A Stochastic Algorithm Based on Reverse Sampling Technique to Fight Against the Cyberbullying

RUIDONG YAN, Inspur Electronic Information Industry Co., Ltd and Renmin University of China
YI LI, University of Texas at Tyler
DEYING LI and YONGCAI WANG, Renmin University of China
YUQING ZHU, California State University at Los Angeles
WEILI WU, University of Texas at Dallas

Cyberbullying has caused serious consequences especially for social network users in recent years. However, the challenge is how to fight against the cyberbullying effectively from the algorithmic perspective. In this article, we study the *fighting against the cyberbullying* problem, i.e., identify an initial witness set with a budget to spread the positive influence to protect the users in a specific target set such that the number of cybervictim users in the target set being activated by the seed set of cyberbullying is minimized. We first formulate this problem and show its NP-hardness. We further prove that the objective function is submodular with respect to the size of witnesses set when we convert the original problem into the maximal version. Then we propose a stochastic approach to solve this maximal version problem based on the *Reverse Sampling Technique* with a constant factor guarantee. In addition, we provide theoretical analysis and discuss the relationship between the optimal value and the value returned by the proposed algorithm. To evaluate the proposed approach, we implement extensive experiments on synthetic and real datasets. The experimental results show our approach is superior to the comparison methods.

CCS Concepts: • **Networks** → *Network algorithms*; • **Theory of computation** → *Design and analysis of algorithms*;

Additional Key Words and Phrases: Social network, seed selection, reverse sampling technique, target users, cyberbullying

This work is partly supported by the National Natural Science Foundation of China under grants 12071478 and 61972404 and the National Science Foundation under grant 1907472.

Authors' addresses: R. Yan, State Key Laboratory of High-End Server & Storage Technology, Inspur Electronic Information Industry Co., Ltd, No. 2 XinXi Road, Beijing, 100085, China and Renmin University of China, No. 59 Zhongguancun Street, Beijing, 100872, China; email: yanruidong@ruc.edu.cn; Y. Li, University of Texas at Tyler, Tyler, Texas, 75799; email: yli@uttyler.edu; D. Li (corresponding author) and Y. Wang, School of Information, Renmin University of China, No. 59 Zhongguancun Street, Beijing, 100872, China; emails: {deyingli, ycw}@ruc.edu.cn; Y. Zhu, California State University at Los Angeles, Los Angeles, California, 90032; email: yuqing.zhu@calstatela.edu; W. Wu, University of Texas at Dallas, 800 W. Campbell Road, Richardson, Texas, 75080; email: weiliwu@utdallas.edu.

1556-4681/2021/03-ART71 $15.00
https://doi.org/10.1145/3441455

ACM Reference format:
Ruidong Yan, Yi Li, Deying Li, Yongcai Wang, Yuqing Zhu, and Weili Wu. 2021. A Stochastic Algorithm Based on Reverse Sampling Technique to Fight Against the Cyberbullying. *ACM Trans. Knowl. Discov. Data* 15, 4, Article 71 (March 2021), 22 pages.
https://doi.org/10.1145/3441455

1 INTRODUCTION

Cyberbullying is a form of bullying or harassment using electronic means like cell phones, computers, and tablets, which is also known as online bullying.[1] It means to attack, humiliate, or disparage some targeted persons over networks, via broadcasting, posting, or sending negative, harmful, offensive information to the targets through cyber-media such as social networks, forums, email, and online gaming communities. In general, there are three roles in the propagation of cyberbullying over a social network [9], [28]: (1) *perpetrators*, i.e., the original attackers who produce offensive remarks; (2) *cybervictims*, i.e., the specific target users which are being attacked by perpetrators, and (3) *witnesses*, i.e., the users in the network who stand up to prevent perpetrators from attacking cybervictims. Cyberbullying can cause extremely bad reputation, mental, even physical damage to the specific target persons. The cyberbullying statistics of the Cyberbullying Research Center reports that in 2016, about 34% of students had been a victim of cyberbullying at some time in their lives.[2]

Social network is a severely afflicted area of cyberbullying for the ease of information propagation. The source set of attackers, *perpetrators*, may activate a set of followers to follow up their attacking behaviours. Although the tongue is boneless, it can break bones. As the influence of perpetrators increases, more and more persons may be activated to be attackers. The *cybervictims*, even though have some resistance capability to the cyber-attacks, they will face much higher probability to be "destroyed" when the group attackers are increasing. For example, a cybervictim has 10 neighbors in a network. If only 1 of 10 neighbors is activated by perpetrators, the probability that the cybervictim activated by cyberbullying is relatively small. However, if more than 8 of these 10 neighbors are activated by perpetrators, the probability that the cybervictim activated by cyberbullying is relatively large. This phenomenon is in line with the fact "Word-of-mouth" that people trust the information obtained from their close social circle. Naturally, how do we take effective measures to fight against the cyberbullying such that the number of cybervictims being activated by cyberbullying on a specific target users is minimized? We call this problem *fighting against the cyberbullying* in social networks.

Note that fighting against the cyberbullying problem is similar to the existing rumor blocking problem [5], [33], i.e., find an optimal seed set such that the expected number of users being activated by rumor is minimized. Inspired by the rumor blocking problem, fighting against the cyberbullying problem can be addressed by employing the analogous methods. For ease of exposition, we first introduce the main methods to solve the rumor blocking problem and we then use a vivid example to illustrate the motivation of this article.

Currently, there are many studies on blocking the rumors spreading over social networks. These studies can be roughly classified into three categories as follows: (1) the first category is removing the global influential nodes from the network such that rumor spreads are minimized [10], [22], [36]; (2) the second one is removing some edges that play a key role in information propagation from the network to limit misinformation propagation [18], [19], [34]; and

[1]See the following website for details: https://en.wikipedia.org/wiki/Cyberbullying.
[2]https://online.maryville.edu/blog/what-is-cyberbullying-an-overview-for-students-parents-and-teachers/.

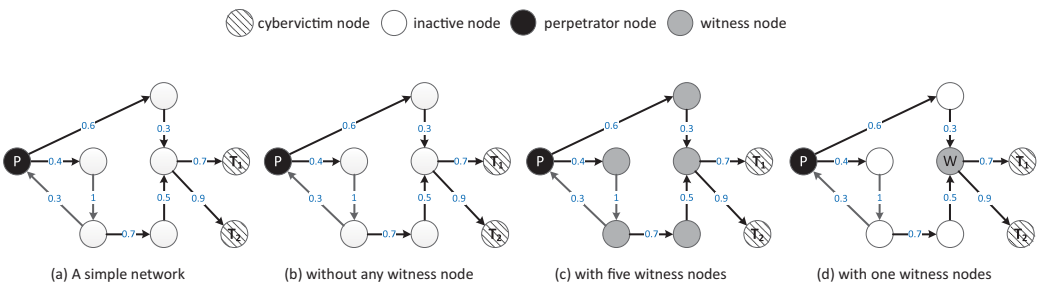


Fig. 1. An illustration of fighting against the cyberbullying problem.

(3) the third one is spreading the positive truth to fight against the rumor such that the positive truth is adopted by as many nodes as possible in the network [33], [5].

Each category has pros and cons. For example, the first two categories need to modify the network structure, which will be costly to be implemented in realistic social networks for protecting a set of people. Take the example of removing nodes on Twitter. Removing some nodes means that the administrators of Twitter will permanently delete some users' accounts. This action will lead to permanent loss of users. On the contrary, the third category provides a very natural way to spread positive truth to stop the rumors without destroying the network structure. In this article, we will learn from the third method to fight against the cyberbullying. In addition, although fighting against the cyberbullying problem and the rumor blocking problem have somethings in common, they are different in the following two aspects: (1) *Target users*. Generally speaking, the former has a specific target user set while the latter is without any special target user set. (2) *Solution strategy*. Most of the existing studies on cyberbullying involves sociology, psychology, and other fields such as [14], [35], and [40]. Solution strategies in these studies focus on case study or statistical analysis, and give corresponding suggestions. However, solution strategies in rumor blocking studies (e.g., [33], [5]) emphasize engineering techniques and provide corresponding algorithms as well as theoretical analysis.

In our study, we show that if there is a small number of witnesses who stand up and stop the cyberbullying or simply spread positive information to support the cybervictims, the negative influence on the target cybervictims could be minimized. An example of fighting against the cyberbullying is shown in Figure 1. In the figure, the directed edges indicate the direction of information propagation and the number embedded on each edge indicates the propagation probability. Nodes in the network are divided into four categories: perpetrator nodes, cybervictim nodes, witness nodes, and inactive nodes. In Figure 1(a), the node P is a perpetrator who aims at attacking the cybervictim set $\mathcal{T} = \{\mathcal{T}_1, \mathcal{T}_2\}$. The other nodes are initially inactive nodes. In Figure 1(b), if there are without any witness nodes in the network, the inactive nodes are activated[3] by P to join the line of attackers, the cybervictim nodes will face higher probability to be activated as well. The activation probability at a cybervictim node can be considered as damaged probability by perpetrators in cyberbullying. In Figure 1(c), if all the inactive nodes will become witness nodes and stand up to propagate the positive information to cybervictim nodes such that they can completely stop attacking, therefore the negative impacts on the cybervictim nodes will be reduced.[4] In Figure 1(c), there are five witness nodes which are expensive and unfeasible in reality. For example, we have

[3]It is worth noting that we assume that the propagation of cyberbullying is similar to that of information propagation which is a stochastic process in social networks. Here we only show one of possible propagation results for ease of explanation.
[4]Here we assume that once an inactive node has already been activated by perpetrator nodes then it cannot be activated by witness nodes any more, and vice versa.

to pay a unit cost when we choose an inactive node as a witness node. However, we can't choose all the inactive nodes as witness nodes in the situation where the budget is limited. In Figure 1(d), if the budget is one, namely, we can only choose one inactive node as a witness node. Which node is the best choice? The answer is obviously to choose W as the witness node. This is because once we choose the W, $\mathcal{T}_1$ and $\mathcal{T}_2$ will be activated by P through W with a very low probability.

Based on the above discussion, we propose the problem of *fighting against the cyberbullying* as follows. Given a directed social network $G = (V, E)$, a set of perpetrators P, a cybervictim set $\mathcal{T}$ and a positive integer budget b, fighting against the cyberbullying problem is to identify a seed set $S_{\mathcal{W}} \subseteq V \backslash \{P \cup \mathcal{T}\}$ as the witness set under the budget restriction b, such that the number of cybervictim nodes in $\mathcal{T}$ being activated by P is minimized.

