

A new method for maximizing influence on social networks based on node membership in communities

Esmail Bagheri

Department of Computer, Dehaghan Branch, Islamic Azad University, Isfahan, Iran

E-mail: Bagheri471@gmail.com

Received 13 September 2020; Accepted 3 November 2020; Published 1 December 2020



Abstract

Influence maximization is one of the fundamental issues in social networks context. In viral marketing which is one of applications of this category, a small group of users are selected to accept a product and influence of these users on other people might result in massive acceptance of this product in social network. The influence maximization problem is choosing a set of k nodes from a social network that maximizes the influence in the network. Various studies have been conducted to find more effective k nodes for influence propagation on social networks. But the main challenges of these studies are the lack of scalability and low speed. Influential nodes must also have local influence and global influence throughout the network so that they can affect the entire network at an acceptable time. Considering the important role of influential nodes in each community for influence propagation in that community and, consequently propagating the influence throughout social network, in this paper, an algorithm is presented that maximizes the influence throughout social network through finding the nodes that have more membership strength to their community. The proposed algorithm is tested on several real and synthetic social networks. Experimental results show that the proposed method can effectively find appropriate seed nodes for influence maximization.

Keywords influence maximization; community detection; social networks.

Network Biology

ISSN 2220-8879

URL: <http://www.iaees.org/publications/journals/nb/online-version.asp>

RSS: <http://www.iaees.org/publications/journals/nb/rss.xml>

E-mail: networkbiology@iaees.org

Editor-in-Chief: WenJun Zhang

Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

Social networks provide a visual representation of individual communications, as well as interesting patterns of behavior in different user communities (Wasserman and Faust, 1994; Zhang, 2018). Analysis of the social network focuses on various domains and has become an important tool for development of intelligent systems in advisory, mass storage services, and so on (Domingos and Richardson, 2001).

Advantage of a social network lies in the power of user interaction, which spreads the influence of individuals throughout a network. Such effects are observed in many real world applications. For example, an influence-based marketing approach can target a small group of influential people and expect to have the most influence on the market by those users. This is one of the general problems of influence maximization in social networks; it is our task to find K influential nodes based on information diffusion models (Kempe et al., 2003).

In ideal conditions, by selecting and activating the K influential nodes in a social network, these nodes will expand the influence. Only if all the selected nodes successfully publish influence, maximum influence will be achieved. However, if some of the nodes do not behave as expected, the problem of influence will be difficult. So, finding the best nodes in order to maximize the diffusion is a fundamental challenge.

In social networks, people are affected by family, relatives, friends, classmates, colleagues, and so on. In fact, people are influenced by their own groups. Usually, each group is known as a community in the social network. People are affected by other people in their community. In each community, people who have more membership strength to their community, have a greater impact on other people in the community (Shang et al., 2018). In this paper, the importance of nodes that have more membership strength to their community, on selecting influential nodes in social networks is examined. In other words, in order to provide an effective solution to the influence maximization problem in social networks, the nodes that have more membership strength to their community are examined.

Since the greedy algorithm cannot provide a fast and scalable solution for influence maximization of big social networks, this paper presents an algorithm for solving two challenges of speed and scalability and first discovers communities within the social network. Then it creates a set of candidate nodes by finding nodes that have more membership strength to their community. Finally, each community's quota of seed nodes is determined, and seed nodes are selected from the candidate nodes based on the quota of each community. Extensive experiments are carried out in several real and synthetic social networks, and performance of our proposed algorithm is shown.

The following paper is organized as follows. In Section 2 background to the community detection algorithms is presented. In Section 3 describes the problem. In Section 4, the related works on influence maximization problem is studied. In Section 5, the proposed algorithm is presented and in Section 6 a set of experiments in several real and synthetic social networks to demonstrate the performance of the proposed algorithm is represented. In Section 7, the proposed algorithm is compared with other algorithms in terms of time complexity, and eventually this study is summarized and some ideas are suggested for further works.

2 Background

Considering the proposed method in this paper, it is necessary to provide background material related to community detection algorithms in this section.

A community is a subset of people who interact with together more than people outside the community (Wasserman and Faust, 1994). The nodes within a community are more interconnected, with more edges between them and fewer edges with other nodes in other communities. Community detection is formulated as a clustering issue. The set of graph nodes $G = (V, E)$ is divided to p sets of C_1, C_2, \dots, C_p using the community detection algorithm so that their community is equal to the whole graph nodes. Community detection algorithms can find these subsets that categorize nodes in different groups (Zhang, 2016a, b). The community detection algorithms are divided into overlapping and non-overlapping groups (Yang et al., 2014). According to Fig. 1, the overlapping community detection algorithms can also find overlapping communities, thus they have more realistic results.

The proposed influence maximization algorithm can use any community detection algorithm to extract communities from the input network. For this purpose, among the community detection algorithms, the algorithm is selected which is faster and quality of its extracted communities is better. In this paper, the SLPA algorithm (Xie et al., 2011) has been used because it is very fast and does not need to know the number of communities and quality of the communities is also high. The SLPA was presented by Xie et al. in 2011. SLPA simulates the information diffusion process using the speaker-listener interaction. This algorithm first

maps labels between nodes. Operations are tagged based on a number of rules that govern interactions between nodes. SLPA takes a memory for each node to store labels received from its neighbors at different steps of the algorithm. The membership degree of a node to a community is computed based on the probability of viewing a community label in that node's memory. For example, if most of the labels stored in memory of node u are v , then the node u is likely to belong to the community v . Because in this algorithm, each node might belong to different communities, so SLPA can also identify overlapping communities (Xie et al., 2011).

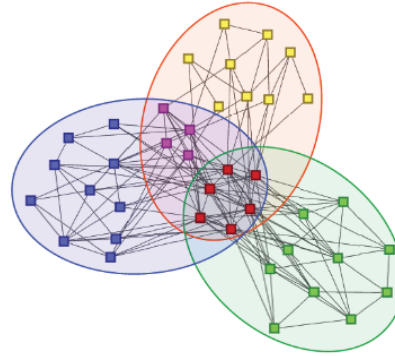


Fig. 1 Overlapping communities.

