

An Influence Maximization Algorithm Based on the Mixed Importance of Nodes

Yong Hua¹, Bolun Chen^{1,2,*}, Yan Yuan¹, Guochang Zhu¹ and Jialin Ma¹

Abstract: The influence maximization is the problem of finding k seed nodes that maximize the scope of influence in a social network. Therefore, the comprehensive influence of node needs to be considered, when we choose the most influential node set consisted of k seed nodes. On account of the traditional methods used to measure the influence of nodes, such as degree centrality, betweenness centrality and closeness centrality, consider only a single aspect of the influence of node, so the influence measured by traditional methods mentioned above of node is not accurate. In this paper, we obtain the following result through experimental analysis: the influence of a node is relevant not only to its degree and coreness, but also to the degree and coreness of the n -order neighbor nodes. Hence, we propose a algorithm based on the mixed importance of nodes to measure the comprehensive influence of node, and the algorithm we proposed is simple and efficient. In addition, the performance of the algorithm we proposed is better than that of traditional influence maximization algorithms.

Keywords: Influence Maximization, social network, mixed importance, coreness.

1 Introduction

With the prosperous development of social network in recent years, the research on social network has attracted attention of many scholars. The social network, such as facebook, twitter and blogs, is composed of social relationships among individuals, as a result the potential business value of social network is enormous. One of the classical applications in social network is viral marketing, in generally, using the ‘word of mouth’ effect [Guille (2013); Goldenberg, Libai and Muller (2001)] to market products online. For instance, a software company plans to sell a software through network. The salesman of the software company will look for k influential initial users and pay the k initial users rewards in the social network. Then, the k initial users promote the software to their acquaintances. The social relationships among acquaintances become the key of viral marketing, because acquaintances are more convinced than strangers [Hill, Provost and Volinsky (2006); Sadovykh, Sundaram and Piramuthu (2015); Schmitt, Skiera and

¹ HuaiYin Institute of Technology, Huaian, 223003, China.

² University of Fribourg, Fribourg, 1700, Switzerland.

* Corresponding Author: Bolun Chen. Email: chenbolun1986@163.com.

Den Bulte (2011); Verbraken, Goethals, Verbeke et al. (2014); Iyengar, Den Bulte and Valente (2011)]. Therefore, the scope of promotion of the software will be expanded through the 'word of mouth' effect. As described in the previous example, the problem above is called the influence maximization problem [Domingos and Richardson (2001)], whose essential problem is how to find the most influential initial k users.

The problem of influence maximization was first proposed by Kempe et al. [Kempe, Kleinberg and Tardos (2003)]. They demonstrated that finding the subset of nodes with optimal influence in social network is an NP-hard problem and an approximate guarantee of optimal solution is obtained by using the simple greedy algorithm. In addition, they put forward using independent cascading model or linear threshold model to simulate the propagation of information in discrete time in social network. The key of selecting influential initial k users in the influence maximization is how to measure the ability of spreading influence of node. A variety of central methods, such as degree centrality, betweenness centrality and closeness centrality, were proposed in the original study of the problem of influence maximization. The ability of node to spread the influence in a social network can be measured directly through the methods above. In the subsequent research in the problem of influence maximization, the property of the submodular function [Nemhauser, Wolsey and Fisher (1978)] was used in many greedy algorithms to approximate the optimal solution through extensive iterations. Chen et al. [Chen, Wang and Yang (2009)] proposed a degree discount heuristic algorithm, compared with CELF algorithm proposed in Leskovec et al. [Leskovec, Krause, Guestrin et al. (2007)], which greatly reduces the time complexity and makes the seed set chosen by this algorithm having more influence. Lee et al. [Lee and Chung (2014)] proposed a 2-hop greedy algorithm based on the phenomenon that the influence of node reach most nodes needing only 2-hop, whose time complexity is less than CELF++ and achieves a more influential seed set [Goyal, Lu and Lakshmanan (2011)]. Chen et al. [Chen, Wang and Wang (2010)] proposed PMIA algorithm, which uses MIA (maximum influence in-arborescence) and MIOA (maximum influence out-arborescence) to model the influence of node. The PMIA algorithm has a lower time complexity than the simple greedy algorithm, but the space complexity of PMIA is higher than the simple greedy algorithm, because PMIA needs to build influence propagation tree for each node. Jung et al. [Jung, Heo and Chen (2012)] proposed a novel algorithm IRIE, which combines the influence ranking algorithm and the influence estimation algorithm, and achieved lower time complexity and more efficient than PMIA. Many novel algorithms with excellent results in measuring the influence of node have been proposed in recent studies. Kitsak et al. [Kitsak, Gallos, Havlin et al. (2010)] found that the most effective influence spreaders are often not the nodes with large degrees, but the nodes are in the core location of the social network. The correlation between the coreness and the propagation ability of node in the social network was analyzed by k-shell decomposition [Carmi, Havlin, Kirkpatrick et al. (2007)], which provided a new solution to the problem of influence maximization. Gao et al. [Gao, Ma, Chen et al. (2014)] presented that the influence of node is related to their local structure. A algorithm based on local structure centrality, which uses the topology and centrality of node and its neighbors to

measure the ability of spreading influence of node, was proposed. This algorithm is more accurate than the previous algorithm when evaluating the influence of node. Xia et al. [Xia, Song, Jing et al. (2018)] construct double-layer network and use markov chains theory to study disease spreading, who proposed that the scale of disease spreading can be reduced by increasing the rate of spreading.

