Summer 6-19-2015

Containment of RumorsunderLimitCost Budget in Social Network

Chunyang Cao
School of Information, Renmin University of China, Beijing,100872,P.R.China

Li Yu
School of Information, Renmin University of China, Beijing,100872,P.R.China, buaayuli@ruc.edu.cn

Yang Hu
School of Information, Renmin University of China, Beijing,100872,P.R.China

Containment of RumorsunderLimitCost Budget in Social Network

Chunyang Cao[1], *Li Yu*[1]**, *Yang Hu*[1]
[1]School of Information, Renmin University of China, Beijing,100872,P.R.China

Abstract: With widely usingof computer and mobile devices available, people share information more frequently on the online social network (OSN) than before, so information spread faster and wider, especially misinformation andrumors. Rumors on the OSN often make E-commerce companiessuffermuch financial losses.Once there is a rumor, the companies always try to control the rumor propagating so that they suffer the least loss by using a certain budget. In this paper,an effective method on information blocking maximization with cost budget (CIBM) is proposed to solve rumor containment problem with cost budget in e-commerce environment. First, CIBM is proved as NP-hardproblem withthe characteristic of sub-modular and monotone.Then a community dividing algorithm based community structure is presented to optimize containment of the rumors. To verify our proposed method, a lot of experiments are conducted on real dataset and random generated datasets. And the experiment results show that our algorithm has advantage over traditional methods.

Keywords: rumor containment, social networks, word-of –mouth, user review, E-commerce

1. INTRODUCTION

Online social network (OSN) has become an integral part in modern life,people sharing information and keeping in touch with friends and families on OSNs, such as Twitter, Facebook, WeChat, etc. Information spreads faster and farther by OSN than traditional media[1]. Unfortunately, rumorsalso spreadin social network, which often make the enterprise to suffer much loss.For example, recently it spreads quickly in many social networks that a lot of fake or inferior commodities are sold in TAOBAO, a famous E-Commerce company. The rumor has made TAOBAO suffer heavy loss. It is urgent that TAOBAO has to take an action to control the rumor so that least loss is suffered. Usually, in order to control the rumor, some influential users in social network are invited to propagate positive information for fighting against rumors. Corresponding, TAOBAO needs to give them reward. To some degree, the reward can be regarded as cost which TAOBAO invest in blocking the rumor. The invested budget which used to control rumor is limited in most instances. So, it is important that companies like TAOBAO design an optimized strategy to block the rumor under limit cost budget. In this paper, it is calledinformation blocking maximization with cost budget (CIBM)problem.

Numerous studies have been carried out on competition betweenpositive and negative informationin viral marketing. Different schemes have been proposed to limit the spread of rumor in earlier works [2,14,15,19,20].However,most studies do not consider the cost to active a node. In fact, it is more reasonable that a company gives reward to spokespersonsinvited by company to spread positive information on the OSNs. In this paper, we aim to find a set of node users on the OSN as a seed set to decrease the rumors' influence under given cost budget. Each user node can be in three statuses: positive, negative or inactive.Most studies have a similar assumption, but they also assume that when rumor and truth arrive at the same time, whether rumor[15]or truth[14] is advantaged. However, it is true that the status of a node depends on his neighbors' current statuses, and a individual isalwayshard to hold a different opinion from most of his friends[14,15,19,20]. Based on this observation, a containing competitive propagating model under Linear Threshold (CCLT) is proposed in this paper.

We also propose a strategy to contain rumors based on the view that community structure is common in

* Corresponding author, Email: buaayuli@ruc.edu.cn (Li Yu)

OSNs. It is notable that individuals' relation within a community is denser than that across communities, information spread faster inside a community while slower across communities accordingly. Our strategy is to isolate communities contain rumor from those do not. Furthermore, we assume that each community have a different value for different content of rumors, which represents the expect money we can spend on it. It is easy to understand that a community full of individuals concerning on fitness is less valuable than one full of musician for a sport company.At last, we conduct experiments on both random networks and real-world network, and results show that propose method has a better performance than traditional algorithms.

