Research on the Influence of Improved K-shell Algorithm on Commodity Profit

Jing Chen¹,²,³ and Jincheng Huang¹

¹ College of Information Science and Engineering, Yanshan University, Qinhuangdao 066004, China
² Key Laboratory for Computer Virtual Technology and System Integration of Hebei Province, Qinhuangdao 066004, China
³ Software Engineering Laboratory in Hebei Province, Qinhuangdao 066004, China
Email: 2935870380@qq.com

Abstract. In order to maximize the influence of commodity profits in e-commerce platforms, designing and improving the K-shell algorithm to select the more influential seed node sets in this paper. The new algorithm improves the number of active nodes by setting node threshold and edge weight attributes. To obtain more commodity profits, a strategy IRDSN (Strategy for Improving Repeat Degree of Seed Nodes) is proposed to select initial seed nodes and improve the repeat degree of seed nodes. The profit maximization based on linear threshold model is realized by setting different propagation modes. The improved algorithm and strategy IRDSN are analysed and verified in real data set and e-commerce platform. The results show that the algorithm effectively improves the profit of commodities.

Keywords. Influence; threshold; K-shell; influence maximization; profit.

1. Introduction

As social networks connect many user groups, they are widely used in marketing, news release, epidemic surveillance, etc. For example, in the application of word-of-mouth marketing and advertisement publishing based on social network. It is necessary to select suitable users for advertisement promotion and make use of the communication influence of social network to maximize the effectiveness and scope of advertisement promotion [1-3]. From a business point of view, if businesses can spread marketing information through social networks and affect potential customer groups, it will bring certain economic benefits. Therefore, using social networking platforms, people tend to obtain a huge impact at a very small cost. However, merchants will encounter the following problems when marketing goods through social networks. Push advertisements cannot be spread to all users, only to as many users as possible according to the relationship between users in social networks. The influence maximizing algorithm is the key factor to solve these application problems.

The problem of influence maximizing was first defined as how to find t initial nodes so that the ultimate spread of information is the widest [4]. Through the study of maximizing the impact of social networks, the essence of social network is to select several seed nodes to spread in the social network graph. The number of affected nodes by the propagation process is the largest and the impact range is the widest [5]. In social network graph, the more the node tends to the center of the graph, the more important role the node has in the whole network. Therefore, the key to accurately identify the influential nodes is to judge the location information of nodes in the graph and classify the node set. Because of the low time complexity of K-shell algorithm, it is widely used to identify influential nodes.
with high accuracy in the process of propagation.

The K-shell method is coarse-grained in partitioning the importance of nodes, it cannot accurately measure the importance of nodes with the same ks value in shell layer, so many researchers put forward improved schemes. On the basis of the above research, improve the resolving power of ks value of nodes and coverage thresholds and consider the commodity profit in e-commerce platform. The edge weight attribute is introduced to identify influential nodes in K-shell algorithm, and the strategy of selecting seed nodes and IRDSN is put forward. Different propagation modes are set by using the linear threshold model to obtain the activation state information of nodes based on the improved K-shell algorithm. The propagation effect and commodity profit on different data sets are verified.

The whole paper is organized as follows. In section 2, the problem of influence maximizing and the related work are introduced; In section 3, the improved K-shell algorithm is described, and the strategy IRDSN to resolve the repeat degree of seed nodes is proposed; In section 4, the transmission mode of node influence in different situations is considered and calculated commodity profit based on linear threshold; In section 5, the improved K-shell algorithm and the results of commercial profit calculation for different experimental data sets are verified. Finally, the full text is summarized and discussed the future work.

2. Related Work

Research on influence maximization problem, the related seed selection algorithm for different information dissemination models was proposed by scholars at home and abroad. The CELF algorithm used the monotonicity and submodularity to optimize Greedy algorithm [6]. The New Greedy algorithm generated influence subgraphs by simulating influence transmission process and added the optimization process of CELF algorithm [7-8]. The IRIE algorithm needs a few iterations to rank the influence of all nodes and select the most influential nodes based on trust propagation [9]. The hybrid algorithm found some nodes with great potential influence in social network, and used the Greedy algorithm to find the rest of the nodes based on heuristic algorithm [10]. The TOPSIS algorithm based on idea similarity was proposed for the problem of maximizing the process of news information dissemination, and verify performance advantages in different data sets [11]. A heuristic influence node mining strategy L_GAUP based on information preference was proposed. The strategy not only has good performance in the index of influence scope ISST and IS, but also efficiency has been greatly improved [12]. The MVCFVS algorithm based on minimum point coverage and feedback point set, and it was better than other algorithm in independent cascade model and weighted cascade model [13].

