# Debunking Rumors in Social Networks: A Timely Approach

Liang Wu
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
699 South Mill Street
Tempe, AZ 85287-8809
liangwu1@asu.edu

Huan Liu
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
699 South Mill Street
Tempe, AZ 85287-8809
huanliu@asu.edu

## ABSTRACT

Social networks have been instrumental in spreading rumor such as fake news and false rumors. Research in rumor intervention to date has concentrated on launching an intervening campaign to limit the number of infectees. However, many emerging and important tasks focus more on *early* intervention. Social and psychological studies have revealed that rumors might evolve 70% of its original content after 6 transmissions. Therefore, ignoring earliness of intervention makes the intervening campaign downgrade rapidly due to the evolved content. In real social networks, the number of social actors is usually *large*, while the budget for an intervening campaign is relatively *small*. The limited budget makes early intervention particularly challenging. Nonetheless, we present an efficient containment method that promptly terminates the diffusion with least cost. To our knowledge, this work is the first to study the earliness of rumor intervention in a large real-world social network. Evaluations on a network of 3 million users show that the key social actors who earliest terminate the spread are not necessarily the most influential users or friends of rumor initiators, and the proposed method effectively reduces the life span of rumors.

## CCS CONCEPTS

• **Mathematics of computing → Graph theory**; • **Networks → Online social networks**;

## KEYWORDS

Graph Mining, Social Network Analysis, Social Media Mining, Classification

**ACM Reference format:**
Liang Wu and Huan Liu. 2019. Debunking Rumors in Social Networks: A Timely Approach. In *Proceedings of 11th ACM Conference on Web Science, Boston, MA, USA, June 30–July 3, 2019 (WebSci '19)*, 9 pages.
DOI: 10.1145/3292522.3326025

## 1 INTRODUCTION

With the blistering expansions in recent years, social media has become an attractive platform for information dissemination. The interconnections between social actors allow for the communication of time-sensitive information. However, social networks have also cultivated the widespread of rumors. For example, a piece of rumor sent by a compromised Associated Press Twitter account wiped out over $136 billion in equity market value within ten minutes[1].

Despite its importance in maintaining "quality" real-time communications, surprisingly little has been studied about the ***earliness*** of intervention, *i.e.,* how quickly the rumor diffusion can be terminated. Existing intervening systems mainly focus on reducing the ultimate number of infectees [7, 35, 38]. However, according to the traditional social and psychological studies, a rumor evolves so rapidly that most details would be altered after 6 transmissions [2]. Therefore, ignoring the lifespan of rumors leads to more variants being generated that may result in greater influence. An example is the urban legend of Ebola: when potential cases appeared in Newark, false rumors of pandemic outbreaks arose in social media. As a common practice of reducing infectees, authoritative and influential sources broadcast to debunk it. However, the rumor evolved into different variants that circulated for a long time, such as the virus can spread through air and salt water cures Ebola[2].

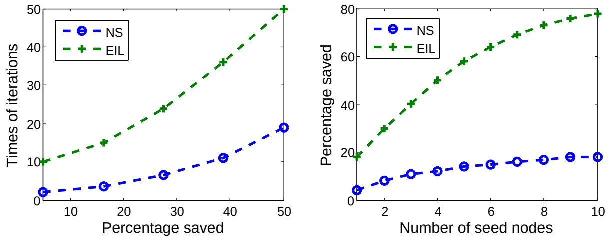
Existing intervention methods that focus on reducing infectees unnecessarily result in an early termination. They model rumor intervention as a multiparty influence maximization problem, and the main intuition is to find those key influential users, given a particular group of rumor initiators, that minimize (or maximize) the influence of rumors (or factual information). Therefore, influential nodes with a higher centrality measure are usually selected for the campaign of factual information, while users that locate remotely from the *centered* nodes are overlooked. Allocating more budget to the center leads to fewer infectees, but rumors may circulate a longer time among less approachable users.

In this work, we postulate an alternative debunking strategy that rapidly terminates the diffusion of rumors, and we also give the justification behind our choices with theoretical analysis and empirical evaluations. The main contributions of the work are summarized below,

- We propose to study the earliness of rumor intervention and formally formulate the problem in the context of social networks;


*WebSci '19, Boston, MA, USA*
 978-1-4503-6202-3/19/06…$15.00
DOI: 10.1145/3292522.3326025

[1]https://www.washingtonpost.com/news/worldviews/wp/2013/04/23/syrian-hackers-claim-ap-hack-that-tipped-stock-market-by-136-billion-is-it-terrorism/
[2]http://time.com/3479254/ebola-social-media/




(a) Earliness of NS and EIL. (b) Costs of NS and EIL.

**Figure 1: We compare NS and EIL in terms of costs and earliness. As shown in Figure 1(a), to achieve similar results, NS is much faster and requires fewer iterations. As shown in Figure 1(b), EIL requires a smaller budget to achieve the same result.**

- Prove the NP-completeness and provide an approximation approach that efficiently captures the key factual information initiators that terminate rumor diffusion at an early stage.
- Conduct extensive experiments with three real-world social network datasets to understand the working of different aspects for the proposed approach.

Rest of the work is organized as follows: in Section 2, we employ exploratory studies to present the motivation of the work. In Section 3, we formulate the problem of early intervention of rumors. In Section 4, we prove the NP-completeness of the problem and provide an approximation method that is theoretically bounded. Experimental results on real-world social networks are described in Section 5. Section 6 presents related research and Section 7 concludes the work.

## 2 EXPLORATORY STUDY

The current containment methods can be classified into immunization and real-time intervention approaches. In the first, *before* a rumor starts spreading, the optimal set of nodes and/or links are immunized (*e.g.,* blocking and removing) to make the social network robust to future attacks [9, 39, 48]. In the second, the optimal set of users are found to launch a debunking campaign *after* a rumor starts spreading [7, 17, 20, 27–29, 35, 37, 38]. Although the term "early" has been mentioned in previous work [28], however, all existing approaches only optimize and evaluate the number of infectees. In order to explore the possible way to early intervention, we compare the earliness and influence of two frameworks.

Without loss of generality, we choose generic methods in each category. We choose NetShield (NS) [9, 39] for immunization methods, which is the state-of-the-art approach. It first exploits the spectral property [14] of graphs for large-scale immunization and it is efficient to optimize. The problem of real-time rumor intervention has been first introduced by Budak *et al.*, and they prove that it is an NP-complete problem and they offer a near-optimal solution named EIL [7]. Both EIL and NS have been well studied and extended to various intervention tasks [9, 17, 27]. We conduct experiments with real-world Twitter data, consisting of 19,240 nodes (users) and 3,933,718 edges. The dataset is publicly available through the Social Computing Data Repository[3].

