationship in social networks. However, in reality, there are often many relationships among users of social networks, and these relationships jointly affect the propagation of network information and its final scope of influence. Based on the classical linear threshold model and combined with various relationships between network nodes, in this paper MRSN-LT propagation model is proposed to model the influence propagation process between nodes in multiple relationships social networks. Then, MRSN-RRset algorithm based on reverse reachable set is proposed to solve the problem of low computational performance caused by greedy algorithm in the research process of traditional influence maximization. Finally, the experimental results on real data sets show that the proposed method has better influence propagation scope and greater computational performance improvement.

KEYWORDS

Big Data, Complex Network, Greedy Algorithm, Influence Maximization, Multi-Functional Complex Networks Model, Propagation Model, Reverse Reachable Set, Social Network

1. INTRODUCTION

With the continuous development of social networks, the influence of social networks has attracted more and more attention. In the field of social network influence research, influence maximization is its research hotspot. The problem is actually an optimization problem based on a specific propagation model to find a set of initial propagation nodes in social network, through which the final propagation influence scope is the largest. Richardson et al. first proposed the influence maximization problem, and then many researchers conducted in-depth research on the problem from the aspects of influence propagation model and optimization algorithm (Nekovee et al., 2008; Wang et al., 2012; Yang et al., 2020; Yuan et al., 2022).

Influence propagation model can be used to describe the dynamic propagation process of influence in social networks. It is the premise and basis of studying the influence maximization in social networks. At present, the popular information propagation models include: infectious diseases model (Hethcote, 2000) independent cascade (IC) model (Goldenberg et al., 2001) and linear threshold (LT) model (Zhou et al., 2015). Although these common models describe the process of information

*Corresponding Author

propagation and diffusion in social networks to a certain extent, due to some limited conditions, these models are not very consistent with the influence propagation in real social networks. Therefore, many researchers proposed a large number of improved influence propagation models based on these models and combined with the real influence propagation in real social networks. Bazgan et al. extended the IC model and LT model (Bazgan et al., 2014), and made the two extended propagation models convert to each other. Wan et al. defined a dynamic function to represent the influence relationship between users, and proposed a fully cascaded information propagation model based on the influence relationship between users (Wan et al., 2019). Tian et al. started with the content of information propagation, introduced the topic of information propagation into the information propagation model, and pointed out that the influence among users is obviously different for different topics (Tian et al., 2020). Li et al. proposed the influence maximization model based on competition and promotion (Li et al., 2020). Chen et al. Proposed M-TLT propagation model based on topic perception (Chen et al., 2020).

Cristina et al. studied the influence maximization problem in social networks based on IC model and LT model, and proved that the problem is an NP-hard problem (Cristina et al., 2013), and a greedy algorithm with approximate solution of $1-1/e$ is designed to approximately solve the problem by selecting the nodes with the greatest marginal influence in the network. However, due to the low efficiency of greedy algorithm, it is not suitable for large-scale social networks. So researchers proposed many optimization algorithms for the computational performance of the algorithm. Leskovec et al. proposed Cost-Effective Lazy Forward (CELF) algorithm by using the submodular characteristics of the influence function between nodes. The algorithm reduces a lot of unnecessary calculations and greatly reduces the time complexity of greedy algorithm (Li et al., 2020). Goyal et al. improved the efficiency of CELF algorithm by using the submodularity and special data storage structure. Experiments show that the optimized CELF algorithm has a performance improvement of more than 17% (Goyal et al., 2011).

Most of the existing researches on influence maximization in social networks focused on single relationship social networks, that is, it is assumed that there is only one social relationship between network users. However, in reality, there are often a variety of relationships between users in social networks. For example, in the microblog system, according to the behavior of microblog users, there are at least four explicit relationships among microblog users: attention, reply, forwarding and reading. If the microblog content and the interactive behavior between microblog users are further analyzed, the interests and preferences of users would be detected, so as to various implicit relationships between users would be found, which shows that the social network in reality is a complex network with multiple relationships. In addition, there is a little interaction between these relationships. These relationships will jointly affect the propagation process of network information and the scope of its final influence. In fact, some researchers have paid attention to this problem. Phan et al. discussed the influence maximization problem when there are different relationships between mutual promotion and mutual inhibition of propagation in social network (Phan et al., 2016). Zhang et al. studied the influence maximization problem in multiple social networks (Zhang et al., 2016). However, social networks considering multiple relationships will be much more complex than single relationship social networks in the research process, and because there is no better network model that can describe the topology structure of multiple relationship networks, researchers often focus on the information propagation with specific relationships when solving the influence maximization of multiple relationships. For example, there is only competitive influence or only promotional influence. However, the actual network relationships are diverse, and the influence of different relationships is also different.

In terms of influence maximization problem of multiple social networks, existing studies (Singh et al., 2019; Lan et al., 2019) focus on the impact of parallel propagation among multiple social networks, these methods focus on the independent cascade model, and believe that the influence propagates in parallel between multiple channels, and the propagation process of nodes in multiple networks is not cross, which is inconsistent with the propagation of influence in multiple networks.

While in this study, propagation is carried out across multiple networks. Therefore, our study is more in line with the actual situation of influence propagation in the real Internet.

Based on the linear threshold model, combined with the various relationships between network nodes, and based on the multi-functional complex network model, in this paper the MRSN-LT propagation model is proposed to model the influence propagation between nodes. In this model, it is not necessary to consider whether there must be some specific interaction between different relationships of nodes. The model can be used to model whether various relationships promote each other's propagation (positive correlation) or inhibit each other's propagation (negative correlation), or even have no interaction. Finally, through the experiment on the real data set s, it is proved that the influence maximization in social network considering the joint influence of multiple relationships is more in line with the actual situation of network information propagation.

