Effective Social Network Quarantine with Minimal Isolation Costs

Huanyang Zheng and Jie Wu
Department of Computer and Information Sciences, Temple University, USA
Email: {huanyang.zheng, jiewu}@temple.edu

***Abstract*—Nowadays, the notion of diseases has been extended from real human diseases to general epidemic information propagations, such as the rumors in distributed systems. Controlling the spread of a disease is usually done through quarantine, where people that have, or are suspected to have, a disease are isolated from having interactions with others. As a tradeoff, normal human interactions are inevitably degraded by the quarantine. This motivates us to explore a robust quarantine strategy that can eliminate epidemic outbreaks with minimal isolation costs. Our problem is shown to be NP-hard. A bounded algorithm with an approximation ratio of two is proposed, through utilizing the feasibility and minimality properties. Finally, real data-driven experiments demonstrate the efficiency and effectiveness of the proposed algorithms in real-world applications.**

***Index Terms*—Social network, epidemic outbreak, effective quarantine strategy, minimal isolation costs.**

I. Introduction

Nowadays, the notion of diseases has been extended from real human diseases to general epidemic information propagations, such as the rumors in distributed systems. Controlling the spread of a disease in a population (human communities and distributed systems) is usually done through quarantine where people that have, or are suspected to have, a disease are restricted from having interactions with others. However, the human interactions are inevitably degraded by the quarantine. This motivates us to explore an effective quarantine strategy that can maximally preserve the human interactions.

This paper uses the classic Susceptible-Infected-Susceptible (SIS) model to simulate the epidemic spreading, where each person has a state of being either susceptible or infected [1]. People transfer their states through a cycle in which their susceptibility (S) causes them to become infected (I), and they return to being susceptible (S) by recovery. The epidemic breaks out when the average infection rate becomes larger than the average recovery rate. In such a case, a large portion of people in the social network will eventually become infected. Consequently, it is necessary to isolate a set of people as a quarantine strategy to depress the infection rate. Epidemic outbreaks could be controlled once the infection rate is cut down by isolations, which are usually costly.

Our objective is to explore an effective quarantine strategy that can eliminate epidemic outbreaks with minimal isolation costs. In other words, we want to isolate a set of people with minimal costs to eliminate epidemic outbreaks. Our problem is extremely challenging, since eliminating epidemic outbreaks and preserving social connections cannot be simultaneously achieved. Intuitively, the isolation of an arbitrary person can help in the elimination of epidemic outbreaks. Moreover, social networks are structurally heterogenous, meaning that the impacts of isolations are very hard to quantify. Should we simply isolate people who have lots of normal social connections? Or should we isolate people who only have a few, but important connections? The network structural heterogeneity should be considered within the quarantine strategy design.

Currently, social network epidemic outbreaks [1] have been well-studied with very rich literatures. The novelty of this paper lies in the quarantine strategy with minimal isolations. Our work casts some new light on real-world quarantines. For example, in an infectious global epidemic (SARS or Ebola), we can avoid epidemic outbreaks while isolating a minimal number of people. Moreover, our work has broader impacts, since it can be applied to situations aside from real social networks. Typical applications of our work can include:

- In computer networks, epidemics are worms. Some computers are turned off to eliminate worm spreading. Our work points out a strategy that can resist worms with maximally-preserved computers.
- In online social networks (e.g., Facebook and Twitter), epidemics are rumors (malicious Facebook posts) that can be shared among friends. The service provider can refer to our work to control rumors with minimal user blocks.

Our main contributions are summarized as follows:

- We address a novel problem on the effective quarantine that can restrict epidemic outbreaks with minimal isolations. It has broader impacts on real-world applications.
- An approximation algorithm, which guarantees a constant ratio to the optimal quarantine strategy, is proposed. The properties of feasibility and minimality are explored for eliminating epidemic outbreaks.
- Real data-driven experiments are conducted to evaluate the proposed algorithms. Evaluation results are shown from different perspectives to provide insightful conclusions for real-world applications.

The remainder of this paper is organized as follows. Section II surveys related works. Section III formulates the problem and describes the SIS epidemic model with discussions on the properties of isolations. Section IV studies the effective quarantine strategy with minimal isolations. Section V includes real data-driven experiments. Finally, Section VI concludes the paper and suggests future research directions.

II. Related Work

Epidemic spreading models have been extensively explored over the past two decades. As one of the most popular epidemic models, the SIS model divides a given population into two compartments: susceptible and infected. People transfer their states through a cycle of being infected after being susceptible, and going back to susceptible by recovery [1]. The SIS model in social networks that have power-law degree distributions was summarized by Lee et al. [2]. There are many other epidemic models. For example, the Susceptible-Infected-Recovered (SIR) model [1] adds one more compartment to represent vaccinated individuals who are no longer susceptible to the infection. The SIS and SIR models were also used to stimulate worm spreadings in distributed systems [3]. A state-of-the-art review on epidemic models is available in [4].

