Article

Substitute Seed Nodes Mining Algorithms for Influence Maximization in Multi-Social Networks

Xuli Rao [1,2]**, Jiaxu Zhao** [1,]*** , Zhide Chen** [2] **and Feng Lin** [1]

[1]   Department of Computer Science, Fuzhou Polytechnic, Fuzhou 350108, Fujian, China;
     xulirao@fvti.edu.cn (X.R.); linfeng-fvti@foxmail.com (F.L.)

[2]   College of Mathematics and Informatics, Fujian Normal University, Fuzhou 350007, Fujian, China;
     zhidechen@fjnu.edu.cn

*   Correspondence: jiaxuzhao@foxmail.com; Tel.: +86-18046070354

check for updates

Abstract: Due to the growing interconnections of social networks, the problem of influence maximization has been extended from a single social network to multiple social networks. However, a critical challenge of influence maximization in multi-social networks is that some initial seed nodes may be unable to be active, which obviously leads to a low performance of influence spreading. Therefore, finding substitute nodes for mitigating the influence loss of uncooperative nodes is extremely helpful in influence maximization. In this paper, we propose three substitute mining algorithms for influence maximization in multi-social networks, namely for the Greedy-based substitute mining algorithm, pre-selected-based substitute mining algorithm, and similar-users-based substitute mining algorithm. The simulation results demonstrate that the existence of the uncooperative seed nodes leads to the range reduction of information influence. Furthermore, the viability and performance of the proposed algorithms are presented, which show that three substitute node mining algorithms can find suitable substitute nodes for multi-social networks influence maximization, thus achieves better influence.

Keywords: multi-social networks; influence maximization; substitute mining algorithm

---

1. Introduction

In recent years, various social networks [1] offering different services have appeared on the Internet, e.g., Facebook, Twitter, YouTube and Foursquare. For the purpose of enjoying different services, some users register accounts on different social networks simultaneously. Different from the traditional connections with other users, they have not only social links among users within one social network, but also crossing links connecting accounts between different social networks. Multi-social networks are connected by these shared users [2]. Due to their obvious advantages in spreading information, as exemplified by the high speed, low cost and wide influence, social networks attract more and more attention of researchers, which results in many hot topics, involving social information spreading [3], public opinion [4], Internet marketing [5] and so on. The influence maximization in a social network has been extensively studied in the academic community [6–10], which is the problem of finding a small subset of nodes in a social network that could maximize the spread of influence [7]. However, the existing influence maximization research for the single social network cannot be directly adapted to the multi-social networks due to their heterogeneous architecture and complex connections among users. The users with multiple social network accounts can spread information from social network A to social network B. An example is given in Figure 1, a user who is registered on Twitter and Facebook has the ability to forward the information on Twitter to Facebook, which demonstrates the scope of information spreading no longer confined within the single social network. This factor

leads to the different characteristics of information spreading as before and needs to be considered in the influence of maximization research.
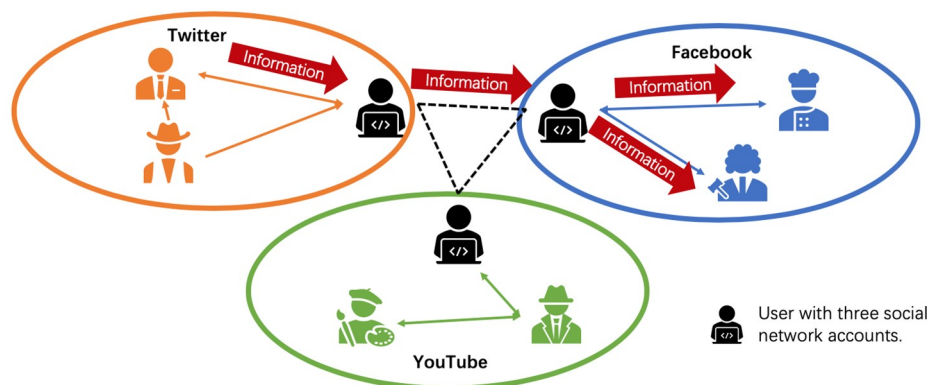

Figure 1. User with multiple accounts forwards information across social networks.

To maximize the influence spreading in multi-social networks, the researchers in References [11–16] focus on finding k ($k \in N^*$) users to constitute the initial active seed node set S to spread information, the expected number of active nodes $R(S)$ at the end of a diffusion process in multi-social networks can be maximized. Due to budgetary constraints, the size of S is usually fixed in existing research, i.e., k is fixed. However, when we get S through complex theoretical calculations, some initial seed nodes in the set S may be difficult to activate for some reasons. In the end, a broken S cannot get the maximum influence, resulting in the low influence spreading performance. For example, in social network advertising marketing, one user has high influence in social networks, but for some reasons, he is unwilling to post advertisements, which may lead to advertising marketing failure. Therefore, when some users in S cannot be activated, it is very necessary to find substitutes for these nodes to reduce the loss of influence.

Some researchers have already done some work on this subject. Li et al. [17] proposed the idea of finding "successors". When a seed node cannot be activated, the neighbor node is selected as a substitute. Ma et al. [18] named this problem as the Substitutes Discovery in Influence Maximization and proposed three solving algorithms based on the Greedy algorithms. However, these two works aim at solving the substitute mining problem in a single social network, which cannot adapt to the information spread in multi-social networks directly. In this paper, we discuss the information spread across social networks and focus on solving the problem of multi-social networks influence maximization. Specifically, the contributions of this paper are summarized as follows:

(1) We analyze the characteristics of information spreading in multi-social networks and define the special use with multiple social network accounts as *Bridge User*, which plays an important role in forwarding information from one social network to another.