The main contributions of this study are as follows:

(1) Based on multi-functional complex network model, a multiple social network composited network consistent with the actual network environment is constructed, and the impact of various relationships in the composited network on influence propagation is studied.
(2) Based on the linear threshold model, combined with the various relationships between network nodes, a new propagation model is proposed to model the influence propagation process between nodes in multiple relationships networks.
(3) A new algorithm based on reverse reachable set is proposed to solve the problem of low computational performance caused by greedy algorithm in the computation process of traditional influence maximization.

2. MULTI-FUNCTIONAL COMPLEX NETWORK MODEL

In the existing influence maximization research, considering the influence maximization of multiple relationships among users in the actual network, the research progress is slow, and the main reason is that there is no suitable complex network model that can describe the multi relationship network. The multi-functional complex network model is just such a complex network model that can well describe the various relationships between different kinds of individuals in network.

2.1 Model Definition

Definition 1 (Feature Attribute) For multi-functional complex networks, feature attributes are used to describe some attributes of nodes in the network, it is denoted as P_h. S_i represents feature attribute set of node $v_i \in V \left(\text{i} = 1,2,...,|V| \right)$, $P = U_{\text{i}=1}^{|V|} S_i = \{P_1, P_2, ..., P_m\}$ is feature attribute set of all nodes in the network, m is the number of feature attributes in the network. The value range of feature attribute P_h is marked as $\text{dom}\left(P_h\right)$, the value of node v_i under feature attribute P_h is marked as $p_{ih} \epsilon \text{dom}\left(P_h\right)$.

Definition 2 (Feature Attribute Mapping Function) If a certain association is established between nodes v_i and v_j based on the feature attribute P_h, it can be realized by the feature attribute mapping function $f_h\left(v_i, v_j\right)$.

Definition 3 (Feature Attribute Set Mapping Function) If some association is established between nodes v_i and v_j based on multiple feature attributes (feature attribute set P^* ($P^* \subseteq P$)), it can be realized by feature attribute set mapping function $F\left(v_i, v_j\right)$. $F\left(v_i, v_j\right) = \xi f_1, f_2, ..., f_t, \quad t \le m$

represents the mapping function corresponding to the given feature attribute set P^* in the multi-functional network.

Definition 4 (Multi-Functional Complex Network, MFCN) Multi-functional complex network is composed of nodes with some feature attributes and mapping function of feature attribute set, it is denoted as $G(V, P, F)$, where V is node set of multi-functional network, $V = \underset{i=1}{\overset{|V|}{U}} v_i$, $|V|$ is the rank of the set, it represents number of nodes in the network. For $\forall v_i \in V(i = 1, 2, ..., |V|)$, $p_i = (p_{i1}, p_{i2}, ..., p_{ih}, ..., p_{im})$ is m-dimensional feature attribute vector of node v_i, where $p_{i1}, p_{i2}, ..., p_{im}$ represents feature attribute component of node v_i, they represent the value of node v_i under m feature attributes. F represents the mapping function of the multi-functional network based on the selected feature attribute set. P is the feature attribute set of all nodes in the network. For a multi-functional network with m feature attributes, it can be described by a $|V| \times m$ matrix as follows:

2.2 Matrix Representation of Multi-Functional Complex Networks

In Multi-functional complex network $G(V, P, F)$, for $\forall v_i, v_j \in V(i = 1, 2, ..., |V|)$, $< v_i, v_j >_h$ represents the association established by feature attribute P_h, it can be an ordered pair or a disordered pair. An ordered pair indicates that the association has a direction, and a disordered pair indicates that the association has no direction. Its association is realized through the mapping function $f_h(v_i, v_j)$. Based on the feature attribute mapping function, an adjacency matrix $A_h = (a_{ij})_h$ whose rank is $|V| \times |V|$ can be usedl to represents the association relationship between nodes under this network structure, a_{ij} represents the element in row i and column j of adjacency matrix A_h:

$$a_{ij} = \begin{cases} w_h, & if < v_i, v_j >_h \ exist \\ 0, & else \end{cases}, 1 \le i, j \le |V|, 1 \le h \le m$$

where w_h represents the association weights of nodes v_i and v_j under the mapping function f_h of feature attribute P_h. If the weight is not considered, then

$$a_{ij} = \begin{cases} 1, & if < v_i, v_j >_h \ exist \\ 0, & else \end{cases}, 1 \le i, j \le |V|, 1 \le h \le m$$

When the network node has m feature attributes, the multi-functional network G under the mapping function of all feature attributes has m feature matrices whose rank is $|V| \times |V|$: $A_1, A_2, ..., A_m$, where $A_1 = (a_{ij})_1$ represents the adjacency matrix under the mapping function f_1 of feature attribute P_1, $A_m = (a_{ij})_m$ represents the adjacency matrix under the mapping function f_m of feature attribute P_m.

When determining the feature attribute set mapping function F of the corresponding function, $F(v_i, v_j) = \frac{3}{4} f_1, f_2, ..., ..., f_t, \quad t \in [1, m]$, based on the selected feature attribute set mapping function,

the network topology of multi-functional network would be obtained, then $G\left(V,P,F\right)$ can be represented by a matrix under the selected feature attribute mapping function, its node relevance under the mapping function of the corresponding feature attribute set is expressed as matrix A, $A = \tilde{\frac{3}{4}}A_1, A_2,\ldots,\ldots,A_t, t \in \left[1, m\right]$. Especially, $F\left(v_i, v_j\right) = \frac{3}{4}f_t, t \in \left[1, m\right]$, $G\left(V,P,F\right)$ can be represented by the adjacency matrix A_t under the feature attribute mapping function f_t, namely, $A = \tilde{\xi}A_t, t \in \left[1, m\right]$.

3. INFLUENCE MAXIMIZATION OF MULTIPLE RELATIONSHIPS SOCIAL NETWORKS BASED ON MRSN-LT MODEL

3.1 MRSN-LT Propagation Model

LT model is one of the classical propagation models. However, it can only describe the network propagation between social network nodes with a single relationship. In order to better model the influence maximization of real multi relationship network, a multiple relationship MRSN-LT model based on LT model is proposed in this paper.