Figure 1(a) shows the earliness of the two methods. Since they both focus on the number of "saved" infectees, we vary the percentage of saved nodes and observe the time needed by each method. A node is *saved* if it would be infected in the absence of the intervention [7]. Time is represented by iterations of transmissions which is discrete. We can see that to save the same number of nodes, NS terminates the rumor spread at the earlier stage. Since the required budget (number of seed nodes) is a main concern in real applications, in Figure 1(b), we vary the budget size and observe the percentage of saved nodes. Given the same number of seed nodes, EIL significantly outperforms NS in saving more nodes.

Based on the results, we draw the following observation: though NS seems to be a better tool to achieve the goal of early intervention, the required large budget makes it less practically useful. NS aims to find the optimal set of nodes for decreasing the vulnerability of the global graph in the context of network design, such as computer network intrusion. The size of the optimal set of nodes linearly increases with the number of nodes in a graph. Whereas in digital rumor intervention, the budget for a particular piece of rumor is usually limited; and it is also impractical to permanently block a large number of nodes for immunization purposes. In this work, we will seek to early intervene rumor spread with least cost. Our main intuition is that given the locality of the initial infection, many nodes that lead to a global robustness are redundant to a particular piece of rumor.

## 3 PROBLEM DEFINITIONS

Traditional information diffusion models study the roles of certain nodes during the process of information being viral. Since they mainly focus on a single-party campaign, information diffusion models cannot be used to model the interaction between multiple campaigns. Though several methods have been proposed to model multi-party information diffusion, the factor of earliness, which is of practical significance in real applications, has been ignored. In this work, we try to model earliness directly.

### 3.1 Information Diffusion Model

In this subsection, we introduce the diffusion model of rumors in a social network. In the diffusion model, a social network is considered as a directed graph $G = (N, E)$, consisting of nodes (people) $N$ and edges (social relationships) $E$. Let $M$ and $F$ denote the diffusion of a *rumor* and *factual* information respectively. For each node in a network, it can be infected by either information or be inactive as initialized. The initial set of nodes influenced by $M$ and $F$ are denoted as $P(M)$ and $P(F)$. We assume that people highly rely on social influence in forming their opinions [19], as a result, the infection can be caused by initialization or sequentially by one of the neighbors.

Considering the characteristics of rumors, we introduce three properties of the diffusion model. First, for the brevity of presentation, we let the diffusion of $M$ and $F$ start at the same time. The setting is generalizable to cases where $M$ starts earlier than $F$ since all infected nodes before $F$ starts can be regarded as initialized

[3]http://socialcomputing.asu.edu/

rumor infectees in another instance of the same problem. Second, we assume people are more likely to believe the factual information than a rumor [21], so when an inactive individual is influenced by $M$ and $F$ at the same time, the individual will be influenced by $F$. Third, in order to make the problem computationally tractable, we assume the diffusion to be *progressive*, *i.e.,* an inactive individual would be influenced by either information, and once being active, it cannot turn back to inactive or switch to the other status.

The diffusion starts with the network $G$, and two initial sets of nodes $P(M)$ and $P(F)$. If an individual node $u$ becomes infected by rumor $M$ or immunized by the factual information $F$ at timestamp $t$, it attempts to influence all its inactive neighbors at timestamp $t+1$. In this work, we assume the time to be discrete as timestamps. Therefore, an early intervention means fewer timestamps and fewer times of rumors to be transmitted. The diffusion happens only once for every edge, and it operates in discrete timestamps. The diffusion keeps running until no more inactive nodes are infected.

### 3.2 The Computational Problem

Based on the conceptual definitions, we introduce our problem and formally define the computational problem of the real-time early intervention of rumors.

As discussed earlier, immunization methods are faster in containing rumors. Immunization methods like NS exploit the structural and spectral properties of social graphs. In particular, the structural property implies that nodes in a graph are usually scattered into different clusters (communities), and the spectral property indicates that nodes between different clusters have a higher value in the first eigenvector. Therefore, the main idea of NS is to block nodes between communities, and when the infection starts, only the local area will be influenced and the global graph is immunized. To better illustrate the idea, we show the process of a rumor spread under different intervention methods in Figure 2. The x-axis represents the progress of the diffusion, where 30% means 30% of all timestamps have passed. The y-axis represents the inter-group entropy. We first calculate the community membership of social actors with the CESNA method in SNAP toolkit [25], which can efficiently detect communities in big graphs. Given the community membership $C \in \mathbb{R}^{|N| \times K}$, where $K$ is the number of communities, and the probability of a community is $p_i = \frac{\sum_{j=1}^{|N|} C_{i,j}}{|N|}$, and the inter-group entropy can be calculated as,

$$e = -\sum_{i=1}^{K} p_i \log(p_i), \tag{1}$$

which measures how well nodes from different communities are blended. It can be seen that, under EIL, rumors are viral in more clusters of nodes, while NS constrains rumors in fewer communities. Containing rumors in fewer communities not only leads to early intervention, but may also be more practically useful: as investigated in recent work [12], online communities are formed with the coherence of beliefs [1, 5], and they may quickly turn into an *echo chamber* of rumors due to the biased self-confirmation and their selective narrative.

Our main intuition to reduce the budget of NS is that many nodes selected by NS are "covered" by others given a particular (set of) seed node(s). Through empirical studies, we show that intuitive measures such as distance to the seed cannot be used to determine the relationship of being covered, and it is actually an NP-complete problem. Here, we denote the set of nodes that best immunize the graph as the immunization set $I$ [39, 48]. Next, we will formally define the problem of Early Intervention of rumors with Minimum Cost (EIMMC) as follows.

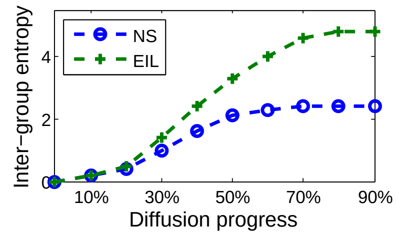



**Figure 2: The change of inter-group entropy with the progress of rumor diffusion. A higher entropy value indicates more groups are infected.**

*Definition 3.1.* Early Intervention of rumors with Minimum Cost (EIMMC) : Given a social network $G = (N, E)$, and rumor originators $P(M) \in N$ and the immunization set $I \in N$, the target of EIMMC is to find the minimum budget $k$ and the corresponding optimal set of factual information originators $P(F) \in N, |P(F)| = k$, so that all nodes in $I$ are guaranteed to be immunized before being reached by the rumor.

## 4 EARLY INTERVENTION OF RUMORS

In this section, we introduce how we conduct early intervention of rumors. We first prove the NP-completeness of the proposed approach. In order to provide an efficient solution, we introduce a greedy strategy that is theoretically bounded.