In addition to the study of influence maximization in traditional social network, many scholars have extended and deformed the problem of influence maximization according to the practical application. The influence maximization algorithm based on cost control used the dynamic programming method to select the seed nodes [14]. Be extended the linear threshold model and studied the problem of influence maximization in the competitive environment based on subject preference, a new algorithm can make the information spread widely [15]. Considering that the accuracy of probability of information dissemination was improved by the location classification method, the solution of influence maximization problem to combine user interest and location promotion was proposed [16]. For practical application scene, the group dissemination model GIC focuses on the problem of combined influence of focus on area or population [17]. The influence maximization strategy IMOOP of Two-Level propagation scheme for online and offline application in social network, and the transmission scheme has a high information coverage in real data set [18]. The seed selection algorithm of profit maximizing defines profit measurement combines the effective of influence diffusion in viral marketing and the cost of seed selection, and was verified in real data set [19].

For now, the research of influence maximization focused on heuristic algorithm with low time complexity and similar influence effect to greedy algorithm, few studies of influence maximization results applied to profit maximizing of small e-commerce platform. The improved K-shell algorithm
and strategy IRDSN were used to calculate profit in e-commerce platform. Then, the multi-propagation mode algorithms based on linear threshold model had used to according the range of historical pricing for a certain kind of commodity to determine its price, and to maximize the number of activate nodes and profit.

3. The Improved K-shell Algorithm

3.1. Description of Algorithm
The K-shell algorithm was proposed by Kitsak in 2010 [20]. It described the hierarchy of network structure and estimates the influence of nodes in the network. The algorithm was as follows:

1. Set the ks value of all nodes with a zero degree to zero in the network:
2. Calculated the degree value of all nodes in the network (assumed that the minimum degree value of all nodes was k), deleted the nodes and connections whose degree value was k. After deletion, the nodes whose degree value was d≤k may reappear again. Continued to delete these nodes and connections until the degree value of all nodes in the network was greater than k, and the deleted nodes and connections jointly constitute the shell layer. The value of k is equal to that of the shell layer, and the ks value of all nodes was equal to k.
3. Repeat step 2, until all nodes have corresponding ks values.

In the K-shell algorithm, the node with a large ks value is at the core of the network. The connectivity between these nodes is strong, so their influence is greater. It shows the hierarchy of network structures in figure 1. The degree value of blue nodes is same as the yellow nodes, but the ks value of the former is greater than the ks value of the latter. Therefore, the influence of the blue nodes is greater than that of the yellow nodes.

Figure 1. The hierarchical structure of social network.

Therefore, the classification of shell layer in social network graph is obtained after processing with K-shell algorithm. The improved K-shell algorithm need to create a set of node information container, and the container information used is shown in table 1.

| Container | Type | Description                      |
|-----------|------|----------------------------------|
| A= [ ]   | List | Store the out-degree value of node |
| B= { }   | Dictionary | Store node number and corresponding ks value |
| C= [ ]   | List | Store the ks value                |
| K= [ ]   | List | Store set of nodes for each shell layer |
| H= { }   | Dictionary | Store the influence of each node |

Because the K-shell algorithm has the same value of ks, which results in the inaccurate measurement of node influence, an improve K-shell algorithm by introducing edge weight attributes was proposed in this paper. The improved K-shell algorithm calculates the influence value of each node (the sum of the ks value of the node and the adjacent weight) by assigning the edge weight function between nodes in the social graph. In the shell layer with the largest value of ks, the k nodes with the greatest influence are selected as the initial seed node set, and the propagation process is carried out based on initial seed node set. The improved K-shell algorithm is as follows:
Algorithm 1. The improved K-shell.