Firstly, through the multi-functional complex network model, multiple single relationship social networks can be combined into a multiple relationship composited network. An example of multiple relationship network construction is shown as Figure 1:
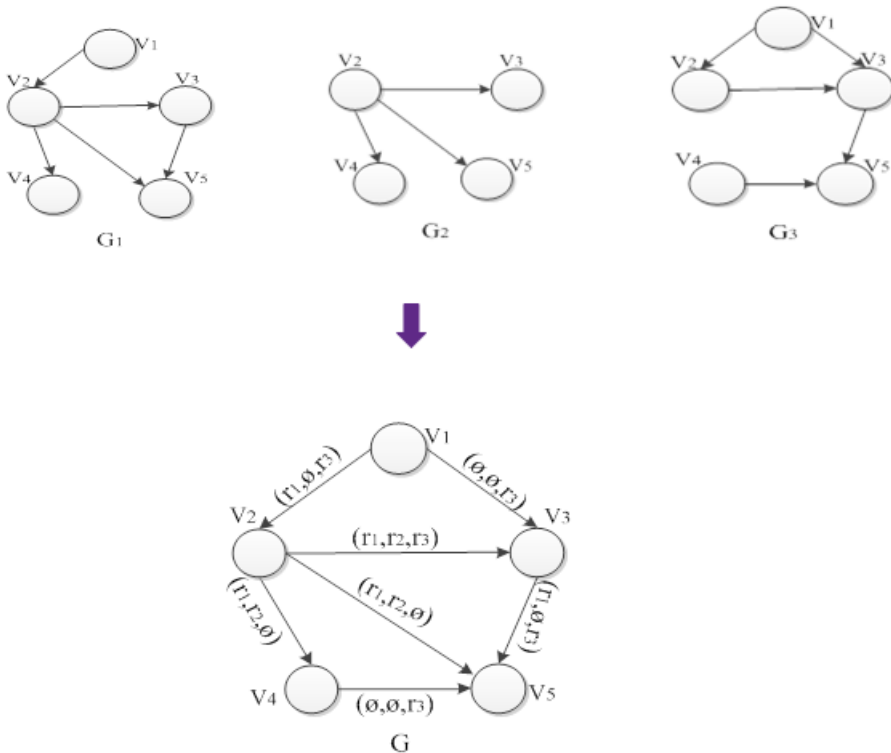
Figure 1. An example of multiple relationship network construction

Figure 1 shows the process of combining three single relationship network G_1, G_2 and G_3 into a multiple relationship network G. The three single relationship networks have r_1, r_2 and r_3 relationships, respectively. The relationship intensities of the three relationships are $\mathrm{dom}\left(r_1\right)$, $\mathrm{dom}\left(r_2\right)$ and $\mathrm{dom}\left(r_3\right)$.

In MRSN-LT model, social network is represented as a directed graph $G = \left(V, E\right)$, V is the set of all nodes in the network, $E \subseteq V \times V$ is the set of all edges in the network. All nodes in Figure 2 have two initial states: active state and inactive state, and the state transition of nodes is determined by the transition probability on the edge.

Figure 2. Schematic diagram of node state transition in MRSN-LT model
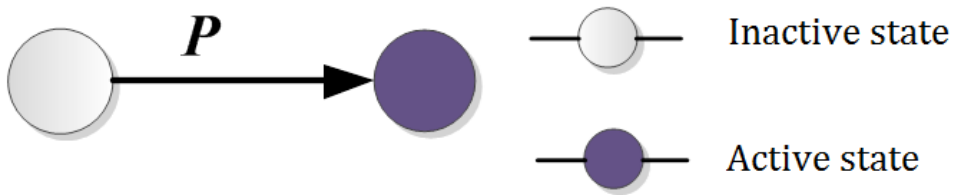
In Figure 2, P represents state transition probability, it can be calculated as follows:

$$P = C^{-1}\mathcal{P}$$

where C is a parameter matrix that can describe the intensity of influence propagation caused by different relationships, $\mathcal{P} = \left(p_1, p_2, \ldots p_i, \ldots, p_n\right), 1 < i < n, \; p_i \in \left[0,1\right]$ is influence probability matrix of various relationships, p_i represents the influence propagation probability in network G_i with only a single relationship r_i. In the process of propagation, the importance of different relationships to network propagation is different, and there will be some interaction between different relationships. That is, each relationship has different influence on the scope of influence propagation. As shown in Figure 1, there are three different relationships in the multiple relationship network G, that is, $i = 3$, then $C = \begin{pmatrix} \alpha & \beta & \gamma \end{pmatrix}^T$, the following formula can be obtained:

$$P = C^{-1}\mathcal{P} = \begin{pmatrix} \alpha & \beta & \gamma \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

Like LT model, MRSN-LT model is also an influence accumulation model. If INv is used to represent the set of active nodes in all neighbor nodes pointing to node v. Then the condition for node v to be activated is $INv \neq \varnothing$ and $\sum_{u \in INv} P_{u\,v} \geq \theta_v$. Combined with the state transition probability, the conditions for the successful activation of node v can be obtained as follows:

$$INv \neq \varnothing \quad \text{and} \quad C^{-1} \sum_{u \in INv} \mathcal{P}_{u\ v} \geq \theta_v$$

Therefore, assuming that the initial seed node set is S, the propagation process of MRSN-LT model can be simply described as follows:

(1) Multiple single relational networks are combined into a multiple relationships composited network through a multi-functional complex network modal.
(2) A threshold $\theta_v \in [0,1]$ is selected randomly for all nodes.
(3) At each time step $t\,(t > 0)$, for all nodes $vv \in V \setminus S$ that have not been activated after time step t -1, if $INv \neq \varnothing$ and $C^{-1} \sum_{u \in INv} \mathcal{P}_{u\ v} \geq \theta_v$, then node v is activated, otherwise node v is not activated.
(4) When no new inactive node in the network is activated, the information propagation ends.