### 4.1 NP-Completeness of EIMMC

In order to prove the NP-completeness of EIMMC, we introduce the definition of the Set Cover problem below,

*Definition 4.1.* Set Cover problem(SC): Given a set of elements $U = \{1, 2, \cdots, n\}$ and a set of $m$ subsets of $S = S_1, S_2, \cdots, S_m$, the Set Cover problem is to find the minimum number of sets covering all elements in $U$ such that the union of $S$ is $U$.

SC is an NP-complete problem and cannot be solved in polynomial time. Later we will use SC to help prove the NP-completeness of EIMMC. As it may be easier to convince a user of the factual information, we assume the high effectiveness property of $F$, meaning that an individual in $P(F)$ will activate its neighbors in the next timestamp in a deterministic manner. The assumption alleviates the difficulty of solving EIMMC, however, we will show that the simplified problem is also difficult.

Theorem 4.2. *EIMMC is NP-complete even with the high effectiveness property.*

*Proof*: Let $S = \{S_1, S_2, \cdots, S_m\}$ be the sets of an SC problem and $S_1 \cup S_2 \cup \cdots S_m = \{a_1, a_2, \cdots, a_n\}$, we construct an EIMMC problem below.

- We create a graph $G'$ with two kinds of nodes $u$ and $v$, and create a node $u_i$ for a set $S_i$, and a node $v_j$ for each element $a_j$. We build an edge from $u_i$ to $v_j$ if $a_j \in S_i$.
- Introduce an infected node set $P'(M)$, and construct an edge from any node in $P(M)$ to $v_1, v_2, \cdots, v_n$.
- Let the immunization set $I'$ include all $v$, i.e., $I' = \{v_1, v_2, \cdots, v_n\}$.

Therefore, the SC problem is reduced to an EIMMC problem with the graph $G'$, the rumor originators $P'(M)$ and the immunization set $I'$. Since the problem is NP-complete, we will provide an approximate solution that is theoretically bounded.

## 4.2 Approximate Solution for EIMMC

THEOREM 4.3. *There is no $o(\ln(n))$-approximation solution for the problem of EIMMC.*

*Proof*: Since there is no $o\ln(n)$-approximation for the SC problem according to the result of inapproximability [13], following the proof of Theorem 4.2, there is no $o\ln(n)$-approximation for the problem of EIMMC.

THEOREM 4.4. *There exists $O(\ln(n))$-approximation solution for the problem of EIMMC.*

*Proof*: Given the rumor originators $M$, using *Breadth First Search* (BFS) method, we can estimate the earliest time of each individual in $I$ that will be infected. Then the problem of EIMMC is to find the minimum set of nodes in the graph that can influence individuals in $I$ before being infected. In particular, we define these nodes as the *roots* of the search trees named $R = r_1, r_2, \cdots, r_l$, and $l$ is the least number of roots that protects $I$ given a particular $P(M)$. Therefore, each node in $R$ immunizes (covers) a certain number of nodes in $I$. So EIMMC can be transformed to an instance of the SC problem within polynomial time. Since there is an $O(\ln(n))$-approximation for SC, there is also an $O(\ln(n))$-approximation for the problem of EIMMC.

Details of the approximation algorithm can be found in Algorithm 1. In line 1, we use BFS to search nodes that need to be immunized in the immunization set $I$ given a particular $M$. In line 2, we find nodes that can reach $I'(M)$ before they are infected, and $I'(M)_i$ is the $i^{th}$ element of $I'(M)$. $V$ are the candidates for $P(F)$. From line 3 to line 11, we find the covered nodes $H$ of each candidate seed in $V$. Therefore, the problem is transformed to finding the least number of sets ($H$) that cover all units ($I'(M)$). From line 12 to line 15, we search for the optimal sets in a hill-climbing scheme, and the approximation threshold is tightly bounded by $(1 - 1/e)$ [13]. Line 16 returns the optimal set of nodes for the factual information campaign with the least budget. The candidate immunization set $I(M)$ can be obtained beforehand in an offline manner.

## 4.3 Submodularity of EIMMC

Due to the NP-completeness of the EIMMC problem, we proposed an approximation algorithm to greedily search for a near-optimal solution. However, an approximation method is bounded by a margin of $(1-\frac{1}{e})$ only if the outcome function is submodular [30, 41]. We define the outcome function as $O(P(F))$, and it denotes the users that are influenced by a certain set of factual information initiators. We need to prove that $O(\cdot)$ is submodular by showing it has a diminishing return. That is, given $P(F)$ and $P'(F)$ where $P(F) \subseteq P'(F)$, the marginal benefit of adding another initiator $i$ to $P(F)$ is always greater or equal to adding it to $P'(F)$: $|O(P(F) \cup \{i\}) - O(P(F))| \geq |O(P'(F) \cup \{i\}) - O(P'(F))|$.

In our work, we aim to cover as many nodes in immunization set as possible. If immunization set can be better covered, the rumor spread can be terminated in an early stage. Since it is extremely difficult to accurately estimate the exact number of infected nodes, for simplicity, we denote $O(P(F))$ by the nodes in the immunization set that can be covered by immunizing $P(F)$. Following conventional practice, we assume the diffusion process is happened in discrete timestamps [18, 35]: active nodes infect their neighbors at each timestamp, and the timestamp will be tagged with the edge between the active node and the infectee. Note that every edge will be assigned at most one timestamp since a node can be activated only once.

Given a graph $G = (V, E)$ with $P(F)$ and $P(M)$ simultaneously starting their campaign, we derive two activation graphs $G_F$ and $G_M$ for the cascade of factual information and rumors, respectively. From the first timestamp, for each node $i \in P(F)$ and node $j \in PM$, we randomly select $i, u$ and $(j, v)$ and activate the corresponding node $u$ and $v$ and assign the timestamp to the edge. The selection is based on the node degree $\frac{1}{d_u}$, where $d_u$ is the degree of node $u$.

By repeatedly infecting more nodes, some nodes in immunization set will be influenced by the factual information campaign. The expected outcome in terms of coverage of immunization set can be denoted by $O(P(F)) = |E_(G_F)(I)|$, which is the number of edges that are linked to immunization set nodes activated by factual information. We prove the diminishing return by presenting the following lemma.