Social networks have also been extensively studied over the past decade. Structures of social networks were confirmed to be scale-free [5], where the node degree follows power-law distribution. The triadic closure phenomenon (a friend-of-a-friend is likely to become a friend) is identified [6]. Social networks are considered to have small-world structures [7]. Compared to random networks, social networks have smaller network diameters and a larger clustering coefficient [8].

Although epidemic models and social network properties have been well-studied, to the best of our knowledge, this paper is the first interdisciplinary study on a quarantine strategy that minimizes isolations without epidemic outbreaks. Our approximation scheme is based on the classic solutions to the knapsack problem [9–11] and the set cover problem [12–15].

III. Problem Formulation and Epidemic Model

A. Problem Formulation

Our social network model is based on a directed graph $G = (V, E)$, where V is a set of nodes (persons), and $E \subseteq V^2$ is a set of directed edges (social relationships). Let $|\cdot|$ denote the cardinality of the corresponding variable. For example, $|V|$ and $|E|$ are the total number of nodes and edges, respectively. To control epidemic outbreaks, some nodes are isolated by the quarantine strategy. A node v is isolated, if all the incoming and outgoing edges of v are removed. An isolated node can no longer interact with its neighbors, but it remains in the network. The isolation cost of v is C_v. The set of nodes isolated by the quarantine strategy is denoted by Q. The objective is to explore a quarantine strategy that eliminates epidemic outbreaks with minimal $\sum_{v \in Q} C_v$.

B. Epidemic Outbreak Model

Our epidemic spreading model is based on the classic SIS model [2]. Nodes have states of either being susceptible or infected. Nodes in the susceptible state are people who do not have the disease, but can potentially catch it. Nodes in the infected state are people who have the disease and can spread the disease to their neighbors in G. Infected nodes can go back into the susceptible state upon recovery, and then can be reinfected. We consider that the infection rate of a given node depends on its infected incoming neighbors. For the node v, each of its infected incoming neighbors independently brings a constant infection rate (the infection probability per time unit) of λ to v. Meanwhile, the recovery rate is set to be a constant of r, as used in many existing models [1, 16, 17].

Let $f(t)$ denote the average fractions of nodes that are infected at time t. To capture the structural heterogeneity of the social network, let $p(d)$ denote the fraction of nodes with in-degree d, and let $f_d(t)$ denote the fraction of infected nodes with in-degree d at time t. By definition, we have $f(t) = \sum_d p(d) \cdot f_d(t)$. Then, $\Theta(f(t))$ is the probability that a uniform-randomly selected edge comes from an infected node at the time t. It can be calculated as:

$$\Theta(f(t)) = \frac{\sum_d d \cdot p(d) \cdot f_d(t)}{\sum_d d \cdot p(d)} \tag{1}$$

The fraction of susceptible nodes with in-degree d at time t is $[1 - f_d(t)]$. Each of these nodes has an incoming degree of d, meaning that it is expected to have $d \times \Theta(f(t))$ infected incoming neighbors. Since each infected incoming neighbor brings an infection rate of λ, the total infection rate is:

$$1 - (1 - \lambda)^{d \cdot \Theta(f(t))} \approx \lambda \cdot d \cdot \Theta(f(t)) \tag{2}$$

λ is assumed to be small. Otherwise, the epidemic is not controllable due to an overly-large infection rate. We have:

$$\frac{\partial f_d(t)}{\partial t} = \lambda d \Theta(f(t))[1 - f_d(t)] - r f_d(t) \tag{3}$$

The first term of $\lambda d[1 - f_d(t)]\Theta(f(t))$ indicates the fraction of newly infected nodes that have in-degrees of d. The last term of $r f_d(t)$ shows the recovery. If we consider a stable epidemic state of $\frac{\mathrm{d} f_d(t)}{\mathrm{d}t} = 0$, then Eq. 3 can be solved as:

$$f_d(t) = \frac{\lambda d \Theta(f(t))}{r + \lambda d \Theta(f(t))} \tag{4}$$

According to Eq. 4, Eq. 1 can be rewritten as:

$$\Theta(f(t)) = \frac{1}{\sum_d dp(d)} \sum_d dp(d) \frac{\lambda d \Theta(f(t))}{r + \lambda d \Theta(f(t))} \tag{5}$$