An example of influence propagation on MRSN-LT model is shown in Figure 3.

Figure 3. Example of influence propagation process on MRSN-LT model
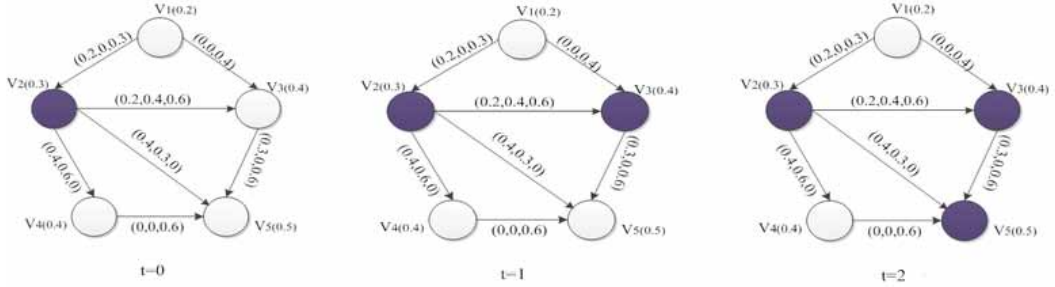
3.2 Influence Maximization Algorithm Based on MRSN-LT Model

σ_S represents the influence scope of the seed set S, then the influence maximization of MRSN-LT model can be described as: For a given social network graph $G = (V, E)$ and a given constant $k \leq |V|$, find a set S so that when $|S| = k$, σS is the largest. The influence maximization problem of social networks based on IC model and LT model had been proved that its influence propagation function $\sigma(\cdot)$ is monotonic and submodular. For any set $S \subseteq T \subseteq V$, and node $u \in V \setminus T$, the inequality $\sigma\left(S \cup \{u\}\right) - \sigma\left(S\right) \geq \sigma\left(T \cup \{u\}\right) - \sigma\left(T\right)$ holds, and a greedy algorithm with approximate solution of $1 - 1/e$ is designed. The algorithm is defined as follows:

Algorithm 1: Greedy(G, k)
Input: $G = V, E, P, k$
Output: Seed node set S ($|S| = k$)
$S \leftarrow \varnothing$
$\sigma(S) \leftarrow \varnothing$
while $|S| \leq k$ do

$$v \leftarrow \arg\max_{u \in V-S} \left(\tilde{A}(S+u) - \sigma(S) \right)$$
$$S \leftarrow S + v$$

return S

MRSN-LT model considers the influence of multiple relationships on propagation and the relationship strength of multiple relationships. The propagation mechanism is not much different from LT model. Therefore, MRSN-LT model is still an influence accumulation model. The conditions for successfully activating a node in the process of influence propagation are the same as those of ordinary LT model. Therefore, it can be proved that under MRSN-LT model, the influence propagation function $\sigma(S)$ also satisfies monotonicity and submodularity. Therefore, greedy algorithm can also be used to approximately maximize the influence of multiple relationships network under MRSN-LT model.

Algorithm 2: MRGreedy(G, k)

Input: G_1, G_2, ...,G_i ,k

Output: Seed node set S ($\left|S\right| = k$)

Generate multiple relationships network G, and obtain the parameter matrix C

$S \leftarrow \varnothing$

$\sigma(S) \leftarrow \varnothing$

while $\left|S\right| \leq k$ do

$\qquad v \leftarrow \arg\max_{u \in V-S} \left(\sigma(S+u) - \sigma(S) \right)$

$\qquad S \leftarrow S + v$

return S

Greedy algorithm is the most popular algorithm in the research of influence maximization, but greedy algorithm also has some shortcomings. The most typical problem is the time complexity of the algorithm. Because the greedy algorithm needs to calculate the marginal benefits of all nodes every time it selects the most influential node, the time efficiency of the algorithm is very low. In the actual network, there is often a huge amount of data. In this way, the greedy algorithm is not a good choice. In this paper, MRSN-RRset algorithm is proposed to maximize the influence of MRSN-LT model based on reverse reachable set method. MRSN-RRset algorithm is divided into two steps:

Step 1: Randomly select n nodes to generate their reverse reachable set, reverse reachable set RRv of node v represents the set composed of all nodes that can reach node v in a multiple relationships network, and n sets together constitute the final reverse reachable set RR.

Step 2: The seed node set is selected by the maximum coverage method, because if node u appears in the reverse reachable set of node v, there must be an reachable path $\text{Path} u \rightarrow v$ from node u to node v. Therefore, the more nodes in the reverse reachable set at the same time means that the node can activate more other nodes.

Algorithm 3: MRSN-RRset algorithm

Input: G_1, G_2, ...,G_i ,k

Output: Seed node set S ($\left|S\right| = k$)

Generate multiple relationships network G, and obtain the parameter matrix C

$S \leftarrow \varnothing$

$RR \leftarrow \varnothing$

$For\ i = 1 2 \ldots, n\ do$

$\qquad$ Select node v_i randomly from G

$\qquad$ Obtain reverse reachable set RRv_i of node v_i

$$RR = RR \cup RRv_i$$
$$For \ i = 1 \, 2 \, ..., k \ \ do$$

```
        Select the node v with the largest coverage RR
          S ← S + v
        Delete the set containing node v
return S
```

MRSN-RRset algorithm mainly includes two parts: generating reverse reachable set and selecting seed nodes. When generating reverse reachable sets, the time complexity is $O(n)$ because n nodes are randomly selected. For any selected node v_i, the time complexity of finding RRv_i through breadth first algorithm or depth first algorithm is $O(m)$, where m represents the number of edges pointing to node v_i. If the expected number of edges pointing to the nodes in the generated RR set is ESV, the time complexity of generating the reverse reachable set RR is $O(n) * O(ESV) = O(n * ESV)$. The method of selecting seed nodes through maximum coverage uses the idea of typical greedy algorithm, which has linear time complexity. Therefore, the time complexity of MRSN-RRset algorithm is $O(n * ESV)$. The time complexity of greedy algorithm is $O(k * |V| * |E| * M)$, where M represents Monte-Carlo times, generally more than 10000 times. It can be seen that MRSN-RRset algorithm has great advantages in time complexity compared with greedy algorithm.

