A Study of Rumor Control Strategies on Social Networks

Rudra M. Tripathy
Department of CS&E, I I T Delhi
New Delhi, India
tripathy@cse.iitd.ac.in

Amitabha Bagchi
Department of CS&E, I I T Delhi
New Delhi, India
bagchi@cse.iitd.ac.in

Sameep Mehta
IBM India Research Lab
New Delhi, India
sameepmehta@in.ibm.com

ABSTRACT

In this paper we study and evaluate rumor-like methods for combating the spread of rumors on a social network. We model rumor spread as a diffusion process on a network and suggest the use of an "anti-rumor" process similar to the rumor process. We study two natural models by which these anti-rumors may arise. The main metrics we study are the belief time, i.e., the duration for which a person believes the rumor to be true and point of decline, i.e., point after which anti-rumor process dominates the rumor process. We evaluate our methods by simulating rumor spread and anti-rumor spread on a data set derived from the social networking site Twitter and on a synthetic network generated according to the Watts and Strogatz model. We find that the lifetime of a rumor increases if the delay in detecting it increases, and the relationship is at least linear. Further our findings show that coupling the detection and anti-rumor strategy by embedding agents in the network, we call them beacons, is an effective means of fighting the spread of rumor, even if these beacons do not share information.

Categories and Subject Descriptors

K.4.2 [**COMPUTERS AND SOCIETY**]: Social Issues—*Abuse and crime involving computers*; H.2.8 [**INFORMATION SYSTEMS**]: Database applications—*Data mining*

General Terms

Design, Experimentation, Measurement

Keywords

Anti-rumor, Diffusion, Rumor, Social networks

1. INTRODUCTION

In social networks, rumors are mainly spread through media, internet or through relationships between individuals. One of the individuals plays the role of the spreader, or the individual who passes the rumor on by telling an acquaintance. The role of the listener is simply that of the individual who hears the rumor from the spreader and makes his/her decision about whether or not they believe it. There are several well studied models for rumor spread in social networks found in the literature: SIS model, SIR model [9], Daley Kendall (DK) SIS model, Maki-Thompson (MT) model [6], Voter model [1, 5], linear threshold model, independent cascade model [3] etc. Some of these models are primarily used for epidemic spread. In our study we have used variants of the independent cascade model for rumor spread. We have used this model because it is able to ensure the spread of the rumor through the entire network in a finite time, making it a robust adversary. If we can claim that our methods are able to combat this model then weaker models should be more easily contained.

As an aside we note that the problem of combating rumor spread is different from the problem of combating virus spread, although the two of them share the basic property that the contagion spreads from person to person through a form of contact. The fundamental difference is that rumor, which moves in the form of messages, can be combated by other messages, which we call *anti-rumors*, and these *anti-rumors can also be spread from person to person* unlike vaccines for viruses which can only be administered to individuals. In some sense this makes our problem more tractable but also it means it has to be studied with different ends in mind.

Rumor's potential for causing harm has led to a long history of efforts to understand it and devise mechanisms to control it (see e.g. [2]). The primary mechanism that governments and other authorities have used in the past is broadcasting messages to the entire population attempting to debunk the rumor (see e.g. [7] for examples from China and USA). Debunking rumor this way is a difficult exercise in situations where people perceive that the authority has a vested interest in debunking the rumor, for the maintenance of law and order, e.g., the case where Nigerians declared a mass boycott of Polio vaccination in 2003 [4]. Nigerian officials took around two years to control it by undertaking a variety of anti-rumor campaigns. Social networking sites present an even more complicated scenario for centralized strategies since they cross national borders. An authoritative source for one user may be an untrusted source from another user's point of view. In such a situation we believe that strategies that do not rely overly on centralized control are the only feasible strategies to follow and therefore in this paper we study a suite of such strategies. Our key

CIKM'10, October 26–30, 2010, Toronto, Ontario, Canada.
Copyright 2010 ACM 978-1-4503-0099-5/10/10 ...$10.00.

