An Evaluation Algorithm of the Importance of Network Node Based on Community Influence

Gongzhen He^(✉), Junyong Luo, and Meijuan Yin

State Key Laboratory of Mathematical Engineering and Advanced Computing,
Zhengzhou 450001, China
he_gz_study@163.com

Abstract. Identifying nodes in social networks that have great influence on information dissemination is of great significance for monitoring and guiding information dissemination. There are few methods to study the influence of communities on social networks among the existing node importance evaluation algorithms, and it is difficult to find nodes that promote information dissemination among communities. In view of this reason, this paper proposes a node importance evaluation algorithm based on community influence (abbreviated as IEBoCI algorithm), which evaluates the importance of the nodes based on the influence degree of the nodes on the communities and the ability to disseminate information the communities to which the nodes are connected. This algorithm firstly calculates the activation probability of nodes to other nodes, which is used to divide communities and evaluate influence. Secondly, the network is divided into communities based on LPA algorithm. Finally, the importance of the node is calculated by combining the influence of the community itself and the influence of the node on the community. Experiments are carried out on real social network data and compared with other community-based methods to verify the effectiveness of the algorithm.

Keywords: Complex network · Social network · Node importance · Community detection · Diffusion model

1 Introduction

With the rapid development of Internet and information technology, social networks such as Weibo, Facebook, Flickr, Twitter, etc. have developed rapidly. Social networks have become one of the main platform for human beings to spread information. Identifying nodes with great influence on information dissemination in social networks is helpful for in-depth analysis of information dissemination and evolution in social networks. Finding the guider or pusher in network public opinion is of great significance for controlling and guiding network public opinion, cracking down on network information crimes, and realizing viral marketing and word-of-mouth communication.

There are two main methods for evaluating the importance of nodes, methods based on centralities and methods based on information dissemination scale. The centrality-based method evaluates the centrality of nodes depends on the network structure, which

represents the degree of nodes in the center of the network. These centrality methods mainly include Degree Centrality [1], Betweenness Centrality [2], Closeness Centrality [3], Eigenvector Centrality [4], etc. This kind of method is suitable for finding the important nodes in the network structure, but not for evaluating the influence of nodes.

The method based on information dissemination scale is to use information dissemination model to simulate the information dissemination process, calculate the information dissemination scale of nodes, and find out the nodes with great influence. A greedy algorithm for calculating the propagation scale of nodes was first proposed by Kempe et al. [5], which is very time consuming and only suitable for small networks. In order to reduce the computational complexity, Leskovec et al. [6] proposed CELF algorithm according to the submodules of influence diffusion, avoiding redundant calculation of activation range. References [7–11] used heuristic strategy instead of Monte Carlo simulation to estimate propagation scale for improving time efficiency. This kind of method finds influential nodes by directly measuring the information dissemination scale of nodes, but it is not suitable for evaluating the importance of nodes in the network that indirectly disseminate influence.

The methods of Cao [15], Wang [16], Shang [17], Zhang [18] and other teams assume the independence between communities. After dividing the network into communities, they find the node with the greatest local influence in each community, and then find the node with the greatest influence in the whole network. M. M. Tulu [19] et al. calculated the node's Shannon Entropy as the node's importance by using the number of nodes outside the community and the number of nodes inside the community after the community was divided. Zhao [20] measured the importance of nodes by the number of communities which the nodes connected to after dividing the network into communities. These methods either do not focus on the association between the communities, or do not consider the relationship between the nodes and the different communities, or do not consider the influence of the communities themselves.

In order to deal with the above problems, we propose a node importance evaluation algorithm based on community influence (abbreviated as IEBoCI algorithm). Its basic assumption is that the stronger the ability of the community connected by nodes to disseminate information and the greater the influence of nodes on the community, the higher the importance of nodes. The algorithm first calculates the activation probability of nodes to other nodes; Secondly, the network is divided into communities based on LPA algorithm; Thirdly, calculate the influence of each community and the influence degree of nodes on the connected communities; Finally, the importance of the node is calculated by combining the influence of the community itself and the influence of the node on the community.

2 IEBoCI Algorithm Framework

If a person has many friends in different societies in social networks, this person does not necessarily directly disseminate a large amount of important information, but he can indirectly disseminate the information in the community through contacts with other community members. From this we can see that this person has a wide influence on information dissemination and plays a more important role in the network.