(2) We discuss the problem of substitutes mining for multi-social networks influence maximization (SMMNIM), which aims to find substitutes and to reduce the loss of information spread caused by the uncooperative seed nodes.

(3) Three algorithms are proposed to solve SMMNIM in this paper for different application scenarios.

The rest of the paper is organized as follows. Section 2 introduces the special type of nodes named as *Bridge Users*, and the multi-social networks influence maximization problem. Section 3 presents the substitutes mining for multi-social networks influence maximization. In Section 4, substitutes seed nodes mining algorithms are proposed. The viability and performance of the proposed algorithms are evaluated in Section 5. Finally, we conclude this paper in Section 6.

2. Influence Maximization in Multi-Social Networks

2.1. Definition and Characteristics of Bridge User

Suppose there are two social networks represented as G_1 and G_2. As shown in Figure 2, in a previous social network, the scope of influence is limited in this social network. However, in today's social networks, as we discussed earlier, there exists special social network users who register for accounts on multiple social networks can break this limit, we define this type of user as a *Bridge User*. The red node in Figures 2 and 3 represents a *Bridge User*, he/she can forward the information cross-platform. The detailed process is shown in Figure 3, after receiving the information in G_1, the *Bridge User* forwards the information to G_2 with a certain probability $p_{\to G_2}$, which makes information flow from G_1 to G_2.
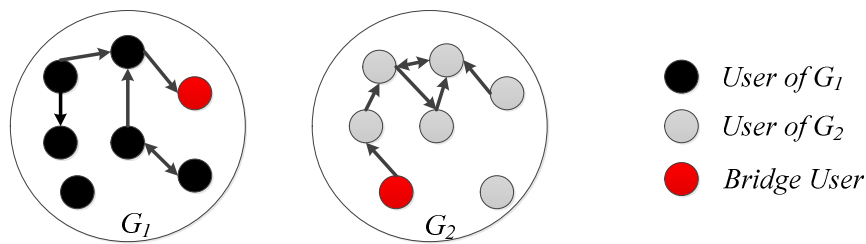

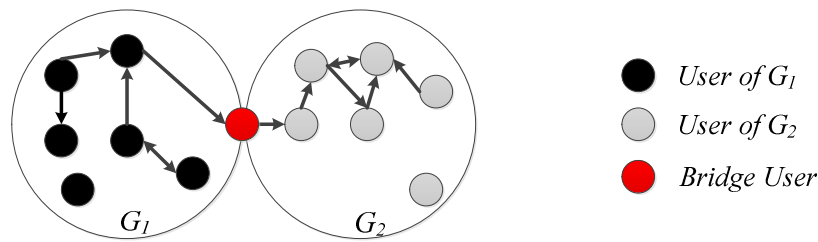**Figure 2.** Information spread within a single social network.

Figure 3. *Bridge User* forwards the information from G_1 to G_2.

Assuming we have multiple social networks $G_1(V_1, E_1)$, $G_2(V_2, E_2)$, $G_3(V_3, E_3)$, ..., $G_n(V_n, E_n)$, which are made of m social network users $U = \{u^1, u^2, u^3, \ldots, u^m\}$ and their relationships. We use u_i^j to indicate that user u^j participates in G_j (owns an account of G_j), so we have $V_i = \{u_i^1, u_i^2, u_i^3, \ldots, u_i^{|V_i|}\}$. BU denotes the set of all *Bridge Users*.

Definition 1. *Self-spread of Bridge User: In multiple social networks, if $u^j \in BU$ and participates in G_p and G_q at the same time, when u_p^j is activated by the information, it has a certain probability of self-spread $p_{\to G_q}^{u^j}$ to spread the information to G_q and activate u_q^j.*

Definition 2. *Self-spread probability matrix of Bridge User: $P_{m \times n}$ indicates the self-spread probability matrix of Bridge User u^i, where P_{ij} is the probability of self-spread $p_{\to G_j}^{u^i}$. The value of P_{ij} depends on the historical self-spread actions, calculated as follows:*

$$P_{ij} = \frac{N_{ij}}{\sum_{k=0}^{n} N_{ik}} \tag{1}$$

where N_{ij} denotes the number of self-spread actions that u^i spread the information to G_j, and $\sum_{k=0}^{n} N_{ik}$ represents the total number of self-spread actions.

2.2. Multi-Social Networks Influence Maximization

In the background of various social networks coexisting on the Internet, information can spread between multiple social networks $G_1(V_1, E_1)$, $G_2(V_2, E_2)$, $G_3(V_3, E_3)$, ..., $G_n(V_n, E_n)$. Each node in G_i has two states: activated or inactivated. The activated nodes are those who have affected by the information and will spread the information to their neighbors in the next time period. The inactivated nodes are those who have not heard of the information or rejected to adopt it [19]. A node whose state is inactive can be activated by activated nodes, and a node that is in an activated state cannot be changed into an inactivated state, that is, the activation process is irreversible.