4. EXPERIMENTAL ANALYSIS

In this paper MRSN-RRset algorithm based on MRSN-LT model is proposed. In the experiment, the proposed algorithm and two commonly used influence maximization algorithms CELF algorithm and IRIE algorithm (Wu et al., 2021) are compared. Since CELF algorithm and IRIE algorithm are only applicable to the influence maximization problem of single relational network, a multi-functional complex network model is used to integrate multiple relationships into a composited relationship in the network, select different seed sets, simulate the propagation process in the multiple relationships network, and evaluate the performance of the algorithm from the two aspects of running time and influence scope.

4.1 Experimental Data Set

The influence maximization analysis in multiple relationships network needs to consider the influence of various relationships between nodes on propagation. In the experiment, the collected Douban data set is used to construct a multiple relationships network. The nodes in the network represent the users in Douban, which is an innovative social networking site integrating taste system (reading, film, music), expression system (I read, I watch, I listen) and communication system (same city, group, friends and neighbors). The relationships between users include attention relationship (AR), friend relationship (FR) and comment similarity relationship (CSR). The statistical information of the data set is shown in Table 1.

Table 1. Statistical information of Douban multiple relationships dataset

The number of nodes	The number of edges	AR	FR	CSR
58691	757621	632651.	572341	93278

In the multiple relationships network constructed by using the experimental data set, it can be seen that the propagation probability is jointly determined by three relationships. In order to quantitatively determine the relationship strength of the three relationships in the propagation influence, 300 groups of sample data can be selected in the data set according to a specific topic or label and the chronological order of the posts published by the user as the basis for information propagation, the relationship strength coefficient matrix corresponding to the three relationships is calculated by linear regression analysis method.

For the attention relationship r_1, friend relationship r_2 and comment similarity relationship r_3 in the constructed multiple relationships network, relational strength coefficient matrix $C = \begin{pmatrix} \alpha & \beta & \gamma \end{pmatrix}^T$, the state transition probability P can be defined as follows:

$$P = C^{-1}\mathcal{P} = \begin{pmatrix} \alpha & \beta & \gamma \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$$

Therefore, the ternary linear regression model is constructed as follows:

$$P = \alpha p_1 + \beta p_2 + \gamma p_3$$

where $\pm$, β and γ are regression coefficients, The least square method is used for parameter estimation, and the equations are established as follows:

$$\sum P = \alpha \sum p_1 + \beta \sum p_2 + \gamma \sum p_3$$
$$\sum p_1 P = \alpha \sum p_1^2 + \beta \sum p_1 p_2 + \gamma \sum p_1 p_3$$
$$\sum p_2 P = \alpha \sum p_1 p_2 + \beta \sum p_2^2 + \gamma \sum p_2 p_3$$
$$\sum p_3 P = \alpha \sum p_1 p_3 + \beta \sum p_2 p_3 + \gamma \sum p_3^2$$

According to 300 sets of sample data, the average value of parameter estimation obtained by solving the equations is $\pm = 0.616$, $\beta = 1.185$ and $\gamma = 0.982$.

4.2 Analysis of Experimental Results

In order to compare different influence maximization algorithms, according to the selected initial seed set, based on MRSN-LT model, Monte Carlo simulation (Wang et al., 2022) is used to evaluate the influence propagation scope. For each selected initial seed set, 5000 times of Monte Carlo is used to simulate the propagation process of information in multiple relationship network, and the average value is used as the propagation scope of influence.

Firstly, under the condition of fixed initial seed set, the influence propagation scope of the three influence maximization algorithms are compared when the values of threshold θ_v in MRSN-LT model are 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3 respectively. Figure 4 shows the comparison of influence propagation scope of CELF algorithm, IRIE algorithm under LT model and MRSN-RRset algorithm under MRSN-LT model under different thresholds when the number of nodes in the initial seed set is 100.

When the number of nodes in the initial seed set is 200, the comparison of influence propagation scope of the three algorithms is shown in Figure 5.

Figure 4. The relationship between the influence propagation scope of the three algorithms and different thresholds (The number of seed set nodes is 100)
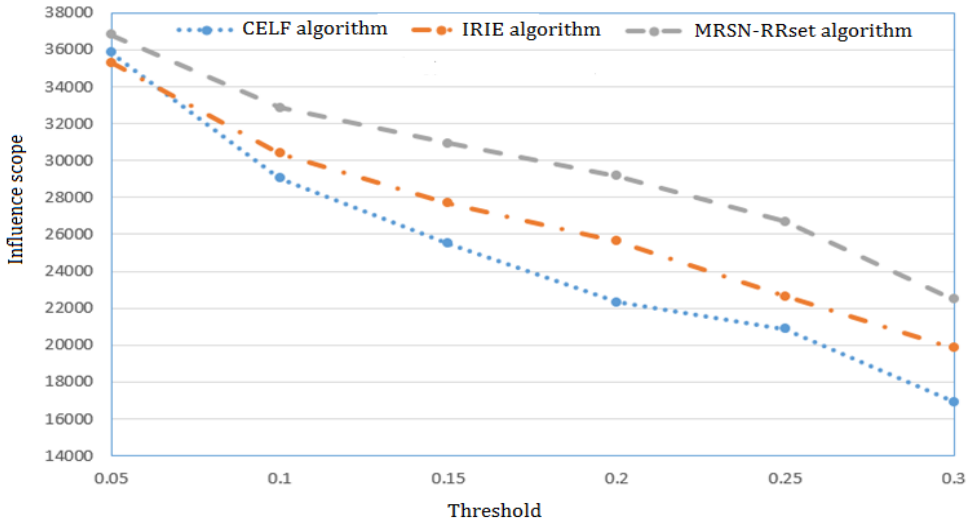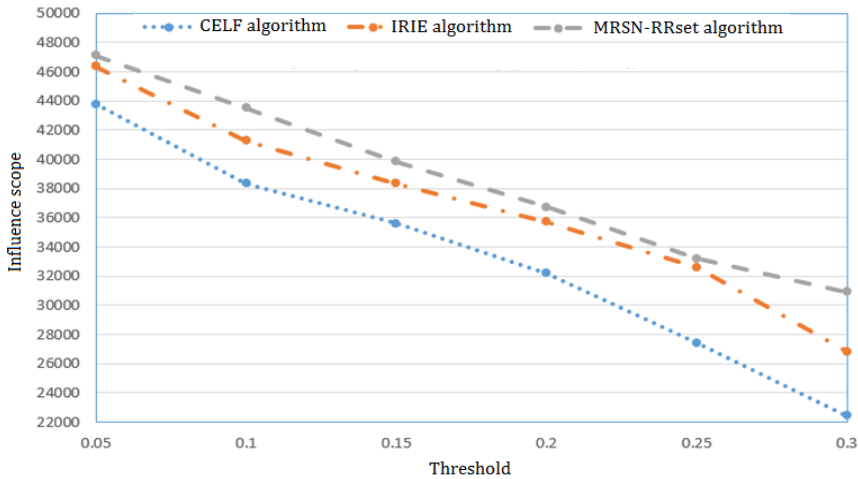