To address this problem, we firstly employ an information cascade model named *Competitive Independent Cascade* (CIC) [5], [33], [39] to simultaneously characterize two opposing information dissemination processes in a social network. Second, we elaborate a stochastic algorithm based on *Reverse Sampling* technique to select witness nodes. This algorithm is proposed to maximize the number of cybervictim nodes being activated by the witness nodes. Therefore we can indirectly minimize the number of cybervictim nodes that are activated by the perpetrators. The main contributions are summarized as follows:

- —We first propose a *fighting against the cyberbullying* problem (Problem 1) in social networks, i.e., we aim at finding an optimal seed set as witnesses such that the number of nodes in cybervictim set being activated by perpetrators is minimized.
- —We then convert the Problem 1 to an equivalent maximal version, i.e., we want to seek an optimal witness seed set such the number of nodes in cybervictim set being activated by witnesses is maximized (Problem 2). And we show the Problem 2 is NP-hard and the objective function is submodular.
- —To effectively address the Problem 2, we propose a stochastic approach (Algorithm 4) based on the *Reverse Sampling* technique instead of Monte Carlo simulation that is time-consuming. In addition, we provide theoretical analysis and discuss the relationship between the optimal value and the one returned by the algorithm.
- —In order to evaluate the proposed method, we use a synthetic and three real-life social networks in experiments. The extensive simulations validate that our method is superior to comparison approaches.

Organizations: We first begin by recalling some existing related work in Section 2. Then we introduce the information propagation models in Section 3. And we show the preliminaries and problem statement in Section 4. Algorithms are presented in Section 5. We analyze and discuss the results of the experiments in Section 6. Finally, we draw conclusions in Section 7.

2 RELATED WORK

In this section, we review related work from the following five aspects: (1) studies on cyberbullying; (2) studies on single influence propagation; (3) studies on multi-influences propagation; (4) studies on rumor blocking; and (5) studies on reverse sampling technique.

2.1 Studies on Cyberbullying

The study of cyberbullying can be tracked back to [27] where authors explore cyberbullying and examining its potential to become as problematic as traditional bullying, particularly with society's increasing reliance on technology. Researchers mainly study the following key issues for cyberbullying [15]: (1) How to identify the cyberbullying? (2) What's differences from the

traditional bullying? (3) How to prevent the cyberbullying? Recently, Reynolds et al. [28] and Dadcar et al. [8] use the methods based on machine learning or deep learning to address the issue (1). In [28], Reynolds et al. can detect language patterns used by bullies and their victims, and develop rules to automatically detect cyberbullying content through machine learning. Dadvar et al. [8] investigate the findings of a recent literature [1] and validate their findings using the same datasets. They further expand the work by applying the developed methods on a new datase. Their rusults show that the deep learning based models outperform the machine learning models previously applied to the same dataset. For the issues (2) and (3), Hinduja et al. [15] and Slonje et al. [29] analyze the differences between cyberbullying and traditional bullying, and give suggestions and measures to prevent cyberbullying in schools.

2.2 Studies on Single Influence Propagation

Research involving the social influence has already been one of the hottest spots in the field of social network analysis over the past decade. Kempe et al. [17] first formulate the problem of *Influence Maximization* (IM) for the single influence propagation. To describe influence spreading over the network, they propose two influence propagation models: *Independent Cascade* (IC) and *Linear Threshold* (LT), respectively. In addition, they also prove that the objective functions are monotone increasing submodular under these two propagation models and consequently the classic *Hill Climbing* algorithm provides a $(1 - 1/e)$-approximation. However, the bottleneck of Kempe's approach is that performing Monte-Carlo simulations is very inefficient. After that, there have been substantial efforts in improving calculation influence spreads such as CELF [20], CELF++ [11], CGA [38], and LDAG [7].

2.3 Studies on Multi-Influences Propagation

Note that Kempe's previous study is based on single influence, i.e., they assume that there are only one kind of influence spreading in a network. However, many scenarios in reality are multiple influences simultaneously spreading in a social network. For example, a customer wants to buy a laptop and he can buy one from MAC, DELL, or ThinkPad. Here MAC, DELL, or ThinkPad can be viewed as multiple influences and they advertise to attract customers to buy their own products. Currently, multiple influences propagation has attracted wide attention from academia and industry such as [5], [39], [41], [33], [21], [3], [13], and [6].

Lu et al. [21] propose the *Comparative Independent Cascade* (Com-IC) model that covers the full spectrum of entity interactions from competition to complementarity. The novelty of this model is that users' adoption decisions depend not only on edge-level information propagation, but also on a node-level automaton. They study two optimization problems, *Self Influence Maximization* and *Complementary Influence Maximization* under the Com-IC model. Borodin et al. [3] extend original LT model to competitive setting. They show that for a broad family of competitive influence models, it is NP-hard to achieve an approximation that is better than a square root of the optimal solution. He et al. [13] study competitive influence propagation under the *Competitive Linear Threshold* (CLT) model and focus on the *Influence Blocking Maximization* (IBM) problem. They show that IBM's objective function is submodular under the CLT model. Instead of performing inefficient Monte Carlo simulations, they design a CLDAG algorithm to address this problem. Zhu et al. [41] present CIC model to characterize how different influences competing with others in a social network. And they propose *Minimum Cost Seed Set problem* to answer the question how an influence uses the minimum cost to choose seeds such that its influence spread can reach a desired threshold under the competitive environment.

2.4 Studies on Rumor Blocking

As a special case of multiple influences (rumor vs. truth) propagation models, the rumor blocking problem is one of the important applications. Budak et al. [5] study the notion of competing campaigns in a social network and address the problem of influence limitation where a "bad" campaign starts propagating from a certain node and use the notion of limiting campaigns to counteract the effect of misinformation. The problem it to identify a subset of individuals that need to be convinced to adopt the competing (or "good") campaign so as to minimize the number of people that adopt the "bad" campaign at the end of both propagation processes. However, they do not consider the influence to the specific target nodes. Tong et al. [33] study the following problem: given a budget k, the rumor blocking problem asks for k seed users to trigger the spread of a positive cascade such that the number of the users who are not influenced by rumor can be maximized. They present a randomized approximation algorithm which is efficient with respect to the running time.

2.5 Studies on Reverse Sampling Technique

A lot of existing literature shows that Monte Carlo simulation-based algorithms are inefficient for computing influence spreads. Fortunately, Borgs et al. [2] first propose a sampling technology named *Reverse Sampling* technique, which is used to estimate the objective function of IM problem. Their sampling technology mainly consists of two steps. In the first step, they randomly and uniformly select a node from the network as a starting node and traverse the other nodes on the transpose network[5] with a certain probability. Consequently, this operation generates a hypergraph, that is, a *Reverse Reachable* (RR) set. Repeat the operations sufficient times. In the second step, they develop a simple greedy strategy to pick the node that can cover the most RR sets in each iteration as the seed node. Their sampling technology reveals the fact that a node has a higher probability of being selected a seed node if the node appears in the multiple RR sets at the same time. However, the shortcoming of [2] is to determine an appropriate size for RR set. In other words, how many times of reverse sampling will be made such that the error as small as enough.

To make the RR set approach practically efficient, many followers have proposed a series of improved methods such as [32], [31], [37], [25], and [16]. For example, Tang et al. [32] propose a method named TIM which improves over sampling technology by a better analysis on the number of RR sets required to ensure the same theoretical bound in [2]. In addition, by improving the parameter estimation procedure, they also propose the TIM+ algorithm in [32]. After that, Tang et al. [31] employ a martingale analysis and design a better algorithm named IMM to improver over TIM/TIM+. Wang et al. [37] solve the problem of high memory consumption of TIM, TIM+, and IMM. They propose a lazy sampling technique and estimate a lower bound of optimal RR set size. They show their method speeds up IMM two orders of magnitude empirically. Nguyen et al. [25] propose two novel frameworks SSA as well as D-SSA and show that they are up to 1,200 times faster than IMM of [31]. Their frameworks work well since they are based on an innovative *Stop-and-Stare* strategy which they stop at exponential check points to verify (stare) if there is adequate statistical evidence on the solution quality. Huang et al. [16] solve inaccuracies in previously reported technical results on the accuracy and efficiency of SSA and D-SSA [25]. Furthermore, they also reveal anomalies in some other results and shed light on the behavior of SSA and D-SSA in settings not considered previously.

In this article, we study fighting against the cyberbullying problem. On the surface, this article is similar to [5] and [33], but they are inherently different. For example, differences include the following two levels between our article and [5]: (1) influence propagation model level.

[5]We say a G^T is the transpose network of G if and only if for an edge $(u, v) \in E$ in $G = (V, E)$, there is an edge $(v, u) \in E^T$ in $G^T = (V, E^T)$.

Budak et al. [5] propose two influence propagation models named MCICM and COICM, respectively. They assume a so-called *high-effectiveness property* and a delay parameter r. However, in this article, we assume that it can be detected as quickly as possible once the perpetrators spread negative influence. (2) Algorithm level. Instead of using *Reverse Sampling Technique*, they calculate the objective function through Monte Carlo simulation. In addition, they develop three heuristic strategies but do not provide corresponding theoretical analysis. As for the literature [33], on one hand, the difference of the information propagation model lies in which influence takes effect when two opposing influences (positive vs. negative) reach a certain node at the same time. They think negative influence works in [33]. It's exactly opposite in this article. On the other hand, the rumor blocking problem in [33] has no specific target node set.

In summary, we adopt a elaborate design based on reverse sampling technique and propose a stochastic algorithm. Furthermore, we try to provide theoretical analysis to ensure the effectiveness and efficiency of the porposed algorithm.

3 INFORMATION DIFFUSION MODEL

A social network is denoted by a directed graph $G = (V, E, p)$, where V is node set, $E \subseteq V \times V$ is edge set, and $p_{uv} \in p$ of the edge $e = (u, v) \in E$ is the probability that node u activates v. We call a node *active* if it adopts the information from other nodes, *inactive* otherwise. For the convenience of follow-up discussion, we first briefly introduce the IC model [17] and *CIC* [5], [33], [39], respectively.

IC model: Information (influence) propagation process unfolds in discrete time steps. The initial seed set is S_0. Let S_t denote the active nodes in time step t, and each node v in S_t has single chance to activate each inactive neighbor u through its out-edge (v, u) with probability p_{vu} at time step $t + 1$. But whether or not v succeeds, it cannot make any further attempts in subsequent rounds. Repeat this process until no more new nodes can be activated. Note that a node can only switch from inactive to active, but not in reverse direction.