LEMMA 4.5. *In the activation graph $G_M$, there exists at least one path from a node in $P(M)$ to a node that is both in the immunization set and $G_M$. In the activation graph $G_F$, there exists at least one path from a node in $P(F)$ to a node that is both in the immunization set and $G_F$. For both paths, we denote the timestamp as $t_i^M$ for node $i \in G_M$ and $t_j^F$ for node $j \in G_F$.*

In order to prove $|O(P(F) \cup \{i\}) - O(P(F))| \geq |O(P'(F) \cup \{i\}) - O(P'(F))|$, we need to show that there exists a node $v \in I$, such that $v \in O(P(F))$ and $v \in O(P'(F) \cup \{i\}) \backslash O(P'(F))$. To this end, we present the following two lemmas to prove equivalence:

LEMMA 4.6. *The sufficient conditions for $v \in O_(P(M))$ are*

- *$v \in G_F(P(F))$ and $v \in G_M(P(M))$;*
- *There exists a timestamp $t_v^F$ for any edge of $v$ in $G_F(P(F))$ that is smaller than or equal to the smallest timestamp of any edge of $v \in G_M(P(M))$.*

LEMMA 4.7. *The sufficient conditions for $v \in O(P'(F) \cup \{i\}) \backslash O(P'(F))$ are as follows*

- *$v \in G_M(P(M))$ and $v \in G_F(P(F) \cup \{i\})$;*
- *There exists a timestamp $t_v^F$ for any edge of $v$ in $G_F(P(F) \cup \{i\})$ that is smaller than or equal to the smallest timestamp of any edge of $v \in G_M(P(M))$;*

**Table 1: Notations and corresponding descriptions**

| | |
|---|---|
| $G = \{N, E\}$ | A graph $G$ with nodes $N$ and edges $E$ |
| $M, P(M)$ | rumor and its initiators |
| $F, P(F)$ | Factual information and its initiators |
| $t$ | Timestamp |
| $I$ | Immunization set |
| $u, v, w$ | Nodes in a graph |

- *For all $j \in P'(F)$, the smallest time stamp among all edges in $G_F(P'(F))$ is larger than the smallest timestamp in $G_M(P(M))$.*

Theorem 4.8. *The outcome function $O(\cdot)$ is submodular.*

It is easy to prove that $O(\cdot)$ is monotonic. Therefore, the marginal gain of $O(\cdot)$ is diminishing and the greedy approximation is tightly bounded by $(1 - \frac{1}{e})$.

**Algorithm 1** Algorithm of Solving EIMMC

**Input:** $G = N, E, P(M), I(M), D = \emptyset, P(F) = \emptyset$
**Output:** $P(F)$

1: For $n \in P(M)$, use BFS to construct paths from n to nodes in $I(M)$. Denote the connected nodes (leaves of BFS paths) as $I'(M)$

2: For $n \in I'(M)$, reversely use BFS to find nodes $v^i$ that can reach $n$, $i$ is the length of the shortest path, *i.e.,* $v^0$ = n, $\forall w \in P(M), w = v^j, j \geq i$; Denote the set of $v^i$ as $V_n$
3: For $i$ = 1 to $|I'(M)|$
4: For $u \in V_i \cup_{k=1}^{i-1} V_k$
5: Add edge from $u$ to $I'(M)_i$
6: For $j = i + 1$ to $|I'(M)|$
7: If $u \in V_j$, add edge from $u$ to $I'(M)_j$;
8: End for
9: Denote nodes that connect by $u$ as $H_u$
10: End for
11: End for
12: While $|D| < |I'(M)|$
13: Select $w = \arg\max_{u \in \cup_{i=1}^{|I'(M)|} V_i} |H_u \setminus D|$
14: $P(F) = P(F) \cup w$, $D = D \cup H_u$
15: End while
16: Return $P(F)$

## 5 EXPERIMENTS

In this section, we evaluate our approach, compare the results with established baselines, and discuss insights gained. In particular, we aim to evaluate the proposed approach from the following aspects: (1) effectiveness at different networks with different distributions; (2) effectiveness at different rumor originators. Here the effectiveness means the earliness of intervention and the required budget size for intervention.

### 5.1 Datasets

We use three real-world networks, including the Twitter social network[4], the contact network of Flickr[5], and the collaboration network of DBLP[6]. Twitter and Flickr datasets have been widely used in social network studies. The academic network is a relatively sparse network that captures key features of social networks [34].

- *Twitter Social Network*: This dataset is extracted from the Twitter social network. A node is a Twitter user, and a directed edge from $i$ to $j$ means $i$ is followed by $j$. The dataset contains 19,240 nodes and 3,933,718 edges with an average node degree of 204.
- *Flickr Contact Network*: This network covers all the contacts of Flickr users. A directed edge is established from $i$ to $j$ if $i$ contacts $j$. This dataset consists of 20,809 nodes connected by 390,629 edges with an average node degree of 18.
- *DBLP Collaboration Network*: The DBLP network is extracted from the DBLP computer science bibliography that covers the co-authorship. A connection is established between two nodes (authors) if they publish at least one paper together. The network contains 317,080 nodes connected by 1,049,866 edges with an average degree of 3.

The Twitter and Flickr datasets are obtained through the Social Computing Data Repository[7], and the collaboration network is also publicly available[8].

### 5.2 Comparison Results

**Baselines**: In order to compare with competitive methods that tackle the problem of early intervention, we compare with the real-time intervention method EIL. Since the immunization method such as NS requires a budget significantly greater than that of real-time intervention methods, NS is not adopted for the first experiment on real-time intervention. In order to prove the necessity of the proposed framework and test whether simple heuristics can effectively terminate rumor spread and reduce the budget for NS, we construct two baseline methods below.

- PROXIMITY: Given $P(M)$ and a budget $k$, we select nodes according to the increasing order of the shortest *distance* to $P(M)$. Here, *distance* is the length of the shortest path between two nodes and we adopt the shortest distance between the node and any node of $P(M)$. This method is to test the heuristic that *neighbors of rumor initiators can quickly intervene the spread.*
- NS+PROXIMITY: Given $I(M)$ and a budget $k$, we select nodes according to the increasing order of the shortest distance to $P(M)$. This method is to test the heuristic that *the immunization set nodes that are closer to the rumor initiators can quickly terminate the spread.*

**Earliness of intervention**: To the best of our knowledge, this paper is the first work that aims to deal with the earliness issue. To evaluate earliness, we conduct experiments as follows. Intuitively,

[4]https://www.twitter.com/
[5]https://www.flickr.com/
[6]http://dblp.uni-trier.de/
[7]http://socialcomputing.asu.edu/
[8]https://snap.stanford.edu/data/com-DBLP.html

the number of timestamps can be reduced by increasing the budget (the number of factual information initiators). For example, an extreme case is that all neighbor nodes of the rumor initiators are selected to be immunized, then the diffusion only happens in zero timestamp. Hence, comparing the earliness is equivalent to comparing the needed budget for each method to contain rumors within certain time.

