Identifying Rumors and Their Sources in Social Networks

Eunsoo Seo[a], Prasant Mohapatra[b] and Tarek Abdelzaher[a]

[a]University of Illinois at Urbana-Champaign, Urbana, IL, USA;
[b]University of California at Davis, Davis, CA, USA

ABSTRACT

Information that propagates through social networks can carry a lot of false claims. For example, rumors on certain topics can propagate rapidly leading to a large number of nodes reporting the same (incorrect) observations. In this paper, we describe an approach for finding the rumor source and assessing the likelihood that a piece of information is in fact a rumor, in the absence of data provenance information. We model the social network as a directed graph, where vertices represent individuals and directed edges represent information flow (e.g., who follows whom on Twitter). A number of monitor nodes are injected into the network whose job is to report data they receive. Our algorithm identifies rumors and their sources by observing which of the monitors received the given piece of information and which did not. We show that, with a sufficient number of monitor nodes, it is possible to recognize most rumors and their sources with high accuracy.

Keywords: Social Networks, Rumor Spreading, Epidemics

1. INTRODUCTION

Social networks are popular media for sharing information. Online social networks enable large-scale information dissemination in a very short time, often not matched by traditional media.[1,2] Mis-information and false claims can also propagate rapidly through social networks. This is exacerbated by the fact that (i) anyone can publish (incorrect) information and (ii) it is hard to tell who the original source of the information is.[3]

In this paper, we focus on two problems related to mitigation of false claims in social networks. First, we study the question of identifying sources of rumors in the absence of complete provenance information about rumor propagation. Second, we study how rumors (false claims) and non-rumors (true information) can be differentiated. Our method is based on an assumption that rumors are initiated from only a small number of sources, whereas truthful information can be observed and originated by a large number of unrelated individuals concurrently. Our approach relies on utilizing network *monitors*; individuals who agree to let us know whether or not they heard a particular piece of information (from their social neighborhood), although do not agree to let us know who told them this information or when they learned it. Hence, all we know is which of the monitors heard a particular piece of information. This, in some sense, is the most challenging scenario that offers a worst-case bound on accuracy of rumor detection and source identification. Additional information can only simplify the problem. We show, that even in the aforementioned worst case, promissing results can be achieved.

The rest of the paper is organized as follows. Section 2 describes the problem of identifying rumors and their sources and explains our approach. We also explain various ways of selecting monitors. Section 3 presents a case study of a real social network crawled from Twitter. We describe the details of the dataset and compare different monitor selection methods in terms of accuracy of rumor and source identification. In Section 4, we provide related work and conclude the paper in Section 5.

Further author information: (Send correspondence to Eunsoo Seo.)
Eunsoo Seo: E-mail: eseo2@cs.illinois.edu, Telephone: 1 217 265 6793
Prasant Mohapatra: E-mail: prasant@cs.ucdavis.edu, Telephone: 1 530 754 8016
Tarek Abdelzaher: E-mail: zaher@cs.illinois.edu, Telephone: 1 217 265 6793

2. IDENTIFYING RUMORS AND THEIR SOURCES

2.1 Identifying Rumor Sources

The first question we are studying in this paper is as follows: if a rumor is initiated by a *single* source in a social network, how can the source be identified?

A social network is modeled as a directed graph $G = (V, E)$ where V is the set of all people and E is the set of edges where each edge represents information flow between two individuals. We assume that a set of k pre-selected nodes M ($M \subseteq V$) are our monitors. For rumor investigation purposes, given a specific piece of information, a monitor reports whether they received it or not. We denote the set of monitor nodes who received the rumor by M^+, and the set of monitor nodes who have not received it by M^- (where $M^+, M^- \subseteq M$). We call the former set *positive* monitors and the latter *negative* monitors.

To identify the souce of a rumor, we use the intuition that the source must be close to the positive monitors but far from the negative monitors. Hence, for each node x, our algorithm calculates the following four metrics:

(1) Reachability to all positive monitors We first calculate how many positive monitors are reachable from each node. For a node x to be the rumor source, x must have paths to all monitors in M^+. If those paths do not exist, x cannot be a rumor source.

(2) Distance to positive monitors Among those nodes that can reach all positive monitors, nodes that are closer, on average, are preferred. In other words, for each node x, we calculate the total distance

$$\sum_{m \in M^+ \text{ and } m \text{ is reachable from } x} d(x, m).$$

where $d(x, m)$ is the distance from x to m, and sort the suspected sources by increasing total distance from positive monitors.

(3) Reachability to negative monitors Among the nodes that can reach all nodes in M^+ and have the same total distance to these positive monitors, we use reachability to negative monitors as a third metric. For each such node x, we count of monitors in M^- that are not reachable from x and prefer larger counts.

(4) Distance to negative monitors As a last metric, we also use the distance to negative monitors. For each node x, we calculate the total distance

$$\sum_{m \in M^- \text{ and } m \text{ is reachable from } x} d(x, m).$$

It is more natural that negative monitors are far from the rumor source, so nodes with large values of total distance are preferred.

Using the above four metrics – number of reachable positive monitors, sum of distances to reachable positive monitors, number of reachable negative monitors, sum of distances to reachable negative monitors –, all nodes in the network are sorted lexicographically. That is, i-th metric is used only when there is a tie in all metrics before it. Note that, for the first and last metric, large numbers are preferred while small numbers are preferred for the second and third. Our implementation converts the sign of first and fourth metrics to make sorting easy.

In the sorted list, the top suspect is the first node. For best accuracy, it is important to choose monitors wisely. In this paper, we compare the following six monitor selection methods.