All in all, our contribution is as follows: First, the CIBM problem is identified and the aim is to find a set S to spread positive information under given cost budget so that the expected number of decontamination nodes in the whole network is maximized. Also, we build a containing competitive propagating model under linear threshold(CCLT) considering uses' attitude is influenced by their friends. Thirdly, a containment method based on community structure is proposed, and experiments are conduct to study the performances of our method.

The remaining parts of the paper will be organized as follows. Related worksare surveyed in next section. In section 3, CIBM problem is formally defined, itscharacteristics be proved. In section 4, containing strategies for CIBM are proposed. In section 5, experiments are conducted on real datasets. Finally, conclusionsaremade.

2. RELATED WORKS

The information propagation problem is first studied by Domingos and Richadson, they also proposed the well-known influence maximization problem[4].After that many researchers have been studied this problem.Some of them focus on the propagate model, in 2003, Kempe et al. proposed the linear threshold(LT)model and the independent cascade(IC) model[5],which are widely used in studies in viral marketing. Researchers extend the basic LT and IC model to solve different problem. eg. Borodin et.al and Heet al. propose several different models and CLT model to study the competitive influence diffusion problem respectively[19]. Yu etal. proposed a game-theoretic model for competitive information dissemination[9]. Wang et al. studied a containing competitive propagating model (IC-CCPM)[11].

Other follow-up researchers study the competitive influence diffusion. For instance, Li et al. studied a $\gamma-$k rumor restriction problem which extend theβ_T^I node protectors problem proposed by Nguyen et al. Fan et al. Proposed least cost rumor blocking problem.Some attempts have been made in limiting the spread of virus or malware in computer networks because it is similar with competitive influence diffusion problem[3,8,12,13]. But in these problems, positive information cannot be spreadafter a node accepts it.

Few studied has been focused on influence maximization problem with limit cost.Wang et al.propose a IMLC problem and prove it a NP-complete problem[11], Kotnis&Kuri[21] study the cost effective rumor containment in social networks and propose some strategy for government to control the outbreak size of rumor.However, they did not consider the condition when positive information as well as rumors spread in the network and they assume the network scale-free, which cannot always characterize the OSNs.

3. INFORMATION BLOCKING MAXIMIZATION UNDER COST BUDGET (CIBM)

In this section,information blocking maximization under cost budget (CIBM) problem is formally defined. Also, its characteristics are proved.

3.1 Definition of Problem

As is mentioned before, Kempe et al. proposed the linear threshold model in [5]. In this model,in the diffusion process of LT model, each node v has a threshold θ_v. One of v's neighbors u will influence v with a weight $w(u,v)$,$\sum_{w\in N(v)} b_{w,v} \leq 1$, where N(v) denotes the set of v's neighbors. Node v becomes active only if $\sum_{w \in N_{\text{active}}(v)} b_{w,v} \geq \theta_v$.

We now extend the LT model by incorporating competitiveinfluence diffusion. In our model, each node has three statuses: inactive, positively activated and negatively activated. A node in the network will stay inactive until it is positive (negative) activated.Each node has a trust value P_v, which denote the current status of a node. In general, if $P_v < 0$ and $|P_v| > \theta_v$, it means that the node is negatively activated; if $P_v > 0$ and $|P_v| > \theta_v$, it is positively activated; else it is confused and turn to(or remain) inactive. It make sense that if an individual has more friends trust negative (positive) information, he will accept it, and remain to trust it. When all of his friends are divided into two groups, he may be confused and stay inactive.The definition of containing competitive propagating model under linear threshold is described as follows:

Definition 1Containing competitive propagating model under linear threshold (CCLT): Given a social networkG $= (V, E, w)$, where $w(u, v)$ denotes the relationship between node u and node v, the higher $w(u, v)$ is, the closer node u and v is. Also, it is assumed that if a node u trust positive (negative) information it will influence all of his neighbors with a positive (negative) weight w, while a node v will be positive active if the positive weight is larger than the negative weight in all of his neighbors $N(v)$and the sum of them exceeds its threshold θ_v and vice versa. What is more , it will remain it as trust value P_v.

