Article

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

Esmaeil 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).

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).

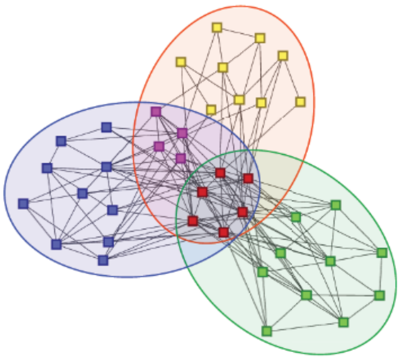


Fig. 1 Overlapping communities.

3 Problem Description

We consider directed graph G = {V, E} with |V| = N vertices and |E| = M edges. For each edge (v, u) ∈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 σ (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), σ (S) which is the influence diffusion function, is equal to the number of active users in the network after the diffusion process is stopped.

$$S^{*}= argmax_{S}\, \sigma(S) \tag{1}$$

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 ← Ø;
for i = 1 to k do

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 betweeness 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-Baghmolaei et al. (2015) have presented trust-based latency-aware influence maximization algorithm called TLIM (Mohamadi-Baghmolaei 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

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.

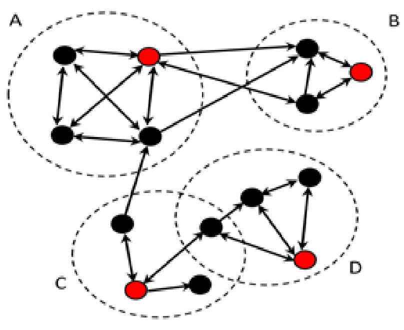
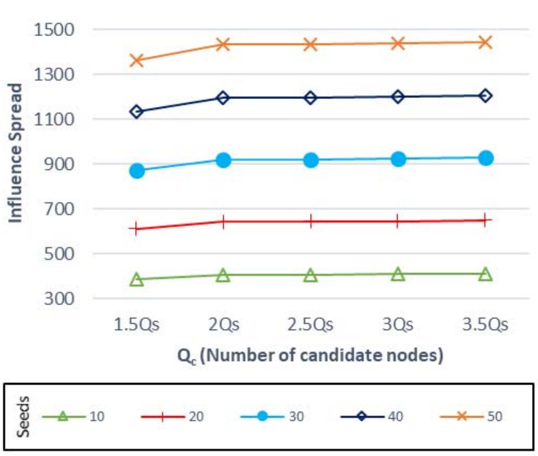


Fig. 2 A sample network with four communities.

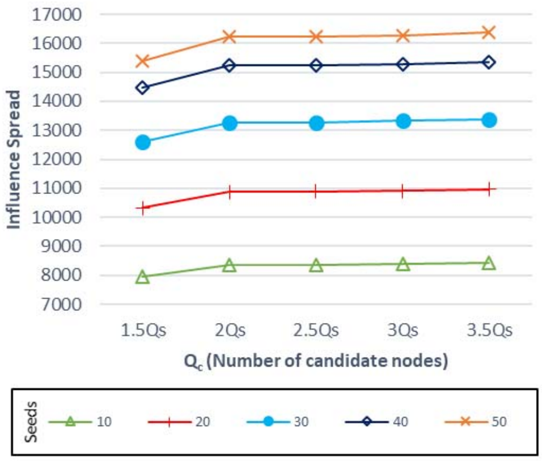
If O_{uc} is the outgoing membership degree of the node u to the community c and I_{uc} is the incoming membership degree of the node u to the community c, then the membership degree of the node u to the community c is determined by (2):

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

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).