CIC model: This model describes the situation where two competitive cascades (e.g., negative influence vs. positive influence, or rumor vs. truth) disseminate simultaneously in a social network. Let $\mathcal{N}$ (for negative cascade) and $\mathcal{P}$ (for positive cascade) be the two competitive cascades. $S_{\mathcal{N}}$ and $S_{\mathcal{P}}$ represent the sets of initially *active* nodes (seed sets) in cascades $\mathcal{N}$ and $\mathcal{P}$. The way of each node u in $\mathcal{N}$ or $\mathcal{P}$ disseminating influence is similar to the *IC* model [17]. More specifically, if a node u becomes *active* in $\mathcal{N}$ or $\mathcal{P}$ at time step t, it has a single chance to activate each currently *inactive* neighbor v through the directed edge (u, v) with probability p_{uv}. The state of node v can be determined at time step $t + 1$. The process continues until there is no newly activated node in either cascade. Furthermore, the state of a node in $\mathcal{N}$ or $\mathcal{P}$ can only be switched from the *inactive* to *active*. Note that a node will adopt the influence of arriving first. In particular, when a node is activated by *active* nodes in both cascades $\mathcal{N}$ and $\mathcal{P}$ at the same time, the positive cascade $\mathcal{P}$ takes effect.

Figure 2 illustrates the propagation of two cascades $\mathcal{N}$ and $\mathcal{P}$ in a network. For ease of exposition, we let all the propagation probabilities be 1. In this case, the propagation process is determined[6] for cascades $\mathcal{N}$ and $\mathcal{P}$. At time step $t = 0$, we let $S_{\mathcal{N}}$ and $S_{\mathcal{P}}$ be the initial seed sets of $\mathcal{N}$ and $\mathcal{P}$, respectively. In other words, $S_{\mathcal{N}}$ and $S_{\mathcal{P}}$ are *active* while other nodes are *inactive*. And they attempt to activate their own *inactive* neighbors. At time step $t = 1$, on one hand, $S_{\mathcal{N}}$ and $S_{\mathcal{P}}$ activate their exclusive neighbors successfully since the propagation probabilities are 1. On the other hand, $S_{\mathcal{N}}$ and $S_{\mathcal{P}}$ simultaneously activate the common neighbor. Consequently, the common neighbor is successfully activated by $S_{\mathcal{P}}$ based on the rules in the CIC model. At time step $t = 2$, cascades continue to spread. After time step $t = 3$, no new nodes can become *active*. Therefore the propagation cascades stop.

[6]In practice, the influence propagation is a random process.

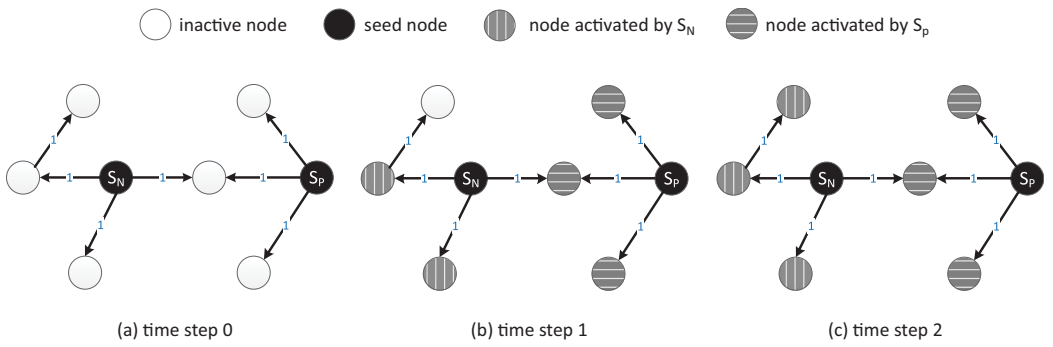


Fig. 2. An illustration of the CIC model.

4 PRELIMINARIES AND PROBLEM STATEMENT

4.1 Preliminaries

Submodular function: Given a ground set V, a function $f : 2^V \rightarrow \mathbb{R}$ is said to be submodular if $f(A \cup \{v\}) - f(A) \geq f(B \cup \{v\}) - f(B)$ for all $v \in V \backslash B$ and sets $A \subseteq B \subseteq V$.

Monotone function: A function f is monotone if $f(A \cup v) \geq f(A)$ for all $v \in V$ and $A \subseteq V$.

Nemahauser et al. [24] show the solution of following optimization problem has a nice theoretical guarantee.

$$\max\{f(A) : |A| = k, A \subseteq V\}. \tag{1}$$

If the function $f(\cdot)$ is non-negative, submodular and monotone, we use the greedy *Hill Climbing* algorithm and repeatedly add the element from V which provides the maximum marginal gain. Solve the following optimization problem until it is satisfied $|A| = k$.

$$\max\{f(A \cup \{v\}) : v \in V \backslash A\}. \tag{2}$$

The greedy *Hill Climbing* yields a $(1 - 1/e)$-approximation.

Let A^* be the optimal solution of Equation (1) and A' is returned by greedy *Hill Climbing* algorithm, then $f(A') \geq (1 - 1/e) f(A^*)$. Furthermore, Kempe et al. [17] extend this result and indicate that we can obtain a $(1 - 1/e - \varepsilon)$-approximation for any ε and there is $\delta > 0$ such that employing a $(1 + \delta)$-approximation of $f(\cdot)$ in Equation (2).

Chernoff Bound: Let X_i be l independent and identically distributed random variables and $\mathbb{E}[X_i] = \mu$. The Chernoff Bound [23] states two inequalities:

$$Pr\left[\sum X_i - l \cdot \mu \geq \delta \cdot l \cdot \mu\right] \leq exp\left\{-\frac{l \cdot \mu \cdot \delta^2}{2 + \delta}\right\}, \tag{3}$$

$$Pr\left[\sum X_i - l \cdot \mu \leq -\delta \cdot l \cdot \mu\right] \leq exp\left\{-\frac{l \cdot \mu \cdot \delta^2}{2}\right\}, \tag{4}$$

where $0 < \delta < 1$ and $Pr[A]$ is the probability of event A.

4.2 Problem Statement

A directed social network is denoted by a graph $G = (V, E, p)$, where V is the node set, E is edge set, and p is propagation probability set. Let P be a seed set for perpetrators and b be a positive integer budget. Let $\mathcal{T}$ denote a cybervictim set and $S_{\mathcal{W}}$ denote the seed set of witnesses.

Problem 1. *Fighting against the cyberbullying problem is to identify a seed set $S_{\mathcal{W}} \subseteq V \backslash \{P \cup \mathcal{T}\}$ as witness set with a budget restriction b such that the number of cybervictim nodes in $\mathcal{T}$ being activated by P is minimized.*

Let $Inf(P)$ denote the influence set of P in the target set $\mathcal{T}$ without $S_{\mathcal{W}}$, that is, the set of nodes in $\mathcal{T}$ that would accept influence introduced by P if there were no limiting. We define a function $\Phi(S_{\mathcal{W}})$ to be the size of the subset of $Inf(P)$ that $S_{\mathcal{W}}$ prevents nodes in $\mathcal{T}$ from adopting influence introduced by P. Inspired by [5], the minimizing the number of cybervictims in $\mathcal{T}$ being activated by perpetrators set P is equivalent to selecting a set $S_{\mathcal{W}}$ such that the expectation of $\Phi(S_{\mathcal{W}})$ is maximized. Therefore, we focus on maximizing version as follow.

Problem 2. *Let $\Phi(S_{\mathcal{W}})$ be the expected number of cybervictims that are successfully activated by witnesses where $S_{\mathcal{W}}$ is the seed set of witnesses. Given a positive integer budget b, the Problem 2 is to find a seed set $S_{\mathcal{W}}$ with at most b nodes such that $\Phi(S_{\mathcal{W}})$ is maximized.*

Regarding Problem 1 and Problem 2, we have the following hardness result.

Theorem 1. *Problem 1 and Problem 2 are NP-hard.*

Proof. The problems of [5], [33] are special cases of our Problem 1 and Problem 2 if the cybervictim set $\mathcal{T} = V \backslash P$. The formers are NP-hard, the latters are also NP-hard. □

5 ALGORITHMS

In this section, we first introduce the definition of *Sample Result* based on *Reverse Sampling Technique*, and estimate the objective function effectively. And then we propose the algorithms and demonstrate their theoretical properties, respectively.

5.1 Sample Result

In this part, we introduce the definition of *Sample Result* which is generated by *Reverse Sampling Technique* and discuss its properties.

Definition 1 (Sample Result). Given a directed social network $G = (V, E, p)$, a *sample result* (samlpe for short) is generated according to the following random rule:

- One sample $\mathcal{X} = (V(\mathcal{X}), E(\mathcal{X}), p(\mathcal{X}))$ is a special subgraph of the original graph $G = (V, E, p)$, where $V(\mathcal{X}) \subseteq V$, $E(\mathcal{X}) \subseteq E$.
- $E(\mathcal{X})$ is generated as follow. For each edge $e \in E$ in the original graph G, we generate a random number $\xi_e \in [0, 1]$. If $\xi_e \leq p_e$ where $e \in E$, then we add this edge e to the $E(\mathcal{X})$ in the sample $\mathcal{X}$ and let $p_e = 1$ for this edge.
- Let Ω be all possible samples space of G. Obviously, there are $2^{|E|}$ samples in Ω.

We now define the probability of a sample $\mathcal{X}$ generated by the above definition as follow:

$$Pr[\mathcal{X}] = \prod_{e \in E(\mathcal{X})} p_e \times \prod_{e \in E \backslash E(\mathcal{X})} (1 - p_e). \tag{5}$$

Notice that one sample $\mathcal{X}$ which is a deterministic IC network since all the $p_e = 1$ for all $e \in E(\mathcal{X})$. However, the IC network is randomized. In other words, the randomness of the IC network is reflected in the process of generating samples. Once a sample has been generated, the information propagating over the sample is deterministic. And This is precisely the key to the success of the reverse sampling technique [2]. So far, many studies such as [17], [33] have clearly pointed out that the following fact is correct.

Fact 1. *Let P be the seed set of perpetrators and $S_{\mathcal{W}}$ be the seed set of witnesses, respectively. Let $\mathcal{X}$ denote a random sample result. We consider the following two propagation processes.*

—*Perpetrators and witnesses perform the random propagation processes simultaneously on the original IC network G with their own seed set P and $S_{\mathcal{W}}$, respectively.*
—*Perpetrators and witnesses perform the deterministic propagation processes simultaneously on the sample result $\mathcal{X}$ with theire own P and $S_{\mathcal{W}}$, respectively.*

The above two propagation processes are equivalent.