The problem of Multi-Social Networks Influence Maximization is to find k (k is an integer and satisfies $k >= 1$) seed users from all the users $U = \{u^1, u^2, u^3, \ldots, u^m\}$ of $G_1(V_1, E_1)$, $G_2(V_2, E_2)$, $G_3(V_3, E_3)$, ..., $G_n(V_n, E_n)$, to constitute the initial seed node set S, such that by activating these users in S, the number of users that are ultimately activated is the most. $R_i(S)$ indicates the number of users that are ultimately activated on G_i. Hence, the problem of multi-social networks influence maximization can be defined as

$$Seed = \underset{S \subseteq U}{argmax} \sum_{i=1}^{n} R_i(S) \tag{2}$$

where *Seed* is expected set of the initial nodes and $|Seed| = k$, $Seed \subseteq U$.

2.3. Multi-Social Networks Aggregation Algorithm based on Bridge Users

In order to solve the problem of multi-social networks influence maximization, this section proposes a multi-social networks aggregation algorithm based on a *Bridge User,* to aggregate multiple social network graphs into one graph, and then we can apply some widely explored algorithms (such as Greedy [6], CELF [20], CELF++ [21] and so on) to obtain the initial seed node set S. The algorithm is described as Algorithm 1.

Algorithm 1. Multi-social networks aggregation algorithm based on Bridge Users.

Input: $G_1(V_1, E_1)$, $G_2(V_2, E_2)$, $G_3(V_3, E_3)$, ..., $G_n(V_n, E_n)$, $U = \{u^1, u^2, u^3, \ldots, u^m\}$, BU, $P_{m \times n}$;
Output: Aggregation graph $G^*(V^*, E^*)$;
(1) for $i = 1$ to n // Traversing n social network graphs
(2) for $j = 1$ to $|V_i|$ // Traversing all nodes of G_i
(3) if $u_i^j \in BU$
(4) if $u_0^j \in V^*$ then
(5) Create edge (u_i^j, u_0^j), (u_0^j, u_i^j);
(6) Add activated probability of edge $p_{u_i^j, u_0^j} = 1$, $p_{u_0^j, u_i^j} = P_{ij}$;
(7) Add edge (u_i^j, u_0^j) and (u_0^j, u_i^j) to E^*;
(8) else
(9) Create node u_0^j;
(10) Create edge (u_i^j, u_0^j), (u_0^j, u_i^j);
(11) Add activated probability of edge $p_{u_i^j, u_0^j} = 1$, $p_{u_0^j, u_i^j} = P_{ij}$;
(12) Add edge (u_i^j, u_0^j) and (u_0^j, u_i^j) to E^*, add node u_0^j to V^*;
(13) end if
(14) end if
(15) end for
(16) end for
(17) Output $G^*(V^*, E^*)$;

---

3. Substitutes Mining for Multi-Social Networks Influence Maximization

Given multiple social networks $G_1(V_1, E_1)$, $G_2(V_2, E_2)$, $G_3(V_3, E_3)$, ..., $G_n(V_n, E_n)$, which are made of *m* social network users $U = \{u^1, u^2, u^3, \ldots, u^m\}$ and the relationship between these users. Using

multi-social networks influence maximization algorithms, we can find the initial user set represented as $S = S_1 \cup S_2$. Assume that the user in the subset S_2 ($|S_2| = t$) cannot be activated at the initial moment, it needs to find t substitute nodes in the set $U - S$ to form a substitute user set S_2', constituting a new initial seed user set $S' = S_1 \cup S_2'$, so that

$$S'^* = \underset{|S'|=k}{argmin}(R(S) - R(S')) \tag{3}$$

e.g., the difference between the new initial set S' and the original set S is smallest, ensuring that the reduction of influence is close to the minimum.

We name this as the problem of substitutes mining for multi-social networks influence maximization (SMMNIM), in the following part, we will propose algorithms to solve this problem.

4. Substitutes Mining Algorithms for SMMNIM

In order to mine the substitutes for influence maximization in multi-social networks, three substitutes mining algorithms for SMMNIM are proposed in this section.

4.1. Greedy-Based Substitutes Mining Algorithm for SMMNIM (G_S)

The literature [13] has proved that substitute nodes mining for influence maximization in a single social network is an NP-hard problem. SMMNIM can be regarded as finding t seed users in the set U/S that can maximize information influence, so it is also an NP-hard problem. The NP-hard problem can be approximated to the optimal based on the greedy algorithm. Therefore, we first design a Greedy-based substitutes mining algorithm for SMMNIM (G_S), which is shown in Algorithm 2.

Algorithm 2. Greedy-based Substitutes Mining Algorithm for SMMNIM (G_S).

Input: $G^*(V^*, E^*)$, seed node set S, the subset of seed nodes that can be activated S_1, number of seeds that cannot be activated t;
Output: New initial seed seeds set S';
(1) $S' = S_1$;
(2) $R(S)$: The number of active nodes obtained by S;
(3) for $i = 1$ to t
(4) for v in $V^*/(S \cup S')$
(5) select $v = argmax(R(S' + v) - R(S'))$
(6) end for
(7) $S' = S' \cup v$
(8) end for
(9) Output S';

4.2. Pre-Selected-Based Substitutes Mining Algorithm for SMMNIM(P_S)

Using G_S, the most suitable set of substitutes can be found. However, G_S needs to simulate and calculate the influence of all users, thus requires a large number of calculations, which leads to low algorithm efficiency. In this section, we propose an algorithm based on the idea of "prepare in advance". When selecting initial seed nodes, additional nodes will be selected, we call these nodes as "pre-selected nodes". When some nodes cannot be activated, pre-selected nodes will be used as substitute nodes. Assuming that k initial seed nodes need to be selected, we select $k + t'$ initial seed nodes form a preselected set S_{pre}. When the number of non-cooperative nodes is t ($t < t'$), the nodes from $k + 1$ to $k + t$ are selected as substitutes. If there are still uncooperative nodes in the substitutes, we will continue to select nodes from the pre-selected set S_{pre}. The amount of calculation will be greatly reduced. This algorithm we named as pre-selected-based substitutes mining algorithm for SMMNIM (P_S), which is shown in Algorithm 3.