The epidemic outbreak elimination indicates that $\Theta(f(t)) = 0$ and $\Theta(f(t))$ will not increase with respect to the time:

$$\frac{\partial}{\partial \Theta(f(t))} \left(\Theta(f(t)) - \frac{\sum_d dp(d) \frac{\lambda d \Theta(f(t))}{r + \lambda d \Theta(f(t))}}{\sum_d dp(d)} \right) \geq 0 \tag{6}$$

When $\Theta(f(t)) = 0$, Eq. 6 should be satisfied to control the growth trend of infected nodes. Based on Eq. 6, we can derive the following prerequisite to control epidemic outbreaks:

$$\frac{\lambda \sum_d d^2 p(d)}{r \sum_d dp(d)} \leq 1 \quad \text{or} \quad \frac{\langle d^2 \rangle}{\langle d \rangle} \leq \frac{r}{\lambda} \tag{7}$$

Let $\langle \cdot \rangle$ denote the mean value of the corresponding variable. Then, we have $\sum_d d^2 p(d) = \langle d^2 \rangle$ and $\sum_d dp(d) = \langle d \rangle$ by the definition. Eq. 7 represents the prerequisite of controlled outbreaks in social networks. The key insight of Eq. 7 is that

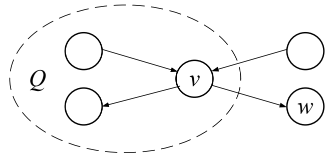


Fig. 1. Proof of Theorem 1.

both a larger average degree and a larger degree variance bring a more vulnerable network with respect to epidemic outbreaks:

$$\frac{\langle d^2\rangle}{\langle d\rangle} = \frac{\langle d^2\rangle - \langle d\rangle^2}{\langle d\rangle} + \langle d\rangle \tag{8}$$

Note that $\langle d\rangle$ is the average in-degree and $\langle d^2\rangle - \langle d\rangle^2$ is the in-degree variance. The ratio of $\langle d^2\rangle$ to $\langle d\rangle$ represents the network vulnerability to epidemics (the larger, the more vulnerable). If we want to control epidemic outbreaks, then we need to control the in-degree distribution though the isolations. Once a node is isolated, its associated incoming and outgoing edges are removed, leading to a degradation on $\frac{\langle d^2\rangle}{\langle d\rangle}$ to control epidemic outbreaks. For simplicity, let $\Delta(Q)$ denote the degradation of $\frac{\langle d^2\rangle}{\langle d\rangle}$, when nodes in Q are isolated by the quarantine strategy. We introduce a constant coefficient of $\delta = \frac{\langle d^2\rangle}{\langle d\rangle} - \frac{r}{\lambda}$ as the degradation threshold to control epidemic outbreaks. At this time, our objective can be reformulated as minimizing $\sum_{v\in Q} C_v$ with the constraint that $\Delta(Q) \geq \delta$. Further analysis is conducted in the next subsection.

C. Feasibility and Minimality

This subsection explores the inherent properties of $\Delta(Q)$ to obtain more insights on the effective quarantine strategy design. We start with the following definition:

***Definition** 1:* A quarantine strategy, Q, is said to be feasible, if the constraint of $\Delta(Q) \geq \delta$ is satisfied.

Basically, a quarantine strategy that can eliminate epidemic outbreaks is defined as a feasible quarantine strategy. In reality, a feasible quarantine strategy usually isolates lots of nodes to control epidemic outbreaks. But for the sake of the theory, we still consider the event that $\{\exists v \in V | \Delta(\{v\}) \geq \delta\}$. It means that isolating only one node of v may be sufficient to control epidemic outbreaks. To facilitate the quarantine strategy design, we make a cutoff on $\Delta(Q)$. If $\Delta(\{v\}) \geq \delta$, we force $\Delta(\{v\})$ to be δ as a cutoff. Note that such a cutoff will not change the feasibility of an arbitrary quarantine strategy. Hence, the optimal quarantine strategy is not changed by this cutoff. We have the following definition:

***Definition** 2:* A feasible quarantine strategy, Q, is said to be minimal, if $Q \setminus \{v\}$ is not feasible for an arbitrary $v \in Q$.

A minimal quarantine strategy means that each node in this quarantine strategy is necessarily isolated. If an arbitrary node in this quarantine strategy is no longer isolated, this quarantine strategy becomes infeasible and can no longer control epidemic outbreaks. Our key observation is that a minimal feasible quarantine strategy has the following property:

***Theorem** 1:* A minimal feasible quarantine strategy, Q, satisfies the property that $\delta \leq \Delta(Q) \leq 2\delta$.