Here is an example for CCLT:

There are six nodes in Figure 1, number on the left side of each line denotes the weight between two nodes, the number in the circle denotes the threshold of a node. At step T, one of them is negative active (node C), two are positive active (A and B), three are inactive (D, E, F).Assume that node C have a trust value of -0.5 at step T. At step T+1, some nodes' status change as follows:

Node C:node A will spread positive information with weight 0.5 next step, it will have a trust value P_C=0.5-0.5=0, and trust value is less than its threshold: $|P_C|$<0.3, it will turn to inactive.

Node D: node A,B will spread positive information with weight 0.4 and 0.4 respectively, node C will spread negative information with weight 0.2, it will have a trust value P_D=0.4+0.4-0.2=0.6, and the trust value is more than its threshold:$|P_D|$=0.6>0, it will be positively activated.

Node F: trust value P_F=0-0.5=-0.5,$|P_F| > 0.4$,it will be negatively activated.

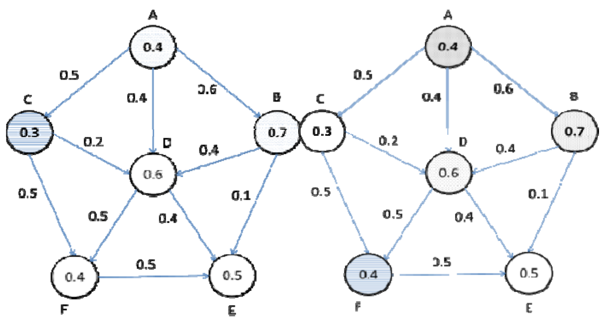


Figure 1. An example of CCLT

Based on model above, information blocking maximization under cost budget (CIBM) problem is defined.

Definition 2Information blocking maximization under cost budget (CIBM) problem: Given a social network represented by a graph $G = (V, E, w, c)$, an diffusion model, a cost budget C and a contaminated set I that holds the rumor(negative information), to find a set $S \subseteq V, \sum_{v \in S} c_v \leq C$, both from the inactive nodes and the negative nodes, so that the except number of negatively activated nodes is minimized, or equivalently,the reduction in the number of negatively activated nodes ismaximized.

We choose protectors both in contaminated nodes and inactive nodes. There is a cost to pick any nodes, in real life a node have more fans will need more money for spreading a certain information (rumor or truth). We assume the cost of each node is directly proportional to the degree. A node with out-degree of p will have a cost of $k \times p$. In fact, individuals in social network often price according to the number of their fans. We also assume that a node will never spread rumor (or negative information) once they get paid. Companies should pay all the money a node charge for, or it won't propagate information.

3.2 Characteristics of CIBM

In order to solve the CIBM, it is important to check its characteristics. In this section, we will firstly prove the information blocking maximization as a NP-hard problem. We will also show it is monotone and submodular.

Let $\sigma^+(S)$ denote the expected number of positive and inactive nodes that will be influenced by set S,$\sigma^-(S)$ denote the expected negative number of nodes that will be influenced by set S,${\sigma_S}^+(C)$ denotes the expected number of positive and inactive nodes that will be influenced by set S with limitation of budget C.

(1) NP-hardness

THEOREM 1 Given a social network represented by a graph $G = (V, E, w, c)$, an diffusion model, a cost budget C and a contaminated set I that holds the rumor(negative information), to find a set $S \subseteq V, \sum_{v \in S} c_v \leq C$,which maximizes ${\sigma_S}^+(C)$ is NP-hard problem.

Proof: Consider a special case of our problem where each node's cost is equal to 1, the budget is equal to k. In this case, we can select k nodes from the graph G, to maximize$\sigma^+(S)$. From above we have successfully reduced the influence blocking maximization problem with give cost budget to the influence blocking maximization problem, which has been proved to be a NP-hard problem in [19].

(2) Monotony

THEOREM 2 Given a social network represented by a graph $G = (V, E, w, c)$,an diffusion model(CCLT) ,a cost budget C and a contaminated set I that holds the rumor(negative information), to find a set $S \subseteq V, \sum_{v \in S} c_v \leq C$,which maximizes${\sigma_S}^+(C)$is monotony.