Algorithm 3. Pre-selected-based Substitutes Mining Algorithm for SMMNIM (P_S).

Input: $G^*(V^*, E^*)$, pre-selected seed node set S_{pre}, the number of additional selected nodes t';
Output: New initial seed seeds set S';
(1) $S_{pre} = \varnothing$, $S' = \varnothing$;
(2) $R(S)$: The number of active nodes obtained by S;
(3) for $i = 1$ to $k + t'$
(4) for v in $V^* / (S_{pre})$
(5) select $v = argmin(R(S_{pre} + v) - R(S_{pre}))$
(6) end for
(7) $S_{pre} = S_{pre} \cup v$
(8) end for
(9) for v in S_{pre}
(10) if v is uncooperative node
(11) $S_{pre} = S_{pre} - v$
(12) end for
(13) for $i = 1$ to k
(14) for v in S_{pre}
(15) $S' = S' \cup v$
(16) end for
(17) end for
(18) Output S';

4.3. Similar-Users-Based Substitutes Mining Algorithm for SMMNIM (S_S)

In this section, we propose another solution to find substitutes based on the idea of "finding the most similar user of uncooperative user". We focus on seeking the most similar users of the uncooperative nodes as the substitutes and propose a similar-users-based substitutes mining algorithm for SMMNIM (S_S), which is shown in Algorithm 4. We integrate structure similarity and neighbor attribute similarity to evaluate the similarity between two users.

Given an aggregation graph $G^*(V^*, E^*)$ of multi-social networks, $(u, v) \in E^*$ indicates the edge from node u to node v. The structure of the node u is determined by the in-edges and out-edges of the node.

$O\_E(u)$ indicates the out-edges and be defined as

$$O\_E(u) = \{v \in V^* \big| (u, v) \in E^*\}. \tag{4}$$

$I\_E(u)$ indicates the in-edges and be defined as

$$I\_E(u) = \{w \in V^* \big| (u, v) \in E^*\}. \tag{5}$$

The structure of the node u is defined as

$$\Psi(u) = (O\_E(u), I\_E(u)). \tag{6}$$

The structural similarity of node u and node v is defined as

$$SS(u, v) = \frac{1}{2}\left(\frac{|O\_E(u) \cap O\_E(v)|}{\sqrt{|O\_E(u)| * |O\_E(v)|}} + \frac{|I\_E(u) \cap I\_E(v)|}{\sqrt{|I\_E(u)| * |I\_E(v)|}} \right). \tag{7}$$

The self-spread of a *Bridge User* makes it more likely to have a wider scope of influence spread than ordinary users. When choosing the most similar user of the uncooperative user, the other factor we need to consider is the similarity of neighbor attributes (*Bridge Users* or ordinary users).

In the neighbor node set of node u, the number of *Bridge Users* is denoted as $B\_U(u)$, and the number of ordinary users is denoted as $N\_U(u)$. The neighbor node attributes of node u are defined as vectors

$$\mathbb{A}(u) = (B\_U(u), N\_U(u)). \tag{8}$$

The Euclidean distance of the neighboring attributes of u and v is

$$ED\_\mathbb{A}(u,v) = \sqrt{(\mathbb{A}(u) - \mathbb{A}(v))(\mathbb{A}(u) - \mathbb{A}(v))^T}. \tag{9}$$

Then we define the neighbor attributes similarity between node u and v is

$$AS(u,v) = \frac{ED\_\mathbb{A}(u,v) - ED\_\mathbb{A}_{min}}{ED\_\mathbb{A}_{max} - ED\_\mathbb{A}_{min}} \tag{10}$$

Finally, the similarity between node u and v is based on the similarity calculated by the structure similarity and the neighbor attributes similarity,

$$Sim(u,v) = SS(u,v) + AS(u,v). \tag{11}$$

Algorithm 4. Similar-users-based Substitutes Mining Algorithm for SMMNIM (S_S).

Input: $G^*(V^*, E^*)$, seed node set S, the subset of seed nodes that can be activated S_1, the subset of seed nodes that cannot be activated S_2;
Output: New initial seed seeds set S';
(1) $S' = S_1$;
(2) $R(S)$: The number of active nodes obtained by S;
(3) for i = 1 to $|S_2|$
(4) for u in S_2
(5) for v in V^*/S
(6) select $v = argmax(Sim(u,v))$
(7) $S' = S' \cup v$
(8) end for
(9) end for
(10) end for
(11) Output S';

---

5. Simulation

5.1. Data Description

The experiments are based on the following three social network datasets:

NetHEPT: This dataset is derived from "High Energy Physics" and is a web-based data about authors of articles. Each node represents an author, and the number of edges between a pair of nodes is equal to the number of papers the two authors collaborated [7].

Epinions: This dataset comes from the social network "Epinions". In the data set, if a user trusts another user, there is an edge between the users.

Slashdot: This dataset is derived from the "Slashdot" website, a consulting technology website and its users can post comments on the website. In this data set, if the user is a friend or an opponent, it is considered that there is a relationship between the users and there is an edge.

The detailed information of nodes and edges is shown in Table 1.

Table 1. Three social network datasets used in the simulation.

