Divide and Conquer: Leveraging Topology in Control of Epidemic Information Dynamics

Jie Wang Wenye Wang
Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, NC 27606
Email: {jwang50@ncsu.edu,wwang}@ncsu.edu

Cliff Wang
Army Research Office
Research Triangle Park, NC 27709
Email: cliff.wang@us.army.mil

***Abstract*—As online social networks grow in both size and connectivity, epidemic information dynamics in such networks is attracting considerable research interests, due to its impact on both the network and individuals. This paper studies control of malicious information (virus) epidemic with replicable antidote information, taking topological characteristics of the underlying graph into consideration. Specifically, we analytically relate the extinction time of the virus to the diameter and giant component size of the remaining graph after the initial antidote distribution. With this *divide and conquer* guideline, topology-based antidote distribution approaches are designed, and then examined through simulations in real world network portions.**

I. Introduction

Advances in communication and networking technologies are allowing humans' need of exchanging information to be satisfied more easily than ever, creating ever-growing, heavily-knit complex networks. The rapid expanding user population and shift of people's information acquisition source from tradition media to the Internet, have accelerated information circulation in Online Social Networks (OSNs), *e.g.* Facebook. Consequently, huge amount of data are flowing through OSNs in every single second. Facebook users alone generate 4 new petabytes of data per day [1]. The generation, forwarding and storage of information has considerable impacts on the participants, when malicious link/malware is hidden in the data, or in other cases, the information itself is a rumor or misinformation. For example, an innocent man was accused to be the Boston Bombing suspect by Reddit users, and this rumor went viral in Twitter, causing inconvenience to the wrongly accused [2]. Impacts are even more severe when the propagation is rapid, and the potential audience is large. In this sense, the dynamic of information propagation is usually described as an *epidemic* process of a piece of virus information. As a powerful tool, single virus epidemic has been studied in contexts of information diffusion in social networks [3], cascading failures in smart grid [4], *etc.*

Similarly as in epidemiology, we are interested in *how to contain hazard of the virus information*. In addition to immunization [5] and quarantine [6], that are also used in epidemiology, more flexible control measures can be applied in OSNs, due to its half-synthetic nature. To be more specific, OSNs differ from the contact network of a herd/flock/population in the following aspects: i) Changes in the underlying topology of the contact network are negligible within the time span of a certain information dynamic, resulting in a more steady and predictable infection pattern among individuals. ii) To alleviate the influence of the virus information, information with counter-value, later referred to as the *antidote information*, can be injected easily into the network. Unlike in epidemiology where only one individual can be cured by one unit of antidote, antidote information in OSNs can be *replicated* and *forwarded*, since itself is a piece of information.

Control of epidemic information propagation is usually modeled as an optimization problem under certain resource constraint. Preciado *et.al.* developed a convex framework [7] to evaluate the optimal allocation of edge control, immunization and non-replicable antidotes resources, in which the network is modeled as a directed graph. Borgs *et.al.* studied the optimal distribution of non-replicable antidote [8] given that the curing rate, proportional to the units of given antidote, is non-uniform. As for replicable antidote, Khouzani *et.al.* formulated the control strategy [9] with both replicable and non-replicable antidote into an optimal resource allocation problem. Chen utilized optimal control theory to determine the optimal distribution time [10] of a replicable antidote for timely control. However in [8]–[10], the information dynamic in the system is described by nonlinear differential equations, which indicates it is a *population dynamic*, rather than a *network dynamic*, or equivalently, topology of the network is not taken into consideration. On the other hand, the influence of network topology is studied in terms of epidemic threshold [11] under immunization and spreading/extinction time [6] under SI or SIS epidemic propagation models, while the dissemination of replicable antidote is not incorporated.

Therefore, this paper intends to explore *efficient distribution strategies of replicable antidote*, taking *topological characteristics* into account. The rest of the paper is is organized as follows. First we introduce the system model and restate our research question in Section II. Then in Section III, we analyze the influence of network topology on the upper bound of the virus extinction time, and utilize such result in designing antidote distribution strategies. Numerical results are presented in Section IV to validate the proposed mechanism and finally the paper is concluded in Section V.

This work is partially supported by NSF CNS1423151, CNS-1527696 and ARO W911NF-15-2-0102.

II. Problem Formulation

In this section, we review basic terminologies, settings and definitions of the Susceptible-Infected-Cured (SIC) epidemic model [12],then formulate our effective antidote distribution strategy problem.

A. System Model

The SIC epidemic model is used to describe the propagation of conflicting information, *i.e.* the propagation of a virus under the presence of an *infectious* antidote.

1) Network Model: The network $\mathcal{G}(\mathcal{V}, \mathcal{E})$ describes the relationship between individuals. For simplicity reasons, it is assumed to be connected, undirected and simple. As mentioned in the previous section, $\mathcal{G}$ remains *static* during the dynamic due to the negligible changes of the "following" relationships during a news/meme cycle [13].

2) Propagation Model: The state transition diagram of a vertex during an SIC epidemic is shown in Fig. 1, where S, I, C corresponds to the *susceptible*, *infected* and *cured* state respectively. During the dynamic, any infected vertex will pass a copy of virus to its susceptible neighbors after a random interval drawn from $Exp(\beta)$, and any cured vertex will pass a copy of antidote after a random interval drawn from $Exp(\gamma)$. β and γ are the *infection* and *curing* rate respectively.

