

Minimizing the Negative Influence by Blocking Links in Social Networks

Qipeng Yao^{1,2}, Chuan Zhou^{2(✉)}, Linbo Xiang¹, Yanan Cao², and Li Guo²

- ¹ Education Ministry Key Laboratory of Trustworthy Distributed Computing and Service, Beijing University of Posts and Telecommunications, Beijing, China
yaoqipeng0706@gmail.com, xianglinbo0523@126.com
- ² Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
{zhouchuan, caoyanan, guoli}@iie.ac.cn

Abstract. In this paper, we address the problem of minimizing the negative influence of undesirable things by blocking a limited number of links in a network. When undesirable thing such as a rumor or an infection emerges in a social network and part of users have already been infected, our goal is to minimize the size of ultimately infected users by blocking k links. A greedy algorithm with accuracy guarantee and two efficient heuristics for finding approximate solutions to this problem are proposed. Using two real networks, we demonstrate experimentally that the greedy algorithm is more effective in terms of minimizing negative influence, while the heuristics based on betweenness and out-degree are orders of magnitude faster than the greedy algorithm in terms of running time.

Keywords: Negative influence · Blocking links · Social networks

1 Introduction

In the past decade, the online social networks are providing convenient platforms for information dissemination and marketing campaign, allowing ideas and behaviors to flow along the social relationships in the effective word-of-mouth manner [3, 4, 10]. From the functional point of perspective, networks can mediate diffusion including not only positive information such as innovations, hot topics, and novel ideas, but also negative information like malicious rumors and disinformation [7, 11]. Take the rumor for example, even with a small number of its initial adopters, the quantity of the ultimately infected users can be large due to triggering a word-of-mouth cascade in the network. Therefore, it is an urgent research issue to design effective strategies for reducing the influence coverage of the negative information and minimizing the spread of the undesirable things.

Previous work studied strategies for reducing the spread size by removing nodes from a network. It has been shown in particular that the strategy of removing nodes in decreasing order of out-degree can often be effective [1, 9, 11]. Here notice that removal of nodes by necessity involves removal of links. Namely, the task of removing links is more fundamental than that of removing nodes.

Therefore, preventing the spread of undesirable things by removing links from the underlying network is an important problem. Along this idea, Kimura et al. [7] aimed to minimize the spread of contaminant by blocking a limited number of links at the expense of lower diffusion capacity.

In this paper, we aim to minimize the spread of an existing undesirable thing by blocking a limited number of links in a network. More specifically, when some undesirable thing starts with some initial nodes and diffuses through the network under the independent cascade (IC) model [6], a widely-used fundamental probabilistic model of information diffusion, we consider finding a set of k links such that the resulting network by blocking those links to minimize the expected contamination area of the undesirable thing, where k is a given positive integer. We refer to this combinatorial optimization problem as the *negative influence minimization problem*. For this problem, we propose a greedy algorithm with accuracy guarantee for efficiently finding a good approximate solution. Using two large real networks include Facebook and Diggers, we experimentally demonstrate that the proposed greedy algorithm significantly outperforms link-removal heuristics that rely on the well-studied notions of betweenness and out-degree method.

The rest of the paper is organized as follows. Section 2 is devoted to related work. Section 3 reviews the IC model and introduces the problem formulation. In Sect. 4, we propose a greedy algorithm and two heuristics to find the approximate solution. In Sect. 5 we verify the performance of proposed algorithms by experiments. Section 6 concludes the paper.

2 Related Work

The research on finding influential nodes that are effective for the spread of information through a social network, namely Influence Maximization Problem, has attracted remarkable attention recently due to its novel idea of leveraging some social network users to propagate the awareness of products [3, 6, 12]. However, the problem of minimizing the negative influence of undesirable things gets less attention, although it is an important research issue.

Some related research work has been made on minimizing the influence of negative information. Previous work studied strategies for reducing the spread size by removing nodes from a network. It has been shown in particular that the strategies of removing nodes in decreasing order of out-degree can often be effective [1, 9, 11]. Kimura et al. proposed a links blocking method to minimize the expected contamination area of the network [7]. However, the fact of part nodes infected is not considered. Yu et al. addressed the problem of finding spread blockers are simply those nodes with high degree [5]. Budak et al. investigated the problem of influence limitation where a bad campaign starts propagation from a certain node in the network and use the notion of limiting campaigns to counteract the effect of misinformation [2]. Different from previous work, our research cares more about a specific contamination scenario in the social network, and how to minimize the negative influence by blocking a small set of links.