In real social network, the larger degrees a node has, the more likely it is to be chosen as a initial node. However, the ability of a node to spread influence is related not only to the degree of node and its neighbors, but also the coreness of node and its neighbors. In other words, the ability of a node to spread influence is also related to the degree and coreness of its neighbors. In this paper, we study the correlation between the mixed importance and the ability of node to spread influence, and we propose an influence maximization algorithm based on the mixed importance of nodes. In contrast to the traditional algorithms, we consider the cost of selecting initial nodes in social network. Moreover, our algorithm is more adaptive [Zhang, Zheng and Xia (2018)] and achieves good performance.

2 Description of the problem

2.1 Influence maximization problem

We define a undirected social network $G = (V, E)$, where V represents the node set in social network G , and E represents the collection of edges between nodes. Moreover, $n = |V|$, where n is the total number of nodes in social network G . If node u and v exist in the social network G , we use $(u, v) \in E$ to represent an edge between the u and v , and use p to represent a propagation probability. We use m to represent the total number of edges in social network G . The influence maximization problem is to find the subset of node set in the graph G , which is a set composed of seed nodes. We use S to represent seed set, $k = |S|$. In addition, the influence of the k seed nodes is maximum.

2.2 Independent cascade model

We use the independent cascade model to simulate the influence of node and measure the ability of node to spread influence [Chen, Fan, Li et al. (2015); Liu, Cong, Xu et al. (2012)]. The principle of the independent cascade model is as follows: in the network G , all nodes have only two states: active or inactive. A_i is the set of nodes activated at time i . The initial phase is the time $t = 0$. $A_0 = S$ represents that the nodes in the seed set S are active and other nodes are inactive at time $t = 0$. At time $t = i$, $u \in A_{i-1}$ and the node v is in the inactive state. u attempts to activate node v with propagation probability p for edge $(u, v) \in E$. If v is successfully activated, v maintains activated state from $t = i + 1$. If v fails to be activated by u , v can not be activated by u at a later time. If node v has multiple neighbors that have already been activated, the neighbors of v will activate v with a probability of $1 - (1 - p)^l$, where l is the number of neighbors that have been activated. When A_i is empty, that is, at time $t = i$, no nodes are activated, the propagation process ends, and the number of activated nodes in the whole process represents the influence of the seed set S [Kim, Kim, Oh et al. (2017); Kimura, Saito, Nakano et al. (2010)].

2.3 Influence function

We denote the influence function as $influence(\cdot)$, where the function $influence(\cdot)$ maps a subset of the node set V to a nonnegative integer. For example, $influence(u)$ is the number of node activated by node u , and $influence(S)$ is the number of node activated by seed set S . We calculate the value of function $influence(\cdot)$ through the independent cascade model.

2.4 Cost

In the initial phase of the influence maximization problem, the initial users need to be selected as seeds, but we need to give the seeds certain rewards. In a real social network, a node with a greater degree has a wider social range. Therefore, a node with large degrees will be given more rewards than a node with small degrees. In this paper, we define the initial cost by Eq. (1), where $degree_u$ is the degree of node u and the function $mean(\cdot)$ is the average function.

$$cost = mean(\{degree_u | u \in S\}) \quad (1)$$

3 Detailed process of the algorithm

In this paper, we use the mixed importance to measure the influence of node. In the mixed importance algorithm, we consider that the influence of node affected by its n -order neighbors. The obvious representation of a node affected by its neighbors is the degree of node, but only using the degree of node to measure the influence of node results in incomprehensive measurement. If node u and v have the same degrees, the influence of the node u is better than node v , when the number of 2-order neighbors of node u is more than node v . Furthermore, if node u and node v have the same degrees, but node u and its neighbors are in the core location of social network, the ability of the node u to spread influence may be better. Many scholars have put forward that the influence of node is limited. Therefore, we limit the propagation scope of node to the 2-order neighbors, ignoring the rest of neighbors, that is, we measure the influence of node by using the mixed importance of the degree and the coreness of the node's 1-order and 2-order neighbors. The mixed importance algorithm show in Tab. 1.

The input of the mixed importance algorithm is social network $G = (V, E)$ and the number of seed nodes k . The output of this algorithm is the seed set S . By traversing the values of α and β , the mixed importance algorithm calculates the mixed importance of node in different α and β . The optimal values of α and β , that is α_{best} and β_{best} , are obtained and put into Eq. (2) by calculating the kendall correlation coefficients. Then, the mixed importance values of all nodes are calculated, and k nodes with the maximum mixed importance values are selected as the seed nodes. Among the mixed importance algorithm, Step 1 is the input of algorithm: the number of social network $G = (V, E)$ and the number of seed nodes k . In Steps 2 and 3, let α and β cycle from 0.1 to 0.9 with an increase of 0.1 per iteration. The key steps are described detailedly in the following subsections.