We believe that the influence of nodes is related to the number and quality of communities connected by nodes based on this assumption. The more communities connected, the greater the influence of the nodes and the higher the importance. Meanwhile, the influence of the nodes is also related to the influence of the connected communities. For the same community, the influence degree of different nodes on the community is also different. If a node has less influence on a community, it is difficult for the node to influence the nodes in the community, and it is not easy to further spread information through the community. Therefore, it is necessary to comprehensively evaluate the influence of the community itself and the influence degree of nodes on the community when evaluating the importance of nodes.

3 Algorithm Steps

Social network is a complex network, which is denoted as directed network $G = (N, E)$ in this paper, among which N is a collection of nodes in a network, E is a set of directed edges in a network. The IEBoCI algorithm proposed in this paper is based on directed network. The algorithm flow is shown in Fig. 1. The steps are as follows:

- (1) calculate the activation probability of nodes activating their reachable nodes based on the information propagation model, which is used to divide communities and calculate the influence range of communities and nodes; (2) divide the network based on label propagation algorithm to obtain the community structure of the network; (3) calculate the influence range of communities according to the activation probability of the nodes; (4) calculate the number expectation of the nodes on communities activated by the nodes, and further obtaining the influence degree of the nodes on communities; (5) calculate the importance of nodes by combining the results of the third and fourth steps.

Fig. 1. Algorithm flow chart

3.1 Node Activation Probability Calculation

This paper calculates the information dissemination scale of nodes and communities based on independent cascade model (IC model). IC model is a probability model, and

there is a probability $p(v_i, v_j) \in [0, 1]$ for all neighboring nodes v_i and v_j in network G . A value between 0 and 1 is randomly assigned, which indicates the probability that the active node v_i successfully directly activates the neighbor node v_j . For non-adjacent nodes v_i and v_j , if v_i to v_j are unreachable, the probability of v_i activating v_j is 0, that is $p(v_i, v_j) = 0$. If v_i is reachable to v_j , the probability that node v_i activates node v_j along the path is the product of one node directly activating another node on each side of the path [22]. If there are m paths between v_i and v_j , one of the paths is $\text{Path}(v_i, v_j)_x = \langle v_j = v_1, v_2, \dots, v_j = v_k \rangle$, the probability $Pp(v_i, v_j)_x$ that node v_i activates node v_j along this path is calculated as follows:

$$Pp(v_i, v_j)_x = \prod_{u=1}^{k-1} p(v_u, v_{u+1}) \quad (1)$$

Where $Pp(v_i, v_j)_x$ is the probability that v_i activates v_j through $\text{Path}(v_i, v_j)_x$ and $p(v_u, v_{u+1})$ is the probability that node v_u activates neighbor node v_{u+1} . If the activation probabilities of different paths are different, the maximum probability is taken as the probability $p(v_i, v_j)$ for node v_i to activate non-adjacent reachable node v_j .

3.2 Community Detection

We divide network into communities based on label propagation algorithm (LPA algorithm) to obtain a collection of communities on the network. LPA algorithm is applicable to undirected and unweighted networks. The social network constructed in this paper is a directed network. Therefore, when calculating the labels to be updated of node v_j , only the in-neighbors of node v_j are calculated, and the labels with the highest activation probability among the neighbors are counted. The steps are as follows:

step1: Label initialization: each node in the network is randomly assigned a unique label l , which represents the community in which the node is located

step2: Determining the node order of the asynchronous updating labels: calculate the degree of the nodes, and arrange the node order of the asynchronous updating labels from large to small according to the degree of the nodes;

step3: Updating the labels of nodes: according to the node order of updating the labels, the labels of nodes are updated one by one, and the label of node v_j is updated to the label with the maximum sum of activation probabilities in its in-neighbor nodes. The label updating formula is as follows:

$$l_{v_j} = \arg \max_l \sum_{i \in IN(v_j)} p(v_i, v_j) \delta(l_{v_i}, l) \quad (2)$$

l_{v_j} represents the label of node v_j to be updated, l_{v_i} represents the label of node v_i , $IN(v_j)$ represents the set of nodes with out-edges to node v_j , $p(v_i, v_j)$ represents the activation probability from node v_i to node v_j , and $\delta(l_i, l)$ is a Kronecker function.

When there is more than one label with the maximum sum of the calculated activation probabilities, one label is randomly selected from them as the new label of the node.

Step4: Termination judgment: it is judged whether the labels of all nodes in the network are the labels with the largest sum of activation probabilities among neighbor nodes. If not, step3 is repeatedly executed, if so, calculation is terminated, and nodes with the same label belong to the same community.