Dataset	NetHEPT	Epinions	Slashdot
Number of nodes	15,233	75,879	77,360
Number of edges	62,796	508,837	905,468

5.2. Analysis of Multi-Social Networks Influence Maximization

In the simulation, CELF++ is used to solve the problem of influence maximization and the propagation model we selected is the LT model. For the edge (u, v) existing in the graph G, the probability p_{uv} the node u to activate its neighbor node v is calculated by the reciprocal of the number of node v's in-degrees, that is $p_{uv} = 1/|in(v)|, v \in out(u)$. Each time the influence of the seed set was calculated, the propagation process was simulated by 10,000 Monte Carlo methods.

To compare with the information spread in multi-social networks, we first show the spread of influence in a single social network, as shown in Figure 4. From the experimental results we can see that, as the number of seed nodes increases, the influence increases.

(a) NetHEPT

(b) Epinions
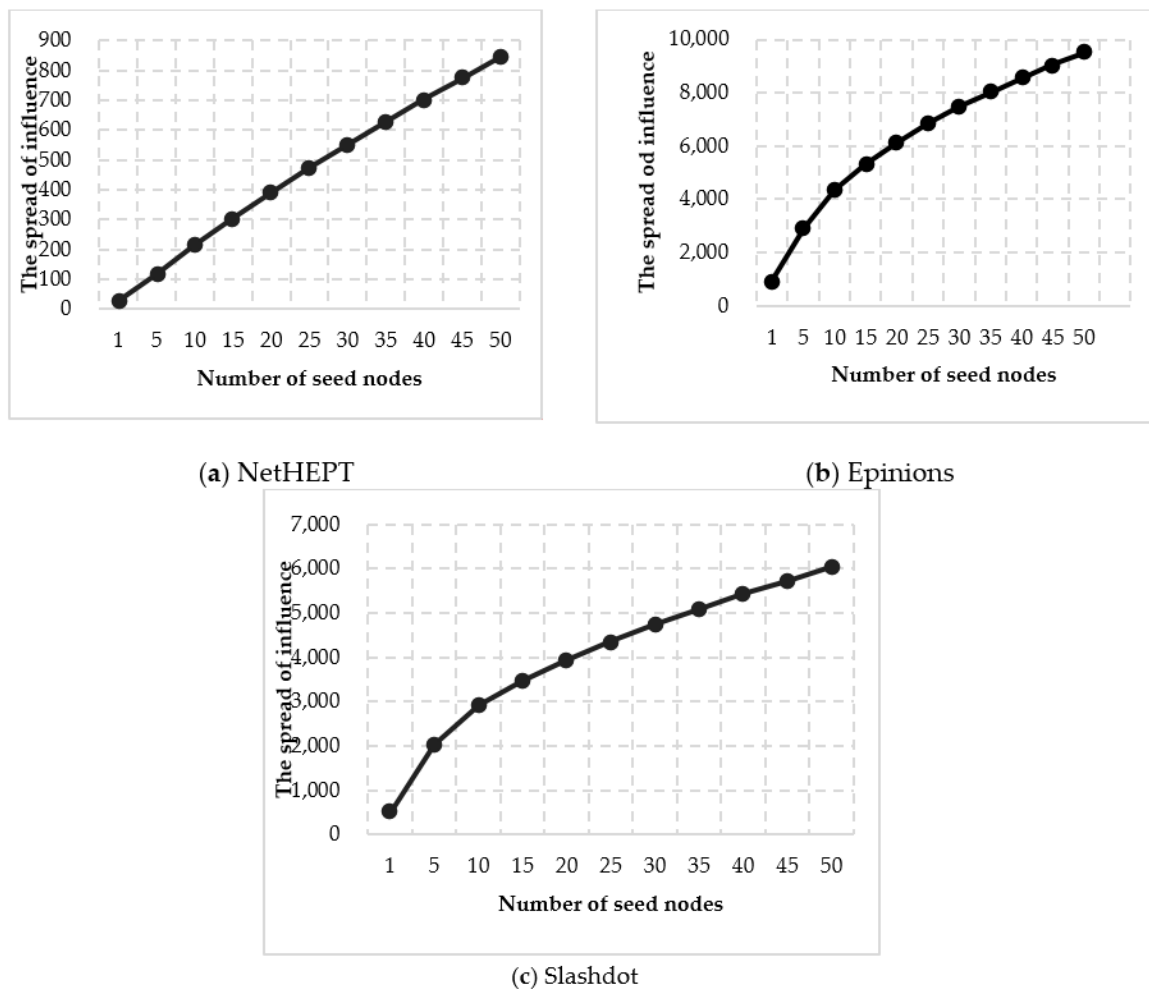
(c) Slashdot

Figure 4. The spread of influence in a single social network. (**a**) The spread of influence in NetHEPT; (**b**) The spread of influence in Epinions; (**c**) The spread of influence in Slashdot.