Table 1: Mixed importance algorithm

Input: social network $G = (V, E)$, the number of seed nodes k
Output: seed set S

- (1) Initializing seed set $S = \emptyset$
- (2) for $\alpha = 0.1 : 0.1 : 0.9$
- (3) for $\beta = 0.1 : 0.1 : 0.9$
- (4) for each vertex $v \in V$
- (5) put α and β into equation (2) to calculate $MI(v)$
- (6) calculate $influence(v)$
- (7) end for
- (8) calculate Kendall coefficients $\tau_{\alpha\beta}$
- (9) end for
- (10) end for
- (11) $\alpha_{best}, \beta_{best} = \max(\tau_{\alpha\beta})$
- (12) for each vertex $v \in V$
- (13) put α_{best} and β_{best} into equation (2) to calculate $MI(v)$
- (14) end for
- (15) selects k nodes with largest $MI(v)$ as seed set S

3.1 Mixed importance

Step 5 calculates the mixed importance value $MI(v)$ of the node v . In the social network G , we set the 1-order neighbor node set of the node u to Γ_1 , where $d_u = |\Gamma_1|$ and d_u are the number of the 1-order neighbors of the node u , which is also the degree of node u . We set the set of 2-order neighbor node set of node u to Γ_2 . Thus, the total number of 1-order and 2-order neighbors of node u is $Ne(u) = |\Gamma_1 \cup \Gamma_2|$. We set the coreness of node u to $Core(u)$. The coreness of node u is small, when node u is at the edge of network. The coreness of node u is high, when node u is at the core location of network. We define the mixed importance of node as shown in Eq. (2).

$$MI(u) = \beta(\alpha d_u + (1-\alpha)Core(u)) + (1-\beta)\left(\alpha \sum_{w \in |\Gamma_1 \cup \Gamma_2|} d_w - Ne(u) + (1-\alpha) \sum_{w \in |\Gamma_1 \cup \Gamma_2|} Core(w)\right) \quad (2)$$

where $MI(u)$ is the mixed importance of node u , $\sum_{w \in |\Gamma_1 \cup \Gamma_2|} d_w - Ne(u)$ is the total degrees of the 1-order neighbors and the 2-order neighbors of node u after node u is removed, and $\sum_{w \in |\Gamma_1 \cup \Gamma_2|} Core(w)$ is the sum of the coreness of the 1-order neighbors and the 2-order neighbors of node u . α is a parameter used to balance the degree and the coreness. β is a parameter used to balance the relationship between a node and its neighbors. Both α and β are both between 0.1 and 0.9.

In Eq. (2), if the value of α is large, it indicates that the degrees of node and its neighbors

in the 2-order scope play a greater role in the influence of node. If the α value is small, it indicates that the coreness of node and its neighbors in the 2-order scope play a greater role in the influence of node. If the value of β is large, it indicates that the degree and coreness of node play a greater role in the influence of node. If the value of β is small, the degrees and coreness of neighbors in the 2-order scope of node play a greater role.

Step 6 is to simulate the influence of node v by using the independent cascade model, and obtain the influence of node v , $influence(v)$.

3.2 Kendall correlation coefficient

Step 8 is to calculate the kendall correlation coefficient under the current value of α and β . The kendall correlation coefficient $\tau_{\alpha\beta}$ [Kendall (1938)] is described as follows: consider the joint observation pairing of two sets of random variables X and Y . If $x_i > x_j$ and $y_i > y_j$ or $x_i < x_j$ and $y_i < y_j$ for the observation pairs (x_i, y_i) and (x_j, y_j) , we say that the pairs are consistent. If $x_i > x_j$ and $y_i < y_j$ or $x_i < x_j$ and $y_i > y_j$, we say that the pairs are inconsistent. If $x_i = x_j$ and $y_i = y_j$, we say that the pairs are neither consistent nor inconsistent. The kendall correlation coefficient is defined by Eq. (3).

$$\tau = \frac{\eta_c - \eta_d}{0.5\eta(\eta - 1)} \quad (3)$$

In Eq. (3), η_c and η_d represents consistent pairs and inconsistent pairs respectively. Because X and Y have the same number of elements, η represents the number of elements in X or Y . In this paper, we calculate the mixed importance $MI(u)$ of every node u in network G and use the independent cascade model to perform multiple influence simulations on node u to obtain the average value, that is, the influence value $Influences(u)$ of node u . Therefore, set X is $\{MI(u)|u \in V\}$ and set Y is $\{Influences(u)|u \in V\}$. Through the calculation of kendall correlation coefficient τ , we can analyze the correlation between the mixed importance of node and the influence of node to more accurately measure the ability of node to spread influence. The value of τ is in the range of $[-1, 1]$. The value of τ is high, when the ability of node to spread influence is great and the mixed importance value of node is high.

The mixed importance value and influence of all nodes can be respectively obtained from Step 5 and Step 6, and $\tau_{\alpha\beta}$ is calculated by Eq. (3). Each pair of α and β values corresponds to an $\tau_{\alpha\beta}$ value, and Step 11 is to select the maximum $\tau_{\alpha\beta}$ value and assigns the corresponding α and β values to α_{best} and β_{best} . From Steps 12 to 14, α_{best} and β_{best} are put into Eq. (2) to calculate the mixed importance value of all nodes. Finally, Step 15 selects k nodes with the maximum mixed importance value to be the seed set S .