insight in studying anti-rumors in a decentralized setting is this: *The propagation of the anti-rumor does not depend primarily on the authoritativeness of the source that issues the anti-rumor but on the trust users place in their friends in the social network.*

The first strategy we study, the *Delayed Start Model*, models a situation where a local authority might discover a rumor n days after it starts and decide to spread an anti-rumor. In the second model, called the *Beacon Model*, we assume that the social network contains a set of vigilant agents, beacons, that are on the look out for the spread of rumors. Once a beacon receives a rumor it immediately starts spreading anti-rumors to combat the rumor. This strategy corresponds to a semi-centralized scenario where coalitions of authorities may proactively decide to seed the network with vigilant users who can both detect rumors and respond to them.

In our study we make one important simplifying assumption. We assume that a person who has accepted the falsity of the rumor once (i.e., has accepted the anti-rumor to be true) will never again believe the rumor.

2. MODELS AND METRICS

Basic Notation.

We assume that the online social network is modeled by a directed graph $\mathcal{G} = \{V, E\}$. For each node V_i, the immediate neighbors are represented by set N_i.

A variable S_i is maintained for each node, where $S_i \in \{0, 1, 2\}$. The nodes having $S_i = 0$ (yet to believe the rumor or anti-rumor) are called *Neutral nodes*. The nodes with $S_i = 1$ (believe the rumor) are called *Infected nodes*. These nodes, after believing the rumor, will spread the rumor in the network. The nodes having $S_i = 2$ (believe the anti-rumor) are classified into two groups, those who believe the anti-rumor before being infected (called *Vaccinated nodes*) and those who believe the anti-rumor after infected (called *cured nodes*). After believing the anti-rumor, these nodes (having $S_i = 2$) will spread the anti-rumor in the network.

$S_i^{(t)}$ denotes the value of S_i at time t. We assume that once a node V_i believes the anti-rumor ($S_i = 2$) never again believes the rumor. For each time t we maintain two variables $R(t)$ and $A(t)$, where $R(t) =$ Number of V_i s.t. $S_i = 1$ and $A(t) =$ Number of V_i s.t. $S_i = 2$. We denote T_i as the infected time of node V_i, i.e., the duration for which V_i remains infected. $T_i = \max\{t : S_i^{(t)} = 1\} - \min\{t : S_i^{(t)} = 1\}$.

2.1 Rumor spread model

We employ a method similar to the *Independent cascade model* as our baseline rumor spread model. In this model each infected node V_i at time t tries to infect each of its uninfected neighbors $w \in N_i$. It succeeds with probability p. Through the course of our experiments we have taken p to be 0.01. If it succeeds then w will become infected at time $t+1$. The process starts with 10 random infected nodes. For more details of the Independent cascade model, the readers are pointed to the article by Kempe et. al [3].

2.2 Anti-rumor spread models

Delayed Start Model.

Here we model the situation that an authority with limited jurisdiction detects the spread of rumor and then combats it by starting an independent cascade from a randomly selected infected node. We contend that there will always be a time lag between the start of rumor and its detection (and hence the start of the anti-rumor). This time lag is referred to as delay time and is represented by n. The process starts from a single infected node V_i, n time units after the rumor started, V_i spreads the anti-rumor messages to its neighbors N_i. Each node $w \in N_i$ accepts the anti-rumor with probability q. Through the course of our experiments we have taken q to be 0.05.

Beacon Model.

Between the time an authority detects the spread of rumor and decides how to combat it, the rumor continues spreading apace. In order to proactively combat rumors, authorities may embed agents in the network that are capable of detecting the spread of rumor and are authorized to start spreading anti-rumors as soon as they detect the spread of rumor. We call these agents *beacons*. In this paper, the beacon node uses the same mechanism as the Delayed start model to spread anti rumor, i.e., it spreads the anti-rumor to all its neighbors with some probability. In our experiments the beacon nodes are selected at random. However, in real networks, the nodes can be selected based on multiple attributes like connectivity, authority, trust etc. Moreover, beacon nodes can also be topic specific. For example, one node may act as a beacon for technology based rumors but not for entertainment based rumors. This selection will involve topic and expertize mining from the network. The Beacon model with one beacon is comparable to the Delayed start model. In the Delayed start model, the starting time of the anti-rumor process is fixed but here it depends upon the time when the beacon is activated.