3 Problem Description

We consider directed graph $G = \{V, E\}$ with $|V| = N$ vertices and $|E| = M$ edges. For each edge $(v, u) \in E$, P_{vu} is the probability of influence diffusion at the edge. The influence maximization problem is to find a set of seeds $S \subseteq V$, $|S| = k$ in such a way that the influence $\sigma(S)$ is maximized by an information diffusion model. We use the linear threshold (LT) model (Kempe et al., 2003), which has been widely used in various researches. In influence maximization, if the diffusion process starts from S , it must be able to maximize influence in the network. According to (1), $\sigma(S)$ which is the influence diffusion function, is equal to the number of active users in the network after the diffusion process is stopped.

$$S^* = \operatorname{argmax}_S \sigma(S) \quad (1)$$

Kempe et al. (2003) proved that the influence maximization problem under LT and IC models is NP-hard and sub-modular. In order to solve this problem, with the mathematical properties of the sub modular functions, they proposed the greedy hill climbing algorithm (Algorithm 1), which starts with an empty set S , and then adds a node to the S repeatedly. As long as $|S| = k$ and the maximum marginal effect is obtained. Theoretically, the greedy algorithm guarantees the optimal solution by 63 percent. In real experiments, the solution provided by the greedy algorithm is quite close to the optimal solution. However, in order to have a good approximation of the target function and seed set S , the greedy algorithm needs tens of thousands of Monte Carlo simulations, which reduces its use in large-scale networks.

Algorithm 1: The hill-climbing greedy algorithm

Input: Network G , number of seed nodes k

Output: Seed set S

Initialize: Let $S \leftarrow \emptyset$;

for $i = 1$ to k *do*

```

 $v = \arg \max \{ \sigma(S \cup \{u\}) - \sigma(S) \};$ 
//  $\sigma(*)$  is computed using Monte-Carlo simulations
 $S \leftarrow S \cup \{v\};$ 
end for
return  $S$ ;

```

So, the goal is to find an algorithm that can select the best seed nodes in the shortest time. On the other hand, despite the useful features of communities in social networks, little attention has been paid to the role of communities in the influence maximization problem. In real social networks, people live in groups that have strong relationships with each other. Information circulation in these groups is high. On the other hand, real networks include a large number of nodes, thus the influence calculation for each node in these networks is costly. The influence computation of each node can be done quickly within the community to which it belongs, and as a result, run time of the algorithm is improved. The nodes that have more membership strength to their community can be candidates for seed nodes. Therefore, in this paper, we intend to present a fast and accurate method for influence maximization in social networks by finding nodes that have more membership strength to their community and then select seeds from them.

4 Related Works

The influence maximization problem has attracted a lot of attention in the past decade. Domingos and Richardson (2001) have pioneered the influence maximization problem in social networks. They examined this issue in a probabilistic framework and solved it using the markov random field. Kempe et al. (2003) formulated this as a discrete optimization problem. They proved that influence optimization problems are NP-Hard and proposed a heuristic algorithm for its approximate solution through frequent selection of nodes. Further research has been developed on the k influential nodes analysis based on the greedy algorithm and the algorithm performance has been improved by reducing complexity of the influence computations. For example, Leskovec et al. (2007) have used the Cost-Effective Lazy Forward (CELF) mechanism to reduce the time periods required to compute the influence rate.

There are many nodes in real social networks. Therefore researchers are trying to provide algorithms that do not need to examine all the nodes. With the help of community detection algorithms, big social networks can be divided into constituent communities, and actually divide the problem into smaller problems and examine the problem of finding influential nodes in communities (Bagheri et al., 2016). In the following, we examine some of the influence maximization algorithms that use the structure of communities.

Chen et al. (2014) examined the influence maximization problem based on heat emission model, and have presented an algorithm called CIM for this purpose. Their method first finds communities in the social network using the H-clustering hierarchical community detection algorithm. Then, it considers a number of communities as important communities using size of the communities and the relationship between them. In the next step, it selects candidate nodes from these important communities. Then, it selects the final seeds set from the candidate nodes based on their position in the communities. Because of the division of the graph into smaller communities and the removal of a large number of non-important nodes in search operation, this algorithm is faster than previous algorithms for large graphs.

In 2014, Ok et al. examined types of graphs in terms of structure and size of their communities, to provide an effective algorithm for influence maximization in different social networks. They have investigated the maximizing diffusion speed of a new invention under a game-based model. They have analyzed three classes of Erdős-Rényi and planted partition graphs and structured topological graphs, and obtained a new topological

view that did not exist in epidemic based models. Their results show that it is not necessary to find the seeds accurately for globally well-connected graphs. In other words, in locally well-connected graphs, it is necessary to find seeds exactly, and the characteristics of the communities must be carefully selected to select better seeds. For graphs that have small and large communities, seeds must be searched between communities and within communities respectively. Their algorithm called PaS is better than previous algorithms for social networks with completely different internal structures, but it is time consuming for large graphs with large communities.

One of the algorithms that uses community structures is the COMPATH algorithm (Rahimkhani et al., 2015), which has been presented by Rahimkhani et al. in 2015. They have provided a linear threshold model based algorithm for influence maximization in social networks, which initially extracts communities from the input graph and selects a limited number of them as important communities by the betweenness centrality measure. Purpose of using the community detection methods and selecting the most important communities in these algorithms is to reduce the time of investigating the graph nodes. Then, it selects some nodes as candidate nodes among the nodes of each important community. Finally, it selects the seeds from the candidate nodes, their algorithm examines different paths with different lengths at the start of candidate nodes to get the final seeds. They have also provided modifications to the linear threshold model for computing the influence diffusion of nodes and they have limited the number of examined nodes. Their algorithm is faster than previous algorithms, but it is not efficient for graphs in which most nodes have high out degree, because it examines different paths with different lengths at the start of candidate nodes to get the final seeds.

Another community detection based influence maximization algorithm called community-based greedy algorithm CGA (Song et al., 2015) is provided by Song et al. in 2015. Their algorithm can find k influential nodes in two phases: First it divides a large mobile social network to several communities according to the information diffusion. In the second phase, it chooses communities to find influential nodes through these selected communities by dynamic programming. To increase performance, they parallelized influence diffusion between communities and among communities. They have also provided an accurate analysis to ensure their approximate model. Their proposed algorithm is more time consuming than earlier methods without using parallelization techniques.

Mohamadi-Baghmolaie et al. (2015) have presented trust-based latency-aware influence maximization algorithm called TLIM (Mohamadi-Baghmolaie et al., 2015) which selects influential nodes in social networks, taking into account time and trust simultaneously. They have first developed the classic IC model based on time and trust simultaneously. Then they got more influential nodes in social networks using time and trust. Instead of concentrating on activated nodes, their model focuses on all of the positive nodes. In a TLIM model, a node can be positively activated with the probability of $P +$ by a trusted neighbor, and again returns to the negative state with a probability P by an untrusted neighboring node.

Bozorgi et al. (2016) proposed INCIM algorithm based on linear threshold model. First, they have used community detection algorithms to find communities, then they have considered communities as nodes and have created a graph from communities. Then, influence of each node has been considered as a combination of local and global influence. Local influence indicates the node's influence in the community and global influence indicates the community's influence of that node in the entire graph. The final influence of each node in the network is derived from the combination of local and global influence. Then they have derived a list of nodes influence based on the CELF idea and obtained seed nodes from the list of each community. Their proposed algorithm has a high degree of complexity because it uses time-consuming methods such as CELF at different steps.