Based on the sample result $\mathcal{X}$, we define the number that a node v in cybervims set $\mathcal{T}$ can be activated by the witness nodes in $S_{\mathcal{W}}$ as follow

$$F_{\mathcal{X}}(v, S_{\mathcal{W}}) = \begin{cases} 1, & \text{if } v \text{ is activated by } S_{\mathcal{W}} \text{ in } \mathcal{X} \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

Now we formulate the objective function of *Problem 2* as follows:

$$\Phi(S_{\mathcal{W}}) = \sum_{\mathcal{X} \in \Omega} \sum_{v \in \mathcal{T}} Pr[\mathcal{X}] \cdot F_{\mathcal{X}}(v, S_{\mathcal{W}}),$$

where $Pr[\mathcal{X}]$ indicates the probability of sample $\mathcal{X}$. Our goal is to maximize $\Phi(S_{\mathcal{W}})$ with restriction of $|S_{\mathcal{W}}| \leq b$, that is,

$$\max\{\Phi(S_{\mathcal{W}}) : |S_{\mathcal{W}}| \leq b\}. \tag{7}$$

For the function $\Phi(S_{\mathcal{W}})$, we have the following theoretical result.

Theorem 2. *The function $\Phi(S_{\mathcal{W}})$ is monotone and submodular with respect to $S_{\mathcal{W}}$ in a sample $\mathcal{X}$.*

Before giving detailed proof, we first provide the following observation.

There a sample $\mathcal{X}$ with a perpetrator set P and a seed set $S_{\mathcal{W}}$ of witness. For two nodes $u, v \in \mathcal{X}$, let $d_{\mathcal{X}}(u, v)$ be the length of the shortest path between u and v. Let $d_{\mathcal{X}}(P, v)$ be the length of the shortest path between P and v. i.e., $d_{\mathcal{X}}(P, v) = \min_{v' \in P} d_{\mathcal{X}}(v', v)$. Let $d_{\mathcal{X}}(S_{\mathcal{W}}, v)$ be the length of the shortest path between $S_{\mathcal{W}}$ and v, i.e., $d_{\mathcal{X}}(S_{\mathcal{W}}, v) = \min_{v'' \in S_{\mathcal{W}}} d_{\mathcal{X}}(v'', v)$.

Observation 1. *For a node v in the sample $\mathcal{X}$, it will be successfully activated by $S_{\mathcal{W}}$ if and only if the $d_{\mathcal{X}}(S_{\mathcal{W}}, v) \leq d_{\mathcal{X}}(P, v)$ and $d_{\mathcal{X}}(P, v) \neq +\infty$.*

Proof of Observation 1. We prove this observation by contradiction. Suppose there is node $v \in \mathcal{X}$ satisfying

$$d_{\mathcal{X}}(S_{\mathcal{W}}, v) \leq d_{\mathcal{X}}(P, v), \tag{8}$$

but v could not be activated by $S_{\mathcal{W}}$. Without loss of generality, suppose a node $v'' \in S_{\mathcal{W}}$ such that $d_{\mathcal{X}}(v'', v) = d_{\mathcal{X}}(S_{\mathcal{W}}, v)$, we have

$$d_{\mathcal{X}}(v'', v) \leq d_{\mathcal{X}}(P, v) \tag{9}$$

for the node $v \in \mathcal{T}$. Then, there must exist a node $v' \in P$ such that the distance between v' and v is much smaller the distance between v'' and v, i.e.,

$$d_{\mathcal{X}}(v', v) < d_{\mathcal{X}}(v'', v). \tag{10}$$

Combining Equations (9) and (10), we have

$$d_{\mathcal{X}}(v', v) < d_{\mathcal{X}}(v'', v) = d_{\mathcal{X}}(S_{\mathcal{W}}, v) < d_{\mathcal{X}}(P, v). \tag{11}$$

The inequality Equation (11) indicates $d_{\mathcal{X}}(v', v) < d_{\mathcal{X}}(P, v)$, which contradicts with the definition of $d_{\mathcal{X}}(P, v)$. Therefore, the above observation immediately holds. □

We now proof the Theorem 2 based on the Observation 1.

Proof of Theorem 2. We first show the *Monotone non-decreasing*. The function $F_{\mathcal{X}}(v, S_{\mathcal{W}})$ is monotonically non-decreasing because $F_{\mathcal{X}}(v, S_{\mathcal{W}})$ does not decrease as the size of the seed node set $S_{\mathcal{W}}$ increases in a sample result $\mathcal{X}$. Therefore $\Phi(S_{\mathcal{W}})$ is also monotone non-decreasing.

We then show the *Submoudularity*. For a sample result $\mathcal{X}$, we only need to prove that the following inequality holds

$$F_{\mathcal{X}}(A \cup \{u\}) - F_{\mathcal{X}}(A) \geq F_{\mathcal{X}}(B \cup \{u\}) - F_{\mathcal{X}}(B) \tag{12}$$

for any $A \subseteq B \subseteq V(\mathcal{X})$ and $u \in V(\mathcal{X}) \backslash B$.

Since function $F_{\mathcal{X}}(A \cup \{u\}$ is either 0 or 1. We discuss the following two cases, respectively.

— *CASE 1.* $F_{\mathcal{X}}(B \cup \{u\}) - F_{\mathcal{X}}(B) = 0$. This case can be further divided into following two subcases:
  1) $F_{\mathcal{X}}(B \cup \{u\}) = 1$ and $F_{\mathcal{X}}(B) = 1$. $F_{\mathcal{X}}(B) = 1$ indicates that a node $v \in \mathcal{T}$ can be successfully activated by B. Since $A \subseteq B$, then $F_{\mathcal{X}}(A \cup \{u\}) - F_{\mathcal{X}}(A) \geq 0$.
  2) $F_{\mathcal{X}}(B \cup \{u\}) = 0$ and $F_{\mathcal{X}}(B) = 0$. $F_{\mathcal{X}}(B \cup \{u\}) = 0$ can derive that $F_{\mathcal{X}}(A \cup \{u\}) = 0$ because $A \subseteq B$. Therefore $F_{\mathcal{X}}(A \cup \{u\}) - F_{\mathcal{X}}(A) = 0$.

  In summary, $F_{\mathcal{X}}(A \cup \{u\}) - F_{\mathcal{X}}(A) \geq F_{\mathcal{X}}(B \cup \{u\}) - F_{\mathcal{X}}(B)$ for *CASE 1*.

— *CASE 2.* $F_{\mathcal{X}}(B \cup \{u\}) - F_{\mathcal{X}}(B) = 1$. This case indicates $F_{\mathcal{X}}(B \cup \{u\}) = 1$ and $F_{\mathcal{X}}(B) = 0$ because $F(\cdot)$ is non-negative. Furthermore, $F_{\mathcal{X}}(B \cup \{u\}) = 1$ and $F_{\mathcal{X}}(B) = 0$ show that the node $v \in \mathcal{T}$ can not be activated by B but it can be activated by u. Therefore, $F_{\mathcal{X}}(A \cup \{u\}) = 1$ if we add the node u into A. On the other hand, $F_{\mathcal{X}}(B) = 0$ indicates that all nodes in B can not activate the node v. Therefore $F_{\mathcal{X}}(A) = 0$ since $A \subseteq B$. Based on the above analysis, we have $F_{\mathcal{X}}(A \cup \{u\}) - F_{\mathcal{X}}(A) = 1$ when $F_{\mathcal{X}}(B \cup \{u\}) - F_{\mathcal{X}}(B) = 1$.

CASE 1 and *CASE 2* reveal the correctness of inequality Equation (12). As we all know, the linear combination of the submodular functions is still a submodular function. Finally, $\Phi(S_{\mathcal{W}})$ is submodular. □

We now propose Algorithms 1 and 2 to randomly generate a sample result $\mathcal{X} = (V(\mathcal{X}), E(\mathcal{X}))$ starting from a node v.

ALGORITHM 1: One-step sampling from v

Input: Reverse network G^T, the perpetrator set P, the cybervictim set $\mathcal{T}$, a start node v, a queue Q.
Output: One-step sampling result $v[\mathcal{X}]$ and Q.

```
1 Q ← ∅;
2 for each edge (v, u) ∈ G^T do
3     if ξ ≤ p_uv and u ∉ P then
4         V(X) ← V(X) ∪ {u};
5         Q ← Q ∪ {u};
6         E(X) ← E(X) ∪ {(u, v)};
7     end
8 end
9 return v[X] = (V(X), E(X)) and Q
```

In Algorithm 1, we maintain a queue Q to record the sampled one-step neighbors set that is from the start node in transpose graph G^T. Initially, we assume that node v is the starting node. Then we perform random sampling in transposed graph G^T and obtain one-step neighbors (lines 2–8). At the same time, we put the obtained one-step neighbors into the queue Q in any order (Line 5). Finally, Algorithm 1 returns the one-step neighbors $v[\mathcal{X}]$ and queue Q.

ALGORITHM 2: Generating a reverse sample $\mathcal{X} = (V(\mathcal{X}), E(\mathcal{X}))$ of the node v

Input: A directed social network $G = (V, E, p)$, the perpetrator set P, the cybervictim set $\mathcal{T}$, a queue Q.
Output: A sample $\mathcal{X} = (V(\mathcal{X}), E(\mathcal{X}))$.

Generate the transpose network of G, that is, G^T;
$V(\mathcal{X}) \leftarrow \emptyset$, $E(\mathcal{X}) \leftarrow \emptyset$, $Q \leftarrow \emptyset$;
Select a node v in $\mathcal{T}$ with probability $1/|\mathcal{T}|$ in G^T;
$V(\mathcal{X}) \leftarrow \{v\}$;
$Q \leftarrow \{v\}$;
while $Q \neq \emptyset$ **do**
  for *each node* $u \in V(\mathcal{X})$ **do**
    if $u \notin P$ *and* u *can continue to traverse in* G^T **then**
      $u[\mathcal{X}] \leftarrow$ Call Algorithm 1 with the starting node u;
      Update $V(\mathcal{X}) \leftarrow V(\mathcal{X}) \cup u[\mathcal{X}]$;
      Update $E(\mathcal{X}) \leftarrow E(\mathcal{X}) \cup u[\mathcal{X}]$;
      Update $Q \leftarrow Q \cup u[\mathcal{X}]$;
    end
    Delete the u from Q, that is, $Q \leftarrow \{Q - u\}$;
  end
end
return $\mathcal{X} = (V(\mathcal{X}), E(\mathcal{X}))$.