**Input:** Social Network Graph $G$; Shell Layer Node List

**Output:** Initial seed node set $S$

1. Traverse all nodes $n$ in social network graph $G$;
2. Take all the successor nodes $n_1$ of $n$;
3. Calculate the weight sum of $(n,n_1)$ and add it to $H[n]$;
4. Take all node sets $n_2$ in shell layer with the largest value of $ks$ in $B$;
5. Compare $H[n_2]$, and sort $n_2$ in descending order according to the nodes influence;
6. The sorting results are added to the initial seed node set $S$;
7. Output initial seed node set $S$;

**END**

3.2. Strategy for Improving Repeat Degree of Seed Nodes

Because different seed nodes will affect different sets of nodes, in addition, the range of nodes affected by different seed nodes also has repeat degree in some cases. In order to improve the commodity profit of e-commerce platform and maximize it, how to select the first seed node accurately and ensure the low repetition rate of seed nodes are further considered.

Set a network $G(V,E)$, where $V$ represents the set of nodes in the network and $E$ represents the set of edges.

**Definition 1.** (Influence on neighboring nodes) For each node $v$ in node set $V$, calculate the influence $inf(v)$ of $v$ on its neighbor's node $(v)$:

$$inf(v) = \sum_{u \in N(v)} p(v,u)$$

(1)

In this formula, $N(v)$ represents the set of neighbor nodes of $v$, and $p(v,u)$ represents the propagation probability of the edge $(v,u)$.

**Definition 2.** (The influence of the node) : For each node $v$ in the node set $V$, calculate the influence of the node $v$ $inf$ $s(v)$:

$$inf s(v)=ks(v)+\lambda inf(v)$$

(2)

In this formula, $ks(v)$ represents the ks value of node $v$, and $\lambda$ is the parameter between the influence of the equilibrium node on its neighbor and the ks value. The $\lambda$ value of the network in experiment of is approximately 0.8.

Equations (1) and (2) can more accurately estimate the influence of all nodes in the network and select the first seed node. In addition, in order to overcome the repetitive influence of seed nodes, strategy IRDSN is adopted, which is described as follows:

1. Calculate the out-degree value of all nodes;
2. Calculate the influence of all nodes in the network with equations (1) and (2);
3. Select the node $v$ with the maximum influence as the first seed node, and mark the node $v$ as covered status;
4. The coverage state is marked V and all nodes whose distance from V is less than 2 and whose activation probability is greater than coverage threshold $\theta$;
5. Update the out-degree value of the nodes according to the seed nodes and the nodes that is marked covered status;
6. Select the node $v$ with the maximum out-degree value after updating as the second seed node;
7. Repeat steps (4)-(6) until the k node is selected.

In order to mark the status of nodes, the definition 3 to calculate the probability of the activation.

**Definition 3 (Nodes activation probability):** If the node $u$ is within 2 distances from the seed
node, the probability of \( u \) being activated is \( \text{pact}(u) \):

\[
\text{pact}(u) = 1 - \prod_{v \in S \cap pr(u)} (1 - p(v,u)) + \prod_{v \in S \cap pr(u)} (1 - p(v,u)) \times \prod_{z \in \text{pr}(u) \cap S \cap pr(z) \cap S} (1 - P(w,z) \times p(z,u))
\] (3)

In equation (3), \( v \) represents the starting point of a path whose path length is 1 from seed node to node \( u \), \( \text{pr}(u) \) represents the precursor set of node \( u \), and \( w \) represents the starting point of a path whose path length is 2 from seed node to node \( u \). \( \prod_{v \in S \cap pr(u)} (1 - p(v,u)) \) represents the probability that node \( u \) is not activated by set \( S \cap \text{pr}(u) \). \( \prod_{z \in \text{pr}(u) \cap S \cap pr(z) \cap S} (1 - P(w,z) \times p(z,u)) \) represents the probability that node \( u \) is activated by set \( S \cap \text{pr}(z) \).

4. Linear Threshold Model Description of Multi-propagation Patterns

4.1. Description of Multi-propagation Mode Algorithms Based on Linear Threshold Model

Node influence propagation is equivalent to seed node set affecting its adjacent and inactive nodes once, and judging whether it is activated or not. Because the whole propagation process is irreversible, a node can be changed from an inactive state to an active state and vice versa. In order to simulate the transaction process in e-commerce platform, this paper differentiates the communication process according to the number of times, and divides the communication process into one time, \( n \) times and the whole mode.