3 Problem Formulation

In this paper, we address the problem of minimizing the spread of undesirable things such as computer viruses and malicious rumors in a network represented by a directed graph $G = (V, E)$. Here, V and E are the sets of all the nodes and edges (or links) in the network, respectively. We assume the IC model to be a mathematical model for the diffusion process of some undesirable thing in the network, and investigate the negative influence minimization problem on G .

3.1 Independent Cascade Model

Consider a directed graph $G = (V, E)$ with N nodes in V and edge labels $pp : E \rightarrow [0, 1]$. For each edge $(u, v) \in E$, $pp(u, v)$ denotes the propagation probability that v is activated by u through the edge. If $(u, v) \notin E$, $pp(u, v) = 0$. Let $Par(v)$ be the set of parent nodes of v , i.e., $Par(v) := \{u \in V, (u, v) \in E\}$.

Given an initially infected set $S \subseteq V$, the IC model works as follows. Let $S_t \subseteq V$ be the set of nodes that are activated at step $t \geq 0$, with $S_0 = S$. Then, at step $t + 1$, each node $u \in S_t$ may infect its out-neighbors $v \in V \setminus \cup_{0 \leq i \leq t} S_i$ with an independent probability of $pp(u, v)$. Thus, a node $v \in V \setminus \cup_{0 \leq i \leq t} S_i$ is infected at step $t + 1$ with the probability $1 - \prod_{u \in S_t \cap Par(v)} (1 - pp(u, v))$. If node v is successfully infected, it is added into the set S_{t+1} . The process ends at a step τ with $S_\tau = \emptyset$. Obviously, the propagation process has $N - |S|$ steps at most, as there are at most $N - |S|$ nodes outside the initially infected set S . Let $S_{\tau+1} = \emptyset, \dots, S_{N-|S|} = \emptyset$, if $\tau < N - |S|$. Note that each infected node only has one chance to infect its out-neighbors at the step right after itself is infected, and each node stays infected once it is infected by others.

Under the directed graph $G = (V, E)$, the negative influence spread of the initially infected set S , which is the ultimately expected number of infected nodes, is denoted as $\sigma(S|E)$ as follow,

$$\sigma(S|E) := \mathbb{E}_E^S \left[\left| \bigcup_{t=0}^{N-|S|} S_t \right| \right] \quad (1)$$

where \mathbb{E}_E^S is the expectation operator in the IC model with the initially infected set S and the graph links set E .

3.2 Negative Influence Minimization Problem

Now we present a mathematical definition for the *negative influence minimization problem*. Assume negative information spreads in the network $G = (V, E)$ with initially infected nodes $S \subseteq V$, our goal here is minimizing the number of ultimately infected nodes by blocking k edges (or links) set D in E , where k ($\ll |E|$) is a given const. It can be represented as the following optimization problem:

$$\min_{D \subseteq E, |D| \leq k} \sigma(S|E \setminus D) \quad (2)$$

where $\sigma(S|E \setminus D)$, defined like Eq. (1), denotes the influence (number of ultimately infected nodes) of S when the edge set D is blocked.

For a large network, any straightforward method for exactly solving the contamination minimization problem suffers from combinatorial explosion. Therefore, we consider approximately solving the problem.

4 Methodology

In this section, we propose a greedy algorithm based on maximum marginal gain rule for the contamination minimization problem on graph $G = (V, E)$. Let k be the number of links to be blocked in this problem. To demonstrate the effectiveness of it, we compare it against two classical centrality based influence evaluation methods.

Greedy Algorithm. To make better use of the greedy algorithm, we consider an equivalent optimization problem of Eq. (2) as follows,

$$\max_{D \subseteq E, |D| \leq k} f(D), \quad (3)$$

where

$$f(D) := \sigma(S|E) - \sigma(S|E \setminus D) \quad (4)$$

is defined as the **decreasing spread** after blocking edges set D when the initially infected set is S . The above alternative formulation has key properties as described in Theorem 1.

