Rumors in the Social Network: Finding the Offenders Using Influence Maximization

Ashis Talukder, Rossi Kamal, Anupam Kumar Bairagi, Md Golam Rabiul Alam, Sarder

Fakhrul Abedin, Md Abu Layek,　Nguyen H. Tran and Choong Seon Hong

Kyung Hee University

Abstract

In recent years, information sharing among the social network users has been increased tremendously. Sometimes false information called 'Rumor' becomes viral and causes different types of penalties including monetary loss in financial organization. On the other hand, Influence Maximization (IM) problem in the social network is based on viral marketing or word-of-mouth effect. In this paper, we have tried to fit IM problem to trace out the sources of rumor in the social network. We have also included a simulation based on synthesized data.

1. Introduction

With the proliferation of smart phones and mobile devices social network has achieved eye-popping popularity in exchanging information in last few years. The information exchanged by the social network users (e.g., status shared by a Facebook user, or a tweet posted by a Twitter user) is not always true. Fact that there may be plenty of false claims termed as "Rumor". Worse case of this risk may include, for example, spread of virus and worms [2]. It may cause various types of penalties as well, such as, monetary loss in financial market [1], even can be a threat to lives. So it would be a potential work to identify the source or sources of rumor.

It has been seen that, unlike true information, false information (rumor) is initiated by a small number of offenders intentionally and can be propagated epidemically through the network in 'word of mouth' effect also called 'viral marketing' [2], [8], [9] [10]. Thus from the nature of Influence Maximization problem, it seems to have relevance with rumor spread. That is our key reason of applying IM technique to identify the offender (s) of the rumor.

In this paper, we have proposed a way of isolating the source of rumor using influence maximization problem. Firstly, we considered a set of rumors disseminated over a social network and determine a set of initial suspected defenders, we named it as a set of spurious defenders. Secondly, we have calculated the influence set or region of each member of the spurious set. Finally, most possible offender and a set of most possible offenders are isolated removing relay nodes. The rest of the paper is as organized: in section two a study of the state-of-the-art has been provided and our system model is unveiled in the section three. Elevation, conclusion and future scope have been described in the next consecutive sections.

2. Related works

Recently, viral spread of rumor in social network has become common and un-avoidable phenomena. It has been seen that for any message to be viral in the social network, the influence of the person acts as a vital catalyst [8], [9], [10]. But generally rumors are created by single or a small set of nodes intentionally [2]. Researchers in [4] have devised a maximum likelihood (ML) estimator and derived its asymptotic performance for regular tree-type networks.

Shah et al. [3] have tried to finding the offenders of rumors by SIR model using a maximum likelihood estimator and claimed linear time complexity to find the rumor centrality for every node. In [5] researchers have solved the problem by using rank and

optimization algorithm and Monte Carlo method. Li et al. [6] have proposed submodular algorithm with (1− 1/e) approximation.

Most of the techniques to solve the IM problem are NP hard. In this paper, we fitted two popular algorithms of influence maximization to this new scenario of the detection of offenders of rumor in social network.

3. Problem formulation and solution process

A) Finding spurious set of offenders

Consider a scenario that there are n nodes in a social network having n users (nodes) $O_1, O_2, \cdots, O_n$ and q assertions (rumors) $r_1, r_2, \cdots, r_q$. The graph is a bipartite graph (See Figure 1) and we have named the graph to be Offender−Rumor Graph, $ORG(n, q)$ where there will be an edge from O_i to r_j if the offender O_i has contribution in the spread of rumor r_j, that is,

$$ORG(i, j) = \begin{cases} 1 & \text{if } O_i \text{ has contribution in spreading rumour } r_j \\ 0 & otherwise \end{cases}$$

Here each column r_i of the matrix ORG represents a vector of n components of rumor i and gives a candidate set of spurious offenders, for example.

$O(r_1) = \{O_1, O_3, O_4\}$ for the graph in the **Figure 1**. These

offenders are initial suspected offenders.

B) Finding Influential set of each spurious offender

After calculating the spurious set of defenders now let us consider another graph, say, Social Relationship Graph $SRG (V, E)$ which is actually a flower−following graph with n users (some of them are offenders of some rumors) and m edges where there is an edge $\{v_i v_j\}$ between two nodes v_i and v_j if and only if v_j follows v_i in social relationship.

Let us consider, for any rumor r_i, an equivalent set of offenders or crooks $C_i = \{v_i = O_i \mid \forall O_i \in O(r_i)\}$. Now considering each spurious defender as v_i as a seed set and $SRG (V, E)$ as a follower−following graph, we find the influence sets, $influence (v_i)$ for all $v_i \in C_i$ using Linear threshold method and Independent cascade method.

Influence Maximization (IM) Problem: Given a small seed set S and a follower−following network graph G (V, E) find a set that will maximize the information
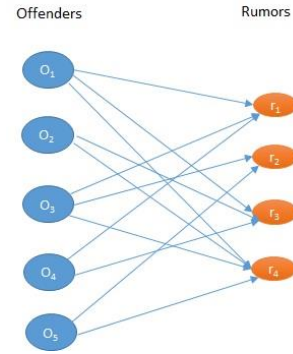
spread in the network.

Figure 1: Offender−Rumor Graph, ORG (n, q) with n = 5, q = 4.