In 2017, Liu et al. presented a method based on the prioritization and heuristic choices used to optimize the

impact on social networks and real world networks that are dynamically changing. Their method has been analyzed on the Facebook and Flickr datasets. Their algorithm has a fairly good execution time, but it does not have a good performance on big social networking issues, due to use of high memory amount and failing to divide the problem properly.

Bagheri et al. (2018) proposed FSIM algorithm based on community detection. Without loss of quality, FSIM reduces the number of nodes that must be examined for finding seeds. Their method first detects communities from the input network and creates a new network from detected communities. The new network has m nodes, where each node represents a community. Hence only a limited number of nodes are examined so it is fast. Within each important community, important nodes are selected. Final seeds are selected after testing initial seeds.

In 2020, Ghanbari et al. proposed C-K-shell algorithm based on K-shell decomposition and community detection. They use SLPA algorithm for community detection and to make a better results use optimizing the decision-making in exploration and extraction of communities. K-shell analysis and community detection are used to choose the more influential nodes, which are proportional to the graph of social networks. C-K-shell reduces number of nodes which should be investigated to find seeds without losing quality. Therefore, only a limited number of nodes are investigated so that speed is increased.

5 Proposed Method

Our proposed method called IMMS (Influence Maximization based on Membership Strength to community) consists of three main steps according to Algorithm2: in the first step, the input network communities are achieved by a community detection algorithm. In this research, we use the SLPA algorithm (Xie et al., 2011), which is a fast and accurate algorithm for community detection. Instead of SLPA, we can use any community detection algorithm. In the second step, the membership strength of each node to its community is computed in order to identify the nodes that have a stronger membership to their community. In the third step, considering the importance of each community, its quota of seed nodes is determined and, finally, seed nodes are obtained based on the membership strength measure and the quota of each community. Experimental results show that our proposed method can effectively find the appropriate seed nodes to maximize influence.

Algorithm 2: $IMMS(G,k)$

Input: Graph $G(V,E)$ and number of seeds k ;

Output: k seeds;

$G =$ Read the input network $G(V,E)$;

$N =$ Number of nodes G ;

Step 1: Community Detection

$C =$ Detect communities of G by SLPA;

$NC =$ Number of nodes in each community $c \in C$

Step 2: Membership Strength Computation

$O_{uc} =$ Compute outgoing membership node u to community c ;

$I_{uc} =$ Compute incoming membership node u to community c ;

$M_{uc} = \alpha.O_{uc} + \beta.I_{uc}$;

Step 3: Community Quota Computation and Seed Selection

$Q_s = k \times (NC / N)$; // community seed quota

$Q_c = 2 \times Q_s$; // community candidate quota

$Candidates =$ Select Q_c number of nodes with the highest M_{uc} within the each community;

Seeds = Select Q_s number of Candidates within the each community by a SimPath-based algorithm;
Output Seeds;

A. Community detection step

At first, communities can be found using the SLPA community detection algorithm. SLPA can quickly locate overlapping communities in social networks. According to the experimental results described in Section 6, our algorithm can use any community detection algorithm and does not depend on the SLPA algorithm.

B. Membership strength computation step

In this step, we determine the membership strength of each node to its community. Obviously, by deleting a node that has higher membership strength to its community, the edges attached to that node are also deleted, so the membership degree of other nodes connected to it in the target community is reduced, thus total membership strength of nodes in the community decreases. In other words, the connectivity level of nodes in the community is reduced. Therefore, the node that has higher membership strength to its community will have a greater impact on the cohesion of that society. For example, in Fig. 2, red nodes in each community have a stronger membership than other nodes in the community, and by eliminating them, community cohesion is reduced.

Node which has the more membership strength to its community can be considered as a candidate for selecting seeds due to having more connections with other nodes in that community. The experimental results (Figs 4 to 6) show that if the nodes with higher membership strength are considered as seeds, the influence spread in social network increases.

Since in a real community, the amount of a person influence on the other person is different from the amount of the other person influence on that person, we have to differentiate between these two categories. First, we compute the outgoing dependencies of the nodes to the communities and the incoming dependencies of the communities to the nodes, and determine the dependence of each node on each community. To compute the outgoing membership degree of the node to the community, the number of output edges from the node to the other nodes of the community is computed. Similarly, to compute the incoming membership degree of the community to the node, the number of input edges from the community nodes to the node is computed.

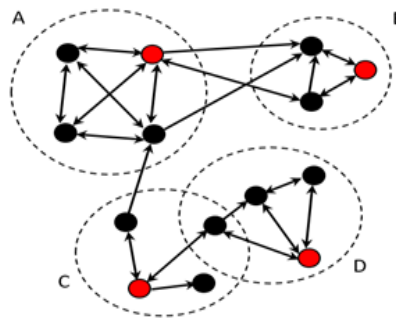


Fig. 2 A sample network with four communities.

If O_{uc} is the outgoing membership degree of the node u to the community c and I_{uc} is the incoming membership degree of the node u to the community c , then the membership degree of the node u to the community c is determined by (2):

$$M_{uc} = \alpha \cdot O_{uc} + \beta \cdot I_{uc} \quad (2)$$

To find the optimal value of α and β , we performed our experiments with different values of α and β . Experimental results (Fig. 3) show the best value for $\alpha = 0.6$ and $\beta = 0.4$.

C. Community quota computation and seed selection step

We determine the quota of each community from seed nodes using the following equation:

$$Q_s = k \times (NC / N) \quad (3)$$

NC is number of nodes in the community, N is the number of nodes in input graph and k is number of required seeds. According to (3), the quota of each community from seed nodes is determined by the number of its nodes, since the final goal is influence maximization in the entire network. The larger is a community, the greater is the share of seed nodes, and vice versa.