Proof: By the definition of feasibility, we have $\Delta(Q) \geq \delta$. Therefore, we focus on proving $\Delta(Q) \leq 2\delta$. Let us start with a special case, where all the nodes in Q do not have outgoing neighbors. In such a case, the isolation of a node in Q will not diminish the in-degrees of the remaining nodes. Let d denote the node in-degree when no node is isolated. d_v is the in-degree of the node v. Then, $\Delta(Q)$ can be calculated as:

$$\Delta(Q) = \frac{\langle d^2\rangle}{\langle d\rangle} - \frac{\langle d^2\rangle - \frac{1}{|V|}\sum_{v\in Q} d_v^2}{\langle d\rangle - \frac{1}{|V|}\sum_{v\in Q} d_v} \tag{9}$$

In Eq. 9, $\frac{1}{|V|}$ results from the fact that each node in Q is a fraction, $\frac{1}{|V|}$, of all the nodes. As another form of Eq. 7, $\langle d^2\rangle$ and $\langle d\rangle$ can also be computed by $\langle d^2\rangle = \frac{1}{|V|}\sum_{v\in V} d_v^2$ and $\langle d\rangle = \frac{1}{|V|}\sum_{v\in V} d_v$, respectively. We assume that $\langle d\rangle \gg \frac{1}{|V|}\sum_{v\in Q} d_v$, leading to the following approximation:

$$\frac{1}{\langle d\rangle - \frac{1}{|V|}\sum_{v\in Q} d_v} \approx \frac{1 + \frac{1}{\langle d\rangle}\frac{1}{|V|}\sum_{v\in Q} d_v}{\langle d\rangle} \tag{10}$$

Back to Eq. 9 with the substitution in Eq. 10, we can obtain:

$$\Delta(Q) \approx \frac{1}{|V|\langle d\rangle}\Big[\sum_{v\in Q} d_v^2 - \frac{\langle d^2\rangle}{\langle d\rangle}\sum_{v\in Q} d_v\Big] + o\Big(\frac{1}{|V|\langle d\rangle}\Big) \tag{11}$$

In Eq. 11, $o(\frac{1}{|V|\langle d\rangle})$ represents the second order term that is relatively ignorable. Eq. 11 implies the following result:

$$\Delta(Q) \leq \Delta(Q \setminus \{u\}) + \Delta(\{u\}) \tag{12}$$

Eq. 12 is obtained through comparing the first and second order terms in Eq. 11 for the left and right parts of Eq. 12. According to the definition of the minimality, we can obtain the result that $\Delta(Q \setminus \{u\}) < \delta$, since $Q \setminus \{u\}$ is not a feasible quarantine strategy. Meanwhile, we have $\Delta(\{u\}) \leq \delta$, since we have made a cutoff on $\Delta(Q)$. Therefore, we have:

$$\Delta(Q) \leq \delta + \delta = 2\delta \tag{13}$$

Eq. 13 concludes that $\delta \leq \Delta(Q) \leq 2\delta$ is true in the special case, where all the nodes in Q do not have outgoing neighbors. The insight of this case is that the isolation of each node is independent to each other. Hence, $\Delta(Q)$ can be decomposed, as shown in Eq. 12, to obtain its upper bound.

Let us go back to the general case, where nodes in Q may have outgoing neighbors. Note that the isolation of a node in Q may diminish the in-degree of a node that is not in Q. An example is shown in Fig. 1, where the isolation of v diminishes the in-degree of w. Let Q' denote the set of nodes that are not in Q, but have diminished in-degrees due to the isolations of nodes in Q. For the node $v \in Q'$, let ρ_v denote its in-degree. Then, for the general case, Eq. 9 is rewritten as:

$$\begin{aligned}\Delta(Q) &= \frac{\langle d^2\rangle}{\langle d\rangle} - \frac{\langle d^2\rangle - \frac{1}{|V|}\sum_{v\in Q} d_v^2 - \frac{1}{|V|}\sum_{v\in Q'}(d_v^2 - \rho_v^2)}{\langle d\rangle - \frac{1}{|V|}\sum_{v\in Q} d_v - \frac{1}{|V|}\sum_{v\in Q'}(d_v - \rho_v)} \\ &= \frac{\langle d^2\rangle}{\langle d\rangle} - \frac{\langle d^2\rangle - \frac{1}{|V|}\sum_{v\in Q\cup Q'} d_v^2 + \frac{1}{|V|}\sum_{v\in Q'} \rho_v^2}{\langle d\rangle - \frac{1}{|V|}\sum_{v\in Q\cup Q'} d_v + \frac{1}{|V|}\sum_{v\in Q'} \rho_v}\end{aligned} \tag{14}$$

Algorithm 1 Marginal Greedy

Input: The social network, G, and the threshold, δ.
Output: The quarantine strategy, Q.

```
Initialize Q = ∅.
while Δ(Q) < δ do
    v = arg min_{v∈V\Q} C_v / (Δ({v}∪Q) − Δ(Q)).
    Q = Q ∪ {v}.
return Q as the quarantine strategy.
```