(1) Random Random selection method selects k monitors randomly. This means that, for any node $x \in V$, the probability that x is selected as a monitor is $\frac{k}{|V|}$.

(2) Inter-Monitor Distance (Dist) This requires any pair of monitors to be at least d hops away. To do that, it first randomly shuffles the list of all nodes. Then, from the first node, it checks whether it is at least d hops away from all already-selected monitors. If it is, the node is selected as a monitor and the next node is checked. Note that the first node is always selected as a monitor. This is repeated until k monitors are selected or it is impossible to select any more monitors. Dist selection method finds the largest d which can choose k monitors. To do that, it starts with a large value of d and decrements it every time it fails to choose k monitors, starting over with the smaller d until it can find k monitors.

(3) Number of Incoming Edges (NI) In this method, the number of incoming edges of each node is counted. Then, the top k nodes which have largest counts are chosen.

(4) NI+Dist This method combines NI and Dist. Nodes with a large number of incoming edges are preferred as monitors, but the algorithm also considers inter-monitor distance. To do that, it first sorts all nodes in the descending order by the number of incoming edges of each node, then Dist is used to choose monitors. In other words, nodes in the sorted list are examined one by one and a node is chosen as a monitor if it is at least d hops away from all previously selected monitors. As in Dist, this method finds the largest d which can choose k monitors.

(5) Betweenness Centrality (BC) This method calculates betweenness centrality[4] for each node v, which is defined as

$$C(v) = \sum_{s \neq t \neq v \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the number of shortest paths from s to t and $\sigma_{st}(v)$ is the number of shortest paths from s to t that pass through v. Then, the k nodes which have the largest betweenness centrality are chosen as monitors.

(6) BC+Dist This method combines BC and Dist. Nodes are first sorted by their betweenness centrality, then Dist is used to choose monitors.

The above six monitor selection algorithms produce different sets of monitors, which result in different accuracy in source finding. Section 3.2 compares these algorithms in detail.

2.2 Identifying Rumors

Given a piece of information, can we determine whether it is a rumor (false claim) or not using the set of monitors that received the information? We use the following two metrics for this classification.

2.2.1 Greedy Sources Set Size (GSSS)

If a rumor is initiated by some person intentionally, it is not independently corroborated by others. Hence, in the absence of collusion, there is only one source of the rumor in the network. If a rumor is initiated by a small colluding group of people, the number of independent sources is just the size of the group. Conversely, if a piece of information is not a rumor, there may be many independent sources of the information. Therefore, it is important to estimate the number of independent sources correctly.

A set of nodes C is a valid source set if the following is satisfied: for all $m \in M^+$, there exists $n \in C$ which satisfies $d(n, m) \neq \infty$. The question is, what is the minimum size of set C? Instead of calculating the exact solution with an exponential algorithm*, we use the greedy approximation algorithm in Figure 1 to get an approximate minimal source set. The algorithm calculates the set of candidate sources (C). For each candidate source x, it also calculates the set of positive monitors (P_x) covered by x that are used in Section 2.2.2.

*For each node m in M^+, we define S_m the set of nodes which have a path to m. Then calculating the minimal C is exactly same as the minimal hitting set problem among $\{S_m\}_{m \in H}$.

$C \leftarrow \{\}$
For each $m \in V$, $P_m \leftarrow \{\}$.
For each $m \in M^+$, $S_m \leftarrow$ the set of nodes which have a path to m.
while $M^+ \neq \{\}$ **do**
  Let x be one of the most frequent elements in all S_m's where $m \in M^+$.
  Add x to C.
  For each $m \in M^+$, add m to P_x and remove m from M^+ if $d(x, m) \neq \infty$.
end while

Figure 1. A Greedy algorithm for calculating an approximate minimal source set.

Initially, C and P_m for all $m \in V$ are initialized to empty sets. At each iteration, a node x which can reach the largest number of elements in M^+ is chosen as a source. Node x is considered as one of the candidate sources and all monitors in M^+ that are reachable from x are assumed to have received the information from x. Then x is added to C. The reachable monitors are removed from M^+ and put into P_x. In the final state, C becomes the Greedy Source Set (GSS), which is an approximate minimal source set. For each node $x \in C$, P_x is the set of monitors that are expected to receive the information from x.

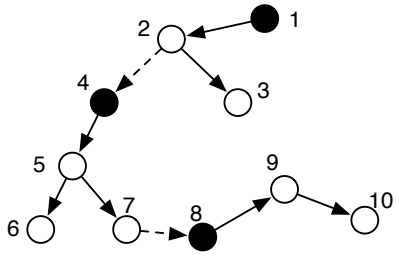

Figure 2. Greedy source selection overestimates rumor propagation distance.

2.2.2 Maximal Distance of Greedy Information Propagation (MDGIP)

The previous greedy algorithm tries to assign as many positive monitors as possible to each source, so the resulting greedy information propagation trees tend to become larger than the real ones. Figure 2 shows an example with three original source nodes (black circles). Information from the sources propagates along the solid arrows. Note that, the dotted edges are *not* used for the actual rumor propagation. Suppose all nodes (black or white) are monitors. The greedy approach in Section 2.2.1 finds that it is possible to cover all positive monitors using only one source node. This means that $C = \{1\}$ and $P_1 = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$. As a result, a new large propagation tree is generated instead of the actual three small trees.