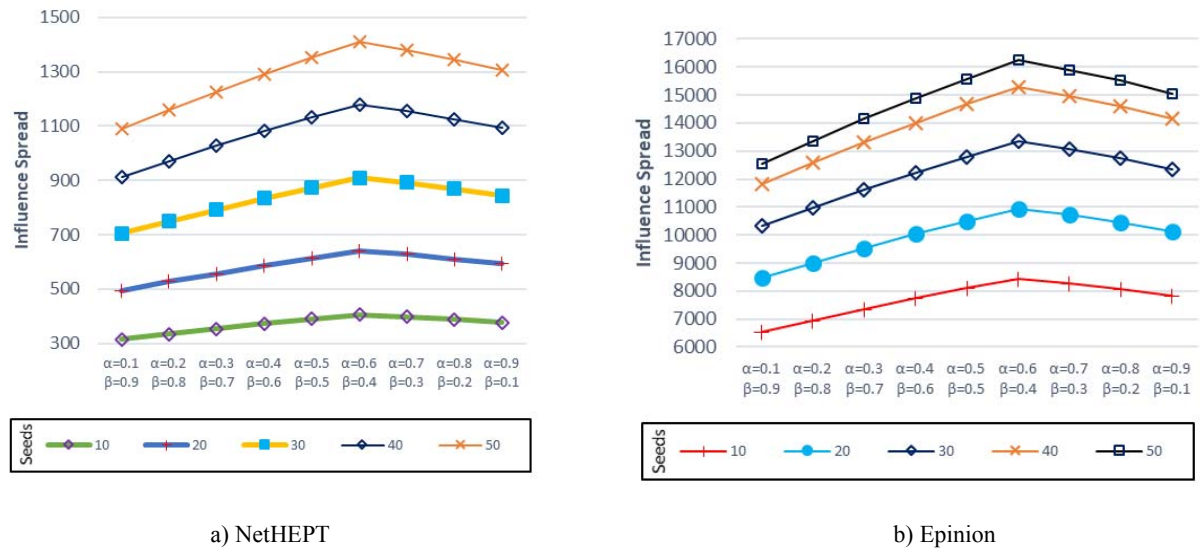


Fig. 3 The comparative results for different values of α and β .

$$Q_c = 2 \times Q_s \quad (4)$$

According to (4), in order to select the best seed nodes, the share of the nodes in each community is twice the share of the seeds of that community. According to the results of the experiments (Fig. 4), there is no need to select more candidate nodes.

In each community, we arrange the nodes according to the membership strength measure of each node to its community (M_{uc}), then, based on the quota of each community, we select candidate nodes in each community. Each node that has higher M_{uc} means that it has a higher membership strength to its community, therefore, it will have a greater influence on members of the community and it can maximize the influence diffusion in the community and the social network as a whole. Experimental results (Figs 5 to 7) clearly illustrate this issue.

The final seed nodes are selected from the candidate nodes. In each community, seed nodes are selected by testing the candidate nodes by a SimPath based algorithm. The SimPath algorithm has a very high accuracy for selecting seed nodes. In order to increase the speed of our proposed algorithm, we only test routes with a maximum length of two starting from the candidate nodes. Since the candidate nodes in each community are selected based on the membership strength of the node to the community, they have a high

quality, then we can reach the best seeds by testing them with a maximum length of two. The number of seed nodes in each community is based on the share of that community (Q_s).

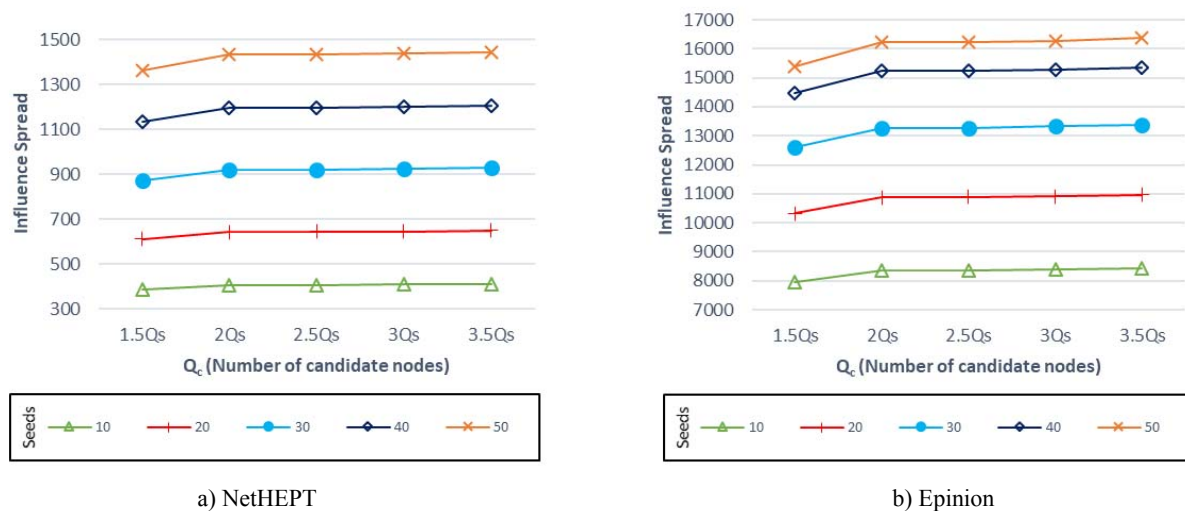


Fig. 4 The comparative results for the different number of candidate nodes.

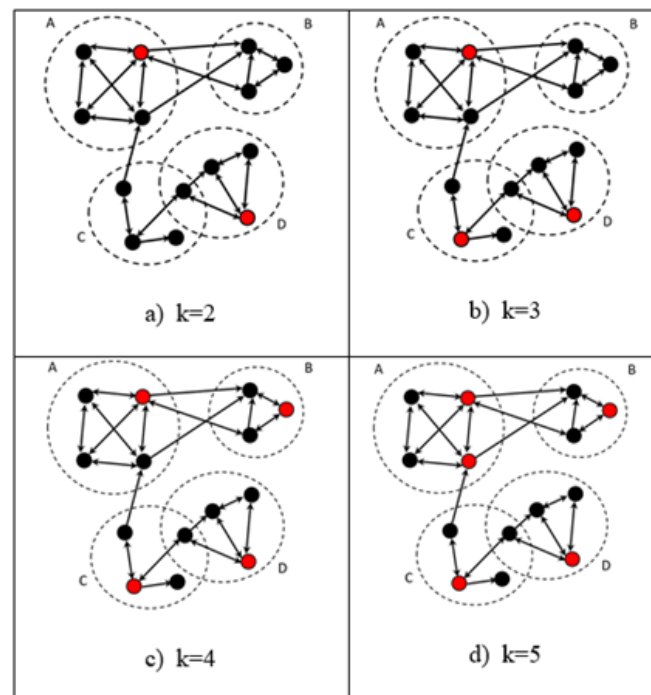


Fig. 5 A sample after running our proposed algorithm for different k .

In Fig. 5, the result of our proposed algorithm is shown on a sample graph for the different number of seeds.

6 Experimental Results

We compared our proposed algorithm with state of the art algorithms LDAG (Chen et al., 2010), CELF (Leskovec et al., 2007) and SIMPATH (Goyal et al., 2011). In addition, because our proposed method is a

community detection based method, we had to compare our algorithm with other community detection based algorithms, including COMPATH (Rahimkhani et al., 2015), CIM (Chen et al., 2014), COFIM (Shang et al., 2016) and INCIM (Bozorgi et al., 2016). We performed our experiments on two real data sets (Table 1) and eight synthetic datasets produced by LFR (Lancichinetti et al., 2008) and on a computer with 3.30 GHz Intel Core i3 CPU and 4GB of memory, and with Windows 10 operating system. Experiments were implemented using MATLABR2015b. Each algorithm is ran 35 times per seed.