Eq. 14 has a similar format with Eq. 9. Through a similar derivation, we find that the analysis in Eq. 12 still holds for the general case. Therefore, we conclude that $\delta \leq \Delta(Q) \leq 2\delta$ is true, which completes the proof. ∎

The key insight behind Theorem 1 is that a minimal feasible quarantine strategy would not lead to excessive isolations. Unnecessary isolations are saved once the epidemic outbreak is controlled. In a minimal feasible quarantine strategy, each node is necessarily isolated. By contradiction, it can be seen that the optimal solution for our problem must be a minimal feasible quarantine strategy. We will describe an effective quarantine strategy through utilizing the minimality property.

IV. Effective Social Network Quarantine

A. NP-hardness and Marginal Greedy Strategy

The objective of this paper is to design an effective quarantine strategy that eliminates epidemic outbreaks with minimal isolation costs. This problem is NP-hard:

Theorem 2: Searching an optimal quarantine strategy of Q, which minimizes $|Q|$ with $\Delta(Q) \geq \delta$, is NP-hard.

Proof: We prove the NP-hardness by a reduction to the partial set cover problem [18] in a special case, where node in-degrees are identical. Note that node out-degrees may not be the same. In such a case, $\frac{\langle d^2 \rangle}{\langle d \rangle} = \langle d \rangle$, meaning that the prerequisite of controlling epidemic outbreaks is reduced to controlling the average node in-degree. In other words, epidemic outbreaks could be eliminated through breaking up $\delta|V|$ edges. Our problem becomes minimizing $\sum_{v \in Q} C_v$ with the constraint that $\delta|V|$ edges are broken. If we correspond an edge to an element, and correspond a node to a set, then our problem reduces to a partial set cover problem that uses the sets with minimal costs to cover $\delta|V|$ elements. Since the partial set cover problem is NP-hard by a reduction to the set cover problem [18], our problem is also NP-hard. ∎

We first present an intuitive greedy solution, as shown in Algorithm 1. It iteratively isolates the node v that can minimize $\frac{C_v}{\Delta(\{v\} \cup Q) - \Delta(Q)}$ (i.e., minimal "cost-to-benefit" ratio). Algorithm 1 will terminate, when the quarantine strategy of Q becomes feasible, i.e., $\Delta(Q) \geq \delta$. The time complexity of Algorithm 1 is $O(V^2)$. This is because it has $O(V)$ iterations, and each iteration takes $O(V)$ to go through all the remaining nodes for the isolation decision. However, Algorithm 1 cannot guarantee an approximation ratio to the optimal solution. The difficulty comes from the fact that the number of isolated nodes in the optimal solution is unknown. Algorithm 1 may

Algorithm 2 Homogeneous Greedy (recursive)

Input: The social network, G, the threshold, δ, and the incomplete quarantine strategy, Q'.
Output: The quarantine strategy, Q.

```
1:  if δ < 0 then
2:      return ∅;
3:  v = arg min_{u∈V\Q'} C_u / (Δ(Q'∪{u}) − Δ(Q')).
4:  Set coefficient ε = C_v / (Δ(Q'∪{v}) − Δ(Q')).
5:  Q' = Q' ∪ {v}.
6:  for each u ∈ V\Q' do
7:      C'_u = ε × Δ({u}).   /* split node cost */
8:      C_u = C_u − C'_u.    /* residual node cost */
9:  Q = Q' ∪ RECURSIVE(G, δ − Δ(Q'), Q').
10: for each u ∈ Q do
11:     if Q \ {u} is a feasible quarantine strategy then
12:         Q = Q \ {u}.
13: return Q as the quarantine strategy.
```

isolate many more, or many fewer nodes than the optimal solution. On the other hand, even if the number of isolated nodes in the optimal solution is known a prior, we cannot guarantee the feasibility of the solution with the same number of isolated nodes. Further explorations are conducted.

B. Homogeneous Greedy Strategy

The key observation of our approach is that a minimal feasible quarantine strategy would not lead to excessive isolations (Theorem 1). Following this intuition, a homogeneous greedy solution is proposed, as shown in Algorithm 2. It is a recursive algorithm that splits the node cost through a homogeneous function. At each recursion level, it isolates the node v that can minimize $\frac{C_v}{\Delta(\{v\} \cup Q) - \Delta(Q)}$ (i.e., minimal "cost-to-benefit" ratio), as implemented in lines 3 to 5. Then, in lines 6 to 9, it splits the node cost through a homogeneous function for a recursive call. Finally, in lines 10 to 12, some nodes in Q are removed to satisfy the minimality property. We claim that Algorithm 2 can guarantee an approximation ratio:

Theorem 3: Algorithm 2 guarantees an approximation ratio of two to the optimal solution for the isolation costs.