On the three datasets, we set the timestamps for rumor intervention to be 2, 5 and 10. We test each method with a large budget and keep reducing it until the time request cannot be fulfilled. We randomly select nodes to be rumor initiators 10 times, and the average least budget is reported to constrain rumors within $T$ timestamps. As depicted in Figure 3, the proposed approach performs the best among all the methods under different settings, *i.e.,* requiring the least budget for early intervention. For most cases (DBLP $|T| = 2, 5, 10$; Flickr $|T| = 2, 5, 10$ Twitter $|T| = 5$), NS+PROXIMITY is the runner-up method. The result shows that nodes that are both close to the rumor initiators and in the immunization set of NS are relatively more effective for early intervention. PROXIMITY, which is another heuristic-based baseline method, is outperformed by NS on most cases (DBLP $|T| = 2, 10$, Flicker $|T| = 2$; Twitter $|T| = 2, 5$), which shows the redundancy of NS in intervening a specific rumor campaign. Since reducing the number of infectees is an NP-complete problem, and EIL is the approximation method that is bounded by $(1 - \frac{1}{e})$, EIL is theoretically the best tractable method. However, according to the empirical results, directly applying EIL does not necessarily produce an early plan. Our method can well complement with existing quantity-based methods by providing an approach toward reducing the timespan of rumor spread.

**A trade-off between earliness and quantity**: To meet the target of early intervention, methods other than EIL suffer from a greater number of rumor infectees. Now we are investigating the trade-off between earliness and the quantity of infected nodes of the proposed method.

We denote the number of rumor initiators as $|M|$. By varying the size of $M$, from (1) the **y-axis** we can observe the quantity of infected nodes; and from (2) the **x-axis** we can observe how early the intervention was. For each setting, we repeat experiments 10 times and report the average results. The $M$ rumor initiators are chosen at random. For each method, the time of intervention may vary in different rounds of experiments, so we also average time length.

As depicted in Figure 4, the proposed approach performs best among all the methods under different settings in terms of earliness. On the other hand, EIL best reduces the ultimate number of rumor infectees. On the Twitter network (from Figure 4(a) to Figure 4(c)), EIMCC reduces over 50% of the rumor's lifespan comparing with the best baseline method, and on the Flickr and DBLP network, the reduction is over 60%. Although we do not focus on the final influence, EIMCC outperforms baselines other than EIL in most experiments (except $|M| = 100$ and $|M| = 500$ of DBLP) while delivering the best of earliness.

We can also see that EIMMC constrains the rumors within the least number of communities and prevents it from further spreading. Through observing the curves in Figure 4, we find that the increase of infectees usually has a big jump, *e.g.,* EIL, Proximity and EIMMC between the second and third timestamp in Figure 4(a). This is because a rumor reaches the "tipping point" and becomes viral in another community. Such "jump" is unfavorable since it shows a community has turned into the "echo chamber" of the rumor, and the rumor may be better trusted, quickly evolved, and more difficult to deal with. Traditional approaches, such as EIL on Flickr dataset with $|M| = 500$, may have more than one "jump", while EIMMC has only one of such jumps in all the experiments.

We see that simple heuristics cannot deal with either the ultimate influence or the earliness of intervention. It would be convenient if nodes that are closer to $P(M)$ or nodes in the immunization set that are closer to $P(M)$ can lead to an early intervention. However, Proximity, which is based on the shortest distance to $P(M)$, result in the largest number of infectees under most settings as well as the longest lifespan. Proximity+NS performs better comparing with Proximity. However, it still leads to a longer timespan than EIMMC. This is because the nodes in the immunization set that are closer to $P(M)$ may be redundant given a particular rumor.

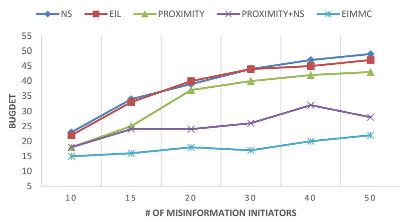
Another interesting finding is that, comparing results on different networks, the rumor diffusion in a denser graph where the average degree is higher is easier to intervene at an early stage. Intuitively, a denser graph makes it easier for rumors to reach more nodes. However, a denser graph also enables the factual information campaign ($F$) to easily influence more users. Therefore, with proper selection of nodes, a denser graph may be easier to protect. In this work, we regard the problem of EIMMC as an instance of the vertex cover problem. A denser graph leads to larger sets in SC, which makes it easier to cover vertices with fewer sets.

**Budget of intervention**: In this work, we postulate the redundancy in the immunization set. Here we will discuss how much EIMMC reduces the budget of intervention given particular rumor campaigns. In order to determine the budget for the proposed method, we first use NS to calculate the immunization set and the budget for NS. Then we apply Algorithm 1 to calculate the budget for EIMMC. The immunization set is greedily expanded until the earliness of the intervention can no longer be improved. Details of NS can be found in [39].

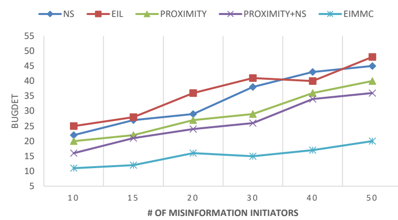
Table 2 illustrates the budget of EIMMC and NS. We vary the size of rumor initiators and observe the corresponding budget. Note that we increase the budget of EIMMC if it leads to a slower intervention. Therefore, the results show how much EIMMC can reduce the budget without any loss of earliness. According to the results, over 70% of the immunization set can be removed, and the required budget of EIMMC linearly increases with $|M|$.

## 6 RELATED WORK

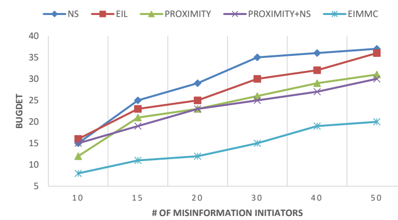
Since the famous study on the psychology of inaccurate and false information [2], *e.g.,* rumors and urban legends, much work has been done in understanding the mechanism of rumors. There is ample previous work on building mathematical models for the spreading of rumors, *i.e.,* describing the growth and decay of the actual spreading process with simulated population and networks [11]. Such models are typically proposed for characterizing dynamics of rumors with stochastic approaches, in specific network structures such as



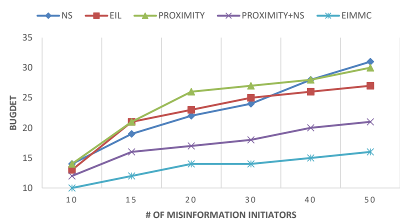

(a) DBLP, $|T|$=2.