3.3 Evaluation of Influence of Communities

In this paper, the number expectation of network nodes activated by a community is taken as the influence of the community. The steps are as follows: firstly, the activation probability between nodes calculated by 3.1 is used to calculate the joint activation probability of all nodes in the community to nodes in the network; then calculate the number expectation of nodes activated by the community according to the joint activation probability to obtain the influence of the community.

In the independent cascade model, whether a node activates another node and whether other nodes activate the node are independent events, so the joint activation probability $Ps(C_l, v_j)$ [23] of the community C_l to a node v_j in the network is calculated according to the probability multiplication of the independent events, and the calculation formula is:

$$Ps(C_l, v_j) = 1 - \prod_{v_i \in C_l} (1 - p(v_i, v_j)) \quad (3)$$

With the joint activation probability $Ps(C_l, v_j)$, the influence scale expectation of the community $EXPs(C_l)$ is calculated as follows:

$$EXPs(C_l) = \sum_{v_j \in N} Ps(C_l, v_j) \quad (4)$$

Where N is the set of all nodes in the network.

3.4 Evaluation of Influence of Nodes on Communities

The number expectation of nodes in a community activated by a node indicates how many nodes in the community a node can successfully activate. The number of nodes in the community indicates the total scale of the community. The greater the proportion of nodes in the community that a node can activate in all nodes of the community, the greater the influence of the node on the community. Therefore, this paper regards the ratio of number expectation of nodes in a community activated by a node to the number of nodes in the community as the influence degree of the node on the community.

Limiting the range of nodes activated by node v_i in the community C_l , the number expectation of nodes $EXPn(v_i, C_l)$ in the community C_l activated by node v_i is obtained, which is equal to the sum of the probabilities of each node in the community C_l being successfully activated by node v_i , and the calculation formula is as follows:

$$EXPn(v_i, C_l) = \sum_{v_j \in C_l} p(v_i, v_j) \quad (5)$$

Where C_l represents the set of all nodes in the community C_l .

$EXPn(v_i, C_l)$ represents the number of nodes that node v_i can activate in community C_l . The ratio of this expectation to the total number of nodes $n(C_l)$ in community C_l is the influence degree $Inf(v_i, C_l)$ of node v_i on community v_i . The formula is:

$$Inf(v_i, C_l) = \frac{EXPn(v_i, C_l)}{n(C_l)} \quad (6)$$

3.5 Node Importance Evaluation

Node v_i can use influence of communities directly connected by node v_i to spread influence indirectly. We calculate the sum of the influence of communities that node v_i can indirectly use, and get the importance $I(v_i)$ of node V_i . The importance $I(v_i)$ of the node v_i is calculated by the following formula using influence of communities and the influence of a node on community:

$$I(v_i) = \sum_{C_l \in Com(v_i)} Inf(v_i, C_l) EXPs(C_l) \quad (7)$$

Where $Com(v_i)$ represents the set of communities in which node v_i and its out-neighbors are located.

4 Experimental Results and Discussions

4.1 Experimental Data and Initial Setup

The experimental data used in this paper are commonly used public social network data sets, which are downloaded from Internet. The name, data scale and description of the network are shown in Table 1.

Table 1. Basic information of datasets

Data set	Number of nodes	Number of edges	Network density	Average out degree	Minimum out degree	Maximum out degree	Description
Facebook ^a	4039	176468	0.0108200	43.691	1	1045	Friends on Facebook
E-Mail ^b	1866	5517	0.0015853	2.9566	0	330	Communication between users

Note: ^a<https://snap.stanford.edu/data/ego-Facebook.html>

^b<http://konect.uni-koblenz.de/networks/dnc-temporalGraph>

The nodes in Facebook are friends relationship, and the formed network is an undirected network. In this experiment, each edge is converted into two directed edges to

convert undirected network into directed network. The nodes in the mail are communication relationship. Each communication forms a directed edge from one node to another node.

In order to simulate the dissemination of information on the network, this paper uses node v_i as the initial activation node, and the scale of the nodes that can be affected by node v_i as a measure of the node's information dissemination capability, which is referred to as the Influence scale. In order to simulate the influence propagation process of nodes and calculate the influence scale of nodes, the commonly used independent cascade model (IC model) [21] is adopted in this paper.