To estimate possible disparty between the actual propagation tree and the one constructed by the above greedy algorithm, we use a second metric. Namely, given a greedy source set C and the set of nodes that receive information from x (P_x) for all $x \in C$, we define Maximal Distance of Greedy Information Propagation (MDGIP), calculated as:

$$\max_{x \in C, y \in P_x} d(x, y)$$

where $d(x, y)$ is the distance from x to y. Note that, when there are many actual sources, the estimated MDGIP tends to become large since many small propagation trees are combined into one greedy propagation tree.

In this section, we suggested two metrics: Greedy Source Set Size (GSSS) and Maximal Distance of Greedy Information Propagation (MDGIP). Both metrics increase as the number of actual sources increases. In the following section, we discuss how they can be used for rumor classification using actual social network data.

3. CASE STUDY

3.1 Data set

To apply our algorithm to a real social network, we extracted a social graph from Twitter. First, we obtained 159271 tweets written in Dec. 2011 containing a special keyword[†]. These tweets were written by 39567 twitter accounts.

In Twitter, tweets are propagated by retweets. When a user y retweets another user x's tweet (or retweet), we assume that there is an edge from x to y. In total, we obtained 102796 edges from the crawled data. The undirected version of this graph has 9243 connected components. In this evaluation, we focus on the largest connected component G which has 30146 nodes and 102608 edges. Maximum in-degree and out-degree among all nodes in G is 193 and 2264, respectively. The undirected version of G has a diameter of 12 hops.

Besides the topology, we also calculated Retweet probability of each edge $x \to y$ as the ratio of "x's tweets retweeted by y" to "all tweets of x." Calculated Retweet probabilities were used to simulate random propagation of rumors.

3.2 Finding Rumor Sources

We first evaluate the accuracy of our approach in finding rumor sources. To simulate random rumor propagation, we did the following: (1) A random rumor source is selected, (2) Using retweet probability of each edge, the rumor is propagated, (3) if the rumor does not reach more than 1% of all nodes, it is considered as a negligible rumor, rumor propagation result is discarded, a new rumor source is selected and the same procedures are repeated. For each rumor propagation, we used our rumor source identification algorithm using different number of monitors: $20, 40, 80, \cdots, 5120$ and we repeated this 200 times.

3.2.1 Rank of True Source

Using the method presented in Section 2.1, all nodes are sorted in the likelihood that they are the actual rumor source. Figure 3 shows the average rank of the actual source in the output. In the ideal case, the rank should be one which means that the top suspect is actually the rumor source. Note that, regardless of the monitor selection method, the rank of the true source generally decreases (i.e., improves by becoming closer to 1) as the number of monitors increases. Dist and NI+Dist generally show a bad accuracy. Random also performs poorly when the number of monitors is small, but it improves as more monitors are added. NI, BC and BC+Dist show better performance than the others. When the number of monitors is very large, the choice of monitor selection does not matter that much anymore, and all algorithms converge.

One of the important factors that affects the accuracy of rumor source identification is the number of positive monitors. Figure 4 shows the ratio of experiments in which no monitor received the rumor. In all monitor selection methods, the ratio decreases as the number of monitors increases. Among the four methods compared, the Dist selection method has the highest ratio. Dist basically maximizes inter-monitor distance, so it tends to choose nodes on the boundary of the graph. Therefore, monitors selected by Dist have low probability of hearing rumors. The Random selection method also has a high ratio of negative monitors when the number of monitors is small. The other methods (NI, NI+Dist, BC, BC+Dist) have small ratio compared to Dist and Random. When no monitor hears the rumor, it is very hard to find the source accurately as shown in Figure 3 (Random and Dist when the number of monitors is 20, for example).

However, a larger number of positive monitors does not always lead to a more accurate result. Figure 5 shows the average number of positive monitors when the number of monitors is 160. Figure 3 shows that BC has best accuracy, NI has second best, and the others are worse. Note that, the number of positive monitors in NI is almost double of that in BC, but BC is more accurate. This means that it is not always helpful to have more positive monitors.

[†]The actual used keyword is "Kim, Geuntae (in Korean)," a Korean politician who died in Dec. 2011. Instead of using Twitter API directly, we downloaded already-crawled tweets from Tweetrend.com. It is a third-party twitter web site which shows the trend of popular keywords and the actual tweets in Korean.

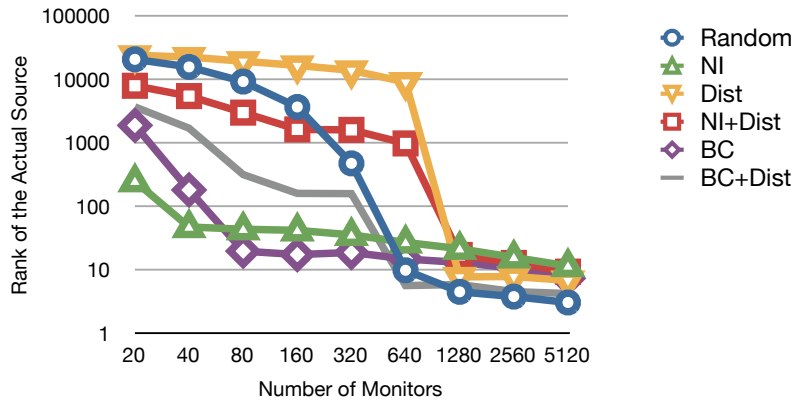

Figure 3. Average Rank of Actual Source in the output (out of 30146 nodes)