Theorem 1. *The decreasing spread function $f : 2^E \rightarrow \mathbb{R}^+$ is monotone and submodular with $f(\emptyset) = 0$. Theoretically, a non-negative real-valued function f on subsets of E is submodular, if $f(D \cup \{e\}) - f(D) \geq f(D' \cup \{e\}) - f(D')$ for all $D \subseteq D' \subseteq E$ and $e \in E \setminus D'$. And f is monotone, if $f(D) \leq f(D')$ for all $D \subseteq D'$.*

Proof. It is trivial that f is monotone with $f(\emptyset) = 0$. By definition of Eq. (4), in order to reach the submodularity of f , we need to prove that

$$\sigma(S|E \setminus D) - \sigma(S|E \setminus (D \cup \{e\})) \geq \sigma(S|E \setminus D') - \sigma(S|E \setminus (D' \cup \{e\})) \quad (5)$$

for all $D \subseteq D' \subseteq E$ and $e \in E \setminus D'$. Let $X(S, E \setminus D)$ be a random activation result (consisting of live edges [6], through which all activated nodes can be reached from S) in the network $(V, E \setminus D)$ with the initially infected set S . Then the influence spread from S can be measured as shown in Eq. (6)

$$\sigma(S|E \setminus D) = \mathbb{E}[|X(S, E \setminus D)|] \quad (6)$$

where $|X(S, E \setminus D)|$ means the number of nodes in activation result $X(S, E \setminus D)$. As we have the following equation

$$|X(S, E \setminus D)| - |X(S, E \setminus (D \cup \{e\}))| \geq |X(S, E \setminus D')| - |X(S, E \setminus (D' \cup \{e\}))| \quad (7)$$

works in all possible results, we finally get Eq. (5). Hence f is submodular.

By following the properties, the problem of finding a set D of size k that maximizes $f(D)$ can be approximated by the greedy algorithm in Algorithm 1. The algorithm iteratively selects a new edge e^* that maximizes the incremental change of $f(D)$ (or equivalently $-\sigma(S|E \setminus D)$) and includes it into the blocking edge set until k edges have been selected. It is shown that the algorithm guarantees an approximation ratio of $f(D)/f(D^*) \geq 1 - 1/e$, where D is the output of the greedy algorithm and D^* is the optimal solution [8].

Algorithm 1. *Greedy*($G = (V, E), S, k$)

```

1: initial  $D_0 = \emptyset$ 
2: for  $i = 1$  to  $k$  do
3:    $e^* = \arg \max_{e \in E \setminus D_{i-1}} \sigma(S|E \setminus D_{i-1}) - \sigma(S|E \setminus (D_{i-1} \cup \{e\}))$ 
4:    $D_i = D_{i-1} \cup \{e^*\}$ 
5: end for
6: output  $D_k$ 
    
```

Comparison Methods. We compared the greedy method with two heuristics based on the well-studied notions of betweenness and outdegree in the field of complex network theory.

To minimize the influence of contaminant, a natural idea is to cut off the edges linking from infected set to uninfected set. Specifically, given the initially infected set S , define the out-edge set $O(S)$ like

$$O(S) := \{(u, v) \in E : u \in S \text{ and } v \in V \setminus S\} \quad (8)$$

We want to block k edges in the set $O(S)$ to minimize the negative influence. Since the set $O(S)$ is usually very large (i.e. $|O(S)| \gg k$), a natural question arises, *how to select k pivotal edges from the set $O(S)$ to block?* In this part, we introduce two scoring methods for the edges in $O(S)$, and then select k edges with the highest scores as the objectives to block.

Betweenness scoring method. Given the initially infected nodes S , the betweenness score $b(e)$ of a link $e \in O(S)$ is defined as follows:

$$b(e) = \sum_{u \in S, v \in V \setminus S} \frac{n(e; u, v)}{N(u, v)} \quad (9)$$

where $N(u, v)$ denotes the number of the shortest paths from node u to node v in G , and $n(e; u, v)$ denotes the number of those paths that pass e . Here we set $n(e; u, v)/N(u, v) = 0$ if $N(u, v) = 0$. We expect that blocking the links with the highest betweenness score can be effective for preventing the spread of contamination in the network. We refer to this method as the betweenness method.