Proof: We prove by induction. For the base case, we have $\delta < 0$ with $Q = \emptyset$ and $\sum_{v \in Q} C_v = 0$. Therefore, the base case holds. For the general case, Algorithm 2 isolates the node v that can minimize $\frac{C_v}{\Delta(\{v\} \cup Q) - \Delta(Q)}$, splitting the corresponding node cost in line 7. The residual node cost is shown in line 8 for a recursive call. We have:

$$\sum_{u \in Q} C_u = \sum_{u \in Q} C'_u + \sum_{u \in Q} (C_u - C'_u) \tag{15}$$

Let Q_1^*, Q_2^*, and Q^* denote the optimal solutions for G with the isolation costs to be C'_u, $C_u - C'_u$, and C_u, respectively. According to Theorem 1, we have $\sum_{u \in Q} C'_u \leq 2 \sum_{u \in Q_1^*} C'_u$ by the minimality property. This is because C'_u scales linearly with respect to $\Delta(\{u\})$. Meanwhile, by induction, we have

$\sum_{u\in Q}(C_u - C'_u) \leq 2\sum_{u\in Q_2^*}(C_u - C'_u)$. Consequently, the following inequality holds:

$$\begin{aligned}\sum_{u\in Q} C_u &= \sum_{u\in Q} C'_u + \sum_{u\in Q}(C_u - C'_u) \\ &\leq 2\sum_{u\in Q_1^*} C'_u + 2\sum_{u\in Q_2^*}(C_u - C'_u) \\ &\leq 2\sum_{u\in Q^*} C'_u + 2\sum_{u\in Q^*}(C_u - C'_u) = 2\sum_{v\in Q*} C_u \quad (16)\end{aligned}$$

The last inequality holds, since Q^* is not the optimal solution for G, with the isolation costs being C'_u or $C_u - C'_u$. The result in Eq. 16 completes the proof that Algorithm 2 guarantees an approximation ratio of two to the optimal solution. ■

The key idea of Theorem 3 is to utilize the minimality property. A minimal feasible quarantine strategy will not have excessive isolations, leading to bounded isolation costs. Algorithm 2 has a complexity of $O(V^2)$. It has a recursion depth of $O(V)$, while each recursion call takes $O(V)$ to select the node and keep the minimality.

V. Experiments

A. Dataset Information and Settings

Our experiments are based on two datasets of Epinions [19] and Wikipedia [20]. Epinions is a general consumer review site, which was launched in 1999. Epinions users can read new and old reviews about a variety of products to help them decide on a purchase [21]. Our approach can be applied to the isolations of users in eliminating online rumor spreading. Wikipedia is a free encyclopedia written collaboratively by volunteers (i.e., users) around the world. A small portion of users are administrators. In order for a user to become an administrator, a request must be issued and then voted upon. If a user (say v) votes for another user (say u), then there exists a directed edge from v to u. Our approach can be applied on isolating users to control disordered votes in Wikipedia. The dataset statistics have been summarized in Table I.

TABLE I
DATASET STATISTICS

	Epinions	Wikipedia
Number of nodes	18,098	7,115
Number of edges	355,754	103,689
Average degree	19.6	14.6
In-degree Variance	3615.8	1006.9
Network Diameter	11	7
Global clustering coefficient	0.138	0.141
Average edge weight	0.0285	0.0076

Note that these two datasets do not include information on node isolation costs. Hence, three cost functions are designed:

- The first cost function uses a constant cost for each node, meaning that all the nodes have identical isolation costs.
- The second cost function determines the node isolation cost based on the node in-degree with a logarithmic mapping. For the node v, we have $C_v = \log(d_v + 1)$.
- The third cost function determines the node isolation cost based on the node in-degree with a square root mapping. For the node v, we have $C_v = \sqrt{d_v}$.

The node costs are normalized for fair comparison. As for the parameters on epidemic spreading, we set a fixed infection rate of $\lambda = 1$. The recovery rate, r, is tuned within the experiments. Note that $\langle d^2\rangle/\langle d\rangle \leq r/\lambda$ shows the prerequisite for controlling epidemic outbreaks. Therefore, a smaller recovery rate means that more isolations (and thus higher isolation costs) are needed to control epidemic outbreaks.