Next, the experiments verify the information spread in multiple social networks. However, we cannot get the social network data sets with the annotated *Bridge Users*. In the experiments, we choose the node with the same ID from the three social networks and assume these nodes (with the same ID) belong to one user, this user is regarded as a *Bridge User*. In the following simulation, the BU is

obtained by this method. Then the above three social networks $G_1(V_1, E_1)$, $G_2(V_2, E_2)$, $G_3(V_3, E_3)$ can be aggregated into one network $G(V, E)$ by algorithm 1. When aggregation, the number of selected *Bridge User* is set as 0, 1000, 2000, 3000, 4000, and 5000 respectively. The information spread with a different number of *Bridge Users* in multi-social networks is shown in Figure 5.
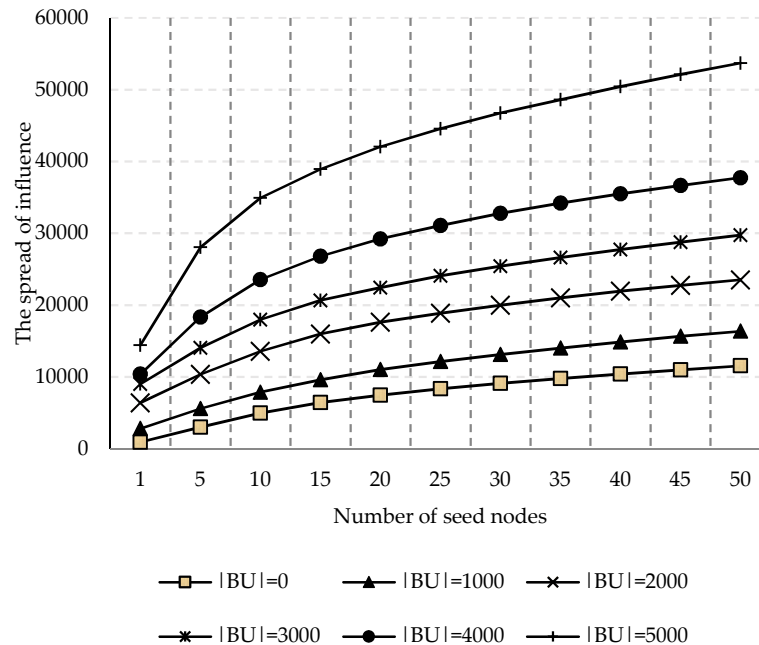

Figure 5. The information spread with a different number of *Bridge Users* in multi-social networks.

As the result shown in Figure 5, we can find that, as the number of seed nodes increases, the spread of influence becomes larger. As discussed before, we have analyzed that *Bridge Users* can spread information from social network G_1 to G_2. When comparing the influence spreading of a different number of *Bridge Users*, we can see that the more *Bridge Users*, the wider influence spreading, which indicates the self-spread of *Bridge User* expands the scope of information spreading. These results confirm that the problem of influence maximization is different from previous, multi-social networks can be connected by *Bridge Users*.

5.3. Comparison Results of Substitutes Mining Algorithms

The uncooperative seed nodes will lead to the failure of influence maximization, it is necessary to select appropriate substitute nodes to replace the uncooperative seed nodes. In this section, assume that the cooperation rate of the seed nodes is 80%, this means that if we expect to select 50 nodes as the initial seed nodes, there may be 10 nodes will be uncooperative seed nodes. The purpose of the three algorithms designed in this paper is to find 10 substitutes to replace the uncooperative seed nodes. To evaluates the performance of each algorithm, this paper compares the algorithm in three aspects: the influence spread range of the new seed set, the loss rate of the influence spreading, and the memory/time consumption of the algorithm running. Each algorithm is executed multiple times to avoid errors (in this experiment, the number of times is five). In the figure of experimental results, init-0 represents the initial set of seed nodes, G_S represents solving algorithms for SMMNIM based on Greedy, P_S represents pre-selected-based substitutes mining algorithm for SMMNIM, and S_S represents similar-users-based substitutes mining algorithm for SMMNIM.

Figure 6 shows the influence spreading of the new seed set obtained by the three algorithms G_S, P_S and S_S. Compare with the influence spreading of original initial seed set (Init_0), we find that the three new seed set get lower influence spreading. Although we have proposed algorithms to find substitutes, it still unable to make up for the losses caused by the uncooperative nodes. It indicates that

the uncooperative nodes lead to the reduction of influence maximization in multiple social networks. In other aspects, the new seed set obtained by G_S algorithm can achieve the closest influence spreading to the original seed set can produce. This is because the G_S algorithm is equivalent to finding the optimal seed set in the rest nodes. Therefore, the new seed set solved by G_S can obtain a wider range of influence than the other two algorithms. Compared with the P_S algorithm, the new seed set obtained by S_S algorithm has a wider range of influence. This is because the S_S algorithm looks for the most similar node of the uncooperative node as the substitute node, the influence that substitute nodes can produce is similar to the original uncooperative node. P_S algorithm just pre-select some nodes for replacing the uncooperative node. Therefore, S_S has a better performance than the P_S algorithm.
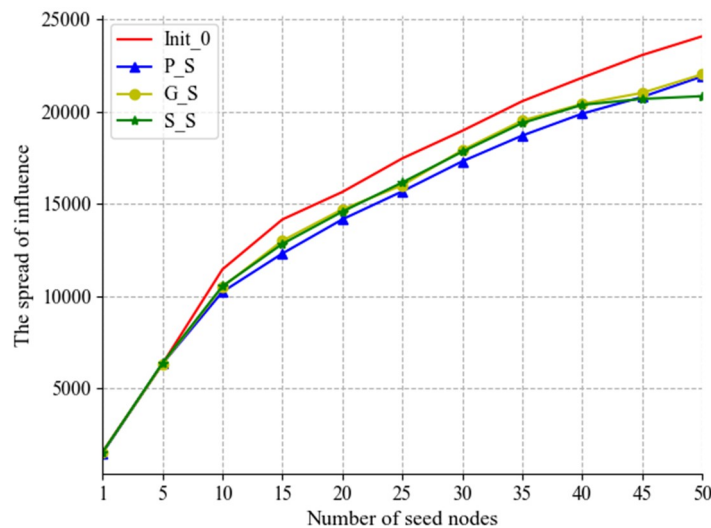
Figure 6. Influence spreading of the new seed set obtained by G_S, P_S and S_S.