4 Experimental results and analysis

We perform the experiments on four real social network datasets and 4 random network datasets. The four real social datasets are email [Yin, Benson, Leskovec et al. (2017)], socfbBowdoin47 [Traud, Mucha and Porter (2012)], hamsterster [Dünker and Kunegis

Table 2: Topological properties of 8 different datasets

Network	n	m	d_{max}	\bar{d}	r	C	D
email	986	16064	345	32.58	-0.0257	0.4071	0.0331
socfbBowdoin47	2252	84387	670	74.94	0.0562	0.2887	0.0333
hamsterster	2426	16630	273	13.70	0.0474	0.5375	0.0057
socfbSmith60	2970	97133	349	65.40	0.0436	0.2835	0.022
random1	752	6930	142	18.43	0.0824	0.1109	0.0245
random2	1446	59589	375	82.41	0.0675	0.3230	0.0570
random3	2000	50080	72	50.08	-0.0003	0.0248	0.0251
random4	2916	77609	601	53.22	0.0928	0.0924	0.0183

(2015)], and socfbSmith60 [Traud, Kelsic, Mucha et al. (2011)]. The data of Email derived from the large European research institutions and consisted of email sent and received by users. If one email at least is sent and received between user u and user v , there is an edge between u and v . SocfbBowdoin47 and socfbSmith60 are extracted from the data of facebook. If there is a friend relationship between user u and user v , there is an edge between u and v . Hamsterster is a collection of friends and loved ones from the hamsterster.com website. The four random datasets are undirected networks generated by Pajek and are denoted as random1, random2, random3, and random4. The topological attributes of all the datasets are shown in Tab. 2, where n is the total number of nodes in the dataset, m is the number of edges, d_{max} is maximum degree, \bar{d} is average degree, r is same coefficient, C is clustering coefficient, and D is network density.

4.1 Selection of the optimal parameters

Firstly, we analyze the value of α and β , because the value of α and β in the Eq. (3) is variable. Therefore we need to calculate the optimal value of α and β in different datasets. Moreover, the value of Eq. (2), that is the value of mixed importance, is positive correlation with the influence of node, so we use kendall correlation coefficient τ to measure the correlation between the mixed importance of node and the influence of node. The optimal value of α and β is obtained when the value of τ is maximum. In the network G , we calculate the mixed importance $MI(u)$ of each node u according to Eq. (2), where the range of α and β ranges from 0.1 to 0.9. We traverse the double circulation constructed by α and β , and the value of α or β increase 0.1 in each circulation. Secondly, we simulate the influence $Influences(u)$ of each node u by using the independent cascade model in each circulation. In this paper, we set the probability of propagation $p = 0.01$ in the independent cascade model. Finally, we calculate τ after the double circulation constructed by α and β , and select the optimal value of α and β when the value of τ is maximum, in detail, the value of τ is calculated by Eq. (3).

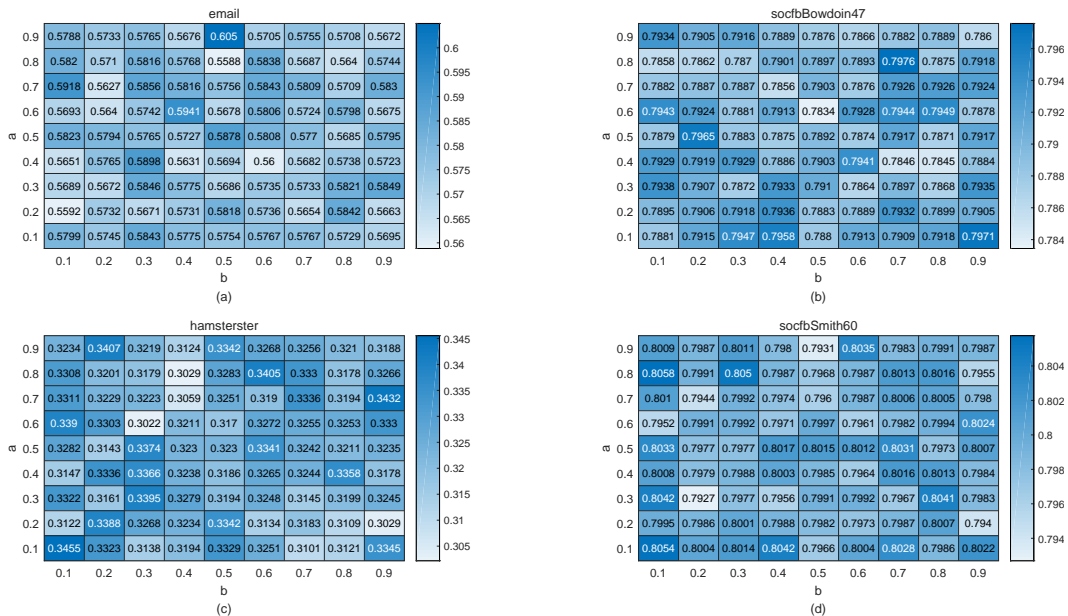


Figure 1: Heat map on four real social network: we analyze the value of α and β in different real social network because α and β are unknown in Eq. (3). We study the correlation between the influence of node and the mixed importance of node by using Kendall correlation coefficient. The maximum τ in the heat map illustrates that the value of Eq. (3) with current α and β accord optimally with the real influence of node