Table 1 Summary of real datasets

Dataset	Directed/ Undirected	Node	Edge
NetHEPT	Undirected	15k	62k
Epinion	Directed	75k	508k

In Fig. 6, the IMMS algorithm is compared with other algorithms based on the number of activated nodes for different number of seeds on the real datasets NETHEPT and EPINION.

Our algorithm does not depend on the SLPA algorithm. To prove this, we run our experiments with the CoDA community detection algorithm (IMMS-CoDA) instead of SLPA (Figs 6 to 9).

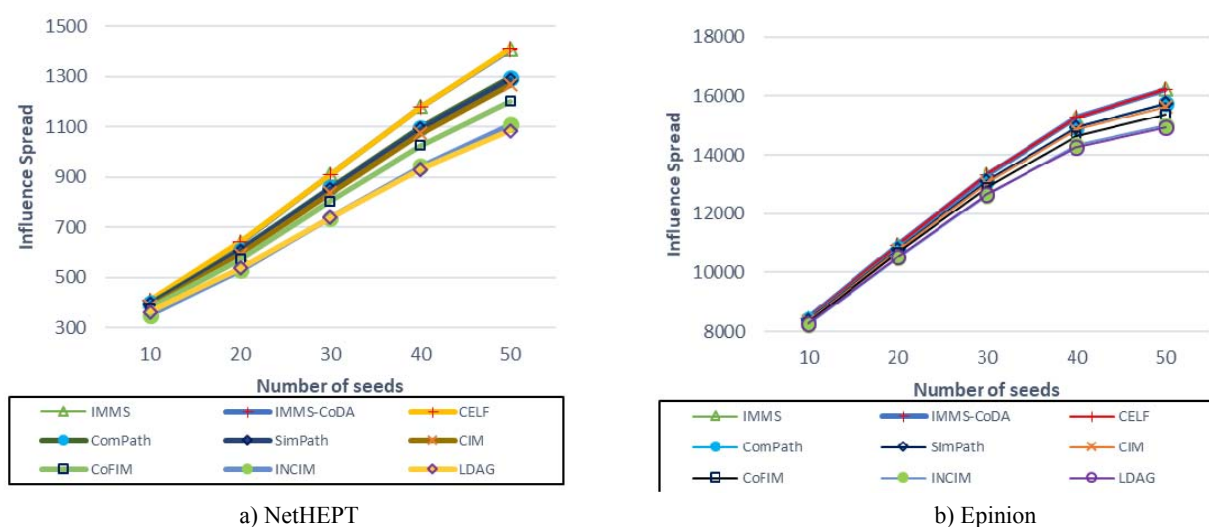


Fig. 6 The influence spread on real world networks.

In Fig. 7 and Fig. 8, our proposed algorithm is compared with other algorithms on the synthetic datasets produced by the LFR. The generated datasets parameters are: the average degree $k = 15$, the maximum degree $\max_k = 50$, the minimum for the community sizes $\min_c = 20$, the maximum for the community sizes $\max_c = 50$, and the mixing parameters m that show the community structures are 0.05 (very strong community), 0.1 (strong community), 0.2 (medium strong community), and 0.3 (weak community). In Fig. 9, algorithms have been compared in terms of Runtime.

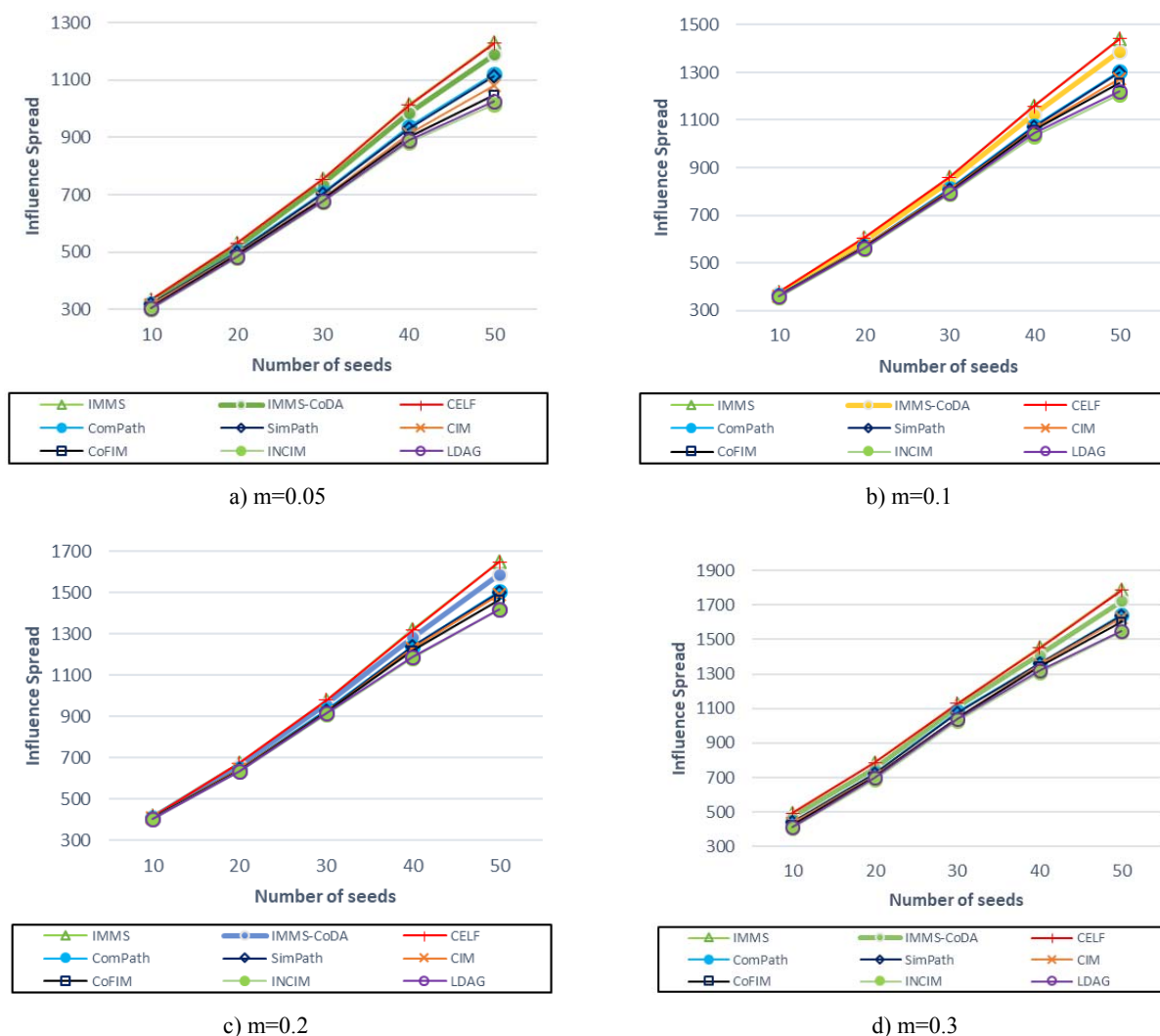
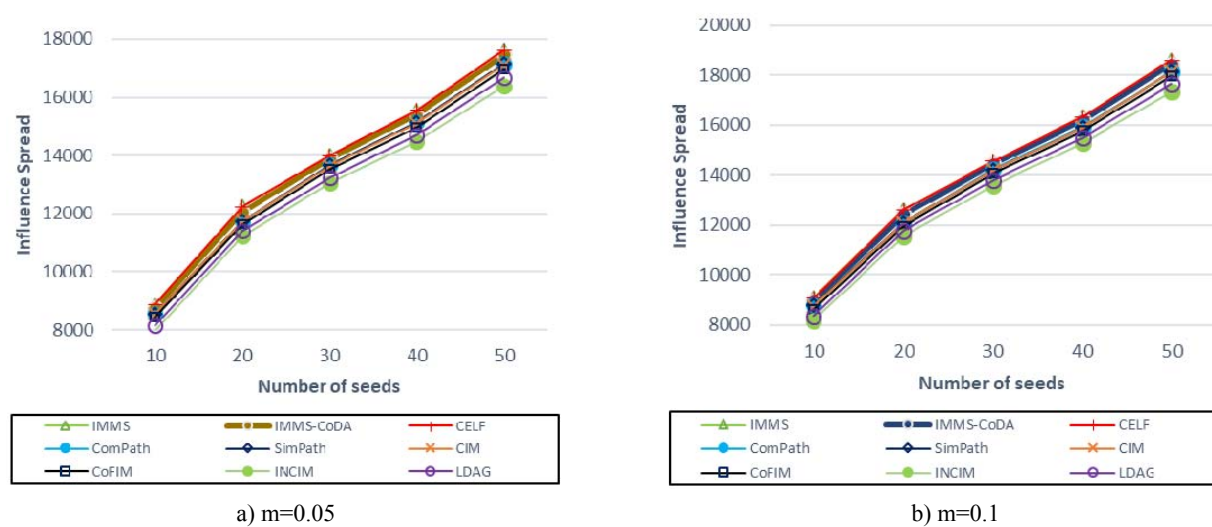