Algorithm 2 is the framework for generating a sample result that includes two phases. In the first phase, we randomly and uniformly select a node v from the cybervictim set $\mathcal{T}$ as the starting node in transpose network G^T. In the second phase, we recursively and iteratively call Algorithm 1 to extend the sample result $\mathcal{X}$. The algorithm stops when it encounters a perpetrator node or it cannot continue sampling. It's worth noting that the essence of a sample $\mathcal{X} = (V(\mathcal{X}), E(\mathcal{X}))$ of a node v reveals the following conclusions: (1) $V(\mathcal{X})$ shows that nodes set that can reach v in original network G; and (2) $E(\mathcal{X})$ shows that corresponding edges set among nodes in $V(\mathcal{X})$. Furthemore, we also need to pay attention to the maintenance and update of queue Q.

Considering a set $S_{\mathcal{W}} \subseteq V \backslash \{P \cup \mathcal{T}\}$, we define

$$\theta_{\mathcal{X}}(S_{\mathcal{W}}, V(\mathcal{X})) = \begin{cases} 1, & \text{if } S_{\mathcal{W}} \cap V(\mathcal{X}) \neq \emptyset \\ 0, & \text{otherwise.} \end{cases} \tag{13}$$

Based on Equation (13), we have following theorem.

Theorem 3. *For a set* $S_{\mathcal{W}} \subseteq V \backslash \{P \cup \mathcal{T}\}$ *and a sample* $\mathcal{X}$, $\Phi(S_{\mathcal{W}}) = |\mathcal{T}| \cdot \mathbb{E}[\theta_{\mathcal{X}}(S_{\mathcal{W}}, V(\mathcal{X}))]$.

Proof of Theorem 3.

$$\begin{aligned} \Phi(S_{\mathcal{W}}) &= \sum_{u \in \mathcal{T}} \sum_{\mathcal{X} \in \Omega} Pr[\mathcal{X}] \cdot F_{\mathcal{X}}(u, S_{\mathcal{W}}) \\ &= \sum_{u \in \mathcal{T}} \sum_{\mathcal{X} \in \Omega} Pr[\exists v \in S_{\mathcal{W}} \textbf{ such that } v \in V(\mathcal{X})] \\ &= \sum_{u \in \mathcal{T}} \sum_{\mathcal{X} \in \Omega} Pr[S_{\mathcal{W}} \cap V(\mathcal{X}) \neq \emptyset] \\ &= \sum_{u \in \mathcal{T}} \mathbb{E}[\theta_{\mathcal{X}}(S_{\mathcal{W}}, V(\mathcal{X}))] \\ &= |\mathcal{T}| \cdot \mathbb{E}[\theta_{\mathcal{X}}(S_{\mathcal{W}}, V(\mathcal{X}))]. \end{aligned}$$

where $\mathbb{E}[\cdot]$ denote the expected operator. □

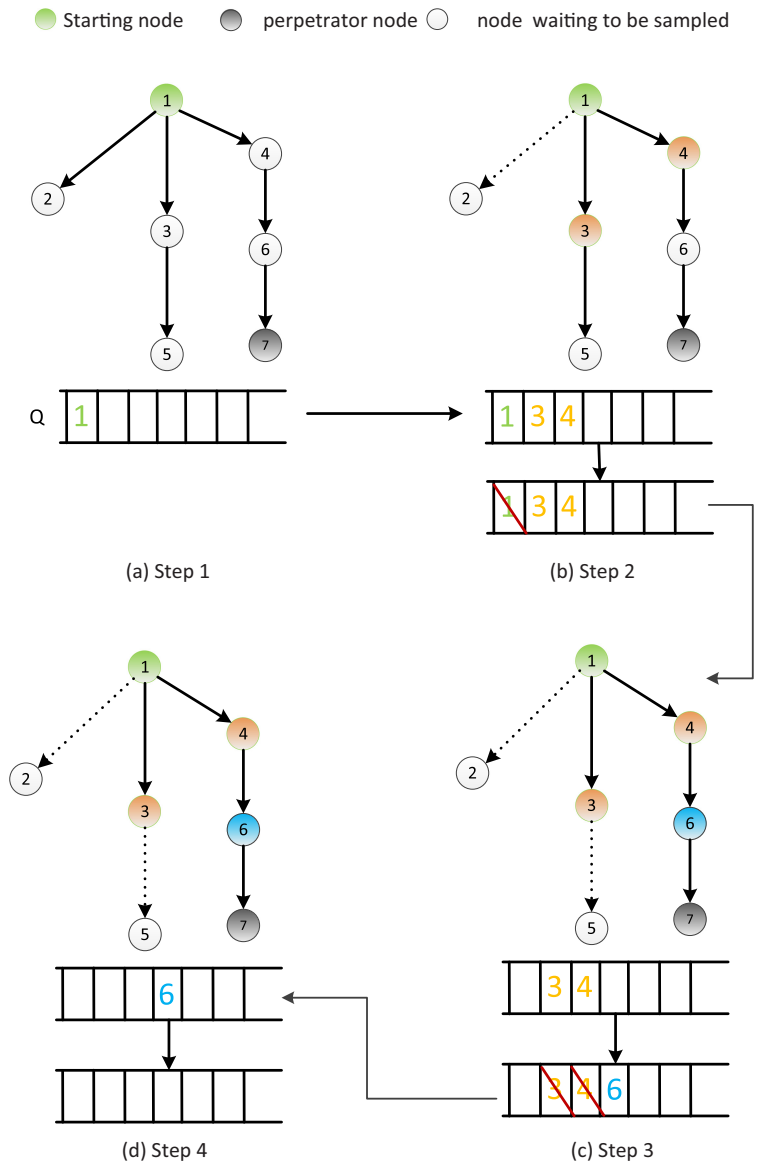


Fig. 3. An illustration of Algorithms 1 and 2.

For ease of explanation and understanding of the Algorithms 1 and 2, we give an example to illustrate their working phases in Figure 3. In the figure, the green and black nodes represent the starting node and the perpetrator node respectively, and other nodes are waiting to be sampled. In Figure 3(a), we assume that the graph contains seven nodes where node 1 is the starting node and node 7 is the perpetrator node. According to the algorithms, we put the node 1 into the queue Q in this time. Then we perform the random reverse sampling based on reverse edges in Figure 3(b). As a result, nodes 3 and 4 are successfully sampled and added them into the queue Q. So far, we have completed the reverse sampling phase of node 1, so node 1 needs to be deleted from queue Q. In Figure 3(c), we recursively perform similar reverse sampling operations on nodes 3 and 4, respectively. Consequently, node 6 is successfully sampled. We put the node 6 into the queue Q

and delete node 3 as well as node 4 from Q. After that, node 6's reverse sampling encounters the perpetrator node 7, so the reverse sampling of node 6 will be stopped. At this time, node 6 is removed from the queue Q such that Q is empty in Figure 3(d). The entire reverse sampling process stops and the sampling result are: (1) $V(\mathcal{X}) = \{1, 3, 4, 6\}$; and (2) $E(\mathcal{X}) = \{(3, 1), (4, 1), (6, 4)\}$).

5.2 Estimating $\Phi(\cdot)$

In this section, we discuss how to estimate the objective function value $\Phi(\cdot)$. Assume that we perform sampling procedures and obtain a set $\mathbb{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_l\}$ including l random samples returned by Algorithm 2. For a set $S_{\mathcal{W}} \subseteq V \backslash \{P \cup \mathcal{T}\}$ and $\mathbb{X}$, let $g(S_{\mathcal{W}}, \mathbb{X}) = \sum_{i=1}^{l} \theta_{\mathcal{X}_i}(S_{\mathcal{W}}, V(\mathcal{X}_i))$. We consider the following problem.

Problem 3. *How to select a seed set $S_{\mathcal{W}} \subseteq V \backslash \{P \cup \mathcal{T}\}$ with restriction of $|S_{\mathcal{W}}| \leq b$ such that $g(S_{\mathcal{W}}, \mathbb{X})$ is maximized?*

Problem 3 is a classic set cover set problem and using greedy *Hill Climbing* algorithm in [24] that can provide a $(1 - 1/e)$-approximation. We propose the seed nodes selection algorithm based on [24] as follow.

ALGORITHM 3: Seed nodes selection

Input: A sample set $\mathbb{X} = \{\mathcal{X}_1, \mathcal{X}_2, ..., \mathcal{X}_l\}$ and b.
Output: Seed set $S'_{\mathcal{W}}$.

```
S'_W ← ∅;
for 1 ≤ i ≤ b do
    Let v denote the node covering the most subsets of X;
    S'_W ← S'_W ∪ {v};
    Remove the v from each X_i;
end
return S'_W.
```

Theorem 4. *Let $S'_{\mathcal{W}} \subseteq V \backslash \{P \cup \mathcal{T}\}$ be the seed set returned by Algorithm 3. For a given sample set $\mathbb{X}$, the following conclusion holds [24],*

$$g(S'_{\mathcal{W}}, \mathbb{X}) \geq (1 - /e) \cdot g(S_{\mathcal{W}}, \mathbb{X}), \tag{14}$$

where $S_{\mathcal{W}}$ is the optimal seed set.

Let $\Phi(S_{\mathcal{W}})$ be the optimal value and $S_{\mathcal{W}}$ be the optimal seed set in (7). According to Theorem 3, $|\mathcal{T}| \cdot \mathbb{E}[\theta_{\mathcal{X}}(S_{\mathcal{W}}, V(\mathcal{X}))]$ is a good way to estimate the $\Phi(S_{\mathcal{W}})$ where $\mathbb{E}[\theta_{\mathcal{X}}(S_{\mathcal{W}}, V(\mathcal{X}))] = \frac{g(S_{\mathcal{W}}, \mathbb{X})}{l}$. On the other hand, by Theorem 4, $\frac{g(S'_{\mathcal{W}}, \mathbb{X})}{l} \geq (1 - 1/e) \cdot \frac{g(S_{\mathcal{W}}, \mathbb{X})}{l}$ that indicates if we choose the $S'_{\mathcal{W}}$ as the seed set with a constant factor guarantee. Therefore, our goal is to estimate the $\frac{g(S'_{\mathcal{W}}, \mathbb{X})}{l}$ and consequently we can estimate $\Phi(S_{\mathcal{W}})$. Based on this idea, the algorithm is proposed in Algorithm 4 (see Algorithm 4 for details).

In Algorithm 4, we search the $\frac{g(S'_{\mathcal{W}}, \mathbb{X})}{l}$ from interval $[1, |\mathcal{T}|]$ since $\mathcal{T} \neq \emptyset$ and one can activate at most all nodes in $\mathcal{T}$. In each iteration, we first call Algorithm 3 to select seed set $S'_{\mathcal{W}}$ and compute the $\frac{g(S'_{\mathcal{W}}, \mathbb{X})}{l}$. Then we compare $\frac{g(S'_{\mathcal{W}}, \mathbb{X})}{l}$ with $temp_i$. Algorithm stops when they are sufficiently close to each other by a parameter δ $(0 < \delta < 1)$. Theoretical results are shown in Theorems 5 and 6, respectively.