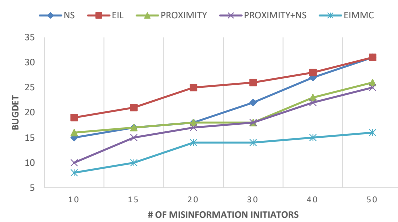

(b) Flickr, $|T|$=2.



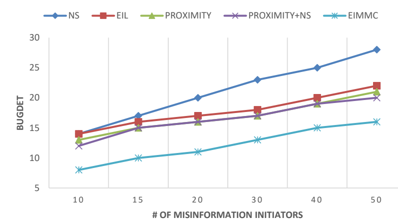

(c) Twitter, $|T|$=2.



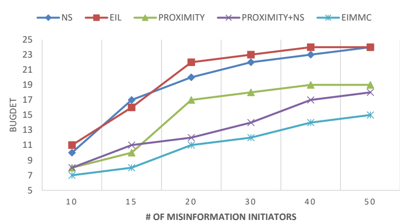

(d) DBLP, $|T|$=5.



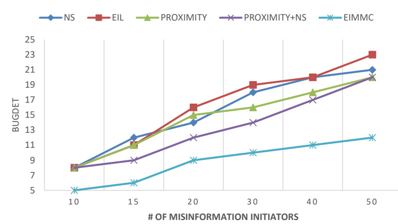

(e) Flickr, $|T|$=5.



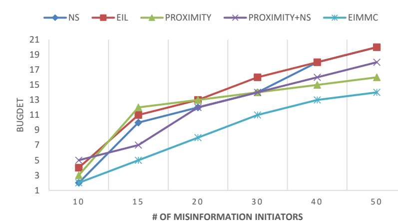

(f) Twitter, $|T|$=5.




(g) DBLP, $|T|$=10.




(h) Flickr, $|T|$=10.




(i) Twitter, $|T|$=10.

**Figure 3: Given a certain social network, we vary the size of rumor initiators $|M|$ and observe the budget of each method needed to terminate rumor diffusion within a target time $|T|$. Methods that require a smaller budget are more effective in early intervening rumors.**

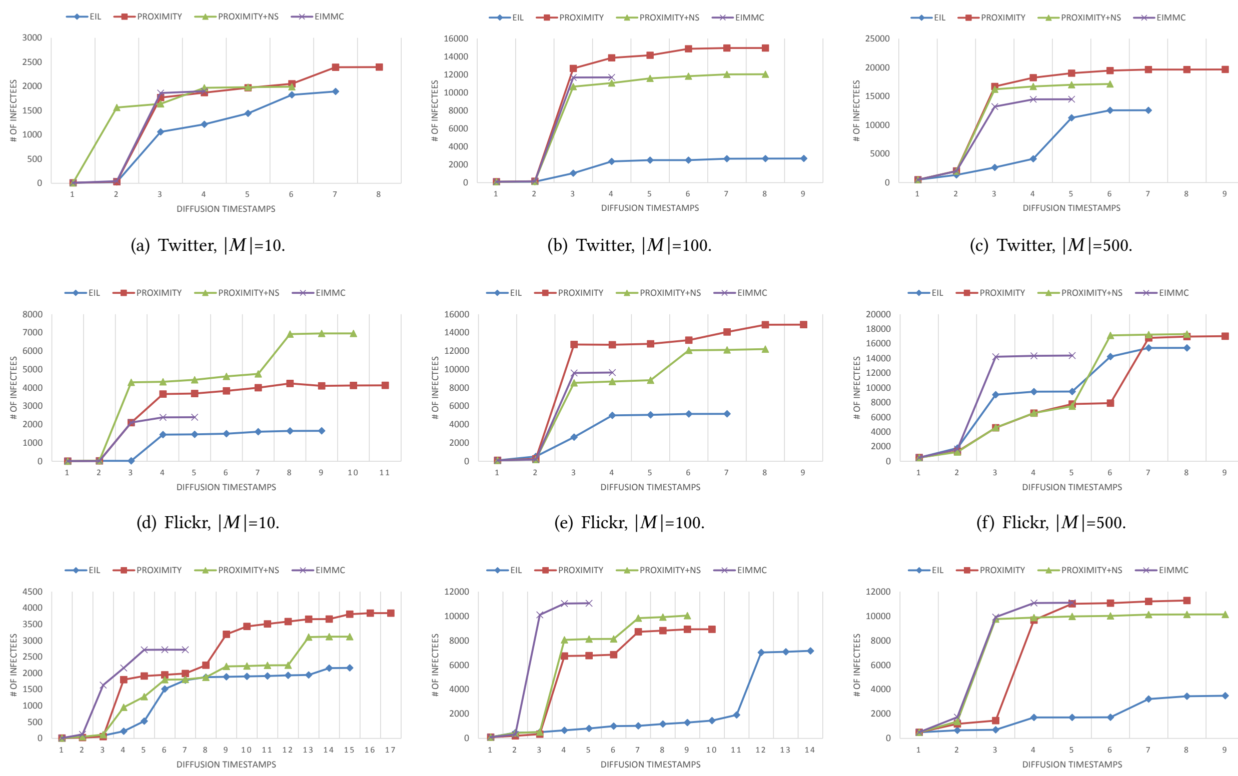
**Table 2: Budget of EIMMC and NS with different datasets and size of $M$.**

| $\|M\|$ | | 10 | 20 | 30 | 40 | 50 | 100 | 500 |
|---|---|---|---|---|---|---|---|---|
| Twitter | NS | 153 | 172 | 179 | 188 | 201 | 220 | 240 |
| | EIMMC | 28 | 33 | 36 | 40 | 55 | 63 | 45 |
| Flickr | NS | 140 | 146 | 155 | 163 | 170 | 182 | 200 |
| | EIMMC | 36 | 41 | 48 | 53 | 51 | 58 | 66 |
| DBLP | NS | 130 | 142 | 147 | 153 | 155 | 170 | 201 |
| | EIMMC | 31 | 33 | 34 | 47 | 49 | 53 | 61 |

scale-free [32]. However, existing information and misinformation diffusion work focuses on estimating the impact of rumors on simulated networks in a small scale [46]. We focus on approaches that directly intervene rumors in real time.