4.1.1 Real network

In this section, we perform the mixed importance algorithm in the 4 real social networks to find the optimal value of α and β in every dataset respectively. In Fig. 1, the vertical axis represents the value of α , the horizontal axis represents the value of β , and (α, β) corresponds to the value of τ calculated by Eq. (3). Fig. 1(a) shows the result of the email dataset. We find the value of τ is the best when $\alpha = 0.9$ and $\beta = 0.5$. In other words, when $\alpha = 0.9$ and $\beta = 0.5$, the value $MI(u)$ of each node u is consistent with the value $Influences(u)$, that is, the mixed importance value of node u is positively correlated with the influence of node u . Similarly, Fig. 1(b) shows that the value of τ of the socfbBowdoin47 dataset is the best, when $\alpha = 0.8$ and $\beta = 0.7$; Fig. 1(c) shows that the value of τ of the hamsterster dataset is the best, when $\alpha = 0.1$ and $\beta = 0.1$; Fig. 1(d) shows that the value of τ of the socfbSmith60 dataset is the best, when $\alpha = 0.8$ and $\beta = 0.1$.

4.1.2 Random network

In this section, we perform the mixed importance algorithm in the 4 random networks to find the optimal value of α and β in every dataset respectively. In Fig. 2, the vertical

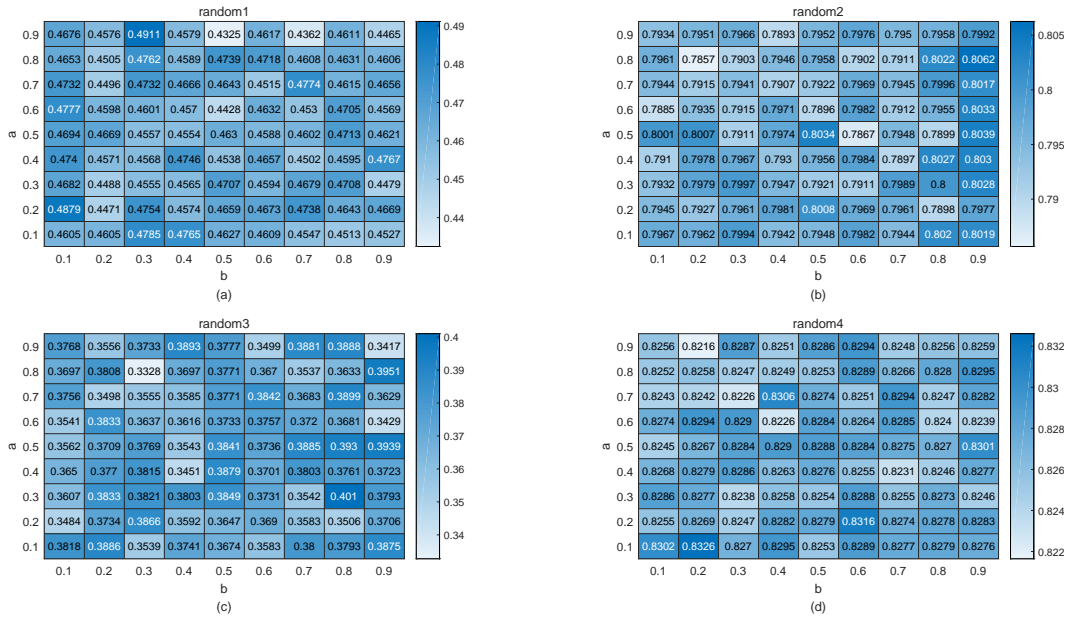


Figure 2: Heat map on four random social network: we analyze the value of α and β in different random social network because α and β are unknown in Eq. (3). We study the correlation between the influence of node and the mixed importance of node by using kendall correlation coefficient. The maximum τ in the heat map illustrate that the value of Eq. (3) with current α and β accord optimally with the real influence of node

axis represents the value of α , the horizontal axis represents the value of β , and (α, β) corresponds to the value of τ . In Fig. 2(a), the τ value of the random1 dataset is optimal, when $\alpha = 0.9$ and $\beta = 0.3$; In Fig. 2(b), the τ value of the random2 dataset is optimal, when $\alpha = 0.8$ and $\beta = 0.9$; In Fig. 2(c), the τ value of the random3 dataset is optimal, when $\alpha = 0.3$ and $\beta = 0.8$; In Fig. 2(d), the τ value of the random4 dataset is optimal, when $\alpha = 0.1$ and $\beta = 0.2$.

4.2 Comparison and analysis of the scope of influence

In the above experiment content, we achieve optimal α and β in the every dataset respectively. Then, we compare the influence of seed set selected by the mixed importance algorithm with the influence of seed set selected by other algorithms, such as Degree Centrality, PageRank [Page, Brin, Motwani et al. (1999)]. We analyze the trend of the influence of the seed set selected by each algorithm, when we change the size of seed set. The experimental results show in Fig. 3 and Fig. 4.