When applying the linear threshold model, it is necessary to set the weight value and cover the threshold \( \theta \) for the nodes and edges, and to judge the propagation mode of influence in the propagation process according to the time value. Since the mode of transmitting the whole mode and once is a special case in the process of transmission, the implementation process of the algorithm of \( n \) times of propagation will be discussed in this paper. The algorithm of \( n \) times propagation can be realized by performing the function of one time propagation in a loop, the algorithm is as follows:

**Algorithm 2.** The function of one time propagation.

**Input:** Social Network Graph \( G \), Seed List \( S \), Influencing propagation times

**Output:** List of all activated nodes

1. Set all_active_nodes as an empty list
2. Add the element of seed node list \( S \) to all_active_nodes in the list of all activated nodes
3. While affects propagation times > 0 and there are still nodes inactivated in the network
4. { Record the number of active nodes in the current network \( G \)
5. The influence propagates once, returning to the active node set and seed list \( S \)
6. Add the active nodes of this propagation to the list of all active nodes
7. If (no new nodes are activated in this influence propagation)
8. Stop Influencing propagation
9. Else(propagation times = times - 1)
10. End If }
11. Returns the list of all_active_nodes that are activated

**END**

4.2. Description of Commodity Profit Calculation

Profit refers to the result of calculating the purchase behavior of a node when the active state of the node is determined after the end of the propagation process. Therefore, the number of active nodes is an important factor to measure influence maximization algorithm; considering from the commercial...
level, the profit brought by the influence of propagation can be used as a reference for the accuracy of the algorithm. The calculation process of commodity profit is described as follows:

**Algorithm 3.** The calculation process of commodity profit.

**Input:** Activate Node  
**Output:** Total Commodity Profit  
1. G.node[n]['History_price'] = random.uniform(x,y)  //Setting historically highest purchase price function  
2. For (i = 1; i ≤ number of active nodes; i++)  
3. { Set the historical maximum purchase price attribute of the activation node;  
4. IF (Historically Highest Purchase Price > Current Price) {User purchases are included in profits.}  
5. Else {Users Abandon Purchase}  
6. End If  
7. i++; }  
8. End For  
END

Profit calculation traverses the nodes in the list of active nodes in the For cycle, and sets the highest price attribute of historical purchase for the nodes. If the highest price of historical purchase is higher than the current price, the user node will choose to buy the goods and calculate the profit, otherwise, it does not need to calculate the profit. (x,y) range should be set according to the commodity category and the season of commodity sales, and should not be set too many, in order to conform to the law of commodity pricing in real life.

5. **Experiment Simulations and Results Analysis**

Aiming at the problem of high repetition of seed nodes, an improved K-shell algorithm is proposed in this paper, two experiments are selected to verify the testing effect of the algorithm in terms of commodity profit. Experiment 1: accurately locate the first seed node and resolve the problem of high overlap of seed node set in the social network, choose three real data sets to verify the effect. Experiment 2: test the improvement strategy by crawling data on the e-commerce platform based on the results of experiment 1. The real data set information in experiment 1 is shown in table 2.

In order to verify the effect of the improved strategy IRDSN in the selection of initial seed nodes and overlap degree of nodes, experiments are carried out to compare IRDSN with Degree and CCA algorithms. In CCA algorithm, d=1, and in improved IRDSN algorithm, the balance parameter λ is set to 0.8. The coverage threshold θ is determined by network and propagation probability. In the experiment, the number of seed nodes ranges from 1 to 50, and the propagation probability of each side of the network is the same. In this paper, the performance of each algorithm is analyzed and compared under different propagation probabilities (p ∈ {0.02,0.04,0.06}). The test results in different experimental sets are shown in figures 2 and 3.

**Table 2.** Real data set.

| Name          | Number of nodes | Number of edges | Average aggregation coefficient |
|---------------|-----------------|-----------------|--------------------------------|
| Wiki-Vote     | 7115            | 103689          | 0.1409                         |
| cit-HepTh     | 27770           | 352807          | 0.3120                         |
| web-NotreDame | 325729          | 1497134         | 0.2346                         |
Coverage threshold $\theta=0.08$, the influence range of 50 seed nodes under different propagation probabilities is shown in figure 2. As shown in figure 2, with the increase of propagation probability, the improved algorithm IRDSN is better than the other two methods. The reason is that when $p=0.04$, the marked uncovered nodes are fixed. Most of these nodes are activated during the propagation process, but a few nodes whose activation probability is close to 0.04 are not activated. When the propagation probability increases, a few nodes are activated, and the covered nodes are almost equal to the activated nodes. The effect is optimal.