The activation probability p between nodes is randomly assigned a value between 0 and 1 when constructing a network using data sets. Due to the randomness of the independent cascade model, the results may be different when calculating the influence scale of nodes. To sum up, each node is taken as the initial activated node to calculate the influence scale for 50 times when calculating the influence scale of nodes, and then the arithmetic average value is taken as the final result.

4.2 The Division and Influence of Communities

The two data sets are divided into communities after the network construction is completed, according to the activation probability between nodes by using the LPA algorithm improved previously. The division results are shown in Tables 2 and 3.

Table 2. Communities of Facebook data sets

Community size	Number of communities	Community size	Number of communities
2	10	40	1
3	1	43	1
4	2	54	1
6	3	72	1
7	1	84	1
8	4	106	1
10	1	189	1
12	1	225	1
14	1	226	1
19	2	266	1
24	1	344	1
25	1	347	1
27	1	467	1
33	1	514	1
38	1	753	1
Total number of communities		46	

Table 3. Communities of mail data sets

Community size	Number of communities
1	490
2	27
3	5
4	3
6	1
7	1
20	1
35	2
537	1
655	1
Total number of communities	532

Fig. 2. Influence of communities

According to the results of community detection, it can be seen that the structural characteristics of Facebook and email are quite different: Facebook has 4039 nodes and 46 communities are divided, with a small number of communities and a large scale of communities, which indicates that the network is relatively close. There are 1866 nodes in the mail network, 532 of which are divided into communities. The number of communities is large, but the number of large-scale communities is small. More communities are 1 node and 2 nodes, which shows that the network is sparse.

The influence of communities is calculated after the community detection is completed, according to the method proposed in this paper, and the calculation results are shown in Fig. 2. Figure 2. (a) shows the influence of communities in Facebook data. The number of communities is small, and the size of communities (the number of members of communities) varies greatly. From the overall trend, the larger the size of communities, the greater the influence of communities. When the influence of the community reaches a certain degree (the number of nodes affected reaches more than 90% of the total number of nodes), the increase in influence becomes less and less obvious, which is consistent with the reality. Figure 2. (b) shows the influence of communities in E-Mail data. Generally speaking, there is also a trend that “the larger the community size, the greater the influence of the community”. As there are a large number of 1-node and 2-node communities in the network, the data in the lower left corner of the image is relatively dense, and the influence of communities is not necessarily the same under the same scale.

4.3 The Importance of Node

According to the method proposed in this paper, the importance I of all nodes in data sets and the Influence scale of nodes are calculated, and compared with a method of indirectly measuring the importance of nodes through communities (the number of directly connected communities V -community [20]). The Degreeout and Betweenness are analyzed as statistical data.

Distribution of Node Importance. After calculating the importance of nodes, numbers of nodes with different values of importance were counted in Facebook and E-Mail data sets. The number of nodes was counted in Facebook data set according to the importance

with 100 as an interval, and the number of nodes was counted in E-Mail data set according to the importance with 20 as an interval. The distribution of statistical results is shown in Fig. 3. The distribution in the two data sets is different. The distribution in Facebook is positively skew distribution and the distribution in mail is power law distribution. The difference between the two results lies in the different characteristics of the two data sets. The network in Facebook data is a directed network transformed from an undirected network, and each node has an out-degree greater than 0, so each node can transmit information to other nodes. In the mail data, the directed network is constructed according to the communication relationship. There are a large number of nodes in the data set that receive mail but do not send mail, which have an out-degree of 0 and do not carry out information dissemination to the outside. Therefore, a large number of nodes with an out-degree of 0 result in a large number of nodes with low importance in the data statistics. So the distribution of importance presents a power law distribution.

Fig. 3. Node importance statistics

Comparison of Node Importance and Number of Directly Connected Communities.

The number distribution statistics of the Influence scale of nodes in the two data sets are shown in Fig. 4.

Fig. 4. Node influence scale statistics

In Fig. 4, (a) Facebook data counts the number of nodes according to the node Influence scale with 100 as an interval, and (b) E-Mail data counts the number of nodes according to the node Influence scale with 20 as an interval.

Most of the nodes in the Facebook dataset have a large scale of influence, with 1746 nodes in the [3900, 4000) interval and 1649 nodes in the [3800, 3900) interval. Facebook dataset has 4,039 nodes, of which three-quarters can affect 95% of the network. This result is also related to the close connection of nodes in the data set. The density of the network is high, the average outdegree of nodes is also large, and the influence spread range of most nodes is large.