Out-degree scoring method. Previous work has shown that simply removing nodes in order of decreasing out-degrees works well for preventing the spread of

contamination in most real networks [1, 9, 11]. Thus, blocking links from contaminated nodes to high out-degrees looks promising for the contamination minimization problem. Here we focus on the contaminated nodes S . We define the out-degree score $o(e)$ of edge $e = (u, v) \in O(S)$ as the number of outgoing links from the node v to non-contaminative nodes. As a comparison method, we employ the method of blocking links $e \in O(S)$ in decreasing order of their out-degrees. We refer to this method as the out-degree method.

5 Experimental Results

We conduct experiments on two real-world data sets to evaluate the performance of greedy algorithm and compare it with that of betweenness and out-degree methods.

5.1 Data Sets

The data we use from Facebook is downloaded from the Stanford Large Network Dataset Collection. Nodes of the network are behalf of people and if a person i have a relationship with the other person j , the graph contains an directed edge from i to j . The Facebook data set contains 4,039 nodes and 88,234 edges. The Digger data set is available at <http://arnetminer.org/heterinf>. Digger is a heterogeneous network, including Digg stories, user actions (submit, digg, comment and reply) with respect to the stories, and friendship relations among users. The Diggers has 8,193 nodes and 56,440 edges.

5.2 Parameter Settings

In the IC Model, we assign a uniform probability p to each edge of the graph. Two propagation probabilities are used in our experiments: $p = 0.05$ and $p = 0.1$. The initially infected set S is chosen in the whole network uniformly with $|S| = 50$. Also we want to cut off 50 edges to minimize the negative influence, i.e., $k = 50$.

5.3 Experimental Results

The results in Fig. 1 show that the *Greedy* algorithm outperforms the *Betweenness* and *Out-degree* methods on both data sets with different probabilities. The Betweenness takes the second place and the Out-degree comes last. By contrast, the performance of Betweenness gets very close to that of out-degree method.

In the first figure in Fig. 1. (a), we can observed that the greedy algorithm reduce the negative spread from 118 to 80 by blocking 50 links in the data set Diggers. Here note that blocking 50 edges means blocking 8.59 % of the links that connected to infected nodes in Diggers. Thus, by appropriately blocking about 8.59 % of the links, the greedy algorithm, betweenness heuristic, and out-degree heuristic reduce the negative spread by about 32 %, 19 % and 15 % respectively. The explanation of the other three figures are the same. Besides, we can draw a

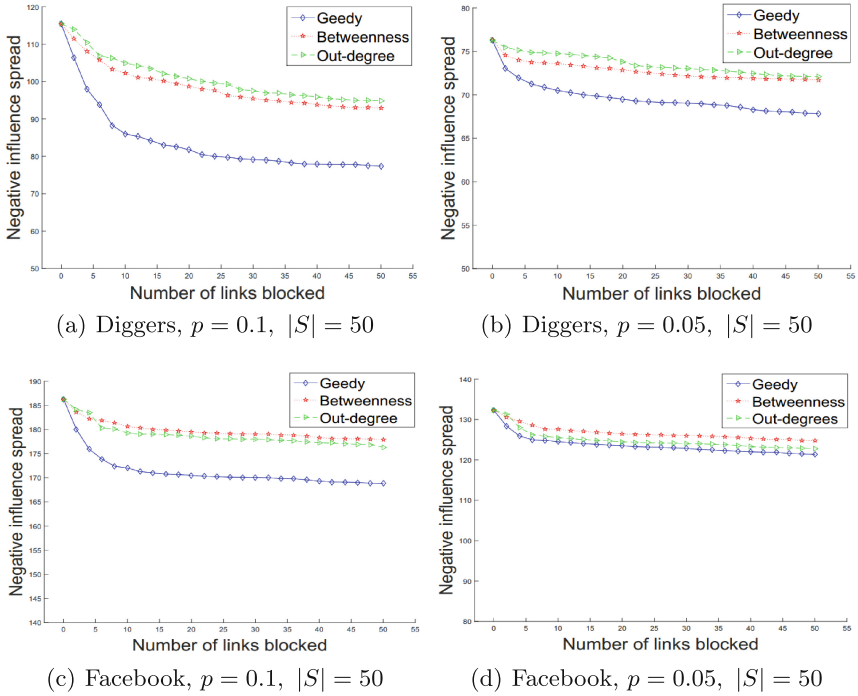


Fig. 1. The experimental results under different data sets and different propagation probabilities.