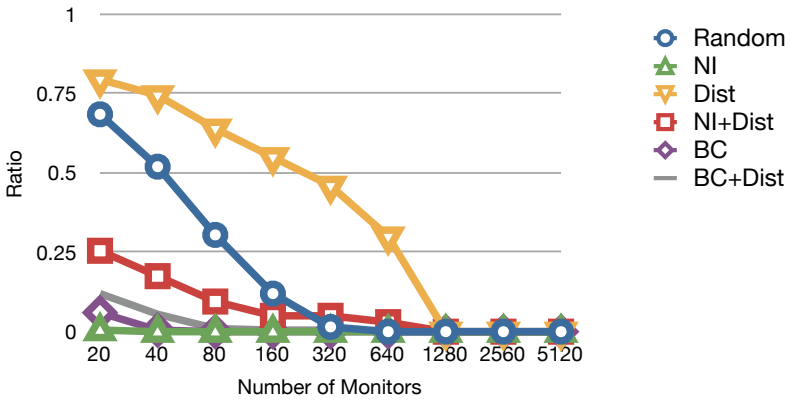

Figure 4. Ratio of experiments in which no monitor receives a rumor (out of 200 experiments)

3.2.2 Distance from Top Suspect to the True Source

Figure 6 shows the distance[‡] between the top suspect and the actual source. In the ideal case, the distance should be zero, meaning the the top suspect is the source. Figure 6 shows a similar tendency as Figure 3. The distance decreases as more monitors are added. Dist shows the largest distance of all monitor selection methods. Random has large distances with a small number of monitors, but the distance decreases drastically as the number of monitors increase. BC and BC+Dist generally show the smallest distance between the top suspect and the actual source.

3.3 Identifying Rumors

As discussed in Section 2.2, rumors are usually initiated by a small number of people. In contrast, true information can be reported by many people independently. In this evaluation, we assumed that rumors have a small number (1 or 10) of sources and non-rumors have a large number (100 or 1000) of sources and show how their propagation features are different.

For each number of sources (1, 10, 100 and 1000), monitor selection method, and number of monitors (20, 40, 80, 160, 320 and 640), we repeated rumor identification 200 times and used the results for rumor identification only when there is at least one monitor that heard the rumor.

[‡]This distance is calculated over the undirected version of the graph.

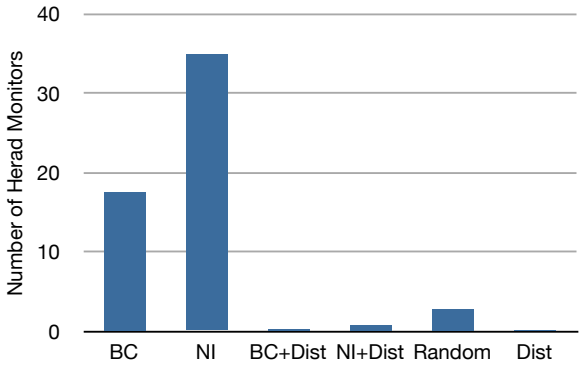

Figure 5. Number of Positive Monitors (Number of Monitors: 160)

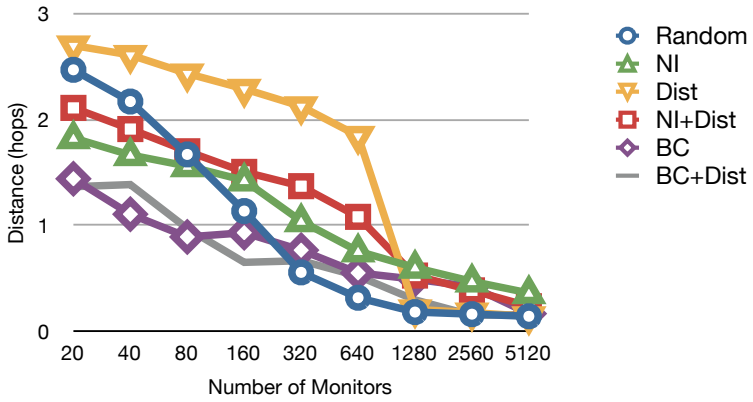

Figure 6. Average Distance between the top suspect and the actual source

Dist Figure 7 shows average GSSS and MDGIP when Dist monitor selection method is used. Left figure shows that, as the number of real sources increases, GSSS also increases. Right figure also shows that, as the number of real sources increases, MDGIP also increases. These two graphs show that GSSS and MDGIP can be used to classify a piece of information as a rumor or not since it is directly related to the number of sources.

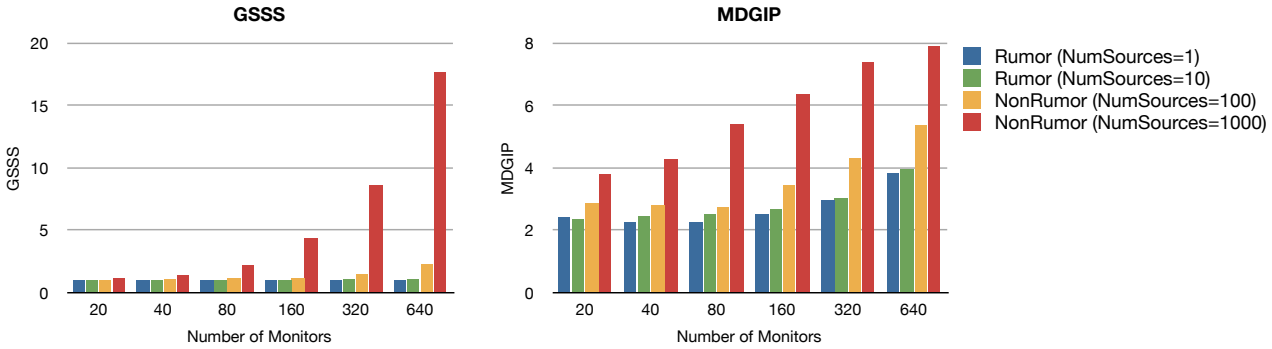

Figure 7. GSSS and MDGIP (Monitor Selection: Dist)