Figure 5. The relationship between the influence propagation scope of the three algorithms and different thresholds (The number of seed set nodes is 200)

As can be seen from Figure 4 and Figure 5, the influence propagation scope of the three algorithms decreases with the increase of the threshold. It is mainly because of the threshold θ_v represents the

difficulty of affecting nodes in MRSN-LT model. The higher the threshold, the greater the influence required to activate a node in an inactive state. In addition, under the same threshold, the influence scope will increase with the increase of the number of nodes in the initial seed node set. At the same threshold, when the initial seed node set is the same, the propagation scope of the proposed algorithm is significantly larger than that of the other two algorithms.

When the values of threshold θ_v in MRSN-LT model are 0.05, 0.1, 0.15, 0.2, 0.25 and 0.3 respectively, the running time of the three influence maximization algorithms are also compared under the condition of fixed initial seed.

Figure 6 and Figure 7 are the comparison of the running times of the three algorithms when the number of nodes in the initial seed node set is 100 and 200, respectively.

Figure 6. The relationship between the running time of the three algorithms and different thresholds (The number of seed set nodes is 100)
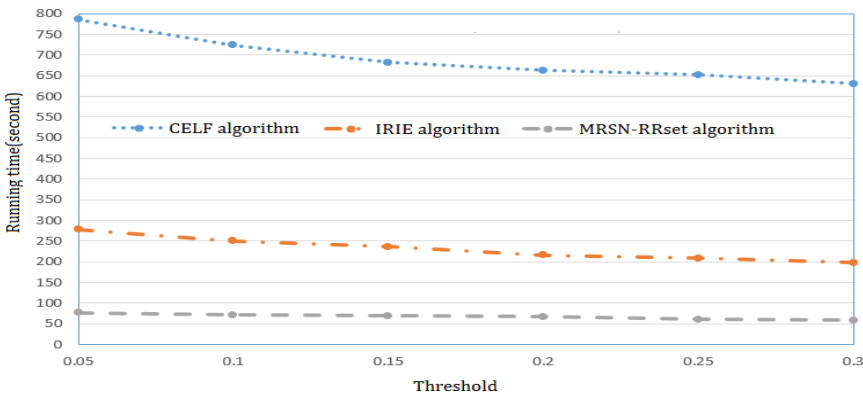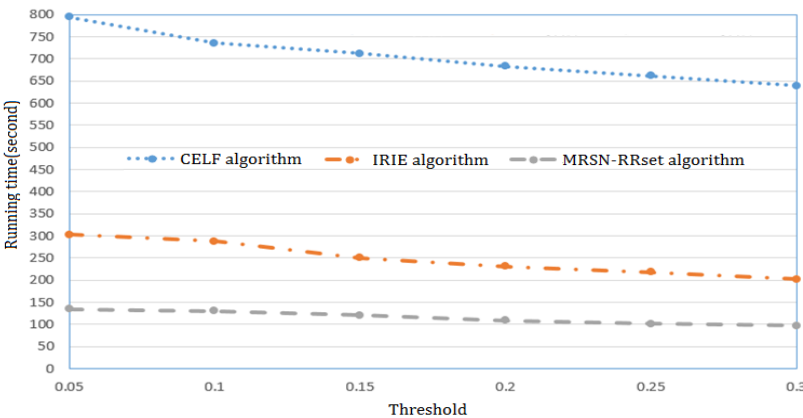

Figure 7. The relationship between the running time of the three algorithms and different thresholds (The number of seed set nodes is 200)

As can be seen from Figure 6 and Figure 7, the running time of the three algorithms decreases with the increase of the threshold. It is mainly because the larger the threshold, the smaller the candidate seed set of greedy algorithm. Under the same threshold and the same initial seed node set, the running time of the proposed algorithm is significantly less than that of the other two algorithms, and has better performance. Moreover, the proposed algorithm is not affected by the threshold, which shows that the seed node set selected by the maximum coverage method is basically not affected by the threshold θ_v. It is proved that the proposed algorithm has good applicability.

Traditional greedy algorithms calculate the influence range of a given node set on the network through multiple Monte Carlo simulations. The proposed algorithm uses the reverse reachability set to select the seed node set. According to the above experimental analysis, it can be seen that the proposed algorithm has obvious improvement over the traditional greedy algorithm in terms of influence propagation scope and algorithm running time.