Figure 7 illustrates the loss rate of influence spreading of the new seed set by three algorithms. While the G_S algorithm solves the better alternative nodes, the loss of influence is smaller. The S_S algorithm takes second place.
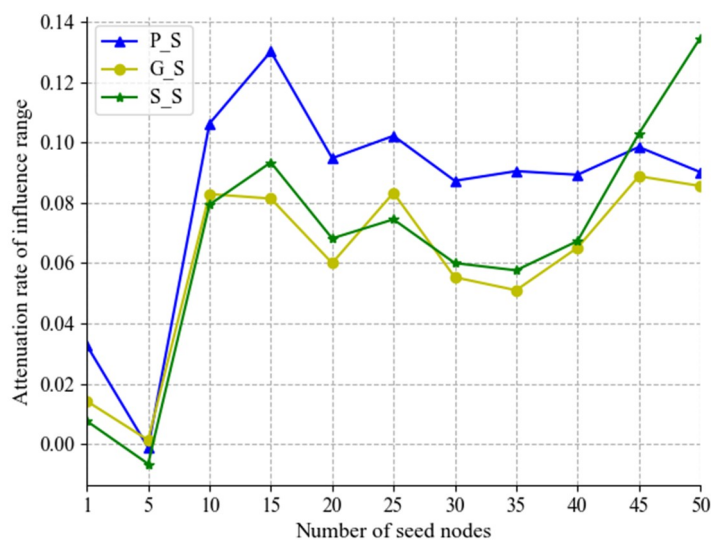

Figure 7. Loss rate of influence spreading of the new seed set obtained by G_S, P_S and S_S.

In terms of algorithm memory/time cost (Figure 8), P_S algorithm costs less memory than G_S algorithm, and the time cost is smaller than G_S algorithm. The main reason is that P_S algorithm will

pre-select some nodes in advance as "standby" nodes. When some nodes do not cooperate, they can select substitute nodes from the "standby" nodes without recalculations. The G_S algorithm needs a long time to recalculate the new seed set, it's about two times that of P_S algorithm, and the cost of memory is also larger than P_S. In addition, the S_S algorithm spends a number amount of calculations on user similarity, so its memory/time cost is the largest.
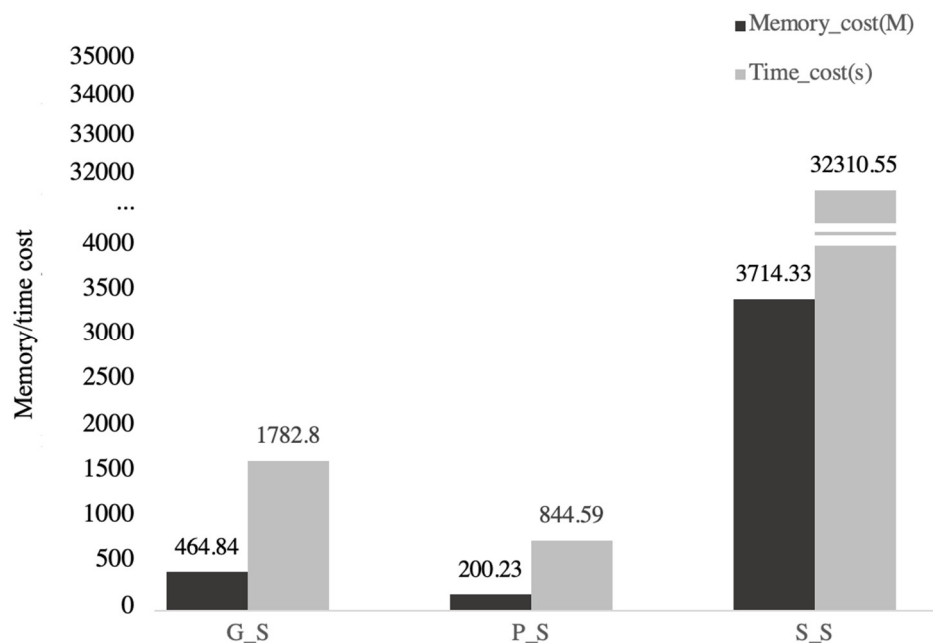

Figure 8. Memory/time cost of G_S, P_S and S_S.

Therefore, all three algorithms can find the substitute nodes and reduce the influence loss caused by uncooperative nodes, we can choose one of them according to different requirements. The substitute nodes found by G_S algorithm can get closer influence to the original seed node set, but it cost more time and memory. That is, G_S is suitable for the scenes that are not sensitive to time or memory but require a wider range of influence; P_S algorithm can get the substitute nodes immediately, so it is more suitable for time-sensitive or memory-sensitive scenes; Due to the large number of calculations on user similarity, the S_S algorithm can be selected when the user similarity is known in advance.