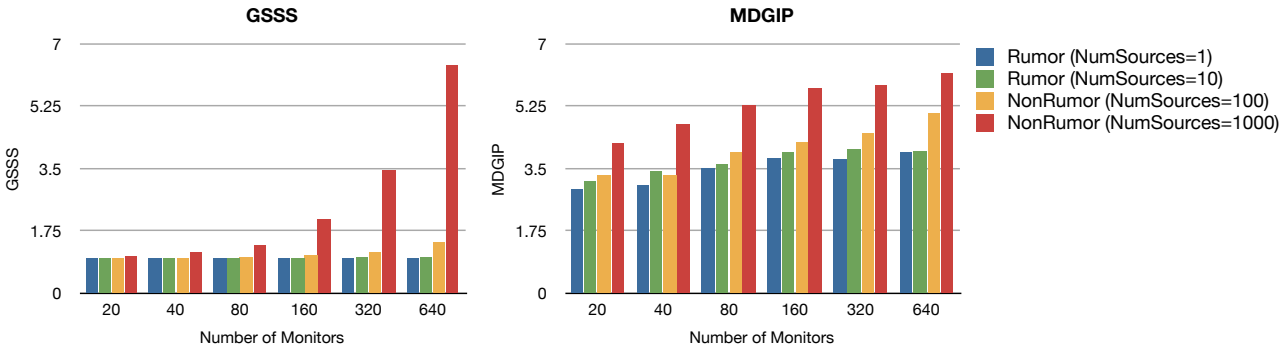

Figure 8. GSSS and MDGIP (Monitor Selection: Random)

Random Figure 8 shows the results from Random monitor section. Overall, Figure 8 looks similar to Figure 7, but the difference of GSSS and MDGIP values with different number of sources is smaller.

NI Figure 9 shows the results from NI monitor selection. Contrary to the previous two monitor selection methods, it shows that GSSS does not change over different number of monitors or different number of sources. In the experiments, GSSS is alway one, which means that only one candidate source can cover all positive monitors. It also shows that MDGIP and the number of sources do not have strictly consistent relation. Overall, GSSS and MDGIP for the NI monitor selection algorithm do not give much information for rumor identification.

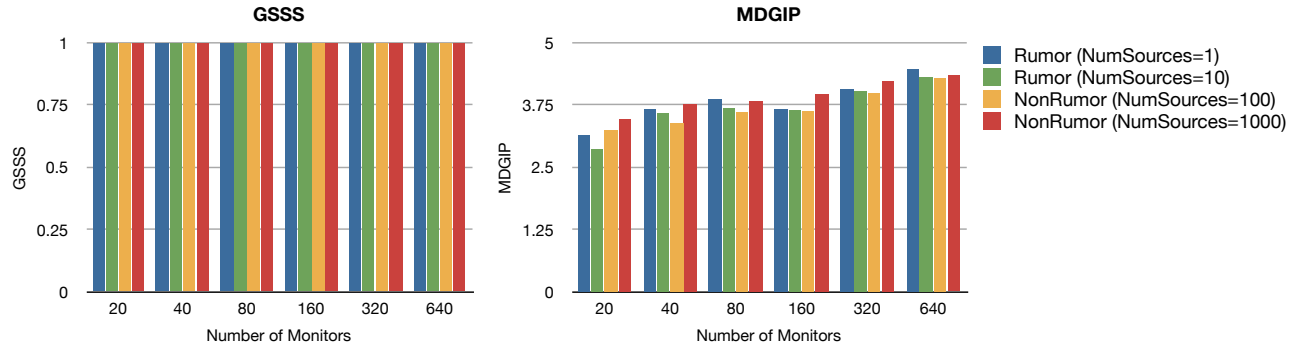

Figure 9. GSSS and MDGIP (Monitor Selection: NI)

NI+Dist Figure 10 shows the results from NI+Dist monitor selection. It shows the overall tendency that GSSS and MDGIP increases with the number of sources.

BC Figure 11 shows the results from BC monitor selection. Similar to NI, GSSS does not change and MDGIP shows inconsistent values with different number of sources.

BC+Dist Figure 12 shows the results from BC+Dist monitor selection. Similar to Dist, Random and NI+Dist, it shows the overall tendency that GSSS and MDGIP increase with the number of sources.

3.3.1 Classification

Previously, we have shown that GSSS and MDGIP can be used to classify a piece of information as rumor or non-rumor. Figure 13 visualizes GSSS/MDGIP and rumor classification in more detail. Four figures compare

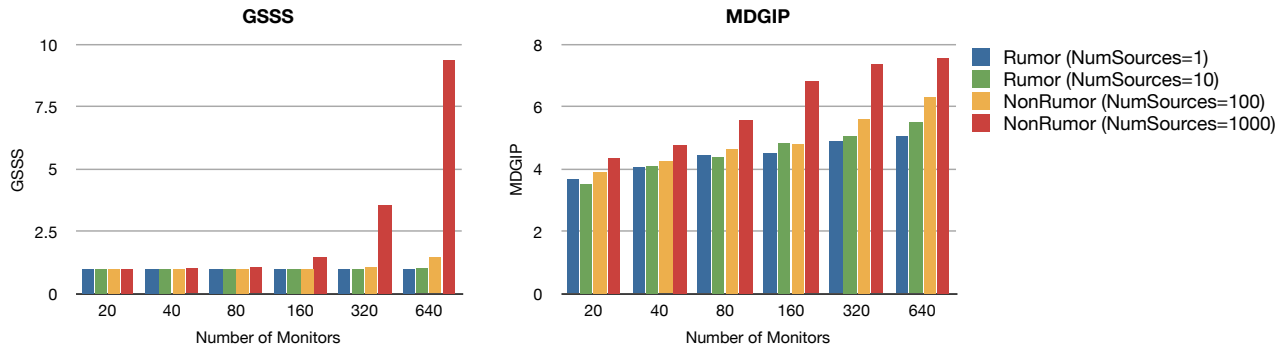

Figure 10. GSSS and MDGIP (Monitor Selection: NI+Dist)