The following four algorithms are involved for evaluations:

- Random. It iteratively and uniform-randomly isolates a node in G, until the epidemic outbreak is eliminated by the quarantine strategy.
- MaxDegree. This algorithm ranks nodes by their degrees. Top ranked nodes are iteratively isolated until the epidemic outbreak is eliminated by the quarantine strategy.
- MargGreedy. This is Algorithm 1, which iteratively isolates the node v that can minimize $\frac{C_v}{\Delta(\{v\}\cup Q)-\Delta(Q)}$ (i.e., minimal "cost-to-benefit" ratio).
- HomoGreedy. This is Algorithm 2, which obtains a bounded result through the minimality property.

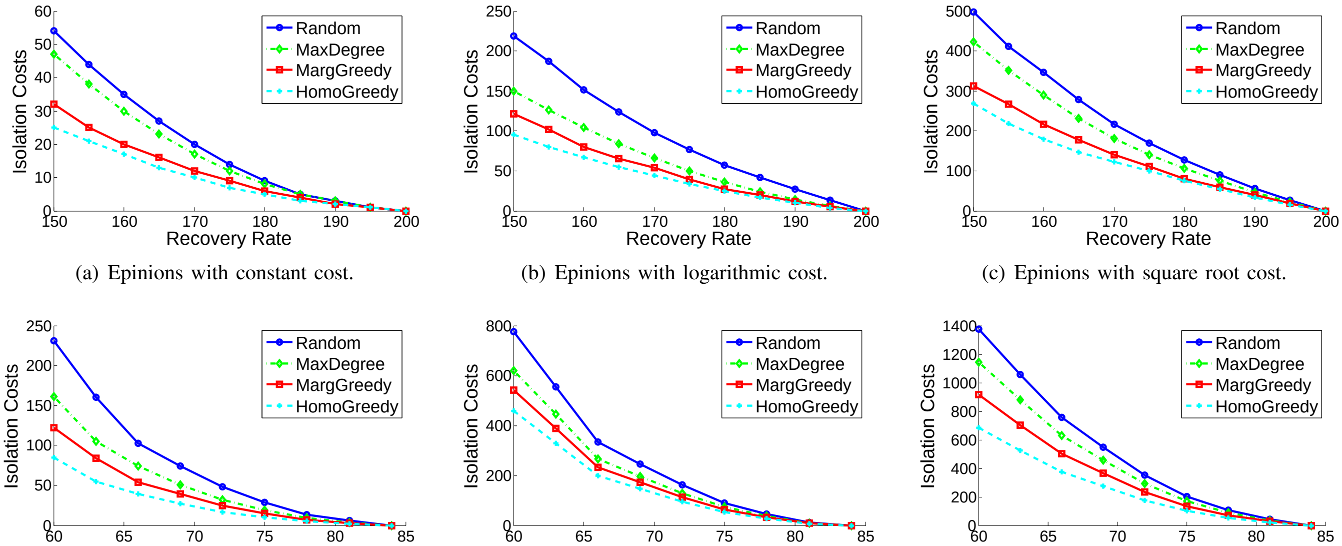
B. Evaluation Results

The evaluation results are shown in Fig. 2, in terms of the relationship between the recovery rate and the isolation cost. Fig. 2 has three columns, each of which corresponds to a different node isolation cost function. Figs. 2(a), 2(b), and 2(c) are the results for the Epinions dataset, while Figs. 2(d), 2(e), and 2(f) are the results for the Wikipedia dataset.

It can be seen that the isolation cost decreases monotonously with respect to the recovery rate. This is because a high recovery rate can resist epidemic spreadings. If the recovery rate is high enough, then epidemic outbreaks can be eliminated without isolations. HomoGreedy always has the lowest isolation costs among the comparison algorithms. Another interesting observation is that, when the cost of a node isolation is a constant, the total isolation cost is the smallest. This is because the quarantine strategy tends to isolate large degree nodes, while their costs are relatively cheap after normalization. On the other hand, when the cost of a node isolation scales with its degree (in a logarithmic manner or a square root manner), the overall isolation costs become very large. This is because the isolations of large degree nodes take relatively large costs after normalization. The last observation is that, the total isolation costs in Epinions are smaller than those in Wikipedia. One reason is that Epinions has a much larger degree variance than Wikipedia, as shown in Table I (355,754 to 103,689). Another reason is that Epinions has more users than Wikipedia (18,098 to 7,115), bringing larger isolation costs.

VI. Conclusion

This paper explores a robust quarantine strategy that can eliminate epidemic outbreaks with minimal isolation costs. This problem is proved to be NP-hard. The classic SIS epidemic model is introduced to model epidemic spreading, where people transfer their states through a cycle of being infected



(a) Epinions with constant cost.
(b) Epinions with logarithmic cost.
(c) Epinions with square root cost.

(d) Wikipedia with constant cost.
(e) Wikipedia with logarithmic cost.
(f) Wikipedia with square root cost.