In order to verify the influence of multiple relationships on influence propagation, the initial seed set size was set to 100 in the experiment, and friend relationship, two relationships (friend relationship and attention relationship), all three relationships in the data set were selected respectively. The MRSN-RRset algorithm based on MRSN-LT model was used to compare the influence propagation scope. Comparison of influence propagation scope of multiple relationships is shown in Figure 8.

Figure 8. Comparison of influence propagation scope of multiple relationships
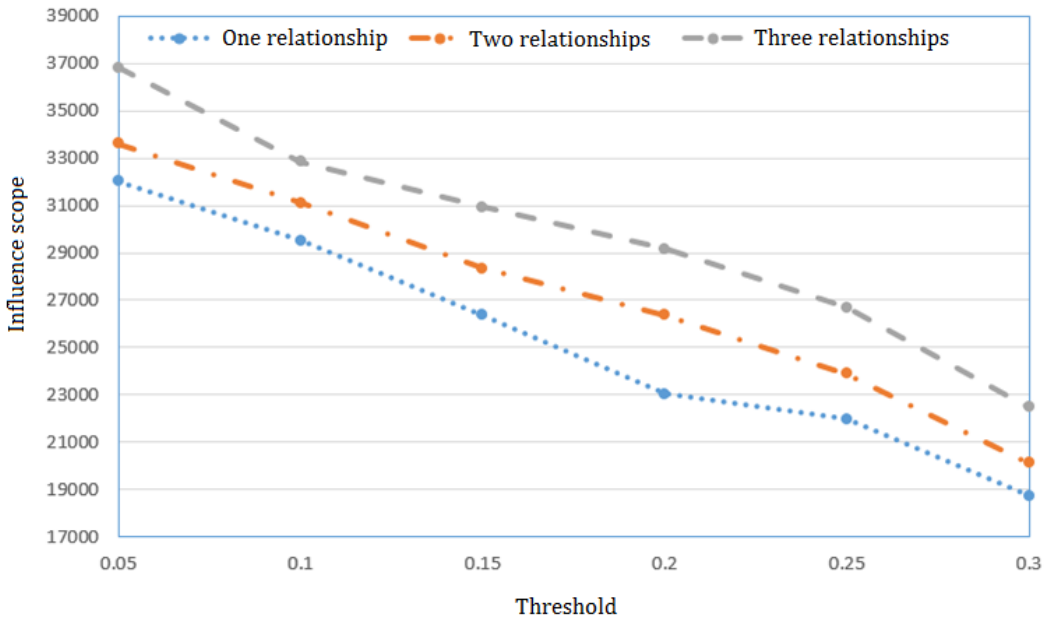

As can be seen from Figure 8, selecting all three relationships in the multiple relationships network and using the MRSN-RRset algorithm based on MRSN-LT model for influence propagation has a larger influence propagation scope than selecting two or one relationship under the same initial seed node set and the same threshold. It is because the three relationships in the data set promote each other and are positively correlated. Therefore, the more relationships selected, the more conducive to the expansion of the scope of influence propagation.

In order to further illustrate the influence of multiple relationships on the propagation of influence, the effects of multiple relationships of mutual promotion (positive correlation) and mutual exclusion (negative correlation) on the propagation of influence are verified by simulation in the experiment.

In the simulation experiment, the total number of nodes in the initial seed set is 50 and the total number of nodes in the whole propagation group is 10000. A network with two positive correlation relationships ($\pm = 1$, $\beta = 1$), a network with a single relationship and a network with two negative correlation relationships ($\partial = 1$, $\beta = -1$) are selected for simulation. The influence propagation scope of MRSN-RRset algorithm based on MRSN-LT model under different thresholds is shown in Figure 9.

Figure 9. Comparison of the correlation of multiple relationships on the scope of influence propagation
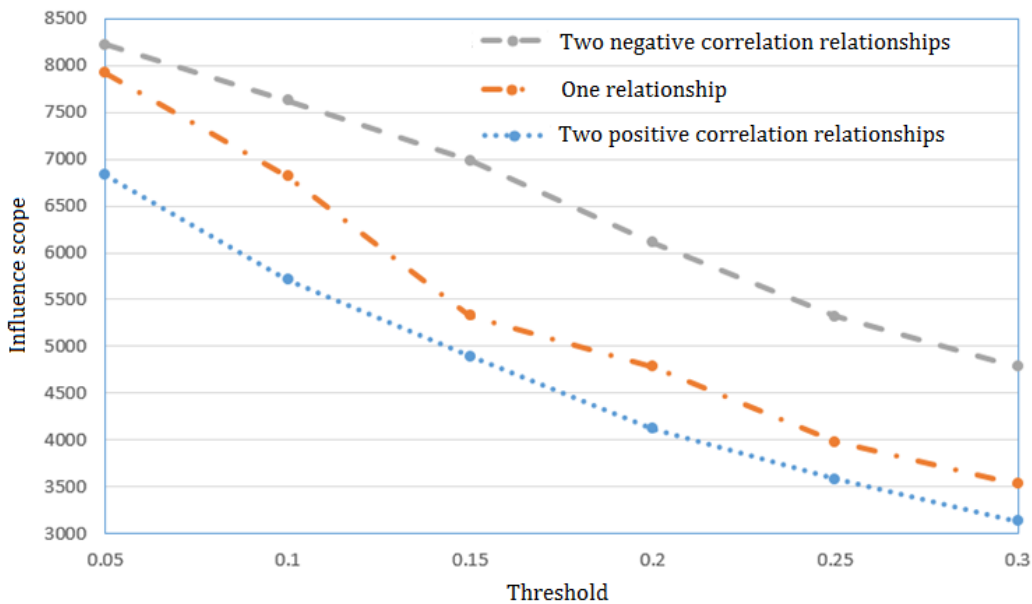
As can be seen from Figure 9, when the two relationships in the network are positively correlated, the scope of influence propagation will be increased, while when the two relationships are negatively correlated, the scope of influence propagation will be weakened, so as to reduce the scope of influence propagation.