Theorem 5. *If $\Phi(S_{\mathcal{W}}) \geq \frac{(1+\delta)^2}{1-1/e} \cdot temp_i$, then the probability of $\frac{|\mathcal{T}| \cdot g(S'_{\mathcal{W}}, \mathbb{X})}{l} \leq (1 + \delta) \cdot temp_i$ is at most $\frac{1}{C \cdot \binom{|V|-|\mathcal{T}-|P|}{b}}$ where $C > 0$, $temp_i = \frac{|\mathcal{T}|}{2^i}$ and $i \in \{1, 2, ..., \log_2(|\mathcal{T}| - 1)\}$.*

ALGORITHM 4: Estimation $\Phi(S'_{\mathcal{W}})$

Input: A sample set $\mathbb{X} = \{\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_l\}$, the budget b and a parameter δ where $0 < \delta < 1$.
Output: $\Phi(S'_{\mathcal{W}})$.

```
for 1 ≤ i ≤ log_2(|T| − 1) do
    Let temp_i ← |T| / 2^i;
    S'_W ← call Algorithm 3;
    Compute g(S'_W, X) / l;
    if |T| · g(S'_W, X) / l ≥ (1 + δ) · temp_i then
        Φ(S'_W) ← |T| · g(S'_W, X) / (l · (1 + δ));
    end
end
return Φ(S'_W).
```

Proof of Theorem 5.

$$
\begin{aligned}
&Pr\left[\frac{|\mathcal{T}| \cdot g(S'_{\mathcal{W}}, \mathbb{X})}{l} \leq (1+\delta) \cdot temp_i\right] \\
\leq\ &Pr\left[\frac{|\mathcal{T}| \cdot (1-1/e) \cdot g(S_{\mathcal{W}}, \mathbb{X})}{l} \leq (1+\delta) \cdot temp_i\right] \\
=\ &Pr\left[\frac{|\mathcal{T}| \cdot g(S_{\mathcal{W}}, \mathbb{X})}{l} \leq \frac{(1+\delta) \cdot temp_i}{1-1/e}\right] \\
\leq\ &Pr\left[\frac{|\mathcal{T}| \cdot g(S_{\mathcal{W}}, \mathbb{X})}{l} \leq \frac{(1+\delta)}{1-1/e} \cdot \frac{1-1/e}{(1+\delta)^2} \cdot \Phi(S_{\mathcal{W}})\right] \\
=\ &Pr\left[\frac{|\mathcal{T}| \cdot g(S_{\mathcal{W}}, \mathbb{X})}{l} \leq \frac{\Phi(S_{\mathcal{W}})}{1+\delta}\right] \\
=\ &Pr\left[g(S_{\mathcal{W}}, \mathbb{X}) - \frac{l \cdot \Phi(S_{\mathcal{W}})}{|\mathcal{T}|} \leq \frac{l \cdot \Phi(S_{\mathcal{W}})}{|\mathcal{T}|} \cdot \frac{-\delta}{1+\delta}\right] \\
\leq\ &exp\left\{-\frac{l \cdot \Phi(S_{\mathcal{W}}) \cdot (\frac{-\delta}{1+\delta})^2}{2}\right\} \\
\leq\ &exp\left\{-\frac{l \cdot \frac{(1+\delta)^2}{1-1/e} \cdot temp_i \cdot (\frac{-\delta}{1+\delta})^2}{2}\right\} \\
\leq\ &\frac{1}{C \cdot \binom{|V|-|\mathcal{T}|-|P|}{b}}.
\end{aligned}
$$

Where the last inequality is true since there are most $\binom{|V|-|\mathcal{T}|-|P|}{b}$ subsets of $V \backslash \{P \cup \mathcal{T}\}$. □

Theorem 6. *Algorithm 4 returns a $\Phi(S'_{\mathcal{W}})$ such that the following inequalities hold with a probability of $1 - \frac{1}{C \cdot \binom{|V|-|\mathcal{T}|-|P|}{b}}$, where $C > 0$.*

$$
\frac{(1-1/e) \cdot \Phi(S_{\mathcal{W}})}{2 \cdot (1+\delta)^2} \leq \Phi(S'_{\mathcal{W}}) \leq \Phi(S_{\mathcal{W}}). \tag{15}
$$

Table 1. Dataset Information

Dataset	Relationship type	# Node	# Node
Synthetic	Synthetic	1,000	5,000
Wiki Vote	Voting Relationship	7,115	103,663
Google+	Friendship	23,628	39,242
Epinions	Trust Relationship	75,897	508,837

Proof of Theorem 6 For the right inequality in Equation (15), it's obviously correct since the approximate value cannot be larger than the optimal value in terms of maximizing problem. We prove the left inequality in Equation (15) as follows.

Assume $|\mathcal{T}|/2^{i+1} \leq \frac{\Phi(S_W)\cdot(1-1/e)}{(1+\delta)^2} \leq |\mathcal{T}|/2^i$ for some i. We consider the following two cases:

— *CASE 1*: When the algorithm stops before the $(i+1)$-th iteration, according to Algorithm 4, then we have $\Phi(S'_{\mathcal{W}}) \geq |\mathcal{T}|/2^i \geq \frac{\Phi(S_W)\cdot(1-1/e)}{2\cdot(1+\delta)^2}$.

— *CASE 2*: When the algorithm executes the $(i+1)$-th iteration, according to Theorem 5, it will stop with a probability larger than $1 - \frac{1}{C\cdot\binom{|V|-|\mathcal{T}|-|P|}{b}}$. This case indicates that $\Phi(S'_{\mathcal{W}}) \geq |\mathcal{T}|/2^{i+1} \geq \frac{\Phi(S_W)\cdot(1-1/e)}{2\cdot(1+\delta)^2}$.

In summary, the inequalities Equation (15) stand immediately. □

6 EXPERIMENT

In this section, we evaluate proposed algorithm on synthetic and real-life networks. First, we describe the datasets and experiment setup. Second, we analyze and discuss experimental results from different perspectives. Finally, we compare with other existing popular approaches.

6.1 Datasets and Experiment Setup

We generate a random network and collect three real-life social networks from SNAP[7] and KONECT[8], respectively. See Table 1 for details.

Synthetic (SYN). We generate a random network that consists of 1,000 nodes and 5,000 edges.

Wiki Vote (WV). This network contains all the Wikipedia voting data from the inception of Wikipedia till January 2008. Nodes represent Wikipedia users and a directed edge (u, v) represents user u votes on user v. It has 7,115 nodes and 103,663 edges.

Google+ (G+). This directed network contains Google+ user-user links. A node represents a user, and a directed edge represents that one user has the other user in his circles. It has 23,628 nodes and 39,242 edges.

Epinions (EP). This directed network is the trust network from the online social network Epinions that includes user-user trust information. A node represents a user of Epinions. A directed edge represents trust between the users. It has 75,897 nodes and 508,837 edges.

Experiment Setup: Given a directed social network $G = (V, E, p)$, 3% of nodes are selected randomly and uniformly from V as the seed set for perpetrators P. And we then randomly and uniformly select various size of nodes in $V \backslash P$ as the cybervictim set. In our all experiments, we adopt the CIC model over reverse sampling results. Due to the lack of parameter p in G, we set propagation probability p in following ways: (1) $p = 0.5$ for each edge on network, and (2) $p = TRI$,

[7] http://snap.stanford.edu/data.
[8] http://konect.uni-koblenz.de.

i.e., we uniformly choose a value from $TRI = \{0.1, 0.5, 0.9\}$ for each edge. We simply set $\delta = 0.1$ and $C = |V|$. In addition, we set the number of samples l according to [2], [33].

Comparison Methods: We compare with the existing greedy methods (Target-IM (TIM) and Local Greedy Algorithm (LGA)) and heuristic methods (Out-Degree (OD), Betweenness Centrality (BC), and PageRank (PR)), respectively.

- —**TIM** [30] works in two phases: (1) sample the number of weighted RR trees; and (2) select k nodes that cover the most number of weighted RR trees generated in first phase.
- —**LGA** [12] first adds one node as a seed each round such that this node can maximize the marginal influence degree on the target node.
- — **OD** [17]. We select the top k nodes with the maximum out degree as the seed set of positive influence.
- —**BC** [4]. We select the top k nodes with the maximum betweenness as the seed set of positive influence.
- —**PR** [26]. We select the top k nodes with the maximum pagerank score as the seed set.

Evaluation Criteria: The evaluation criteria are the total expected number of the cybervictims and running time, respectively. When we select the same number of seed nodes in different methods, the more the number of cybervictims can activate, the better the method is. We run each algorithm 10 times and take average as the results of experiments.

6.2 Results

Estimating Algorithm 4: To evaluate Algorithm 4, we compare the greedy algorithms (TIM and LGA) and heuristic algorithms (OD, BC, and PR), respectively. We perform simulations according to the parameter settings in Figure 4. The average results are illustrated in Figure 4. In each subfigure, the horizontal axis is the parameter b and vertical axis is the total number of target nodes being activated by witnesses. The RAN method randomly selects a set of nodes as seeds. We have following observations: (1) as the b increases, the total number of target nodes being activated by witnesses increases. This is because function (7) is non-decreasing. It means that if we select more nodes for $S_{\mathcal{W}}$, then the more nodes in $\mathcal{T}$ will be activated by $S_{\mathcal{W}}$. (2) Greedy algorithms (TIM and LGA) are better than heuristic algorithms (OD, BC, and PR) with respect to the same number of seeds. The reason is that greedy algorithms have performance guarantees while heuristic algorithms do not. On one hand, for greedy algorithms, TIM is better than LGA. On the other hand, for heuristic algorithms, PR is the best while OD is the worst. It shows that PR method is more appropriate for measuring the importance of nodes compared to OD and BC methods. Furthermore, RAN method is the worst (the total number of nodes be activated by witnesses in $\mathcal{T}$ is minimized) compared to all methods. This is because randomly selecting the seeds cannot activate the target nodes with a higher probability. (3) Our Algorithm 4 is the best compared to all greedy algorithms and heuristic algorithms. Compared with TIM, our algorithm's sampling process is more efficient. Specifically, we reverse sample from the target nodes instead of arbitrary nodes in networks, which makes the seed set better. As for LGA, it only considers the local information of target nodes, which leads to smaller results. Compared with heuristic algorithms, our Algorithm 4 is much better. The reason is that the global influential nodes returned by heuristic algorithms sometimes cannot activate the target nodes if there is no paths between them.