Definition of Influence Set of a node v: The set of nodes that found by solving the IM problem taking the seed set to be $S = \{v\}$. It is also called the influence region of v.
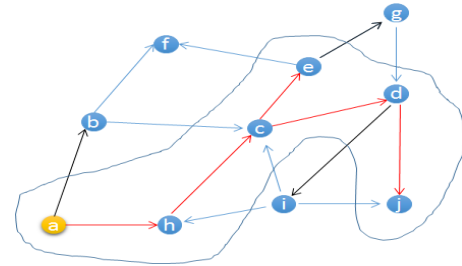

Figure 2: Finding Influence set. influence ({a}) = {c, d, e, h, j}.

C) Finding actual offenders

After finding the influence set for each node of the spurious offenders, we have checked whether there are any offenders in the influence set of other offenders. Since sometimes some nodes are flowers of some offenders and just passed the rumor to their followers only and they are not actual crooks. They just act as relay nodes of rumors. So these relay nodes should be removed from the spurious offenders list to get the offender set. Now the node with maximum influence region or k nodes with maximum influence region is considered to be actual (set of) offender (s) of the rumor. We perform similar calculation for each rumor r_i.

4. Evaluation

We performed simulation by a program in C on a machine (Core i7 2GHz, 2.5GHz, 4GB RAM, Windows 8) with synthesized data.

Nodes	50	60	70	80	90	100
Threshold	7.14	9.34	13.19	15.93	19.78	24.18
Cascade	9.89	13.74	18.13	23.63	28.57	35.71

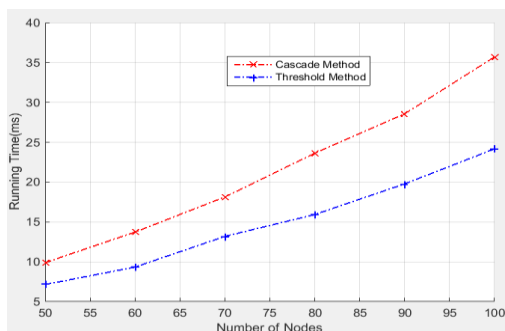**Table 1:** Running time comparison

Figure 3: Running time comparison

We have executed the both the algorithms (Linear threshold method and Independent cascade method [10]) except that seed set had only one node at a time and then we filtered the list by deducting relay nodes from the influence region. We have compared the running time (ms) of both the algorithms executing them hundred times and taking the average time. The result has been stated in the Table 1 and Figure 3. Here we have seen that the threshold method over performs the cascade method.

5. Conclusion

In this paper, we have seen that the Influence Maximization (IM) technique can be applied to find the source of rumor. Our main contribution in this paper is that to fit the IM problem to a different scenario of finding the source of rumor and filtering the relay nodes to find the possible source of rumor. There are other graph theoretic approaches but IM can also be a good candidate to solve this problem. The result would be more attractive if the algorithms would be run using real dataset.

Acknowledgement

6. References

[1]. J Kostka, Yvonne Anne Oswald, Roger Wattenhofer, *"Word of Mouth: Rumor Dissemination in Social Networks",* Structural Information and Communication Complexity. Springer Berlin Heidelberg, 2008. 185–196.

[2]. Eunsoo Seo, Prasant Mohapatra and Tarek Abdelzaher, *"Identifying Rumors and Their Sources in Social Networks",* Proceedings of SPIE, Volume8389, Ground/Air Multi-sensor Interoperability, Integration, and Networking for Persistent ISR III, 83891I (May 1, 2012); doi:10.1117/12.919823

[3]. Devavrat Shah and Tauhid Zaman, *"Rumors in a Network: Who's the Culprit?",* IEEE Transactions on Information Theory, Volume: 57 ,Issue: 8, 2011.

[4]. Wenxiang Dong, Wenyi Zhang and Chee Wei Tan, *"Rooting out the Rumor Culprit from Suspects",* IEEE International Symposium on Information Theory Proceedings (ISIT), 2013.

[5]. D.T. Nguyen, N.P. Nguyen, and M.T. Thai, *"Sources of Misinformation in Online Social Networks: Who to Suspect?,"* Proceedings of Military Communications Conference (MILCOM), 2012.

[6]. S. Li, Y. Zhu, D. Li, D. Kim, and H. Huang., *"Rumor restriction in online social networks,"* IEEE International Performance Computing and Communications Conference (IPCCC), 2013.

[7]. D. Wang, T. Abdelzaher, H. Ahmadi, J. Pasternack, D. Roth, M. Gupta, J. Han, O. Fatemieh, and H. Le. *"On bayesian interpretation of fact-finding in information networks",* In 14th International Conference on Information Fusion (Fusion 2011), 2011.

[8]. Goyal, F. Bonchi, and L. V. S. Lakshmanan, *"Discovering Leaders From Community Actions",* Pages 499–508, ACM, Proceeding of the 17th ACM Conference on Information and Knowledge Management, 2008.

[9]. Nguyen, H., and Zheng, R. *"On Budgeted Influence Maximization in Social Networks",* IEEE Journal on Selected Areas in Communications 31, 6 (2013), 1084–1094.

[10]. D. Kempe, J. Kleinberg, and E. Tardos. *"Maximizing the spread of influence through a social network",* In Proc. of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2003.