Proof: Suppose that$C_1 \leq C_2$, then withC_1 *and* C_2, we can select S and T respectively. It is obvious we have $S \leq T$ here, with more money we can choose more nodes as spokesman.To be more clear, let R denote the maximum set of nodes can be selected withcost$C_2 - C_1$. As a result, we get $S \cup R \leq T$,and$S \leq T$.Let $\sum^+ P$ denotes all the trust value in the whole graph, it is easy to know that if $\sum^+ P$ increase the expect number of positive and inactive number in the graph will increase. With more nodes in the network to spread positive information, the expected number of inactive and positive nodes will increase, which implies that if $S \leq T$, we have$\sigma^+(S) \leq \sigma^+(T)$, therefore ${\sigma_S}^+(C)$ on C is monotone.

(3) Submodularity

THEOREM 3 Given a social network represented by a graph$G = (V, E, w, c)$, an diffusion model(CCLT), a cost budget C and a contaminated set I that holds the rumor(negative information), to find a set $S \subseteq V, \sum_{v \in S} c_v \leq C$,which maximizes${\sigma_S}^+(C)$ is submodular .

Proof: Suppose that$C_1 \leq C_2$, which also means that C_1 is subset of C_2.Then with C_1 *and* C_2, we can select S and T respectively. In order to prove ${\sigma_S}^+(C)$ is submodular, we need to verify that ${\sigma_S}^+(C)$ satisfies the diminishing return condition. Assume that α>0 is a real number, we need to prove that:

$${\sigma_S}^+(C_1 + \alpha) - {\sigma_S}^+(C_1) \geq {\sigma_S}^+(C_2 + \alpha) - {\sigma_S}^+(C_2) \qquad (1)$$

Let S, T, R represent maximum sets can be selected with cost of C_1, C_2,α respectively. So if we can prove (2),${\sigma_S}^+(C)$ is submodular.Here $S \subseteq T$, if $R \cap T = \emptyset$, we have(3);if $R \cap T \neq \emptyset, R \cap S = \emptyset$,we have(4);if $R \cap T \neq \emptyset, R \cap S \neq \emptyset$,we have (5).Since all the condition has been considered, ${\sigma_S}^+(C)$ is submodular. Figure 2 is helpful for understanding all the conditions.

$$\sigma^+(S \cup R) - \sigma^+(S) \geq \sigma^+(T \cup R) - \sigma^+(T) \qquad (2)$$

$$\sigma^+(S \cup R) - \sigma^+(S) = \sigma^+(T \cup R) - \sigma^+(T) (3)$$

$$\sigma^{+}(S \cup R)-\sigma^{+}(S)=\sigma^{+}(R) \geq \sigma^{+}(T \cup R)-\sigma^{+}(T)(4)$$
$$\sigma^{+}(S \cup R)-\sigma^{+}(S)>0=\sigma^{+}(T \cup R)-\sigma^{+}(T)(5)$$

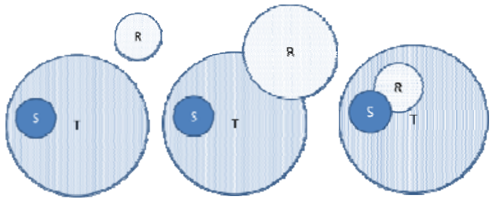


Figure 2.Relationships between sets S, R, T

4. CONTAINING STRATEGIES FOR CIBM

In this section we propose an algorithm called community based (CB) for the CIBM problem, this algorithm is based on the knowledge that a social network is denser inside than outside a community.Its key view is to isolate communities with rumor in it and communities without rumor.

First, we detect community in the whole graph and divided it into p communities, the community detection method is proposed in[16].The cost budget is allocated to each community according to its size. Here we assume that all the nodes in the graph have equal purchasing power.