Fig. 7 The influence spread on four LFR synthetic networks with 10000 nodes and different community structures.



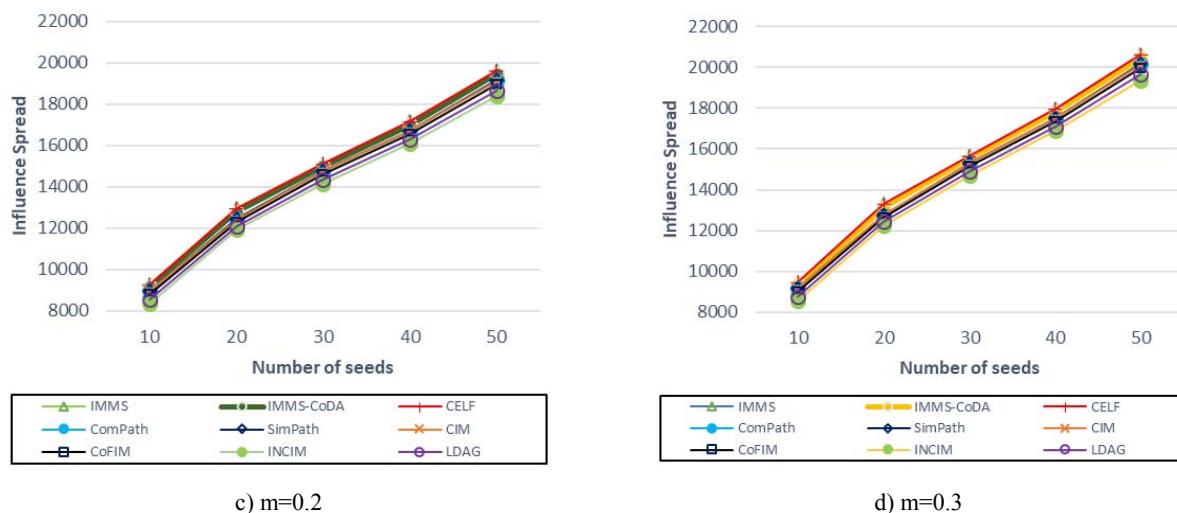


Fig. 8 The influence spread on four LFR synthetic networks with 50000 nodes and different community structures.

As shown in Fig. 6 to 8, increasing the number of initial required seeds, the difference in the number of activated nodes in different algorithms increases, and the difference between algorithms is quite clear. Our proposed algorithms IMMS and CELF activate more nodes than other algorithms. IMMS chooses the nodes in each community that have higher membership strength to their community. On the other hand, seeds are selected from different communities, so seeds are selected across the network and are distributed and can affect the entire network. When the required number of seeds is low, most algorithms have the same function, but by increasing the number of seeds, the results of the CELF algorithm and our proposed algorithm are better than other algorithms. The results show that our proposed method has similar results with CELF. Our algorithm outperforms CELF only in terms of runtime. COMPATH and SIMPATH algorithms affect the number of nodes less than IMMS and CELF. COMPATH and SIMPATH investigate different paths with different lengths from the initial seeds at the final seed selection step, so they select the most influential nodes, thus more nodes are activated. CIM and COFIM algorithms are ranked fifth and sixth in terms of number of activated nodes because they choose their final seeds in comparison with most of the candidate nodes in important communities, so they select good nodes as seeds. The INCIM algorithm activates less nodes compared to IMMS, CELF, CIM, COMPATH, SIMPATH and COFIM algorithms. Because it ignores some of the influential communities in the influence process. In fact, INCIM is effective on some networks and is not effective for a variety of networks with different structures. LDAG activates the smallest nodes compared to other algorithms because it computes the information diffusion locally within DAGs (Directed Acyclic Graph) and assumes that each node can affect a limited number of its neighbors. In other words, for selecting seeds, it does not consider the impact of each node on the information diffusion across the entire network.