In the context of rumor intervention, Budak *et al.* explore augmenting the traditional influence diffusion techniques to find the optimal nodes to launch an intervening campaign [7], by introducing a competitive party into the traditional *Influence Maximization* (IM) approaches. There are two classical models for IM, *i.e., Linear Threshold* (LT) and *Independent Cascade* (IC). Borodin *et al.* study competitive influence diffusion under the extension of the LT model [6]. Bharathi *et al.* study the IC model and provide an approximation algorithm to maximize the spread of the influence of a single factual information spreader. In addition, Kostka *et al.* regard rumor intervention as a game theoretical problem and analyze the effect of the time lag of a delayed factual information campaign [23]. Since the budget in real-time intervention is usually limited, Nguyen *et al.* aim to limit the spread of rumors to a predefined rate with the least possible set of nodes [35]. However, the existing approaches emphasize on reducing the ultimate number of infected nodes. By contrast, we address the problem of early intervention. As a result, our work first introduces the earliness as a metric to evaluate intervention approaches, which is important in dealing with rumors due to the quick spread and evolved content [2].

Tong *et al.* first propose to immunize a certain number of nodes to increase the robustness of the network to future attacks [39]. Their idea is to contain the rumors or virus in a small group locally, and they provide a greedy algorithm to efficiently find the optimal set of nodes via leveraging the structural and spectral properties of graphs. The method has been well studied and extended to solve node immunization problems [9] such as group immunization.




(a) Twitter, $|M|$=10.

(b) Twitter, $|M|$=100.

(c) Twitter, $|M|$=500.

(d) Flickr, $|M|$=10.

(e) Flickr, $|M|$=100.

(f) Flickr, $|M|$=500.

(g) DBLP, $|M|$=10.

(h) DBLP, $|M|$=100.

(i) DBLP, $|M|$=500.

**Figure 4: The diffusion progress with different number of rumor initiators and on different datasets. A curve stops progressing to the next timestamp if the rumor is terminated then. We report the lifespan of rumors (horizontally) and the number of infectees (vertically). For the task of early intervention, a shorter lifespan is more favorable.**

However, given a particular set of rumor initiators, we postulate the redundancy of the immunization set given a specific rumor campaign, and the proposed approach can be applied to more problems such as the epidemic containment and the group-level immunization that share strong contagion patterns.

Our work is also related to studies on information diffusion. There are various models which are designed to abstract the pattern of information diffusion, such as *SIR Model* [22], *Tipping Model* [8], *Independent Cascade Model* [21] and *Linear Threshold Model* [21]. Traditional information diffusion models ignore the interaction between multiple campaigns, which cannot be directly applied here. In order to accelerate the diffusion models, more scalable methods have been proposed. SIMPATH [15] reduces the computational time through filtering out paths without enough confidence, and Maximum Influence Arborescence (MIA) [10] was also proposed to accelerate the computation of independent cascade model. Various other acceleration algorithms are also available [16, 31]. Our work can also be accelerated by integrating with MIA.

Another stream of research focuses on spreaders of rumors. They aim to find patterns revealing a malicious account, and use these patterns to block potential spreaders. For example, profiles of automatic generated accounts may look similar, duplication of profiles between malicious accounts can be used to detect rumor spreaders [33, 40, 42, 43]. Lee and Kim propose to reveal the patterns hidden behind malicious accounts [24], so as to filtering them in an early stage [44]. Agglomerative hierarchical clustering is adopted in their work, where the likelihood of two names being generated by an identical Markov chain is used for measuring distance, and characters are used as features. After obtaining clusters of similar account names, a supervised method is adopted to classify whether a name cluster is a group of malicious accounts. Account names have also been quantitatively examined [47]. More features including behavioral ones are further incorporated in such algorithms [4]. In contrast to existing work that focuses on the spreader of rumors [3, 36], network properties of rumor propagation [26, 45] or the content of rumors, we concentrate on finding the key users in a network to early intervene the propagation.

## 7 CONCLUSION

An early intervention is especially crucial in containing widespread of devastating rumors, such as fake news and false rumors. This work is the first to study the earliness of intervening rumors spread in a real-world social network. We focus on social interactions between people, and we aim to find the key social actors that can terminate the spread of rumors at an early stage. Although the budget is relatively small compared with the nodes in a social network, we can identify the least number of nodes that lead to early intervention. We achieve this by proving the NP-completeness, which is a precondition for finding a near-optimal approximation, and developing a hill-climbing method that is theoretically bounded.

## ACKNOWLEDGMENTS

This material is based upon work supported by, or in part by, the National Science Foundation (NSF) grant 1614576, and the Office of Naval Research (ONR) grant N00014-16-1-2257.

## REFERENCES


[1] Luca Maria Aiello, Alain Barrat, Rossano Schifanella, Ciro Cattuto, Benjamin Markines, and Filippo Menczer. 2012. Friendship prediction and homophily in social media. *ACM Transactions on the Web (TWEB)* 6, 2 (2012), 9.

[2] Gordon W Allport and Leo Postman. 1947. The psychology of rumor. (1947).

[3] Despoina Antonakaki, Iasonas Polakis, Elias Athanasopoulos, Sotiris Ioannidis, and Paraskevi Fragopoulou. 2016. Exploiting abused trending topics to identify spam campaigns in Twitter. *Social Network Analysis and Mining* 6, 1 (2016), 1–11.

[4] Fabricio Benevenuto, Tiago Rodrigues, Virgilio Almeida, Jussara Almeida, Chao Zhang, and Keith Ross. Identifying video spammers in online social networks. In *International Workshop on Adversarial Information Retrieval on the Web*. ACM, 45–52.

[5] Alessandro Bessi, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Viral misinformation: The role of homophily and polarization. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 355–356.

[6] Allan Borodin, Yuval Filmus, and Joel Oren. 2010. Threshold models for competitive influence in social networks. In *International Workshop on Internet and Network Economics*. Springer, 539–550.

[7] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the spread of misinformation in social networks. In *Proceedings of the 20th international conference on World Wide Web*. ACM, 665–674.

[8] Damon Centola. 2010. The spread of behavior in an online social network experiment. *science* 329, 5996 (2010), 1194–1197.

[9] Chen Chen, Hanghang Tong, B Aditya Prakash, Charalampos E Tsourakakis, Tina Eliassi-Rad, Christos Faloutsos, and Duen Horng Chau. 2016. Node Immunization on Large Graphs: Theory and Algorithms. *IEEE Transactions on Knowledge and Data Engineering* 28, 1 (2016), 113–126.

[10] Wei Chen, Chi Wang, and Yajun Wang. 2010. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*. ACM, 1029–1038.

[11] Daryl J Daley and David G Kendall. 1964. Epidemics and rumours. (1964).

[12] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences(PNAS)* 113, 3 (2016), 554–559.

[13] Uriel Feige. 1998. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)* 45, 4 (1998), 634–652.

[14] Miroslav Fiedler. 1973. Algebraic connectivity of graphs. *Czechoslovak mathematical journal* 23, 2 (1973), 298–305.

[15] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. 2011. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, 211–220.