2.3 Metrics

In the previous section, we discussed different models to spread anti-rumors for combating rumor in social networks. To evaluate the efficacy of these models, we propose the following three metrics: Maximum infected time ($M(G)$), Average infected time ($A(G)$) and Point of decline ($P(G)$).

The *Maximum infected time* $M(G)$ measures the life time of rumor in the network, i.e., it measures the maximum duration for which any node continues to believe the rumor. Mathematically it is defined as:

$$M(G) = \max\{T_i | \text{for all } V_i \in V[G]\}$$

where $V[G]$ stands for the vertices set of graph G. If the maximum infected time is finite, i.e., $M(G) < \infty$, then eventually the rumor will die out.

The second metric *Average infected time* $A(G)$ finds the average time for users continue to believe the rumor to be true. Therefore it shows the overall performance of anti-rumor spread model in the entire network. Mathematically it is defined as:

$$A(G) = \frac{1}{|V[G]|} \sum_{V_i \in V[G]} T_i$$

The third metric *Point of decline* $P(G)$ shows the inflection point of rumor growth, i.e., the point at which the number of users believing the rumor starts declining. Mathematically we can define this as:

$$P(G) = \min\{t | R(t) > R(t+1)\}$$

3. EXPERIMENTS

In this paper, our main goal is to study the performance of our anti-rumor models for combating rumor spread. We have tested our models on real (derived from the Twitter website) as well as synthetic data set (generated using Watts and Strogatz model [8]). We crawled Twitter using the API provided by Twitter, the final data set contains 49,965 nodes and 10,54,243 edges. The synthetic data contains 50,000 nodes (nearly the same size as that of the Twitter data) and having 22 edges (the average degree of the Twitter data) per node on an average and re-writing probability 0.4. In this section we discuss the results obtained from the above anti-rumor spread models for combating rumors.

3.1 Delayed Start Model

The results for both the Twitter data and the synthetic data are shown in Figure-1. The values shown are the average values over 20 iterations. It is evident in Figure-1 that the $M(G)$ values for the synthetic data are quite small as compared to the Twitter data. This is because the diameter of the synthetic graph is small and clustering coefficient is very high compared to the Twitter graph. The diameter for the Twitter graph is 15 whereas the diameter for the synthetic graph is exactly 6. Therefore, the anti-rumor message in the synthetic graph can reach all the nodes faster than in the Twitter graph. We also observed that, in the Twitter graph there are several nodes having in-degree 1 (Figure is omitted due to space constraint). So, these nodes have a lower probability of taking up the anti-rumor. Therefore, from these experimental results we can say that, in real social networking sites rumor persists for a longer period of time than in the theoretical model. Also we note that in both cases the speed of rumor spread increases as the delay time increases. Hence the role of authority is very crucial because if it can identify the presence of rumor in the network quickly then the maximum infected time can be significantly reduced.

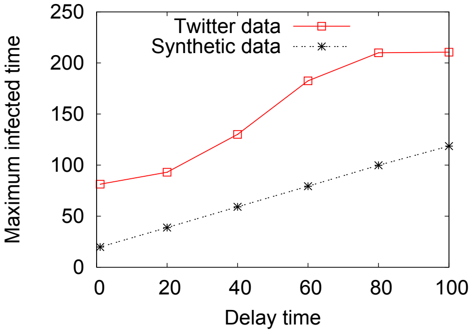


Figure 1: Maximum infected time: Delayed Start Model

The second metric is the average infected time ($A(G)$). The results for average infected time are shown in Figure-2. Unlike the maximum infected time the average infected time for the Twitter data (Figure-2) increases almost linearly with delay time. Comparing the $M(G)$ values with $A(G)$ (Figure-1 and 2) values we have observed that, in the Twitter data, the difference between $M(G)$ and $A(G)$ values is very high compared to the synthetic data. Therefore, there are very few nodes (particularly lower in-degree nodes) in the Twitter data that remain infected for longer time but in the synthetic data all nodes remain infected for similar time period. Comparing the result of the synthetic data with the result obtained using the Twitter data (Figure-2), one can observe that the average infected time is very low compared to the Twitter data, since the anti-rumor growth process is much faster in small world graphs and offsets the rumor growth quickly.