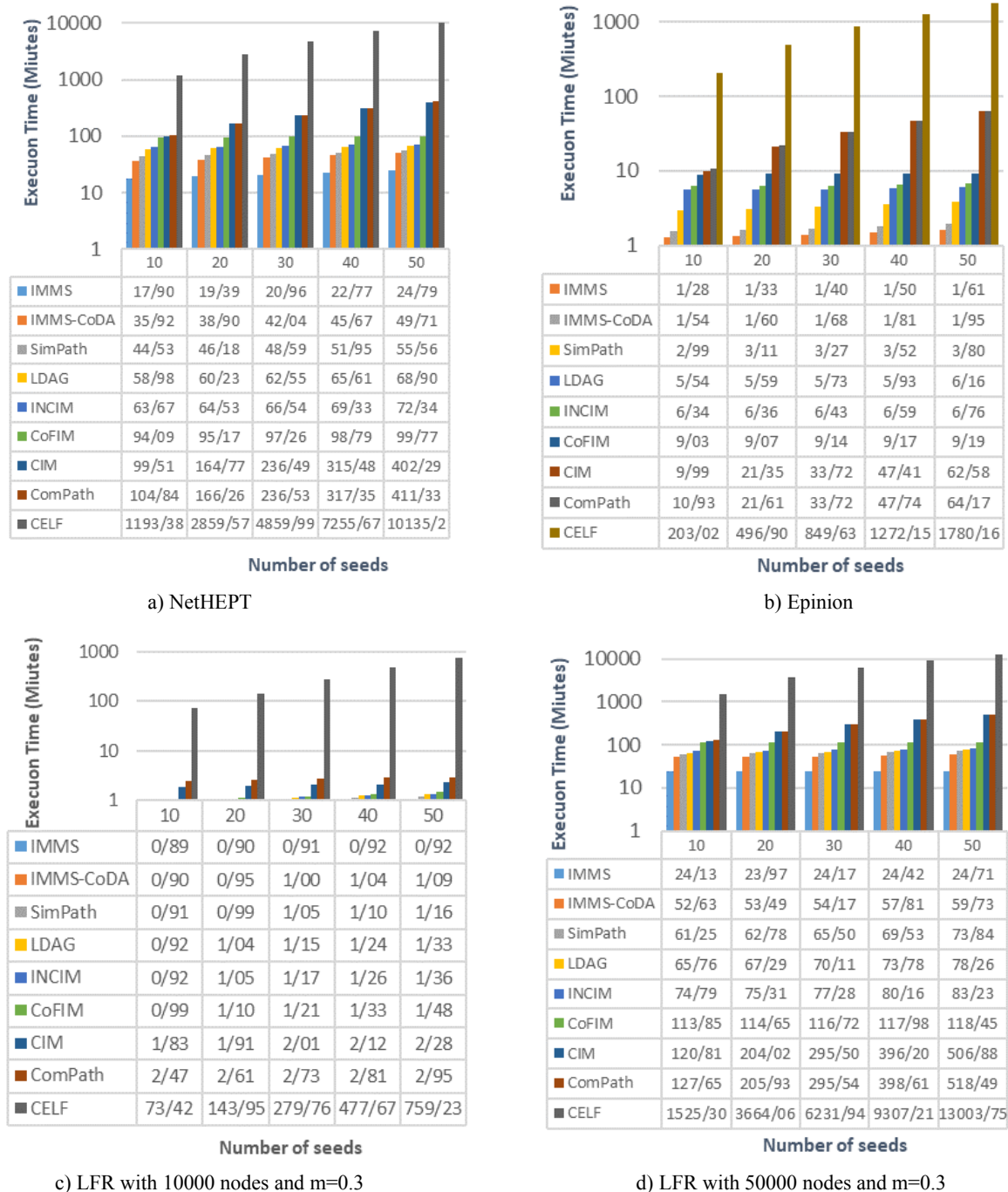


Fig. 9 Comparisons of running time.

In Fig. 9, execution time of various algorithms is computed on NETHEPT, EPINION, and LFR datasets. Our proposed algorithm, IMMS, has a higher rate compared to other algorithms because it first restricts the initial input network that is large by choosing its communities and then selects seed nodes within communities based on membership strength of the node to the community and its quota. Therefore, its speed is higher than other algorithms. After IMMS, SIMPATH and LDAG algorithms, have higher speeds, respectively because they do not consider many sections of the input network in the same way and investigate effect of the nodes locally. The INCIM algorithm is ranked next, because INCIM initially uses community

detection algorithms and eliminates unimportant communities, but its speed is reduced because it uses the SIMPATH algorithm twice and uses the CELF algorithm once. The COFIM algorithm is ranked fifth in terms of speed, since it works the same as the INCIM method, but computing the influence diffusion among communities and then computing the influence diffusion within communities reduces its speed compared to INCIM. After COFIM, the CIM algorithm has a faster rate because it gains important communities only based on size of communities and their relationship, so it can quickly ignore much of the network that has less impact and get the seeds faster than expected. COMPATH has a lower rate than CIM, because it examines different paths with different lengths that are available from the initial seeds at the final seed selection step and increase its runtime. The CELF method is slower than other algorithms, since it creates additional simulation graphs and searches for seed nodes greedily which reduces its speed significantly.

In Table 2 and Table 3, the best value, mean value and variance of the IMMS algorithm results are given. Also, the significance level for all results is less than 0.01.

The results show that our proposed algorithm (IMMS) not only activates more nodes, but also its speed is higher than other algorithms. Results of comparison with other algorithms show that if a large number of seeds are needed, our proposed algorithm can choose high quality seeds at a higher rate. Also, according to the results, our proposed method is more accurate than similar methods. On the other hand, our algorithm uses the community detection algorithm and then it does not consider the inefficient sections of the initial input network, therefore it is faster than other algorithms.

The IMMS algorithm consists of three steps presented in Algorithm 2. Time complexity of the IMMS algorithm is the sum of time complexity of these three steps. Time complexity of the first step (community detection) is equal to time complexity of the SLPA algorithm which is $O(T \times n)$, where n is the number of network nodes and T is the number of repetitions of the algorithm chosen by the user which is usually a small constant. Therefore, time complexity of the first step is $O(n)$. Time complexity of the second step (the membership strength computation of each node into the community) is $O(n)$ because it runs linearly. Time complexity of the third step is the sum of time complexity of the community quota computation and selecting candidate and seed nodes. The community quota computation is done linearly so its time complexity is $O(m)$, where m is the number of detected communities. Time complexity of selecting candidate nodes is $O(n \log n)$. The time complexity of selecting seed nodes is at most $O(k \times d_{\max})$, where k is the number of required nodes and d_{\max} is the maximum degree of nodes. The total time complexity of the algorithm is equal to $O(2n + m + n \log n + k \times d_{\max})$ and, consequently, is equal to $O(m + n \log n + k \times d_{\max})$. Time complexity of the entire algorithm is $O(2n + m + n \log n)$, therefore it is $O(m + n \log n)$. Compared to time complexity of other algorithms presented in Table 4, our proposed algorithm has less time complexity (t is the average time of LDAG calculation for each node and l_{\max} is the maximum size of LDAG).