The influence scale of most nodes in E-Mail data set is very small, 921 nodes are in [0, 20) interval. This result is related to the fact that nodes in the data set are not closely related. Different nodes have different impact sizes. There are 1866 nodes in the data set, of which 800 nodes have an output of 0. These nodes cannot transmit information outward, so nodes with low Influence scale account for the majority.

The relationship between node outdegree, betweenness and node influence scale is shown in Fig. 5. In the chart, the X axis represents the node outdegree and betweenness for the corresponding data set, and the Y axis represents the node Influence scale. In Fig. 5. (a) (b), it can be seen that there is almost no correlation between the node influence scale and the node output. In Fig. 5. (c) (d), it can also be seen that there is almost no correlation between node influence scale and node betweenness.

Fig. 5. Relationship between outdegree, betweenness and node influence scale

The relationship between number of directly connected communitie (V-community), node importance and node influence scale is shown in Fig. 6. In the chart, theX axis represents V-community and node importance for the corresponding data set, and the Y axis represents the node Influence scale. In Fig. 6. (a) (b), it can be seen that there is a certain correlation between the influence scale of nodes and the number of communities directly connected by nodes. Nodes with a large number of directly connected communities have a larger influence scale, but the number of communities connected by nodes

with a larger influence scale is not necessarily large. It can be seen from Fig. 6. (c) (d) that there is a strong correlation between node Influence scale and node importance. The image in Fig. 6. (c) has a larger value range of X axis and a larger image density on the left. for convenience of observation, the distribution image in the range of importance 0 to 5000 is captured, as shown in Fig. 7. As can be seen from Fig. 7, nodes with low importance may have a higher Influence scale, nodes with high importance have a higher Influence scale. In Fig. 6. (d), the correlation between node influence scale and node importance is more obvious, and the image is basically scattered between two oblique lines passing through the origin (oblique lines have been marked in the figure). There is a strong correlation between node influence scale and node importance, which shows that the nodes with high importance we find through this method have high Influence scale.

Fig. 6. Relationship between V-community, node importance and node influence scale

Fig. 7. Node importance and node influence scale in Facebook dataset

The data of the top ten nodes in Facebook data set are shown in Table 4, and the data of the top ten nodes in mail data set are shown in Table 5.

It can be observed that for nodes with high I, the Influence scale is very high from Table 4 and Table 5; for nodes with high I, the Degree_{out} is not necessarily high, and some are even very low. I high node's Betweenness is not necessarily high, some even very low; for nodes with high I, the V-community is not necessarily high, and some are even very low.

In conclusion, the influence of nodes in the network is not related to structural features such as node degree and betweenness. The method in this paper evaluates the importance of nodes based on community influence. Nodes with higher importance

Table 4. Top ten nodes in Facebook data set in node importance

Node number	Degree _{out}	Betweenness	V-community	I	Influence scale
107	1045	0.480518	11	12227.468	3874.939
3437	547	0.236115	14	11849.017	3622.140
563	91	0.062780	7	8008.597	3791.743
1593	32	0.000553	6	7922.979	3495.210
0	347	0.146305	11	7714.364	3700.363
1173	115	0.000942	6	7627.214	3602.769
606	91	0.000997	5	7343.892	3440.250
1687	43	0.000907	5	7064.857	3796.674
1684	792	0.337797	8	6935.021	3478.205
428	115	0.064309	7	6397.974	3582.737

Table 5. Top ten nodes of node importance in E-Mail data set

Node number	Degree _{out}	Betweenness	V-community	I	Influence scale
1957	3	0	3	803.4425465	693.26
1159	155	0.050850	5	778.5791855	703.88
1312	6	0	5	756.1487419	647.72
993	44	0.015875	3	718.0338654	704.08
1882	4	0	4	714.651491	645.12
1669	241	0.128871	3	693.0661212	701.62
869	36	0.003789	3	683.0508954	702.42
1	96	0.025603	3	674.3651542	706.34
1618	17	0	5	667.307672	704.86
585	75	0.008709	2	653.1068147	708.86

have greater influence, which can promote information dissemination, and can find some nodes with less prominent structural characteristics but larger actual influence.

5 Conclusion

Identifying nodes that have greater influence on the dissemination of information in social networks is a hot research field in social networks. However, there are few methods in the existing algorithms for evaluating node importance to study the influence of communities on the dissemination of information in social networks. In this paper, a node importance evaluation algorithm based on community influence is proposed. After the network is divided into communities, the influence of communities and the influence of nodes on communities are evaluated. Finally, the importance of nodes is comprehensively evaluated by combining both. The experimental results show that the nodes with high importance evaluated by the algorithm have high influence in the network, which can promote the dissemination of information, and can find some nodes with high influence but not prominent structural characteristics.