CONCLUSION

In this paper, the influence maximization problem on multiple relationships social networks is studied. Compared with traditional social networks, many kinds of relationships in multiple relationships social networks have different effects on information propagation in the network. According to this characteristic, MRSN-LT propagation model in multiple relationships social networks is proposed. In order to improve the computational efficiency of the influence maximization algorithm, MRSN-RRset algorithm based on reverse reachable set is proposed based on MRSN-LT model. The experimental results on real data sets show that compared with the current popular influence maximization algorithms, the proposed algorithm can achieve ideal results in influence scope and running time.

This paper studies the impact of various relationships between network users in social networks on influence propagation. In the future research, we will focus on the maximization of social network influence when there are multiple information sources spreading simultaneously in the network.

FUNDING AGENCY

ACKNOWLEDGMENT

REFERENCES

Bazgan, C., Chopin, M., Nichterlein, A., & Sikora, F. (2014). Parameterized Approximability of Maximizing the Spread of Influence in Networks. *Journal of Discrete Algorithms*, *27*, 54–65. doi:10.1016/j.jda.2014.05.001

Chen, S., Yin, X., Cao, Q., Li, Q., & Long, H. (2020). Targeted Influence Maximization Based on Cloud Computing Over Big Data in Social Networks. *IEEE Access: Practical Innovations, Open Solutions*, *8*, 45512–45522. doi:10.1109/ACCESS.2020.2978010

Cristina, B., Morgan, C., & André, N. (2013). Parameterized Approximability of Maximizing the Spread of Influence in Networks. *Journal of Discrete Algorithms*, *27*, 54–65. doi:10.1007/978-3-642-38768-5_48

Goldenberg, J., & Muller, L. E. (2001). Talk of the Network: A Complex Systems Look at the Underlying Process of Word-of-Mouth. *Marketing Letters*, *12*(3), 211–223. doi:10.1023/A:1011122126881

Goyal, A., Lu, W., & Lakshmanan, L. V. S. (2011). CELF++: Optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th International Conference on World Wide Web* (pp. 47-48). ACM. doi:10.1145/1963192.1963217

Hethcote, H. W. (2000). The Mathematics of Infectious Diseases. *SIAM Review*, *42*(4), 599–653. doi:10.1137/S0036144500371907

Lan, Y., Zhiwu, L., & Alessandro, G. (2019). Influence minimization in linear threshold networks. *Automatica*, *100*, 10–16. doi:10.1016/j.automatica.2018.10.053

Leskovec, J., Krause, A., & Guestrin, C. (2007). Cost-effective outbreak detection in networks. In *Proceedings of the 13th International Conference on Knowledge Discovery and Data Mining* (pp. 420-429). ACM.

Li, A., Yl, B., & Qz, A. (2020). Targeted influence maximization under a multifactor-based information propagation model. *Information Sciences*, *519*, 124–140. doi:10.1016/j.ins.2020.01.040

Nekovee, M., Moreno, Y., Bianconi, G., & Marsili, M. (2008). Theory of Rumour Spreading in Complex Social Networks. *Physica A*, *374*(1), 457–470. doi:10.1016/j.physa.2006.07.017

Phan, N., Ebrahimi, J., Kil, D., Piniewski, B., & Dou, D. (2016). Topic-Aware Physical Activity Propagation in a Health Social Network. *IEEE Intelligent Systems*, *31*(1), 5–14. doi:10.1109/MIS.2015.92 PMID:27087794

Singh, S., Singh, K., Kumar, A., & Biswas, B. (2019). MIM2: Multiple influence maximization across multiple social networks. *Physica A*, *526*, 15–27. doi:10.1016/j.physa.2019.04.138

Tian, S., Mo, S., Wang, L., & Peng, Z. (2020). Deep Reinforcement Learning-Based Approach to Tackle Topic-Aware Influence Maximization. *Data Science and Engineering*, *5*(1), 1–11. doi:10.1007/s41019-020-00117-1

Wan, J., Chen, X., Du, Y., & Jia, M. (2019). Information propagation model based on hybrid social factors of opportunity, trust and motivation. *Neurocomputing*, *333*, 169–184. doi:10.1016/j.neucom.2018.12.062

Wang, C., Chen, W., & Wang, Y. (2012). Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery*, *25*(3), 545–576. doi:10.1007/s10618-012-0262-1

Wang, J., & Dai, Y. (2022). The Agglomeration Mechanism of Network Emerging E-Eommerce Industry Based on Social Science. *Journal of Organizational and End User Computing*, *34*(3), 1–16. doi:10.4018/JOEUC.291561

Wu, W., Ma, S., Su, Y., & Wu, C. (2021). Double-Layer Learning, Leaders' Forgetting, and Knowledge Performance in Online Work Community Organizations. *Journal of Organizational and End User Computing*, *33*(1), 92–117. doi:10.4018/JOEUC.2021010105

Yang, W., Zhang, Y., & Du, D. Z. (2020). Influence maximization problem: Properties and algorithms. *Journal of Combinatorial Optimization*, *40*(2), 907–928. doi:10.1007/s10878-020-00638-5

Yuan, Y., Yeh, Y., Wu, C., Liu, C., Chen, H., & Chen, C. (2022). The Study of Para-Social Interaction With E-Word-of-Mouth for Influencer Marketing by Complex Computing. *Journal of Organizational and End User Computing*, *34*(3), 1–15. doi:10.4018/JOEUC.287105

Zhang, H., Nguyen, D. T., Zhang, H., & Thai, M. T. (2016). Least Cost Influence Maximization Across Multiple Social Networks. *IEEE/ACM Transactions on Networking*, *24*(2), 929–939. doi:10.1109/TNET.2015.2394793

Zhou, F., Jiao, J., & Lei, B. (2015). A linear threshold-hurdle model for product adoption prediction incorporating social network effects. *Information Sciences*, *307*, 95–109. doi:10.1016/j.ins.2015.02.027