The Parameter b vs. Cyberbullying activated nodes. We evaluate the relationship between budget b and the number of cybervictims being activated by cyberbullying. We set $b \in [0, 30]$, propagation probability $p = 0.5$ and the size of cybervictim set $|\mathcal{T}| = 150$ on Synthetic network. We set propagation probability $p = TRI$ and the size of cybervictim set $|\mathcal{T}| = 500$ on Wiki-Vote network. We execute the Algorithm 4. Figure 5 shows the results.

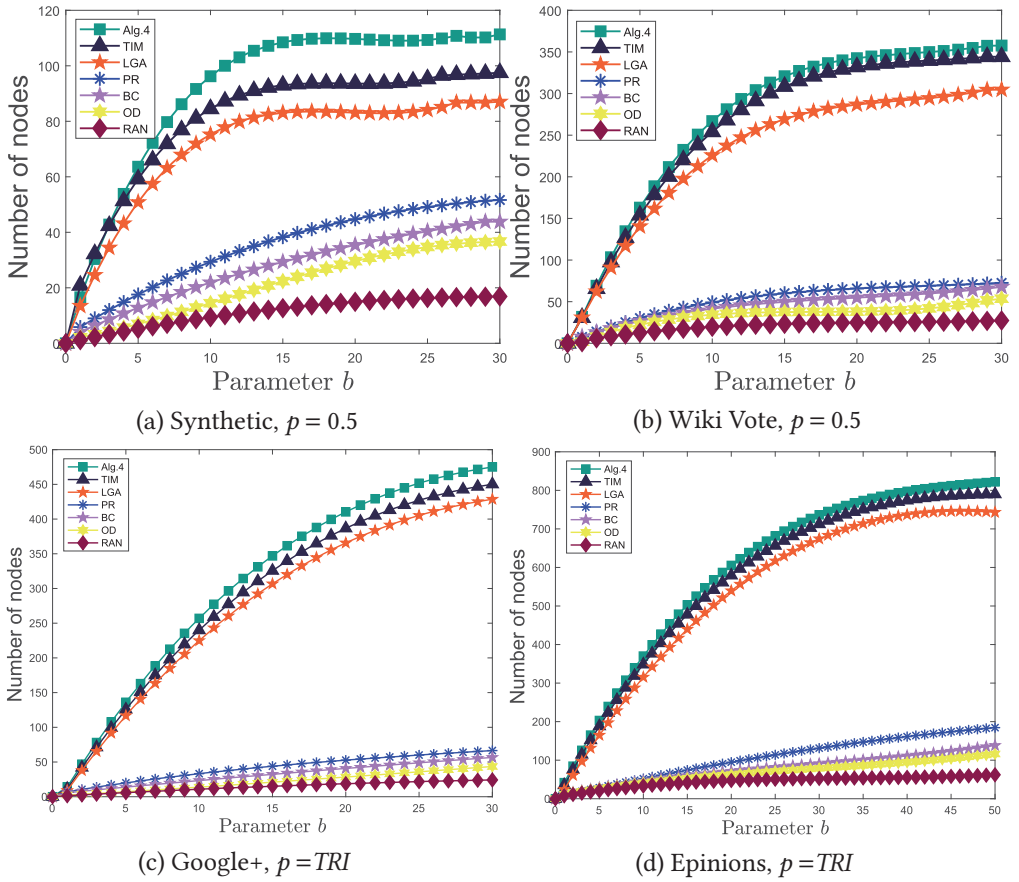
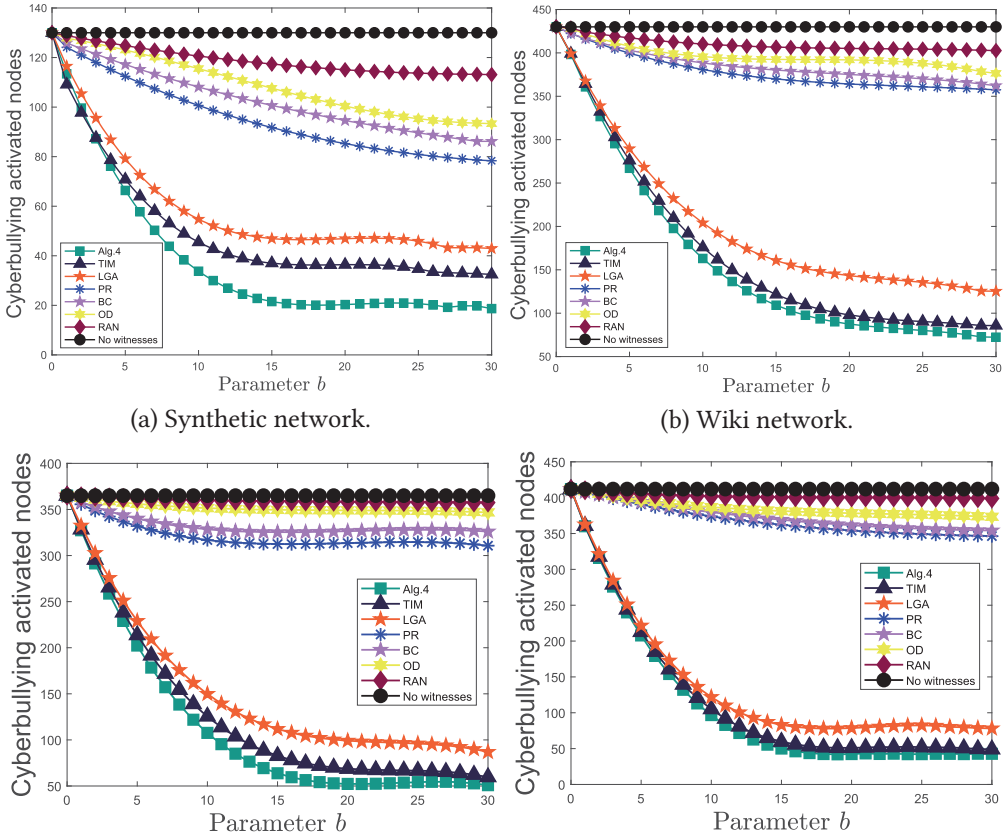


Fig. 4. Parameter b vs. the number of nodes in $\mathcal{T}$ being activated by witnesses : $p = 0.5$, $|\mathcal{T}| = 150$ on Synthetic; $p = 0.5$, $|\mathcal{T}| = 500$ Wiki Vote network; $p = TRI$, $|\mathcal{T}| = 1{,}000$ on Google+ network; $p = TRI$, $|\mathcal{T}| = 1{,}000$ on Epinions network.

In the figure, the vertical axis represents the number of nodes in $\mathcal{T}$ that are activated by cyberbullying. The horizontal axis represents seed size b. These two subfigures are corresponding to Figure 4(a) and (b), respectively. We have the following observations: (1) if there is without any witness nodes, the number of nodes in $\mathcal{T}$ that can be activated by cyberbullying is the maximum; (2) as the b increases, the number of nodes in $\mathcal{T}$ that can be activated by cyberbullying decreases; and (3) the results show that our algorithm minimizes the impact of cyberbullying because the number of nodes in $\mathcal{T}$ activated by cyberbullying is minimized when we choose the same number of seeds.

In addition, we also do similar experiments on the other two datasets: Google+ network and Epinions network. We set seed size $b \in [0, 30]$, propagation probability $p = 0.5$ and the size of cybervictim set $|\mathcal{T}| = 500$ on Google network. We set propagation probability $p = TRI$ and the size of cybervictim set $|\mathcal{T}| = 500$ on Epinions network. Then we execute the Algorithm 4. Figure 5(c) and (d) shows the results respectively. From the experimental results, we mainly have the following two observations: (1) the experimental results of the greedy algorithm are significantly better than the heuristic algorithms; and (2) compared with the greedy algorithms, our algorithm is the best.



(a) Synthetic network. (b) Wiki network.

(c) Google+ network. (d) Epinions network.

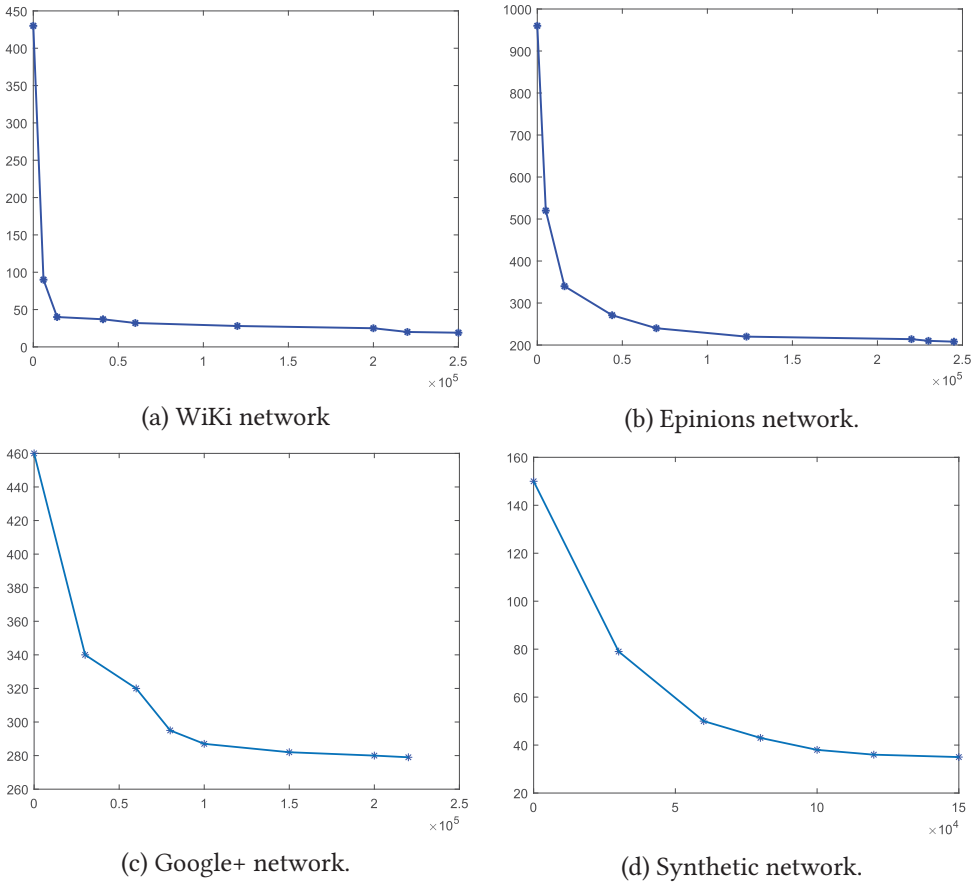
Fig. 5. The Parameter b vs. Cyberbullying activated nodes.