After dividing the graph into different communities, these communities are classifiedinto two groups according to if there are negative acted nodes, those include negative nodes are called infected community, the other is called uninfected community. For each infected community, the algorithm in Table 2 is applied. At each round of the algorithm a node adding the maximalmarginal gain of $(\sigma(F+u)-\sigma(u))$ with the minimumcost, selecting process will end until the cost budget for this community is spent up.

Table 2. Community based Algorithm in infected community

Input: Graph $G=(V, E, P, C)$, $\lvert V\rvert=N$, $V=\{V_1, V_2, V_3 \cdots V_p\}$ Output:Selected nodes：F
1. For each infected community: $S \leftarrow \emptyset$ 2. For each $v \in G_c$:$S \leftarrow \{v\}$ 3. For each$v_i \in S$ 4. while $V_B \geq 0$ 5. If v_i is infected 6. While $(C_i \leq V_B)$ do 7. $v \leftarrow argmin_{u \in S/F}(\frac{C_u}{(\sigma(F+u)-\sigma(u))})$ 8. $F \leftarrow F \cup v, V_B \leftarrow V_B - C_i$ 9. End while 10. End if 11. End while 12. Output F

For each uninfected community, at each round of the algorithm we try to find the nodes connect with different communities and positive act them.Selecting process will end until the cost budget for this community has all spent.The algorithm will stop after all the communities have been computed.

5. EXPERIMENT

In this section we try to test our algorithmthrough experiments. Experiment design, and datasets are introduced, and result is given.

5.1 EXPERIMENT DESIGN

Two experiments are designed to check the proposed method.

The first experiment is made to compare the performance of four different methods as follows:

- Community based algorithm (CB): the algorithm propose in this paper.
- Maximum out-degree first algorithm (MOF): Each round of this algorithm, it will select nodes with most degree in the graph.
- Infected maximum out-degree first algorithm (IMOF): Each round of this algorithm, it will select negative acted nodes with most degree in the graph.
- Random algorithm (RAN): Each round of this algorithm, it will select nodes randomly.

The second experiment is made to find out when information about negative acted nodes is not clear, which algorithm will act better. In this experiment we will compare CB with MOF and RAN, while IMOF is not included considering that we don't know which one in the graph is negative acted.

5.2 EXPERIMENT DATASET

The datasets we used in the experiment are Arxiv HEP-PH (high energy physics phenomenology) citation graph is from the e-print arXiv and covers all the citations(34,546 nodes, 421,578 edges)[17], in this paper it is called Cit-HepPh. In addition to that, two random datasets (1000 nodes ,14323 edges and 1500 nodes , 23421 edges) are generated in this paper. In order to compare results of different datasets, we extract sub-dataset with 6000 user nodes including dense edge from original Cit-HepPh dataset.

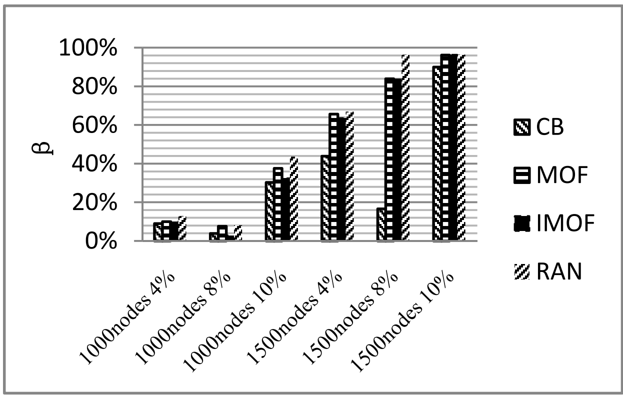
In experiments, we let theinfluence weight $w(u, v)$ be the reciprocal of in-degree of each node v. We also suppose that an individual have same threshold when facing with rumor (negative) and truth (positive).Prevalence of infection denotes the ratio between negative nodes (rumor) and all nodes in it in the dataset, total cost denotes the cost budget and K value denotes how much each out-degree value. All the experiment setup is shown in Table 1 below.