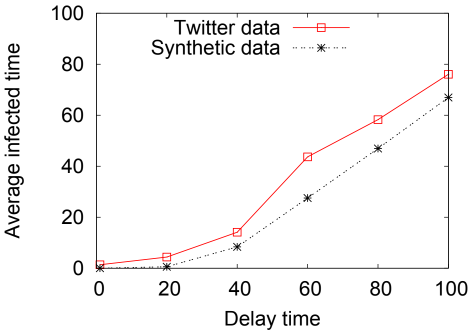
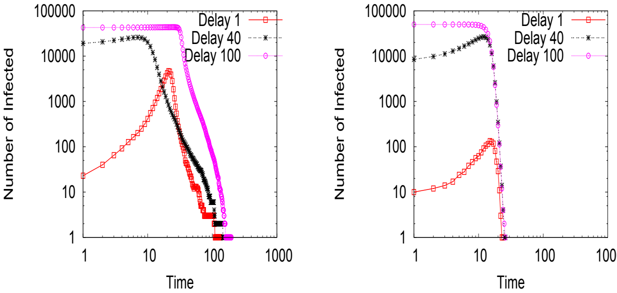


Figure 2: Average infected time: Delayed Start Model

Next, we study the spreading behavior of rumor after the anti-rumor process started. The results for the Delayed start model for both form of the graphs are shown in Figure-3. Let us first look at the results for the Twitter data. In Figure-3(a), we can see that even though the anti-rumor process has already started, the number of infected nodes $R(t)$ still increases with time t. But after a certain time the values of $R(t)$ declines very fast. We can think of this as a game between rumor and anti-rumor process. For some time the rumor process wins ($R(t) \leq R(t+1)$) over the anti-rumor process and but the moment anti-rumor process wins over the rumor process ($R(t) > R(t+1)$), it removes all the infected nodes very quickly. Recall that in our experiments we assume that a node who knows about the anti-rumor never accepts the rumor again. Therefore after the decline starts, eventually there is only one process (anti-rumor process) which is active. Similar observations can be made for the synthetic graph but the growth of anti-rumor spread is faster for the synthetic graph.



(a) Twitter data (b) Synthetic data

Figure 3: Decline process: Delayed Start Model

3.2 Beacon Model

In this section we discuss the efficacy of the Beacon model and also compare it's results to that of the Delayed start model. The maximum infected time for the Twitter data and the synthetic data are shown in Figure-4. For both

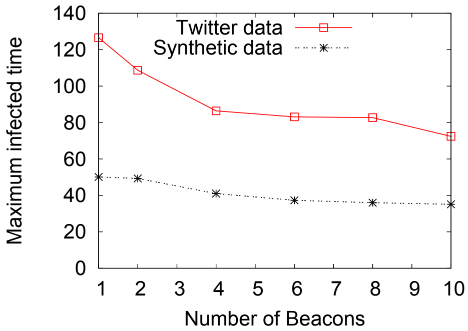


Figure 4: Maximum infected time: Beacon Model

the Twitter data and the synthetic data, we can see that maximum infected time decreases as the number of beacons increases (Figure-4). But the maximum infected times in the synthetic data are lower than the Twitter data. This happens due to the small diameter of the synthetic graph; the beacons are activated much earlier in the synthetic graph. To compare the Beacon model to the Delayed start model, we consider a single Beacon model. First we try to learn the time when the beacon is activated. By repeated runs, we found that the time on which beacon actives is lies between 20 and 35. Therefore, it makes sense to compare the results of the single Beacon model to the Delayed start model with delay time between 20 to 35. Now let us closely look at Figure-4 (single beacon) and Figure-1 (delay time between 20 to 35). We can see that for the Twitter data, the maximum infected time for the Delayed start model is more than the Beacon model.