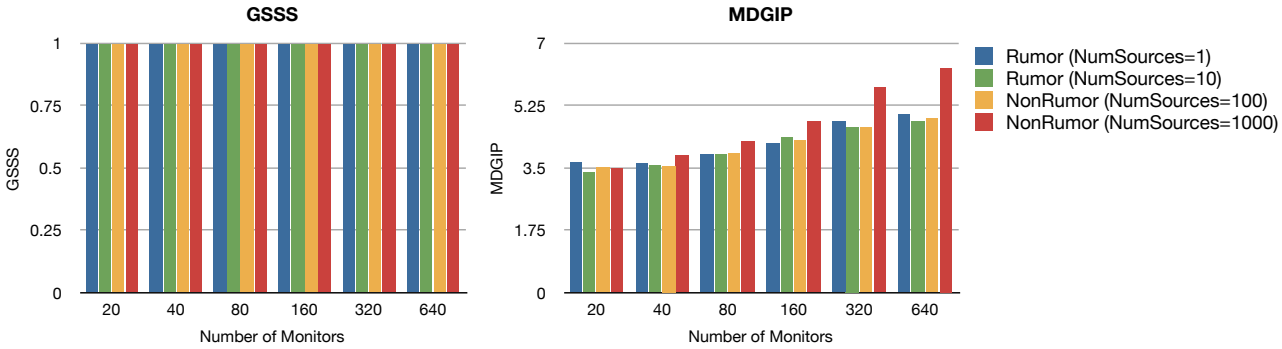

Figure 11. GSSS and MDGIP (Monitor Selection: BC)

the cases in which rumors have 1 or 10 sources and non-rumors have 100 or 1000 sources when Dist monitor selection algorithm is used and the number of monitors is 640. When there is a large difference in the number of sources of rumors and non-rumors (two right figures), it is shown that rumors and non-rumors are clearly separated. However, when the difference is smaller (two left figures), rumors and non-rumors overlap in some cases which causes inaccuracy in classification.

Using the experimental data, we evaluated how accurately logistic regression can classify rumor and non-rumors. We used the first half of experimental data as training set and the second half as test set. Table 1 summarizes the results when the number of monitors is 640. The first column of Table 1 shows the monitor selection algorithm. The second and third columns show the number of sources of a rumor and a non-rumor. Next three columns (4–6) show the classification results of 100 rumors. They are classified as rumors (True Positive, Column 4) or non-rumors (False Negative, Column 5). If no monitor hears the rumor, our classification algorithm cannot work, so it cannot be classified (Column 6). Similarly, last three columns (7–9) show the classification results of 100 non-rumors. They are classified as non-rumors (True Negative, Column 7) or rumors (False Positive, Column 8). If no monitor hears anything, it cannot be classified (Column 9).

From the table, we can observe that, if rumors and non-rumors have very large difference in the number of sources, rumor classification can be done with very high accuracy. As the difference in the number of sources of rumors and non-rumors decreases, it gets harder to classify rumors and non-rumors accurately.

Another observation from the table is that the algorithms which show good results in rumor source identification (BC and BC+Dist) do not always work well in rumor classification. This is because the two tasks have different conditions for best performance. In rumor source identification, it is best to have monitors near the rumor source, so that they receive the rumor and estimate the rumor source based on their locations. In rumor classification, it is best to have monitors in various places in the rumor propagation trees so that GSSS and MDGIP can be estimated accurately. We leave finding a monitor selection algorithm that is good for both tasks as future work.

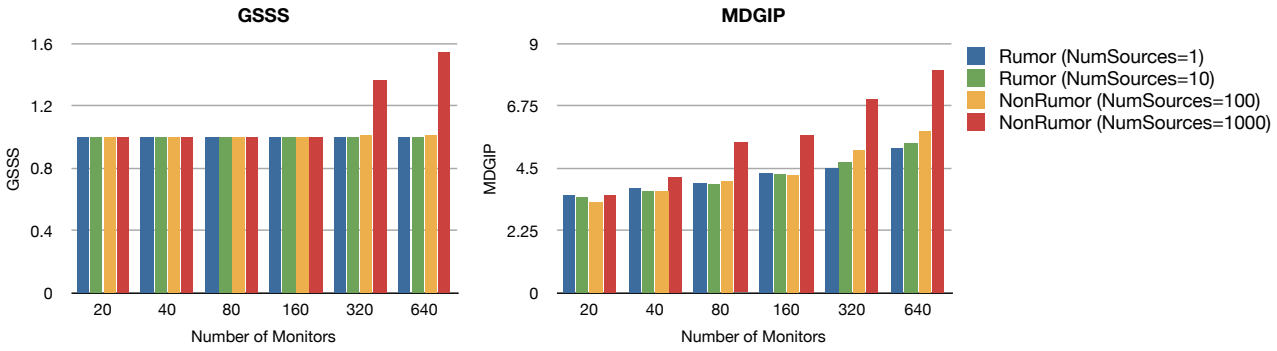

Figure 12. GSSS and MDGIP (Monitor Selection: BC+Dist)