a) NetHEPT



b) Epinion

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

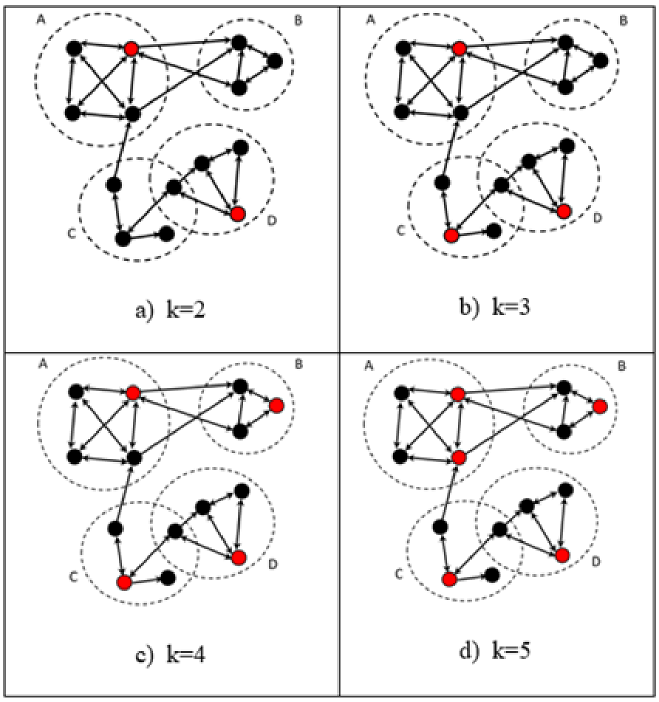
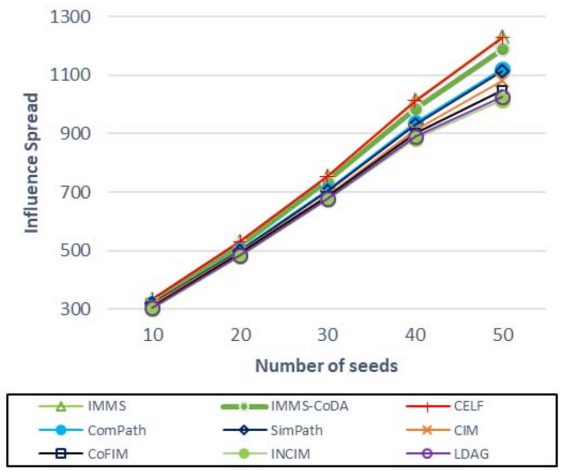


Fig. 5 A sample after running our proposed algorithm for different k.

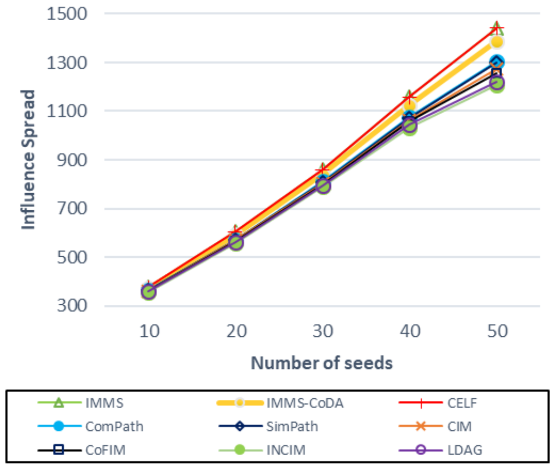
In Fig. 5, the result of our proposed algorithm is shown on a sample graph for the different number of seeds.

6 Experimental Results

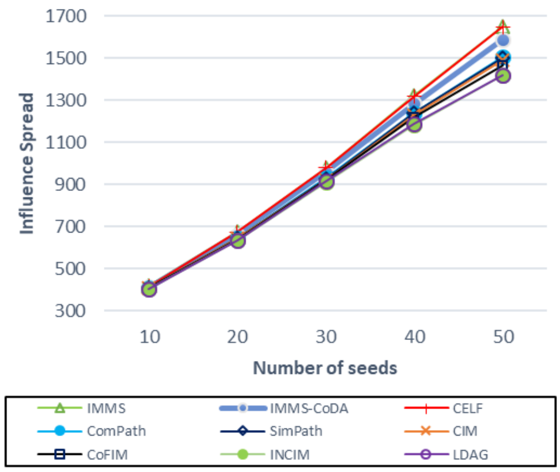
We compared our proposed algorithm with state of the art algorithms LDAG (Chen et al., 2010), CELF (Leskovec et al., 2007) and SIMPATH (Goyal et al., 2011). In addition, because our proposed method is a