Table 3.Two group experiments

Experiment Setup			
Experiment 1			
		Total Cost	Prevalence of Infection
Datasets	Cit-HepPh_6000	300000	4%
	Random_1500	50000	4% 8% 10%
	Random_1000	75000	4% 8% 10%
K value	100		
Threshold	Random between 0.1 and 0.3		
Algorithm	CB MOF IMOF RAN		
Experiment 2			
		Total Cost	Prevalence of Infection
Dataset	Random_1000	50000	4% 10% 20%
	Random_1500	75000	4% 10% 20%
K value	100		
Threshold	Random between 0.1 and 0.3		
Algorithm	CB MOF RAN		

5.3 EXPERIMENT RESULT

In the first experiment, results of Random datasets (Random_1000,Random_1500) are shown in Figure 2. The β here denotes ratio between negative nodes and all nodes in it in the whole dataset. We can conclude that CB is better than other algorithm in most situations. We can see that in Figure 3, which is the result of real dataset Cit-HepPh, CB perform better especially when the number of step increases. Some other interesting phenomenon is indicated here, according to the result, the later we found rumor is spreading, which means the more negative nodes or a higher prevalence rate of infection when we start to contain rumors on the network, the harder to control it (a higher ratio of infection in the end).



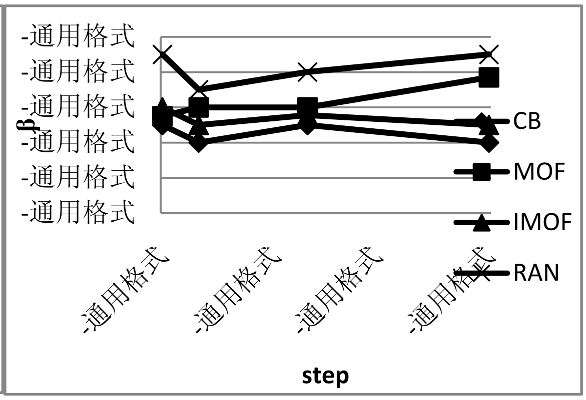


Figure 2.Result of random datasets **Figure 3.Result of Cit-HepPh_6000**

In the second experiment, as is shown in Figure 4, when information of negative nodes is not clearandwe only know there is rumor spreading on the internet, CB also have a lower infected ratio. Result in this experiment reveals the conclusion we proposed in experiment 1 again. The later we take action, the harder controlling rumor will be.

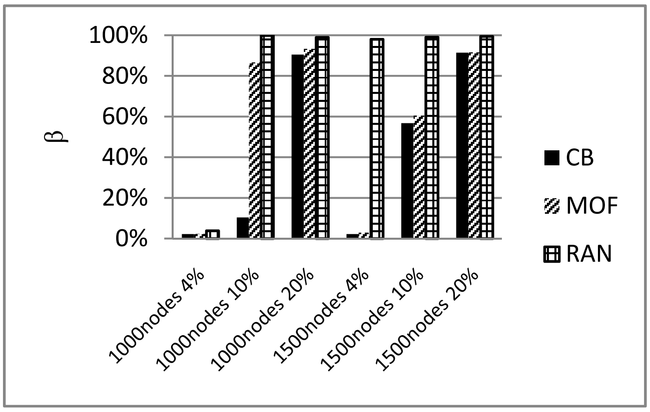


Figure 4.Result of datasets when rumor is unclear

6. CONCLUSIONS

In this paper we focus on the rumor containment problem with cost budget in social network, which is also demonstrated as information blocking maximization with cost budget (CIBM) problem.We first prove CIBM problem as NP-hard, monotone and submodular. Based on the knowledge that a social network is denser inside than outside a community, we propose a community based algorithm which is perform well in experiments we conducted in both real and random generated datasets. Results of experiments show our method isefficient on most of the datasets. In this paper we assumed that cost function is linear in node out-degree k, there may be situations where cost functions are sub-linear or super-linear. Different cost functions may lead to different results in experiments. We leave this interesting problem to the future.

ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China under Grant 71271209，71331007,Beijing Municipal Natural Science Foundation under Grant 4132067, Humanity and Social Science Youth Foundation of Ministry of Education of China under Grant 11YJC630268, Special Fund for Fast Sharing of Science Paper in Net Era by CSTDunder Grant 2013112, Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China.