Table 2 The best values, average values and variances of IMMS algorithm for different seeds on NetHEPT.

Seeds	Maximum	Average	Variance
10	421	406/60	22/20
20	657	639/53	21/80
30	933	909/89	22/70
40	1205	1178/53	22/31
50	1441	1408/64	23/77

Table 3 The best values, average values and variances of IMMS algorithm for different seeds on Epinion.

Seeds	Maximum	Average	Variance
10	8466	8435/71	47/43
20	10978	10946/13	45/53
30	13366	13332/07	46/39
40	15308	15273/67	43/71
50	16269	16230/23	51/81

Table 4 Time complexity of algorithms.

Algorithm	Time Complexity
IMMS	$O(m + n \log n)$
SimPath	$O(k \times n \times d_{\max})$
ComPath	$O(n + k \times n \times d_{\max})$
CoFIM	$O(k^2 \times n \times d_{\max})$
INCIM	$O(k^2 \times n \times d_{\max})$
CIM	$O(n + m \times (k + m))$
LDAG	$O(n \times t + k \times n \times l_{\max} \times (l_{\max} + \log n))$

7 Conclusion

In this paper, an influence maximization algorithm in the social network based on membership strength of the node to the community called IMMS is presented. IMMS initially detects communities by the SLPA algorithm from the input network. Then it computes membership strength of each node to its community, which combines output and input membership strengths of the node to its community. Finally, it determines the quota of each community from the seed nodes based on the number of community nodes and selects the seed nodes based on their membership strength. Experimental results show that our proposed algorithm can quickly select high quality seeds. In other words, the IMMS algorithm can activate more nodes in less time using the selected seeds. If a large number of seeds are needed, this algorithm can choose them at a high speed. For future work, it is suggested that the method presented in this paper is examined in dynamic social networks, and it is suggested to find influential nodes based on characteristics of nodes as an individual in the community.

References

- Bagheri E, Dastghaibafard G, Hamzeh A. 2016. An efficient and fast influence maximization algorithm based on community detection. 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD). Changsha, China
- Bagheri E, Dastghaibafard G, Hamzeh A. 2018. FSIM: A fast and scalable influence maximization algorithm based on community detection. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 26(3)
- Bozorgi A, Haghighi H, Zahedi M, Rezvani M. 2016. INCIM: A community-based algorithm for influence maximization problem under the linear threshold model. Information Processing and Management, 52(6): 1188-1199
- Chen W, Yuan Y, Zhang L. 2010. Scalable influence maximization in social networks under the linear threshold model. The 2010 IEEE International Conference On Data Mining. Washington DC, USA
- Chen YC, Zhu WY, Peng WC, Lee WC, Lee SY. 2014. CIM: Community-based influence maximization in

- social networks. *ACM Transactions on Intelligent Systems and Technology*, 5(2): 25
- Domingos P, Richardson M. 2001. Mining the network value of customers. The 7th International Conference On Knowledge Discovery and Data Mining. 57-66, ACM, New York, USA
- Ghanbari A, Bagheri E. An influence maximization algorithm in social network using K-shell decomposition and community detection. *Network Biology*, 2020, 6(1): 163-168
- Goyal A, Lu W, Lakshmanan L. 2011. SIMPATH: An efficient algorithm for influence maximization under the linear threshold model. The 2011 IEEE 11th International Conference on Data Mining (ICDM). 211-220, Washington, USA
- Kempe D, Kleinberg J, Tardos E. 2003. Maximizing the spread of influence through a social network. The 9th ACM SIGKDD International Conference On Knowledge Discovery and Data Mining. 137-146, ACM, Washington, USA
- Lancichinetti A, Fortunato S, Radicchi F. 2008. Benchmark graphs for testing community detection algorithms. *Physical Review E arXiv*, 78(4): 46110
- Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance N. 2007. Cost-effective outbreak detection in networks. The 13th ACM SIGKDD International Conference On Knowledge Discovery and Data Mining. San Jose, California, USA
- Liu X, Liao X, Li S, Zheng S, Lin B, Zhang J, et al. 2017. On the Shoulders of Giants: Incremental Influence Maximization in Evolving Social Networks. *Complexity*, 2017: 5049836
- Mohamadi-Baghmolaei R, Mozafari N, Hamzeh A. 2015. Trust based latency aware influence maximization in social networks. *Engineering Applications of Artificial Intelligence*, 41: 195-206
- Ok J, Jin Y, Shin J, Yi Y. 2014. On maximizing diffusion speed in social networks: impact of random seeding and clustering. The 2014 ACM International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS '14). ACM, 301-313, New York, NY, USA
- Rahimkhani K, Aleahmad A, Rahgozar M, Moeini A. 2015. A fast algorithm for finding most influential people based on the linear threshold model. *Expert Systems with Applications*, 42(3): 1353-1361
- Shang J, Zhou S, Li X, Liu L, Wu H. 2016. CoFIM: A community-based framework for influence maximization on large-scale networks. *Knowledge-Based Systems*, 117: 88-100
- Shang J, Wu H, Zhou S, Zhong J, Feng Y, Qiang B. 2018. IMPC: Influence maximization based on multi-neighbor potential in community networks. *Physica A: Statistical Mechanics and its Application*, 512: 1085-1103
- Song G, Zhou X, Wang Y, Xie K. 2015. Influence maximization on large-scale mobile social network: a divide-and-conquer method. *IEEE Trans. IEEE Transactions on Parallel and Distributed Systems*, 26: 1379-1392
- Wasserman S, Faust K. 1994. *Social Network Analysis: Methods and applications*. Cambridge University Press, USA
- Xie J, Szymanski BK, Liu X. 2011. SLPA: Uncovering overlapping communities in social networks via a speaker-listener interaction dynamic process. The 2011 IEEE 11th International Conference On Data Mining Workshops. Vancouver, Canada
- Yang J, McAuley J, Leskovec J. 2014. Detecting cohesive and 2-mode communities in directed and undirected networks. *Proceedings of the 7th ACM International Conference On Web Search and Data Mining*. New York, USA
- Zhang WJ. 2016a. A method for identifying hierarchical sub-networks / modules and weighting network links based on their similarity in sub-network / module affiliation. *Network Pharmacology*, 1(2): 54-65
- Zhang WJ. 2016b. *Selforganizology: The Science of Self-Organization*. World Scientific, Singapore
- Zhang WJ. 2018. *Fundamentals of Network Biology*. World Scientific Europe, London, UK