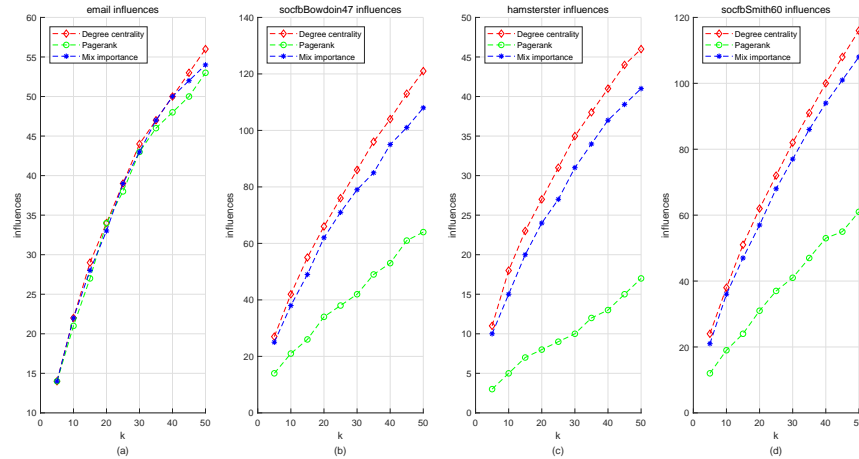


Figure 3: The scatter figure of influence of the seed set selected by mixed importance algorithm on 4 real dataset: every scatter figure has 10 point for every algorithm and we connect 10 points with a dotted line for observing the trend of influence

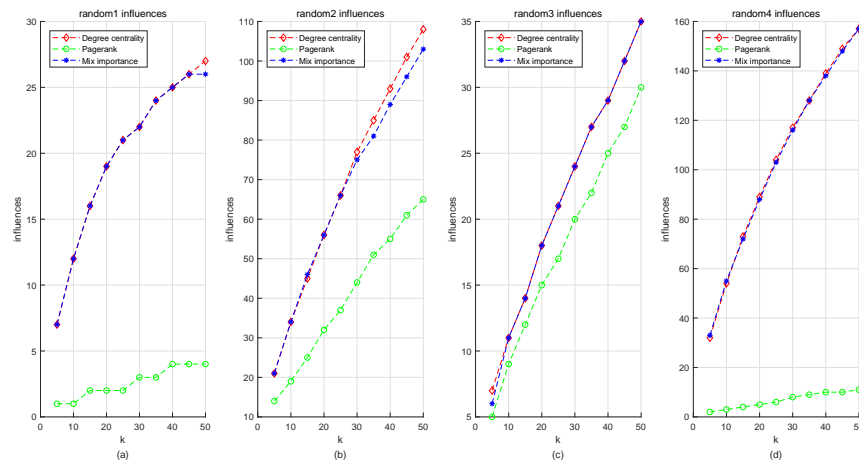


Figure 4: The scatter figure of influence of the seed set selected by mixed importance algorithm on 4 random dataset: every scatter figure has 10 point for every algorithm and we connect 10 points with a dotted line for observing the trend of influence

4.2.1 Real network

In Fig. 3 and Fig. 4, the horizontal axis and the vertical axis respectively represents the size of seed set and the influence of the seed set. Fig. 3(a) shows the influence range of the email. The influence of the mixed importance algorithm is almost the same as that of the degree centrality. The influence of the mixed importance algorithm is generally higher than that of PageRank. The seed set selected by the mixed importance algorithm has a good influence on the email. Fig. 3(b) shows that the influence of the seed set selected by the mixed importance algorithm on the socfbBowdoin47 is less than that of degree centrality, but the differences between the two algorithms are small. Fig. 3(c) is the influence range of the hamsterster, and Fig. 3(d) is the influence range of the socfbSmith60. Similarly, the influence of mixed importance is at least slightly less than the influence of degree centrality, but the scope of mixed importance is much larger than PageRank.

4.2.2 Random network

Fig. 4(a) shows the influence range of the random1 dataset, and the influence of the mixed importance algorithm is almost the same as that of degree centrality. Fig. 4(b) shows the influence range of the random2 dataset. The influence of the mixed importance algorithm exceeds that of degree centrality. In general, the influence range of the mixed importance algorithm is slightly smaller than that of degree centrality. Fig. 4(c) shows the influence range of the random3 dataset. The influence ranges of the mixed importance algorithm and degree centrality are generally the same. Fig. 4(d) shows the influence range of the random4 dataset, where the influence of the mixed importance algorithm and the influence of degree centrality are generally the same.

4.3 Comparison and analysis of impact cost

Finally, we compare the mixed importance algorithm with other algorithms, such as Degree Centrality, PageRank. We analyze the cost of the seed set selected by each algorithm, when we change the size of seed set. The experimental results show in Fig. 5 and Fig. 6.

4.3.1 Real network

In Fig. 5 and Fig. 6, the horizontal axis and the vertical axis respectively represents the size of seed set and the cost of the seed set. Fig. 5(a) shows the cost of the email, where the cost of mixed importance algorithm, degree centrality and PageRank are similar. Fig. 5(b) shows the cost of the socfbBowdoin47, Fig. 5(c) shows the cost of the hamsterster, and Fig. 5(d) shows the cost of the socfbSmith60. The cost of mixed importance algorithm and degree centrality are much higher than that of PageRank, and the cost of mixed importance algorithm is small compared to the Degree Centrality.