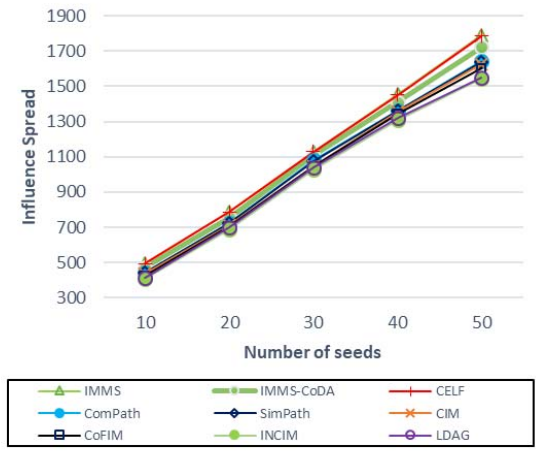
a) m=0.05



b) m=0.1

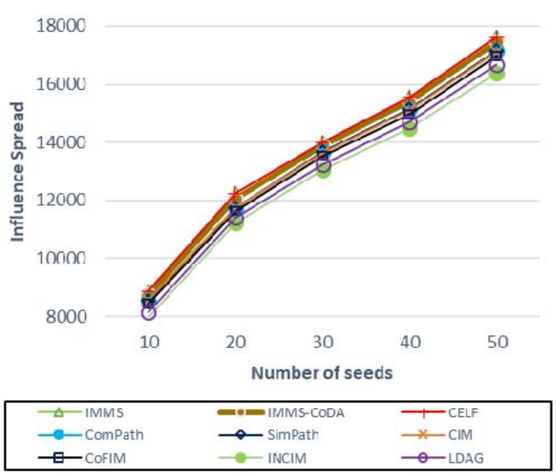


c) m=0.2

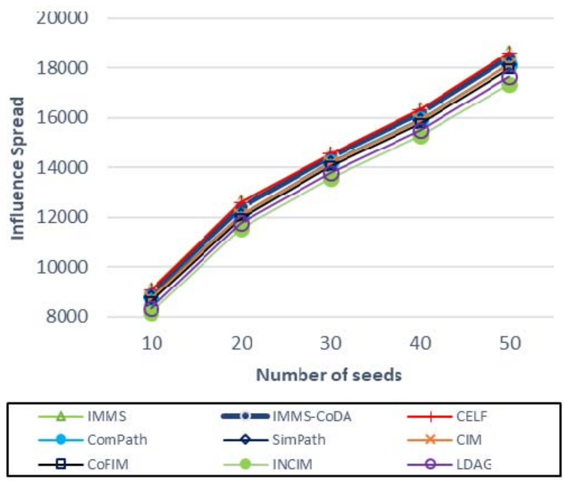


d) m=0.3

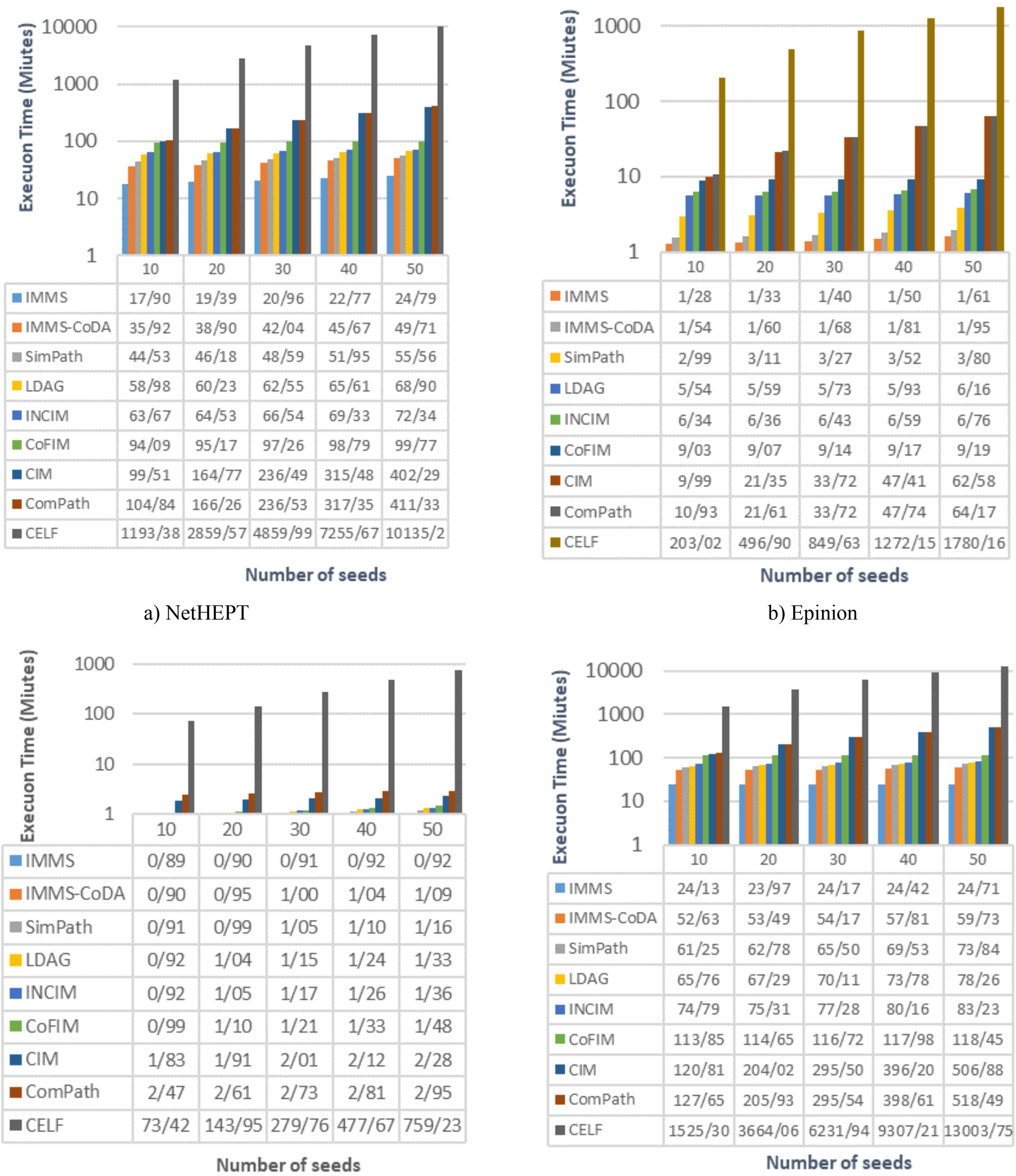
Fig. 7 The influence spread on four LFR synthetic networks with 10000 nodes and different community structures.