6. Conclusions

In this paper, we first studied the problem of multi-social networks influence maximization. By defining the user with multiple social network accounts as a *Bridge User*, we discussed how a *Bridge User* affects the information spreading in multiple social networks. Then we considered a new and significant problem by analyzing that there may be some seed nodes cannot be activated in the process of influence maximization. Hence, it is necessary to find substitute nodes to reduce the losses caused by these uncooperative seed nodes. This brings up the problem of Substitutes Mining for Multi-Social Networks Influence Maximization (SMMNIM). In this paper, three substitute nodes mining algorithms were proposed (G_S, P_S and S_S). The experimental results showed that: (1) In multi-social networks, *Bridge Users* can make information spread across social networks and expand the range of information influence; (2) the uncooperative nodes will reduce the range of information influence; (3) three substitute node mining algorithms can find suitable substitute nodes and construct the new seed set, which makes the information influence as close as possible to the original seed node set; (4) according to different application scenarios, the three algorithms can be selected for mining the substitute seed nodes.

In the future research, we will further consider the attributes of nodes and information, such as the node's interests, the subject of the information, etc., take these factors into the process of multi-social networks influence maximization, and propose more accurate and efficient substitute node mining algorithm.

References

1. Wasserman, S. Social Network Analysis Methods and Applications. *Contemp. Sociol.* **1995**, *91*, 219–220.
2. Zhan, Q.; Zhang, J.; Philip, S.Y. Integrated anchor and social link predictions across multiple social networks. *Knowl. Inf. Syst.* **2018**, 1–24. [CrossRef]
3. Bakshy, E.; Rosenn, I.; Marlow, C.; Adamic, L. The role of social networks in information diffusion. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012; ACM: New York, NY, USA, 2012; pp. 519–528.
4. Zhao, J.H.; Wan, K.W. Research on the Communication Dynamics Model of Social Network Public Opinion Based on the SIS Model. *Inf. Sci.* **2017**, *12*, 34–38.
5. Liu, S.; Jiang, C.; Lin, Z.; Ding, Y.; Duan, R.; Xu, Z. Identifying effective influencers based on trust for electronic word-of-mouth marketing. *Inf. Sci.* **2015**, *306*, 34–52. [CrossRef]
6. Kempe, D.; Kleinberg, J. Maximizing the spread of influence through a social network. In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, 24–27 August 2003; ACM: New York, NY, USA, 2003; pp. 137–146.
7. Chen, W.; Wang, Y.; Yang, S. Efficient influence maximization in social networks. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, 28 June–1 July 2009; ACM: New York, NY, USA, 2009; pp. 199–208.
8. Du, N.; Liang, Y.; Balcan, M.F.; Gomez-Rodriguez, M.; Zha, H.; Song, L. Scalable Influence Maximization for Multiple Products in Continuous-Time Diffusion Networks. *J. Mach. Learn. Res.* **2017**, *18*, 1–45.
9. Tong, G.; Wu, W.; Tang, S.; Du, D.Z. Adaptive Influence Maximization in Dynamic Social Networks. *IEEE/ACM Trans. Netw.* **2017**, *25*, 112–125. [CrossRef]
10. Wang, Y.; Vasilakos, A.V.; Jin, Q.; Ma, J. PPRank: Economically Selecting Initial Users for Influence Maximization in Social Networks. *IEEE Syst. J.* **2017**, *11*, 2279–2290. [CrossRef]
11. Li, G.L.; Chu, Y.P.; Feng, H.J.; Xu, Y.Q. Influence maximization on multiple social network. *Chin. J. Comput.* **2016**, *39*, 643–656.
12. Li, X.; Zhang, X.; Shu, H.; Sun, G. Influence Maximization Acros Multi-Channels in Social Network. *J. Comput. Res. Dev.* **2016**, *53*, 1709–1718.
13. Shen, Y.; Dinh, T.N.; Zhang, H.; Thai, M.T. Interest-matching information propagation in multiple online social networks. In Proceedings of the International Conference on Information and Knowledge Management, Maui, HI, USA, 29 October–2 November 2012; ACM: New York, NY, USA, 2012; pp. 1824–1828.
14. Zhang, H.; Nguyen, D.T.; Zhang, H.; Thai, M.T. Least cost influence maximization across multiple social networks. *IEEE/ACM Trans. Netw.* **2016**, *24*, 929–939. [CrossRef]
15. Nguyen, D.T.; Das, S.; Thai, M.T. Influence maximization in multiple online social networks. In Proceedings of the International Conference on Global Communications Conference, Atlanta, GA, USA, 9–13 December 2013; pp. 3060–3065.

16. Zhao, J.; Chen, Z.; Luo, J. Influence maximization on multi-social networks based on bridge users. *Comput. Syst. Appl.* **2017**, *26*, 199–204.

17. Li, C.T.; Hsieh, H.P.; Lin, S.D.; Shan, M.K. Finding influential seed successors in social networks. In Proceedings of the 21st International Conference Companion on World Wide Web, Lyon, France, 16–20 April 2012; ACM: New York, NY, USA, 2012; pp. 557–558.

18. Ma, Q.; Ma, J. Discovering the substitutes for the seeds in influence maximization problem. *Chin. J. Comput.* **2017**, *40*, 674–686.

19. Wang, Y.; Dong, W.; Dong, X. A novel ITÖ Algorithm for influence maximization in the large-scale social networks. *Future Gener. Comput. Syst.* **2018**, *88*, 755–763. [CrossRef]

20. Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J.; Glance, N.S. Cost-effective outbreak detection in networks. In Proceedings of the 13th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, San Jose, CA, USA, 12–15 August 2007; pp. 420–429.

21. Goyal, A.; Lu, W.; Lakshmanan, L.V.S. CELF++: Optimizing the greedy algorithm for influence maximization in social networks. In Proceedings of the International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; pp. 47–48.