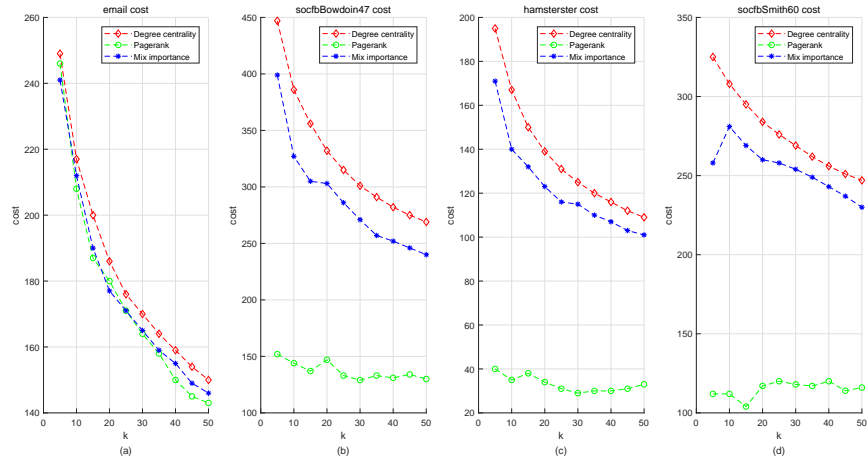


Figure 5: The scatter figure of cost of the seed set selected by mixed importance algorithm on 4 real dataset: every scatter figure has 10 point for every algorithm and we connect 10 points with a dotted line for observing the trend of cost

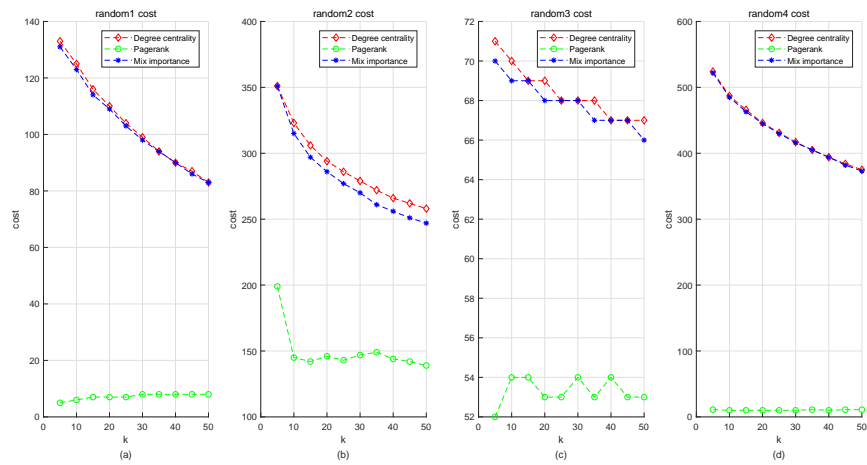


Figure 6: The scatter figure of cost of the seed set selected by mixed importance algorithm on 4 random dataset: every scatter figure has 10 point for every algorithm and we connect 10 points with a dotted line for observing the trend of cost

4.3.2 Random network

Fig. 6(a) shows the cost of the random1, where the cost of mixed importance algorithm is slightly less than the cost of the Degree Centrality. Fig. 6(b) shows the cost of the random2, and Fig. 6(c) shows the cost of the random3. The cost of mixed importance algorithm is small compared to the cost of the Degree Centrality. Fig. 6(d) shows the cost of the random4, where the cost of mixed importance algorithm and Degree Centrality are flat. In Fig. 6, the cost of mixed importance algorithm is greater than the cost of PageRank. Based on the experimental comparison, the cost of PageRank is always less than the cost of mixed importance algorithm and Degree Centrality, but the range of influence of PageRank is also much smaller than that of mixed importance algorithm and Degree Centrality. The influence range of the mixed importance algorithm is equal to that of Degree Centrality, and the cost of mixed importance algorithm is much smaller than the cost of Degree Centrality. When the influence range of mixed importance algorithm is better than that of Degree Centrality, the cost of mixed importance algorithm is also less than the cost of Degree Centrality. The experiments show that the mixed importance algorithm has achieved good results.

5 Conclusion

This paper mainly analyzes how to effectively measure the influence of node in the process of influence maximization. The experiments show that the influence of a node is related not only to its degree, but also related to its coreness, as well as the degree and coreness of its neighbors. Therefore, we proposed an influence maximization algorithm based on the mixed importance of node. In this algorithm, the independent cascade model is used to simulate the influence of nodes. The Kendall correlation coefficient is used to measure the correlation of the mixed importance of node and the influence of node and the optimal values of two important parameters are achieved through calculation. The experimental comparisons of algorithms on four real datasets and four random datasets show that the influence maximization algorithm based on the mixed importance of node we propose in this paper achieved excellent performance.

Acknowledgement: This research was supported in part by the Chinese National Natural Science Foundation under grant Nos. 61602202 and 61702441, the Natural Science Foundation of Jiangsu Province under contracts BK20160428 and BK20161302, the Six talent peaks project in Jiangsu Province under contract XYDXX-034 and the project in Jiangsu Association for science and technology.

References

Carmi, S.; Havlin, S.; Kirkpatrick, S.; Shavitt, Y.; Shir, E. (2007): A model of internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences*, vol. 104, no. 27, pp. 11150-11154.