Now let us consider the second metric, average infected time ($A(G)$), for both the Twitter and the synthetic data. The results are shown in Figure-5. In Figure-5, we can see that the $A(G)$ values for the Twitter data decrease exponentially as the number of beacons increase. Comparing the results of Delayed start model (Figure-2) with single Beacon model (Figure-5) we found that, for the Twitter data, the average infected time of the Delayed start model is high. Similar results are obtained for the synthetic data.

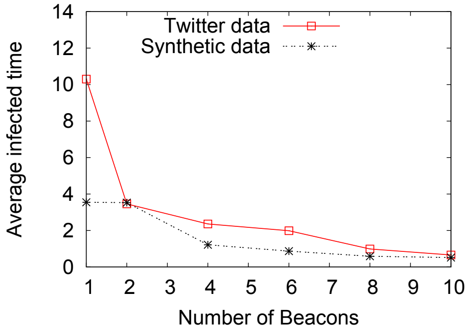
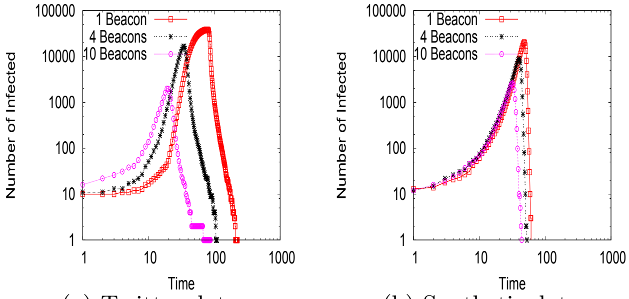


Figure 5: Average infected time: Beacon Model

Next we study the decline process of rumor spread. The results for the synthetic data and the Twitter data are shown in Figure-6. Let us first look at the plots for the Twitter data (Figure-6(a)). Initially, the number of infected nodes $R(t)$ increases with time t but after certain point (when the anti-rumor process wins over the rumor process) it starts decline in a faster rate. Also note that as the number of beacons increase the time required to remove the rumor (completely) increase. Almost similar observations can be made from the synthetic data in Figure-6(b). In our dataset we have found that, it takes more time to remove the rumor from real data as compared to the synthetic data. This is similar to the observations we made for the Delayed start model.



(a) Twitter data (b) Synthetic data

Figure 6: Decline process: Beacon Model

The main contribution of this paper is to study rumor-like strategies of fighting the spread of rumor. We have studied a reactive situation where there is a time lag in the detection of rumor and a local authority's attempt to stop the rumor by starting an anti-rumor. We found that the time lag is an important parameter. The lifetime of the rumor grows at least linearly with the delay in detection. We also studied a proactive situation where beacons embedded in the network detect and fight rumor and found that this is an effective means of fighting the spread of rumor.

4. REFERENCES

[1] C. Castellano, D. Vilone, and A. Vespignani. Incomplete ordering of the voter model on small-world networks. *Europhys. Lett.*, 63(1):153, 2003.

[2] P. Donovan. How idle is idle talk? one hundred years of rumor research. *Diogenes*, 213:59–82, 2007.

[3] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the* 9^{th} *ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

[4] Merck-Article. Polio vaccination fears, 2003. Science in Africa: Africa's First On-Line Science Magazine.

[5] C. M. Mizell and L. M. Sander. A generalized voter model on complex networks. *J. Statist. Phys.*, 136(1):59–71, 2009.

[6] R. Thompson, R. C. Estrada, D. Daugherty, and A. Clintron-Arias. A deterministic approach to the spread of rumors. *Technical report, Mathematical and Theoretical Biology*, 2003.

[7] R. Turner. *Disasters, Collective Behavior and Social Organization*, chapter Rumors as intensified information seeking: Earthquake rumors in China and the United States, pages 244–256. University of Delaware Press, 1994.

[8] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, June 1998.

[9] D. H. Zanette. Dynamics of rumor propagation on small-world networks. *Phys. Rev. E*, 65(4), Mar 2002.