The experimental result on a large-scale social network web-NotreDame is verified in figure 3. We can see from figure 3 that when $p=0.02$, 0.04, 0.06, Degree and IRDSN algorithms are significantly better than CCA algorithms. When $p=0.06$, IRDSN algorithm has the best effect, about 10% higher than Degree algorithm and about 35% higher than CCA algorithm.

The experimental result on cit-HepTh network is shown in figure 4. The coverage threshold of IRDSN algorithm is $\theta=0.16$. We can see from figure 4 that the number of seed nodes ranges from 1 to 30, the effect of IRDSN algorithm and Degree algorithm is similar. As the number of seed nodes increases, IRDSN algorithm has the best influence, 11% better than CCA algorithm and 13% better than Degree algorithm. Therefore, we can see that the improved K-shell and strategy IRDSN achieves better results through solving the problem of initial seed node selection and high repetition of nodes in real data sets of different scales.

Applying the strategy IRDSN to test the selection of initial seed nodes and the coverage range of seed nodes in improved K-shell algorithm, and achieves good experimental results. Avoiding seed coverage of the same node set is the key factor to improve commodity profit of e-commerce platform. Therefore, the above experimental results are a foundation for improving the calculation of total profit of commodities in experiment 2. In order to test the profit of experiment 2, crawling the data information of clothing shirts in 10 e-commerce websites in this paper, and the related parameters are set as shown in table 3.
Table 3. Set Parameter.

| Variable          | Parameter                      | Meaning                                      |
|-------------------|--------------------------------|----------------------------------------------|
| Times             | Integer                        | Propagation times                            |
| Seeds             | List of seed nodes             | Set of initial seed nodes                    |
| History_price     | Real numbers in range (100, 200) | highest price of historical purchase         |
| Price             | 150                            | Current commodity pricing                     |
| Cost              | 100                            | Cost of unit price goods                      |
| Weight            | \(1/\text{in}_\text{degree}(n)\) | Weight value of edge                         |
| Threshold         | 0.5                            | Threshold of node                            |

Ten social networks are produced by using the “shirt” related data information obtained from the e-commerce platform, and an influence maximization algorithm is executed once for each network. The average value is taken as the experimental result. History price, which is generated by random function, has different attributes of the highest historical purchase price for each node in the network. Fifty seed nodes are selected and the K-shell algorithm, degree center algorithm and improved K-shell are influenced by linear threshold model. The number of active nodes, running time and the numerical data of profit gained by propagation are affected as shown in Table 4.

Table 4. Experimental results.

| Algorithm                | Maximum number of activated nodes (number) | Maximum value of running times (s) | Maximum number of propagation times (times) | Maximum value of profit gained (yuan) |
|--------------------------|--------------------------------------------|-----------------------------------|---------------------------------------------|---------------------------------------|
| The K-shell algorithm    | 284                                        | 6.6                               | 3                                           | 7500                                  |
| Degree-centered          | 765                                        | 21.3                              | 5                                           | 19700                                 |
| Improved K-shell algorithm | 930                                      | 22.1                              | 5                                           | 23850                                 |

The comparison results of the propagation times of the three algorithms are shown in Figure 5. We can see from Figure 5 that compared with the degree-centered algorithm, the number of active nodes increases with the increase of propagation times. Among them, the improved K-shell algorithm and degree-centered algorithm are not activated after five times of propagation, which affects the stop of propagation process; while the K-shell algorithm is propagated three times, the propagation process stops. In addition, under the same number of propagation times, the number of nodes activated by the improved K-shell algorithm is more than that of the K-shell algorithm and the degree-centered algorithm after the propagation of the linear threshold model. The reason is that the K-shell algorithm cannot accurately evaluate the importance of nodes in the same layer, while the improved K-shell algorithm has a good performance in evaluating the impact of nodes in the same layer, and the nodes with higher have more influence on their adjacent nodes. With the increase of propagation times and the expansion of propagation range, the number gap of activated nodes gradually narrows, but the improved K-shell algorithm still has an advantage in the number of activated nodes.