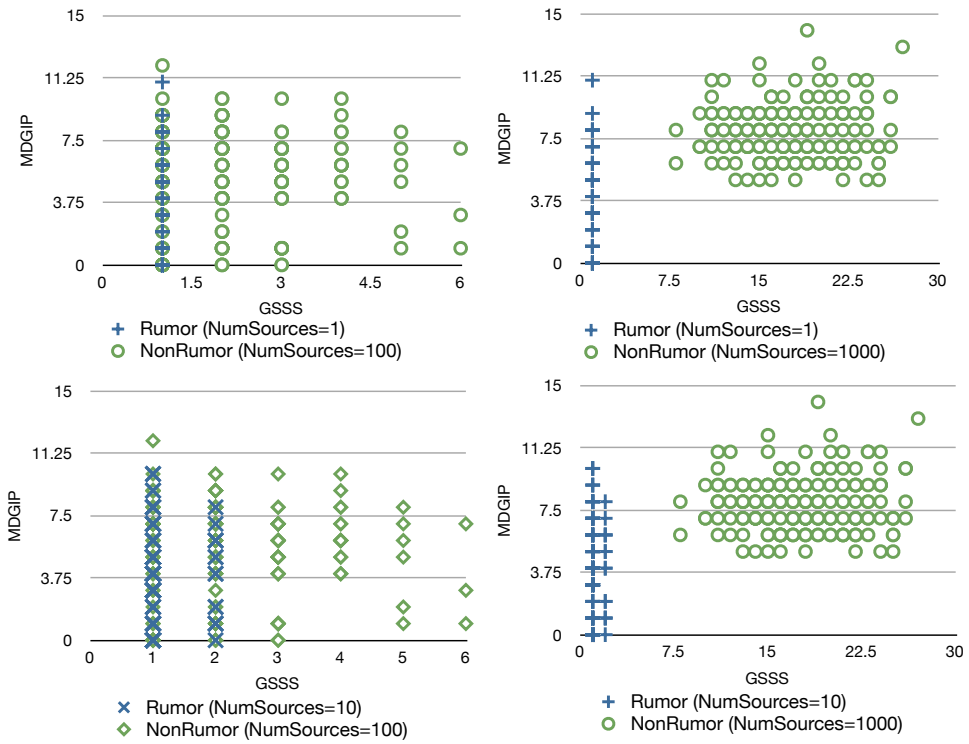

Figure 13. Scatterplot of GSSS and MDGIP (Monitor Selection: Dist, Number of Monitors: 640)

4. RELATED WORK

Online social networks have emerged as a new medium for information sharing.[1,5] Contrary to the traditional media, anyone can share information and it can be delivered to a large number of people in a very short time.[2] Unfortunately, it also carries undesirable information such as rumors.[6]

Shah and Zaman addressed the problem of finding rumor sources.[7] They model rumor spreading with a variant of an SIR model[8] and define rumor centrality as an ML estimator to find the rumor source. Their algorithm makes use of all nodes that hear the rumor. In contrast, our algorithm is based on a small set of monitors which are pre-selected by various methods.

A way to avoid rumors is to subscribe only trustworthy information sources. Adler and Alfaro proposed a content-driven reputation system for Wikipedia authors.[9] For each preserved edit or rollback of a user's edit, reputation is adjusted. Zhao et al. also proposed SocialWiki in which social context including each user's interest

Table 1. Classification of rumors and non-rumors (Number of Monitors: 640).

Monitor Selection	NumSources (Rumor)	NumSource (Non-rumor)	Rumors			NonRumors		
			True Positive	False Negative	Not Classified	True Negative	as False Positive	Not Classified
Dist	1	100	70	0	30	**75**	24	1
Random	1	100	82	18	0	59	41	0
NI+Dist	1	100	**84**	13	3	55	45	0
NI	1	100	37	63	0	67	33	0
BC	1	100	51	49	0	52	48	0
BC+Dist	1	100	64	36	0	43	57	0
Dist	1	1000	70	0	30	**100**	0	0
Random	1	1000	**100**	0	0	99	1	0
NI+Dist	1	1000	97	0	3	**100**	0	0
NI	1	1000	37	63	0	71	29	0
BC	1	1000	67	33	0	59	41	0
BC+Dist	1	1000	87	13	0	82	18	0
Dist	10	100	68	10	22	**73**	26	1
Random	10	100	81	19	0	59	41	0
NI+Dist	10	100	**91**	6	3	43	57	0
NI	10	100	34	66	0	67	33	0
BC	10	100	54	46	0	48	52	0
BC+Dist	10	100	51	49	0	43	57	0
Dist	10	1000	78	0	22	**100**	0	0
Random	10	1000	**99**	1	0	99	1	0
NI+Dist	10	1000	97	0	3	**100**	0	0
NI	10	1000	34	66	0	71	29	0
BC	10	1000	67	33	0	59	41	0
BC+Dist	10	1000	89	11	0	82	18	0

and trust is used to select trustworthy contributors[10] and TrustWiki in which conflicts are resolved by matching compatible editors and readers.[11] Canini et al. proposed a system that ranks users using topical content and social network structure.[12] Nel et al. tackled the problem of rumor detection by monitoring publishing behavior of information sources.[3] They cluster groups of sources that have similar publishing behaviors. Our approach uses social graph topology and monitors to detect rumors and it is orthogonal to user reputation systems.