[16] Daniel Gruhl, Ramanathan Guha, David Liben-Nowell, and Andrew Tomkins. 2004. Information diffusion through blogspace. In *Proceedings of the 13th international conference on World Wide Web*. ACM, 491–501.

[17] Xinran He, Guojie Song, Wei Chen, and Qingye Jiang. 2012. Influence Blocking Maximization in Social Networks under the Competitive Linear Threshold Model.. In *SDM*. SIAM, 463–474.

[18] Zaobo He, Zhipeng Cai, Jiguo Yu, Xiaoming Wang, Yunchuan Sun, and Yingshu Li. 2017. Cost-efficient strategies for restraining rumor spreading in mobile social networks. *IEEE Transactions on Vehicular Technology* 66, 3 (2017), 2789–2800.

[19] Andrew Hiles, This Chapter Covers, and How Rumors Spread. 2011. How Firms Should Fight Rumors. *Reputation Management: Building and Protecting Your CompanyâĂŹs Profile in a Digital World* (2011), 1–44.

[20] Ken-ichi Kawarabayashi. 2014. Optimal budget allocation: Theoretical guarantee and efficient algorithm. (2014).

[21] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 137–146.

[22] William O Kermack and Anderson G McKendrick. 1927. A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, Vol. 115. 700–721.

[23] Jan Kostka, Yvonne Anne Oswald, and Roger Wattenhofer. 2008. Rumor dissemination in social networks. In *International Colloquium on Structural Information and Communication Complexity*. Springer, 185–196.

[24] Sangho Lee and Jong Kim. 2014. Early filtering of ephemeral malicious accounts on Twitter. *Computer Communications* 54 (2014), 48–57.

[25] Jure Leskovec and Rok Sosič. 2016. SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *ACM Transactions on Intelligent Systems and Technology (TIST)* 8, 1 (2016), 1.

[26] Chaz Lever, Platon Kotzias, Davide Balzarotti, Juan Caballero, and Manos Antonakakis. 2017. A Lustrum of malware network communication: Evolution and insights. In *Security and Privacy (SP), 2017 IEEE Symposium on*. IEEE, 788–804.

[27] Hui Li, Sourav S Bhowmick, Jiangtao Cui, Yunjun Gao, and Jianfeng Ma. 2015. Getreal: Towards realistic selection of influence maximization strategies in competitive networks. In *ACM SIGMOD International Conference on Management of Data*. ACM, 1525–1537.

[28] Iouliana Litou, Vana Kalogeraki, Ioannis Katakis, and Dimitrios Gunopulos. 2016. Real-Time and Cost-Effective Limitation of Misinformation Propagation. (2016).

[29] Wei Lu, Francesco Bonchi, Amit Goyal, and Laks VS Lakshmanan. 2013. The bang for the buck: fair competitive viral marketing from the host perspective. In *International Conference on Knowledge Discovery and Data Dining (KDD)*. ACM, 928–936.

[30] Carsten Lund and Mihalis Yannakakis. 1994. On the hardness of approximating minimization problems. *Journal of the ACM (JACM)* 41, 5 (1994), 960–981.

[31] Baharan Mirzasoleiman, Ashwinkumar Badanidiyuru, Amin Karbasi, Jan Vondrák, and Andreas Krause. 2014. Lazier Than Lazy Greedy. *arXiv preprint arXiv:1409.7938* (2014).

[32] Yamir Moreno, Maziar Nekovee, and Amalio F Pacheco. 2004. Dynamics of rumor spreading in complex networks. *Physical Review E* 69, 6 (2004), 066130.

[33] Fred Morstatter, Liang Wu, Tahora H Nazer, Kathleen M Carley, and Huan Liu. 2016. A new approach to bot detection: striking the balance between precision and recall. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 533–540.

[34] Mark EJ Newman. 2001. The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences (PNAS)* 98, 2 (2001), 404–409.

[35] Nam P Nguyen, Guanhua Yan, My T Thai, and Stephan Eidenbenz. 2012. Containment of misinformation spread in online social networks. In *Proceedings of the 4th Annual ACM Web Science Conference*. ACM, 213–222.

[36] Shirin Nilizadeh, François Labrèche, Alireza Sedighian, Ali Zand, José Fernandez, Christopher Kruegel, Gianluca Stringhini, and Giovanni Vigna. 2017. POISED: Spotting Twitter Spam Off the Beaten Paths. *arXiv preprint arXiv:1708.09058* (2017).

[37] Michael Simpson, Venkatesh Srinivasan, and Alex Thomo. 2016. Clearing contamination in large networks. *IEEE Transactions on Knowledge and Data Engineering* 28, 6 (2016), 1435–1448.

[38] Chonggang Song, Wynne Hsu, and Mong Li Lee. 2015. Node Immunization over Infectious Period. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. ACM, 831–840.

[39] Hanghang Tong, B Aditya Prakash, Charalampos Tsourakakis, Tina Eliassi-Rad, Christos Faloutsos, and Duen Horng Chau. 2010. On the vulnerability of large graphs. In *International Conference on Data Mining*. IEEE, 1091–1096.

[40] Steve Webb, James Caverlee, and Calton Pu. 2008. Social Honeypots: Making Friends With A Spammer Near You.. In *CEAS*.

[41] Laurence A Wolsey. 1982. An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica* 2, 4 (1982), 385–393.

[42] Liang Wu, Xia Hu, Fred Morstatter, and Huan Liu. 2017. Adaptive spammer detection with sparse group modeling. In *Eleventh International AAAI Conference on Web and Social Media*.

[43] Liang Wu, Xia Hu, Fred Morstatter, and Huan Liu. 2017. Detecting camouflaged content polluters. In *Eleventh International AAAI Conference on Web and Social Media*.

[44] Liang Wu, Jundong Li, Xia Hu, and Huan Liu. 2017. Gleaning wisdom from the past: Early detection of emerging rumors in social media. In *Proceedings of the 2017 SIAM International Conference on Data Mining*. SIAM, 99–107.

[45] Liang Wu and Huan Liu. 2018. Tracing fake-news footprints: Characterizing social media messages by how they propagate. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM, 637–645.

[46] Liang Wu, Fred Morstatter, Xia Hu, and Huan Liu. 2016. Mining misinformation in social media. *Big Data in Complex and Social Networks* (2016), 123–152.

[47] Reza Zafarani and Huan Liu. 10 Bits of Surprise: Detecting Malicious Users with Minimum Information *(CIKM'15)*. ACM.

[48] Yao Zhang, Sudip Saha, Anil Vullikanti, B Aditya Prakash, and others. 2015. Near-optimal Algorithms for Controlling Propagation at Group Scale on Networks. *IEEE Transactions on Knowledge and Data Engineering* (2015).