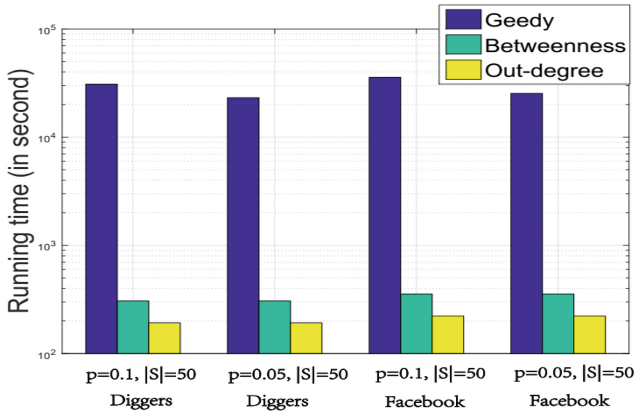


Fig. 2. The time comparison among the three method.

conclusion from the experiment that our proposed methods perform better in the sparse networks, and their performance is unsatisfactory in the dense networks.

In Fig. 2, the running time of betweenness and out-degree heuristics is orders of magnitude faster than the greedy algorithm in terms of running time.

6 Conclusion

In this paper we investigate the problem of minimizing the spread of negative things by blocking links in social networks. This minimization problem provides an alternate approach to the problem of preventing the spread of contamination by removing nodes in a network. We proposed a greedy algorithm to find efficiently an approximate solution to this problem. Meanwhile we introduced two heuristics based betweenness and outdegree to compare with the greedy algorithm. Employing the Facebook and Diggers data sets, we have experimentally demonstrated that the greedy algorithm can effectively work, and the two proposed heuristics can significantly reduce running time.

There are several interesting future directions. First, the diffusion model employed in this paper is the IC model, which is a discrete time model. What can we do when the underlying diffusion model is a continuous time one? Second, how to extend it to a dynamic network when the network structure changes over time is also an interesting question.

Acknowledgments. This work was supported by the Strategic Leading Science and Technology Projects of Chinese Academy of Sciences (No. XDA06030200).

References

1. Albert, R., Jeong, H., Barabási, A.-L.: Error and attack tolerance of complex networks. *Nature* **406**, 378–382 (2000)
2. Budak, C., Agrawal, D., Abbadi, A.E.: Limiting the spread of misinformation in social networks. In: *WWW 2011* (2011)
3. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: *KDD 2009* (2009)
4. Domingos, P., Richardson, M.: Mining the network value of customers. In: *KDD 2001* (2001)
5. Habiba, Yu, Y., Berger-Wolf, T.Y., Saia, J.: Finding spread blockers in dynamic networks. In: Giles, L., Smith, M., Yen, J., Zhang, H. (eds.) *SNAKDD 2008*. LNCS, vol. 5498, pp. 55–76. Springer, Heidelberg (2010)
6. Kempe, D., Kleinberg, J.M., Tardos, É.: Maximizing the spread of influence through a social network. In: *KDD 2003* (2003)
7. Kimura, M., Saito, K., Motoda, H.: Minimizing the spread of contamination by blocking links in a network. In: *AAAI 2008* (2008)
8. Nemhauser, G., Wolsey, L., Fisher, M.: An analysis of the approximations for maximizing submodular set functions. *Math. Program.* **14**, 265–294 (1978)
9. Newman, M.E.J., Forrest, S., Balthrop, J.: Email networks and the spread of computer viruses. *Phys. Rev. E* **66**, 035101 (2002)
10. Richardson, M., Domingos, P.: Mining knowledge-sharing sites for viral marketing. In: *KDD 2002* (2002)

11. Wang, S., Zhao, X., Chen, Y., Li, Z., Zhang, K., Xia, J.: Negative influence minimizing by blocking nodes in social networks. In: AAAI (Late-Breaking Developments) (2013)
12. Zhou, C., Guo, L.: A note on influence maximization in social networks from local to global and beyond. *Procedia Comput. Sci.* **30**, 81–87 (2014)