Morris et al. studied how people feel about the credibility of new tweets.[13] They showed that people use various heuristics to assess the credibility – whether the tweet is retweeted, author's expertise, etc. This can be used to improve the quality of Twitter search engine.

Mendoza et al. studied tweets about 2010 earthquake in Chile and found that rumors and non-rumors are retweeted in a different way.[14] That is, people question more about rumors and affirm more about non-rumors when they retweet. By making use of the comments of users in retweets, they have shown that it is possible to classify rumors and non-rumors. Castillo et al. used a machine leaning technique that makes use of text in tweets, user characteristics and tweet propagation pattern to classify rumors and non-rumors.[15] Our methods could enhance such leaning techniques by exploiting social network graph structure even when the investigator has a limited view on the rumor propagation. Ratkiewicz et all developed a tool that visualizes tweet propagation and can be used to detect abusive behaviors.[16] This tool assumes that full provenance about information propagation is known, which is not used by our method.

5. CONCLUSION

In this paper, we proposed an approach for (i) determining whether a piece of information is a rumor or not, and (ii) finding the source of the rumor. Our approach uses a very small amount of provenance information; namely, which of a set of monitors heard the piece of informaiton at hand. To find the rumor source, our algorithm

evaluates the likelihood of each node to be the source, calculated from node connectivity and shortest path distances. For rumor classification, we proposed two metrics – Greedy Source Set Size (GSSS) and Maximal Distance of Greedy Information Propagation (MDGIP) – and used logistic regression. To evaluate the proposed approach, we performed a case study involving a real social network crawled from Twitter. The algorithm shows good potential to help users in identifying rumors and their sources.

ACKNOWLEDGMENTS

Research reported in this paper was sponsored by the Army Research Laboratory and was accomplished under Cooperative Agreement Number W911NF-09-2-0053 and NSF CNS 09-05014. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

REFERENCES

[1] Kwak, H., Lee, C., Park, H., and Moon, S., "What is twitter, a social network or a news media?," in [*Proceedings of the 19th international conference on World wide web*], *WWW '10*, 591–600, ACM, New York, NY, USA (2010).

[2] Sakaki, T., Okazaki, M., and Matsuo, Y., "Earthquake shakes twitter users: real-time event detection by social sensors," in [*Proceedings of the 19th international conference on World wide web*], *WWW '10*, 851–860, ACM, New York, NY, USA (2010).

[3] Nel, F., Lesot, M.-J., Capet, P., and Delavallade, T., "Rumour detection and monitoring in open source intelligence: understanding publishing behaviours as a prerequisite," in [*Proceedings of the Terrorism and New Media Conference*], (2010).

[4] Newman, M. E. J., [*Networks: An Introduction*], Oxford University Press (March 2010).

[5] Java, A., Song, X., Finin, T., and Tseng, B., "Why we twitter: understanding microblogging usage and communities," in [*Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis*], *WebKDD/SNA-KDD '07*, 56–65, ACM, New York, NY, USA (2007).

[6] Corcoran, M., "Death by cliff plunge, with a push from twitter," *New York Times* (July 12, 2009).

[7] Shah, D. and Zaman, T., "Rumors in a network: Who's the culprit?," *IEEE Transactions on Information Theory* **57**(8), 5163–5181 (2011).

[8] Bailey, N., [*The mathematical theory of infectious diseases and its applications*], Mathematics in Medicine Series, Griffin (1975).

[9] Adler, B. T. and de Alfaro, L., "A content-driven reputation system for the wikipedia," in [*Proceedings of the 16th international conference on World Wide Web*], *WWW '07*, 261–270, ACM, New York, NY, USA (2007).

[10] Zhao, H., Ye, S., Bhattacharyya, P., Rowe, J., Gribble, K., and Wu, S. F., "Socialwiki: bring order to wiki systems with social context," in [*Proceedings of the Second international conference on Social informatics*], *SocInfo'10*, 232–247, Springer-Verlag, Berlin, Heidelberg (2010).

[11] Zhao, H., Kallander, W., Gbedema, T., Johnson, H., and Wu, F., "Read what you trust: An open wiki model enhanced by social context," in [*Proceedings of the 2011 IEEE Third International Conference on Social Computing*], *SocialCom '11*, IEEE Computer Society, Boston, MA, USA (2011).

[12] Canini, K. R., Suh, B., and Pirolli, P. L., "Finding credible information sources in social networks based on content and social structure," in [*Proceedings of the 2011 IEEE Second International Conference on Social Computing*], *SocialCom '11*, 1–8 (2011).

[13] Morris, M. R., Counts, S., Roseway, A., Hoff, A., and Schwarz, J., "Tweeting is believing?: understanding microblog credibility perceptions," in [*Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*], *CSCW '12*, 441–450, ACM, New York, NY, USA (2012).

[14] Mendoza, M., Poblete, B., and Castillo, C., "Twitter under crisis: can we trust what we rt?," in [*Proceedings of the First Workshop on Social Media Analytics*], *SOMA '10*, 71–79, ACM, New York, NY, USA (2010).

[15] Castillo, C., Mendoza, M., and Poblete, B., "Information credibility on twitter," in [*Proceedings of the 20th international conference on World wide web*], *WWW '11*, 675–684, ACM, New York, NY, USA (2011).

[16] Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., and Menczer, F., "Truthy: mapping the spread of astroturf in microblog streams," in [*Proceedings of the 20th international conference companion on World wide web*], *WWW '11*, 249–252, ACM, New York, NY, USA (2011).