Under the condition that the other parameters remain unchanged, the profits between different algorithms are shown in Figure 6. It illustrates whether the improved algorithm can make more profit at the level of commercial platform. The total profit of different algorithms increases with the increase of propagation times, but the improved K-shell algorithm has more active nodes in Figure 6, so more user nodes may purchase to increase the total profit.
6. Conclusion
The improved K-shell algorithm and strategy IRDSN solve the problem that the influence of nodes with the same ks value unable to measure accurately. Influence nodes obtained by improved K-shell algorithm are brought into linear threshold model, the profit of e-commerce platform is higher. The method works well in calculating profit on the small e-commerce platform and provides a foundation for the profit calculation of small networking.

In the network, the influence of each node is different, so the propagation probability of each edge is not equal. Therefore, the propagation probability should be a value varying with the influence of the node, which will make the experimental effect closer to the real situation. Although the number of active nodes of the improved K-shell algorithm is large, the running time also increases accordingly. The future work is to find the appropriate propagation probability, optimize the running time of the algorithm and can be applied to different scale e-commerce platforms.

References
[1] Cao J X, Dong D and Xun S A 2015 K-Core based algorithm for influence maximization in social networks Chinese Journal of Computers 38 238-48.
[2] Zhao J X 2017 Research on Social Network Marketing Strategies Based on Algorithms (Fujian Normal University).
[3] Guo J, Zhang P and Fang B X 2014 Personalized key propagating users mining based on LT model Chinese Journal of Computers (4) 809-18.
[4] Richardson M and Domingos P 2002 Mining knowledge-sharing sites for viral marketing Proc. Int. Conf. on Knowledge Discovery and Data Mining Eighth ACM SIGKDD pp 61-70.
[5] Kempe D, Kleinberg J and ardos É 2003 Maximizing the spread of influence through a social network Proc. Int. Conf. on Knowledge Discovery and Data Mining Ninth ACM SIGKDD pp 137-46.
[6] Leskovec J, Krause A and Guestrin C 2007 Cost-effective outbreak detection in networks Proc. Int. Conf. on Knowledge Discovery and Data Mining ACM SIGKDD pp 420-29.
[7] Chen W, Wang Y and Yang S 2009 Efficient influence maximization in social networks Proc. Int. Conf. on Knowledge Discovery and Data Mining ACM SIGKDD pp 199-208.
[8] Chen W, Wang C and Wang Y 2010 Scalable influence maximization for prevalent viral marketing in large-scale social networks Proc. Int. Conf. International Conference on Knowledge Discovery and Data Mining ACM SIGKDD pp 1029-38.
[9] Jung K, Heo W and Chen W 2012 IRIE: Scalable and robust influence maximization in social networks IEEE International Conference on Data Mining 5 918-23.
[10] Tian J T, Wang Y T and Feng X J 2011 A new hybrid algorithm for influence maximization in social networks Chinese Journal of Computers 34 1956-65.

[11] Zareie A, Sheikhamadi A and Khamforoosh K 2018 Influence maximization in social networks based on topsis Expert Systems with Applications 108 96-107.

[12] Guo J F and Lü J G 2015 Influence maximization based on information preference Journal of Computer Research and Development 52 533-41.

[13] Xun Y G, Pan J Z and Xie H Y 2016 Minimum vertex covering and feedback vertex set-based algorithm for influence maximization in social network Journal of Electronics & Information Technology 38 795-802.

[14] Liu Y Y, Guo J F and Wei L D 2017 Fast influence maximization algorithm in social network under budget control Journal of Computer Applications 37 367-72.

[15] Cao J X, Yan H Y and Wang H R 2019 Self-interest influence maximization algorithm based on subject preference in competitive environment Chinese Journal of Computers 42 1495-1510.

[16] Zhu W Y, Peng W C and Chen L J 2016 Exploiting viral marketing for location promotion in location-based social networks ACM Transactions on Knowledge Discovery from Data 11 25.

[17] Zhang P, Wang L W and Peng Z Y 2017 Group-based method for influence maximization Journal of Software 28 2161-74.

[18] Yang Y, Xu Y and Wang E 2018 Exploring influence maximization in online and offline double-layer propagation scheme Information Sciences 450 96-107.

[19] Tang J, Tang X Y and Yuan J S 2018 Profit maximization for viral marketing in online social networks: Algorithms and analysis IEEE Transactions on Knowledge and Data Engineering 30 1095-1108.

[20] Kitsak M, Gallos L K and Havlin S, et al. 2010 Identification of influential spreaders in complex networks Nature Physics 6 888-93.