For the next step, we plan to introduce content features and user behavior features in social networks to integrate the importance of computing nodes.

References

1. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**(14), 3200–3203 (2001)
2. Freeman, L.C.: A set of measures of centrality based on betweenness. *Sociometry* **40**(1), 35–41 (1977)
3. Sabidussi, G.: The centrality index of a graph. *Psychometrika* **31**(4), 581–603 (1966)
4. Bonacich, P.F.: Factoring and weighting approaches to status scores and clique identification. *J. Math. Sociol.* **2**(1), 113–120 (1972)
5. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington DC, USA, pp. 137–146 (2003)
6. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Costeffective outbreak detection in networks. In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Jose, USA, pp. 420–429 (2007)
7. Kimura, M., Saito, K.: Tractable models for information diffusion in social networks. In: Fürnkranz, J., Scheffer, T., Spiliopoulou, M. (eds.) *PKDD 2006. LNCS (LNAI)*, vol. 4213, pp. 259–271. Springer, Heidelberg (2006). https://doi.org/10.1007/11871637_27
8. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, France, pp. 199–208 (2009)
9. Chen, W., Yuan, Y., Zhang, L.: Scalable influence maximization in social networks under the linear threshold model. In: *Proceedings of IEEE 10th International Conference on Data Mining (ICDM 2010)*, Sydney, Australia, pp. 88–97 (2010)
10. Goyal, A., Lu, W., Lakshmanan, L.V.: SIMPATH: an efficient algorithm for influence maximization under the linear threshold model. In: *2011 IEEE 11th International Conference on Data Mining (ICDM 2011)*, Vancouver, Canada, pp. 211–220 (2011)

11. Kimura, M., Saito, K., Nakano, R., et al.: Extracting influential nodes on a social network for information diffusion. *Data Min. Knowl. Discov.* **20**(1), 70–97 (2010)
12. Wu, X., Liu, Z.: How community structure influences epidemic spread in social networks. *Phys. A: Stat. Mech. Appl.* **387**(2–3), 623–630 (2008)
13. Huang, W., Li, C.: Epidemic spreading in scale-free networks with community structure. *J. Stat. Mech: Theory Exp.* **2007**(01), P01014–P01014 (2007)
14. Chu, X., Guan, J., Zhang, Z., et al.: Epidemic spreading in weighted scale-free networks with community structure. *J. Stat. Mech.: Theory Exp.* **2009**(7), P07043 (18 pp.) (2009)
15. Cao, T., Wu, X., Wang, S., Hu, X.: OASNET: an optimal allocation approach to influence maximization in modular social networks. In: *Proceedings of the ACM Symposium on Applied Computing*, Sierre, Switzerland, pp. 1088–1094 (2010)
16. Wang, Y., Cong, G., Song, G., Xie, K.: Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Washington, pp. 1039–1048 (2010)
17. Shang, J., Zhou, S., Li, X., et al.: CoFIM: a community-based framework for influence maximization on large-scale networks. *Knowl.-Based Syst.* **117**(FEB), 88–100 (2017)
18. Zhang, X., Zhu, J., Wang, Q., et al.: Identifying influential nodes in complex networks with community structure. *Knowl. Based Syst.* **42**, 74–84 (2013)
19. Tulu, M.M., Hou, R., Younas, T.: Identifying influential nodes based on community structure to speed up the dissemination of information in complex network. *IEEE Access* **6**, 7390–7401 (2018)
20. Zhao, Z.Y., Yu, H., Zhu, Z.L., et al.: Identifying influential spreaders based on network community structure. *Chin. J. Comput.* **37**, 753–766 (2014)
21. Goldenberg, J., Libai, B., Muller, E.: Using complex systems analysis to advance marketing theory development: modeling heterogeneity effects on new product growth through stochastic cellular automata. *Acad. Market. Sci. Rev.* **9**(3), 1–18 (2001)
22. Huang, H., Shen, H., Meng, Z.: Community-based influence maximization in attributed networks. *Appl. Intell.* **50**(2), 354–364 (2020)
23. Li, J., Wang, X., Deng, K., et al.: Most influential community search over large social networks. In: *2017 IEEE 33rd International Conference on Data Engineering (ICDE)*, pp. 871–882. IEEE Computer Society (2017)