- Chen, S.; Fan, J.; Li, G.; Feng, J.; Tan, K. L. et al.** (2015): Online topic-aware influence maximization. *Proceedings of the VLDB Endowment*, vol. 8, no. 6, pp. 666-677.
- Chen, W.; Wang, C.; Wang, Y.** (2010): Scalable influence maximization for prevalent viral marketing in large-scale social networks. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1029-1038.
- Chen, W.; Wang, Y.; Yang, S.** (2009): Efficient influence maximization in social networks. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 199-208.
- Domingos, P.; Richardson, M.** (2001): Mining the network value of customers. *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 57-66.
- Dünker, D.; Kunegis, J.** (2015): Social networking by proxy: analysis of dogster, catster and hamsterster. *Proceedings of the 24th International Conference on World Wide Web*, pp. 361-362.
- Gao, S.; Ma, J.; Chen, Z.; Wang, G.; Xing, C.** (2014): Ranking the spreading ability of nodes in complex networks based on local structure. *Physica A: Statistical Mechanics and Its Applications*, vol. 403, pp. 130-147.
- Goldenberg, J.; Libai, B.; Muller, E.** (2001): Talk of the network: a complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, vol. 12, no. 3, pp. 211-223.
- Goyal, A.; Lu, W.; Lakshmanan, L. V.** (2011): Celf++: optimizing the greedy algorithm for influence maximization in social networks. *Proceedings of the 20th International Conference Companion on World Wide Web*, pp. 47-48.
- Guille, A.** (2013): Information diffusion in online social networks. *ACM SIGMOD Record*, vol. 42, no. 2, pp. 17-28.
- Hill, S.; Provost, F.; Volinsky, C.** (2006): Network-based marketing: identifying likely adopters via consumer networks. *Statistical Science*, vol. 21, no. 2, pp. 256-276.
- Iyengar, R.; Den Bulte, C. V.; Valente, T. W.** (2011): Opinion leadership and social contagion in new product diffusion. *Marketing Science*, vol. 30, no. 2, pp. 195-212.
- Jung, K.; Heo, W.; Chen, W.** (2012): Irie: scalable and robust influence maximization in social networks. *IEEE 12th International Conference on Data Mining*, pp. 918-923.
- Kempe, D.; Kleinberg, J.; Tardos, É.** (2003): Maximizing the spread of influence through a social network. *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137-146.
- Kendall, M. G.** (1938): A new measure of rank correlation. *Biometrika*, vol. 30, no. 1/2, pp. 81-93.
- Kim, S.; Kim, D.; Oh, J.; Hwang, J. H.; Han, W. S. et al.** (2017): Scalable and parallelizable influence maximization with random walk ranking and rank merge pruning. *Information Sciences*, vol. 415, pp. 171-189.

- Kimura, M.; Saito, K.; Nakano, R.; Motoda, H.** (2010): Extracting influential nodes on a social network for information diffusion. *Data Mining and Knowledge Discovery*, vol. 20, no. 1, pp. 70.
- Kitsak, M.; Gallos, L. K.; Havlin, S.; Liljeros, F.; Muchnik, L. et al.** (2010): Identification of influential spreaders in complex networks. *Nature Physics*, vol. 6, no. 11, pp. 888.
- Lee, J. R.; Chung, C. W.** (2014): A fast approximation for influence maximization in large social networks. *Proceedings of the 23rd International Conference on World Wide Web*, pp. 1157-1162.
- Leskovec, J.; Krause, A.; Guestrin, C.; Faloutsos, C.; VanBriesen, J. et al.** (2007): Cost-effective outbreak detection in networks. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 420-429.
- Liu, B.; Cong, G.; Xu, D.; Zeng, Y.** (2012): Time constrained influence maximization in social networks. *IEEE 12th International Conference on Data Mining*, pp. 439-448.
- Nemhauser, G. L.; Wolsey, L. A.; Fisher, M. L.** (1978): An analysis of approximations for maximizing submodular set functions-i. *Mathematical Programming*, vol. 14, no. 1, pp. 265-294.
- Page, L.; Brin, S.; Motwani, R.; Winograd, T.** (1999): *The PageRank Citation Ranking: Bringing Order to the Web. (Technical Report)*. Stanford InfoLab.
- Sadovkyh, V.; Sundaram, D.; Piramuthu, S.** (2015): Do online social networks support decision-making. *Decision Support Systems*, vol. 70, pp. 15-30.
- Schmitt, P.; Skiera, B.; Den Bulte, C. V.** (2011): Referral programs and customer value. *Social Science Electronic Publishing*, vol. 75, no. 1, pp. 46-59.
- Traud, A. L.; Kelsic, E. D.; Mucha, P. J.; Porter, M. A.** (2011): Comparing community structure to characteristics in online collegiate social networks. *SIAM Review*, vol. 53, no. 3, pp. 526-543.
- Traud, A. L.; Mucha, P. J.; Porter, M. A.** (2012): Social structure of facebook networks. *Physica A: Statistical Mechanics and Its Applications*, vol. 391, no. 16, pp. 4165-4180.
- Verbraken, T.; Goethals, F.; Verbeke, W.; Baesens, B.** (2014): Predicting online channel acceptance with social network data. *Decision Support Systems*, vol. 63, no. 3, pp. 104-114.
- Xia, L.; Song, B.; Jing, Z.; Song, Y.; Zhang, L.** (2018): Dynamical interaction between information and disease spreading in populations of moving agents. *Computers, Materials & Continua*, vol. 57, no. 1, pp. 123-144.
- Yin, H.; Benson, A. R.; Leskovec, J.; Gleich, D. F.** (2017): Local higher-order graph clustering. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 555-564.
- Zhang, G.; Zheng, Y.; Xia, G.** (2018): Domain adaptive collaborative representation based classification. *Multimedia Tools and Applications*, pp. 1-22.