Fig. 2. Evaluation results with respect to the isolation costs.

from susceptible, and going back to susceptible by recovery. We show that a minimal feasible quarantine strategy will not have excessive isolations. A bounded algorithm with an approximation ratio of two is proposed, through utilizing the feasibility and minimality properties. This algorithm has a time complexity of $O(V^2)$. Finally, real data-driven experiments demonstrate the efficiency and effectiveness of the proposed algorithms in real-world applications.

References

[1] F. D. Sahneh, C. Scoglio, and P. Van Mieghem, "Generalized epidemic mean-field model for spreading processes over multilayer complex networks," *IEEE/ACM Transactions on Networking*, vol. 21, no. 5, pp. 1609–1620, 2013.

[2] H. K. Lee, P.-S. Shim, and J. D. Noh, "Epidemic threshold of the susceptible-infected-susceptible model on complex networks," *Physical Review E*, vol. 87, no. 6, p. 062812, 2013.

[3] S. Wen, W. Zhou, J. Zhang, Y. Xiang, W. Zhou, and W. Jia, "Modeling propagation dynamics of social network worms," *IEEE Transactions on Parallel and Distributed Systems*, vol. 24, no. 8, pp. 1633–1643, 2013.

[4] D. Chen, "Modeling the spread of infectious diseases: A review," *Analyzing and Modeling Spatial and Temporal Dynamics of Infectious Diseases*, pp. 19–42, 2014.

[5] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and analysis of online social networks," in *Proceedings of ACM IMC 2007*, pp. 29–42.

[6] K. Lewis, M. Gonzalez, and J. Kaufman, "Social selection and peer influence in an online social network," *Proceedings of the National Academy of Sciences*, vol. 109, no. 1, pp. 68–72, 2012.

[7] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley, "Classes of small-world networks," *Proceedings of the National Academy of Sciences*, vol. 97, no. 21, pp. 11 149–11 152, 2000.

[8] B. Proulx and J. Zhang, "Modeling social network relationships via t-cherry junction trees," in *Proceedings of IEEE INFOCOM 2014*, pp. 2229–2237.

[9] X. Chen, I. Diakonikolas, D. Paparas, X. Sun, and M. Yannakakis, "The complexity of optimal multidimensional pricing," in *Proceedings of ACM-SIAM SODA 2014*, pp. 1319–1328.

[10] Z. He, J. Cao, and X. Liu, "High quality participant recruitment in vehicle-based crowdsourcing using predictable mobility," in *Proceedings of IEEE INFOCOM 2015*, pp. 2542–2550.

[11] O. Gurewitz, Y. Sandomirsky, and G. Scalosub, "Cellular multi-coverage with non-uniform rates," in *Proceedings of IEEE INFOCOM 2014*, pp. 1330–1338.

[12] A. Deshpande, L. Hellerstein, and D. Kletenik, "Approximation algorithms for stochastic boolean function evaluation and stochastic submodular set cover," in *Proceedings of ACM-SIAM SODA 2014*, pp. 1453–1467.

[13] R. Bar-Yehuda, "Using homogeneous weights for approximating the partial cover problem," *Journal of Algorithms*, vol. 39, no. 2, pp. 137–144, 2001.

[14] S. Khuller, M. Purohit, and K. K. Sarpatwar, "Analyzing the optimal neighborhood: algorithms for budgeted and partial connected dominating set problems," in *Proceedings of ACM-SIAM SODA 2014*, pp. 1702–1713.

[15] S. Funke, A. Nusser, and S. Storandt, "On k-path covers and their applications," *Proceedings of the VLDB Endowment*, vol. 7, no. 10, pp. 893–902, 2014.

[16] J. Zhang and J. M. Moura, "Diffusion in social networks as sis epidemics: beyond full mixing and complete graphs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 537–551, 2014.

[17] L. Hébert-Dufresne and B. M. Althouse, "Complex dynamics of synergistic coinfections on realistically clustered networks," *Proceedings of the National Academy of Sciences*, vol. 112, no. 33, pp. 10 551–10 556, 2015.

[18] R. Gandhi, S. Khuller, and A. Srinivasan, "Approximation algorithms for partial covering problems," *Journal of Algorithms*, vol. 53, no. 1, pp. 55–84, 2004.

[19] J. Tang, H. Gao, H. Liu, and A. Das Sarma, "etrust: Understanding trust evolution in an online world," in *Proceedings of ACM SIGKDD 2012*, pp. 253–261.

[20] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Predicting positive and negative links in online social networks," in *Proceedings of ACM WWW 2010*, pp. 641–650.

[21] S. Ewen, *Captains of consciousness: Advertising and the social roots of the consumer culture*. Basic Books, 2008.