REFERENCES

[1]Subramani, M. R.,Rajagopalan, B. (2003). Knowledge-sharing and influence in online social networks via viral marketing. Communications of the ACM, 46(12), 300-307.

[2] Fan, L., Lu, Z., Wu, W., Thuraisingham, B., Ma, H., & Bi, Y. (2013-07). Least cost rumor blocking in social networks. In Distributed Computing Systems (ICDCS), 2013 IEEE 33rd International Conference on (pp. 540-549).IEEE.

[3] Aspnes, J., Chang, K., &Yampolskiy, A. (2005-01). Inoculation strategies for victims of viruses and the sum-of-squares partition problem. In Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms (pp. 43-52). Society for Industrial and Applied Mathematics.

[4] Domingos, P., & Richardson, M. (2001-08). Mining the network value of customers.In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 57-66).ACM.

[5] Kempe, D., Kleinberg, J., &Tardos, É. (2003-08). Maximizing the spread of influence through a social network.In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 137-146).ACM.

[6]Li, Y. M., Lin, C. H., & Lai, C. Y. (2010). Identifying influential reviewers for word-of-mouth marketing. Electronic Commerce Research and Applications, 9(4), 294-304.

[7] Trusov, M., Bodapati, A. V., & Bucklin, R. E. (2010). Determining influential users in internet social networks. Journal of Marketing Research, 47(4), 643-658.

[8] Bharathi, S., Kempe, D., &Salek, M. (2007). Competitive influence maximization in social networks. In Internet and Network Economics (pp. 306-311). Springer Berlin Heidelberg.

[9] Yu, L., Li, Q., &Xun, L. (2012). Research on Viral Marketing Propagating Oriented to Marketing Context.

[10] Arthur, D., Motwani, R., Sharma, A., & Xu, Y. (2009). Pricing strategies for viral marketing on social networks. In Internet and Network Economics (pp. 101-112). Springer Berlin Heidelberg.

[11] Wang, Y., Huang, W., Zong, L., Wang, T., & Yang, D. (2013). Influence maximization with limit cost in social network. Science China Information Sciences, 56(7), 1-14.

[12] Nicol, D. M., &Liljenstam, M. (2005). Models and analysis of active worm defense.In Computer Network Security (pp. 38-53).Springer Berlin Heidelberg.

[13]Tanachaiwiwat,S.,Helmy, A. (2009). Encounter-based worms: Analysis and defense. Ad Hoc Networks, 7(7), 1414-1430.

[14]Budak, C., Agrawal, D., & El Abbadi, A. (2011-03). Limiting the spread of misinformation in social networks. In Proceedings of the 20th international conference on World wide web (pp. 665-674). ACM.

[15] Chen, W., Collins, A., Cummings, R., Ke, T., Liu, Z., Rincon, D., ...& Yuan, Y. (2011-04). Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate. In SDM (Vol. 11, pp. 379-390).

[16] Reichardt, J., &Bornholdt, S. (2006). Statistical mechanics of community detection.Physical Review E, 74(1), 016110.

[17] Leskovec, J., Kleinberg, J., &Faloutsos, C. (2005-08). Graphs over time: densification laws, shrinking diameters and possible explanations. In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining (pp. 177-187).ACM.

[18] Wang, N., Yu, L., Ding, N., & Yang, D. (2014). CONTAINMENT OF MISINFORMATION PROPAGATION IN ONLINE SOCIAL NETWORKS WITH GIVEN DEADLINE.

[19] He, X., Song, G., Chen, W., & Jiang, Q. (2012). Influence Blocking Maximization in Social Networks under the Competitive Linear Threshold Model.In SDM (pp. 463-474).

[20] Nguyen, N. P., Yan, G., Thai, M. T., &Eidenbenz, S. (2012-06). Containment of misinformation spread in online social networks. In Proceedings of the 4th Annual ACM Web Science Conference (pp. 213-222).ACM.

[21] Kotnis, B., &Kuri, J. (2014). Cost Effective Rumor Containment in Social Networks.arXiv preprint arXiv:1403.6315.