a) m=0.05



b) m=0.1



	10	20	30	40	50
IMMS	17/90	19/39	20/96	22/77	24/79
IMMS-CoDA	35/92	38/90	42/04	45/67	49/71
SimPath	44/53	46/18	48/59	51/95	55/56
LDAG	58/98	60/23	62/55	65/61	68/90
INCIM	63/67	64/53	66/54	69/33	72/34
CoFIM	94/09	95/17	97/26	98/79	99/77
CIM	99/51	164/77	236/49	315/48	402/29
ComPath	104/84	166/26	236/53	317/35	411/33
CELF	1193/38	2859/57	4859/99	7255/67	10135/2

a) NetHEPT

	10	20	30	40	50
IMMS	1/28	1/33	1/40	1/50	1/61
IMMS-CoDA	1/54	1/60	1/68	1/81	1/95
SimPath	2/99	3/11	3/27	3/52	3/80
LDAG	5/54	5/59	5/73	5/93	6/16
INCIM	6/34	6/36	6/43	6/59	6/76
CoFIM	9/03	9/07	9/14	9/17	9/19
CIM	9/99	21/35	33/72	47/41	62/58
ComPath	10/93	21/61	33/72	47/74	64/17
CELF	203/02	496/90	849/63	1272/15	1780/16

b) Epinion

	10	20	30	40	50
IMMS	0/89	0/90	0/91	0/92	0/92
IMMS-CoDA	0/90	0/95	1/00	1/04	1/09
SimPath	0/91	0/99	1/05	1/10	1/16
LDAG	0/92	1/04	1/15	1/24	1/33
INCIM	0/92	1/05	1/17	1/26	1/36
CoFIM	0/99	1/10	1/21	1/33	1/48
CIM	1/83	1/91	2/01	2/12	2/28
ComPath	2/47	2/61	2/73	2/81	2/95
CELF	73/42	143/95	279/76	477/67	759/23

c) LFR with 10000 nodes and m=0.3

	10	20	30	40	50
IMMS	24/13	23/97	24/17	24/42	24/71
IMMS-CoDA	52/63	53/49	54/17	57/81	59/73
SimPath	61/25	62/78	65/50	69/53	73/84
LDAG	65/76	67/29	70/11	73/78	78/26
INCIM	74/79	75/31	77/28	80/16	83/23
CoFIM	113/85	114/65	116/72	117/98	118/45
CIM	120/81	204/02	295/50	396/20	506/88
ComPath	127/65	205/93	295/54	398/61	518/49
CELF	1525/30	3664/06	6231/94	9307/21	13003/75

d) LFR with 50000 nodes and m=0.3

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

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 + nlogn)
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, Dastghaibyfard 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, Dastghaibyfard 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