This is because under the condition of selecting the same number of seeds, the number of nodes being activated by cyberbullying is the minimized.

The size of reverse sampling: we experimentally test how the quality of the seed sets varies with the increase of used samples. In particular, we are interested in that whether or not the number of samples used by Algorithm 4 is proper. For each dataset, we increase the size of samples l until the quality of the produced seed set tends to converge. The results are given by Figure 6.

In Figure 6, the horizontal axis represents the number of reverse samples and the vertical axis represents the number of seeds. According to the figure, Algorithm 4 generates sufficient number of samples in practice for most of the considered datasets. For example, on Wiki network, Algorithm 4 totally generates 250K samples as shown in Figure 6(a), the quality of the seed set does not significantly increases when more than 250K samples have been used. For this case, 250K samples are sufficient and it cannot help improve to the quality anymore. We have the similar conclusions on the other three datasets. For example, on Google+ network, the number of reverse samples close to 200k is sufficient. However, on Synthetic network, this number is about 100K.

Running time: We have the following observations through experiments: (1) RAN method has the minimum running time. (2) The running time of heuristic algorithms (within 30 seconds) is much less than the running time of the greedy algorithms (hundreds of minutes). This is because



(a) WiKi network

(b) Epinions network.

(c) Google+ network.

(d) Synthetic network.

Fig. 6. The number of reverse sampling.

heuristic algorithms only select influential nodes based on certain heuristic criteria. However, the quality of the seed sets returned by the heuristic algorithms are very poor. (3) The running time of Algorithm 4 (within a few minutes) is between greedy algorithms and heuristic algorithms. Making a balance between efficiency and effectiveness, our Algorithm 4 is more appropriate.

7 CONCLUSIONS

In this article, we introduce a stochastic approach based on reverse sampling technique to maximize the number of target nodes being activated by witnesses, which automatically minimizes the negative influence of cyberbullying on the target node set. We show that problem is NP-hard and the objective is submodular. In addition, we provide theoretical analysis and discuss the relationship beween the optimal value and the one returned by the algorithm. At last, extensive experiments have been implemented to evaluate our proposed algorithms. The results demonstrate that our algorithm finds the seeds of witnesses in different networks with high quality. Furthermore, we also compare with several state-of-the-art methods. The results indicate our approach is superior to the comparison methods.

REFERENCES

[1] Sweta Agrawal and Amit Awekar. 2018. Deep learning for detecting cyberbullying across multiple social media platforms. In *Proceedings of the European Conference on Information Retrieval*. Springer, 141–153.

[2] Christian Borgs, Michael Brautbar, Jennifer Chayes, and Brendan Lucier. 2014. Maximizing social influence in nearly optimal time. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 946–957.

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

[4] Ulrik Brandes. 2008. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks* 30, 2 (2008), 136–145.

[5] 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.

[6] Tim Carnes, Chandrashekhar Nagarajan, Stefan M. Wild, and Anke Van Zuylen. 2007. Maximizing influence in a competitive social network: A follower's perspective. In *Proceedings of the 9th International Conference on Electronic Commerce*. ACM, 351–360.

[7] Wei Chen, Yifei Yuan, and Li Zhang. 2010. Scalable influence maximization in social networks under the linear threshold model. In *Proceedings of the 2010 IEEE International Conference on Data Mining*. IEEE, 88–97.

[8] Maral Dadvar and Kai Eckert. 2020. Cyberbullying detection in social networks using deep learning based models. In *Proceedings of the International Conference on Big Data Analytics and Knowledge Discovery*. Springer, 245–255.

[9] Maral Dadvar, Dolf Trieschnigg, and Franciska de Jong. 2014. Experts and machines against bullies: A hybrid approach to detect cyberbullies. In *Proceedings of the Canadian Conference on Artificial Intelligence*. Springer, 275–281.

[10] Lidan Fan, Zaixin Lu, Weili Wu, Bhavani Thuraisingham, Huan Ma, and Yuanjun Bi. 2013. Least cost rumor blocking in social networks. In *Proceedings of the 2013 IEEE 33rd International Conference on Distributed Computing Systems (ICDCS'13)*. IEEE, 540–549.

[11] Amit Goyal, Wei Lu, and Laks VS Lakshmanan. 2011. Celf++: Optimizing the greedy algorithm for influence maximization in social networks. In *Proceedings of the 20th International Conference on World Wide Web*. ACM, 47–48.

[12] Jing Guo, Peng Zhang, Chuan Zhou, Yanan Cao, and Li Guo. 2013. Personalized influence maximization on social networks. In *Proceedings of the 22nd ACM International Conference on Information & Knowledge Management*. ACM, 199–208.

[13] Xinran He, Guojie Song, Wei Chen, and Qingye Jiang. 2012. Influence blocking maximization in social networks under the competitive linear threshold model. In *Proceedings of the 2012 SIAM International Conference on Data Mining*. SIAM, 463–474.

[14] Sameer Hinduja and Justin W. Patchin. 2008. Cyberbullying: An exploratory analysis of factors related to offending and victimization. *Deviant Behavior* 29, 2 (2008), 129–156.

[15] Sameer Hinduja and Justin W. Patchin. 2014. Cyberbullying. Cyberbullying Research Center. Retrieved 7 September, 2015.

[16] Keke Huang, Sibo Wang, Glenn Bevilacqua, Xiaokui Xiao, and Laks VS Lakshmanan. 2017. Revisiting the stop-and-stare algorithms for influence maximization. *Proceedings of the VLDB Endowment* 10, 9 (2017), 913–924.

[17] David Kempe, Jon Kleinberg, and Éva Tardos. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 137–146.

[18] Elias Boutros Khalil, Bistra Dilkina, and Le Song. 2014. Scalable diffusion-aware optimization of network topology. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1226–1235.

[19] Masahiro Kimura, Kazumi Saito, and Hiroshi Motoda. 2008. Minimizing the spread of contamination by blocking links in a network. In *Proceedings of the 23rd AAAI Conference on Artificial Intelligence*. Vol. 8. 1175–1180.

[20] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne VanBriesen, and Natalie Glance. 2007. Cost-effective outbreak detection in networks. In *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 420–429.

[21] Wei Lu, Wei Chen, and Laks V. S. Lakshmanan. 2015. From competition to complementarity: Comparative influence diffusion and maximization. *Proceedings of the VLDB Endowment* 9, 2 (2015), 60–71.

[22] Ling-ling Ma, Chuang Ma, Hai-Feng Zhang, and Bing-Hong Wang. 2016. Identifying influential spreaders in complex networks based on gravity formula. *Physica A: Statistical Mechanics and its Applications* 451 (2016), 205–212.

[23] Rajeev Motwani and Prabhakar Raghavan. 1995. *Randomized Algorithms*. Cambridge University press.

[24] George L. Nemhauser, Laurence A. Wolsey, and Marshall L Fisher. 1978. An analysis of approximations for maximizing submodular set functions–I. *Mathematical Programming* 14, 1 (1978), 265–294.

[25] Hung T. Nguyen, My T. Thai, and Thang N. Dinh. 2016. Stop-and-stare: Optimal sampling algorithms for viral marketing in billion-scale networks. In *Proceedings of the 2016 International Conference on Management of Data*. ACM, 695–710.

[26] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1998. The PageRank citation ranking: Bringing order to the web. In *Proceedings of ASIS*. 161–172.

[27] Justin W. Patchin and Sameer Hinduja. 2006. Bullies move beyond the schoolyard: A preliminary look at cyberbullying. *Youth Violence and Juvenile Justice* 4, 2 (2006), 148–169.

[28] Kelly Reynolds, April Kontostathis, and Lynne Edwards. 2011. Using machine learning to detect cyberbullying. In *Proceedings of the 2011 10th International Conference on Machine Learning and Applications and Workshops (ICMLA'11)*. Vol. 2. IEEE, 241–244.

[29] Robert Slonje, Peter K. Smith, and Ann Frisén. 2013. The nature of cyberbullying, and strategies for prevention. *Computers in Human Behavior* 29, 1 (2013), 26–32.

[30] Chonggang Song, Wynne Hsu, and Mong Li Lee. 2016. Targeted influence maximization in social networks. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. 1683–1692.

[31] Youze Tang, Yanchen Shi, and Xiaokui Xiao. 2015. Influence maximization in near-linear time: A martingale approach. In *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*. ACM, 1539–1554.

[32] Youze Tang, Xiaokui Xiao, and Yanchen Shi. 2014. Influence maximization: Near-optimal time complexity meets practical efficiency. In *Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data*. ACM, 75–86.

[33] Guangmo Tong, Weili Wu, Ling Guo, Deying Li, Cong Liu, Bin Liu, and Ding-Zhu Du. 2017. An efficient randomized algorithm for rumor blocking in online social networks. In *Proceedings of the IEEE Conference on Computer Communications*.

[34] Hanghang Tong, B. Aditya Prakash, Tina Eliassi-Rad, Michalis Faloutsos, and Christos Faloutsos. 2012. Gelling, and melting, large graphs by edge manipulation. In *Proceedings of the 21st ACM International Conference on Information and Knowledge Management*. ACM, 245–254.

[35] Michel Walrave and Wannes Heirman. 2011. Cyberbullying: Predicting victimisation and perpetration. *Children & Society* 25, 1 (2011), 59–72.

[36] Senzhang Wang, Xiaojian Zhao, Yan Chen, Zhoujun Li, Kai Zhang, and Jiali Xia. 2013. Negative influence minimizing by blocking nodes in social networks. In *Proceedings of the 17th AAAI Conference on Late-Breaking Developments in the Field of Artificial Intelligence*. 134–136.

[37] Xiaoyang Wang, Ying Zhang, Wenjie Zhang, Xuemin Lin, and Chen Chen. 2016. Bring order into the samples: A novel scalable method for influence maximization. *IEEE Transactions on Knowledge and Data Engineering* 29, 2 (2016), 243–256.

[38] Yu Wang, Gao Cong, Guojie Song, and Kunqing Xie. 2010. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1039–1048.

[39] Chen Wei, Laks V. S. Lakshmanan, and Carlos Castillo. 2013. *Information and Influence Propagation in Social Networks*. Morgan & Claypool.

[40] Elizabeth Whittaker and Robin M. Kowalski. 2015. Cyberbullying via social media. *Journal of School Violence* 14, 1 (2015), 11–29.

[41] Yuqing Zhu, Deying Li, and Zhao Zhang. 2016. Minimum cost seed set for competitive social influence. In *Proceedings of the 35th Annual IEEE International Conference on Computer Communications*. IEEE, 1–9.

Received May 2020; revised October 2020; accepted December 2020