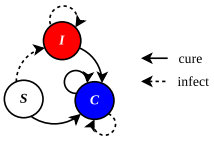


Fig. 1. State transition diagram of a vertex in an SIC epidemic.

Let $\mathcal{I}(t)$ and $\mathcal{C}(t) \subset \mathcal{V}$ denote the set of infected and cured vertices at time t, respectively. The antidote is distributed to the initial cured set $\mathcal{C}(0)$ at time 0, which we also refer to as the *initial distribution*. At $t = 0$, the virus is already present in the network, *i.e.* $\mathcal{I}(0) \neq \phi$. There is no interference from the outside after the initial distribution, hence the system evolve on their own along the edges of $\mathcal{G}$ after time 0.

B. Effectiveness of the Antidote Distribution

Since *cured* is the only absorbing state for each vertex, and $\mathcal{G}$ is connected, the virus will be eliminated in finite time. So we define the *extinction time* and use it as the indication of the effectiveness of a certain initial distribution.

DEFINITION 1. *The extinction time τ_e of the virus is defined as the first time after the initial distribution that none of the vertices are in infected state any more, i.e.*

$$\tau_e = \sup\{t > 0 | \mathcal{I}(t) = \phi\}.$$

In addition to when the virus is eliminated, we are also interested in how much time it takes an initial distribution to keep the infected set $\mathcal{I}(t)$ below a manageable size. Therefore, we introduce the *half-life time* of the virus epidemic.

DEFINITION 2. *The half-life time of the virus epidemic, denoted as $\tau_{\frac{1}{2}}$, is defined as the last time that the size of the infected set $\mathcal{I}(t)$ is larger than half of its original size $|\mathcal{I}(0)|$, i.e.*

$$\tau_{\frac{1}{2}} := \sup\{t \in [0, \tau_e] : |\mathcal{I}(t)| \geq \frac{1}{2}|\mathcal{I}(0)|\}.$$

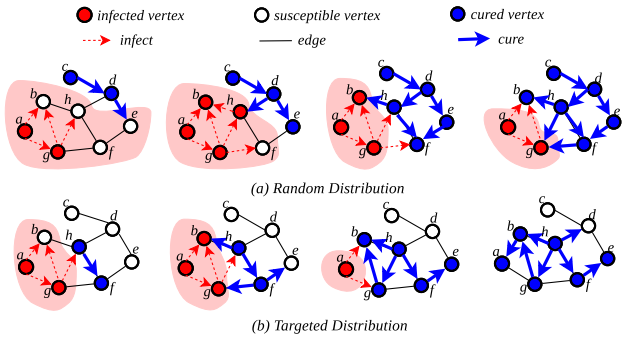


Fig. 2. An example of (a) random distribution; (b) targeted distribution.

Apparently, both the extinction time and the half-life time depends on the initial distribution $\mathcal{C}(0)$. Fig. 2 shows an example of the evolution of an SIC epidemic with different initial distributions. As can be seen, after the initial distribution, the potential hazard zone (pink shaded region) in the random distribution case is larger than that of the targeted distribution.

The effect of the replicable antidote is two-fold: on individual bases, it cures infected vertices which decreases the infection count; on the other hand, the expanding cured set composes a structure to retain the potential hazard zone of the virus. In the latter case the influence of network topology is more evident because the dynamics is changing a topological property of the system.

Provided the SIC epidemic model, this paper studies *the effective initial antidote distribution strategy*, that is, *the choice of the initial cured set $\mathcal{C}_0$ such that the mean extinction time $\mathbb{E}(\tau_e)$ and $\mathbb{E}(\tau_{\frac{1}{2}})$ of the virus can be shortened.*

III. Topology-based Antidote Distribution

In this section, we first relate the upper bound of expected extinction time $\mathbb{E}(\tau_e)$ to graphical characteristics of the remaining graph $\mathcal{G}^*(t)$ as the theoretic basis of the *divide and conquer* guideline, then propose several operable approaches to achieve the goal of shortening $\mathbb{E}(\tau_e)$ and $\mathbb{E}(\tau_{\frac{1}{2}})$.

A. Reiteration of the Expected Extinction Time $\mathbb{E}(\tau_e)$

As shown in the previous example in Fig. 2 (b), the potential hazard zone of the virus is restricted to a limited region of the network, *i.e.* a non-empty set of susceptible vertices are "quarantined" by the initial antidote distribution $\mathcal{C}(t)$, that will never be infected during the dynamic. We are interested in such *locking* condition, that effectively restrained the propagation of the virus. First we introduce the *initial locking time* to characterize such effect, eventually relating to the extinction time τ_e.

DEFINITION 3. *Let $\mathcal{G}^*(t)$ be the induced subgraph of $\mathcal{G}$ by removing the cured vertices $\mathcal{C}(t)$. We write $\mathcal{G}^*(t) = \cup_{1 \leq i \leq k(t)} G_i$, where $G_i(V_i, E_i)$ are components of $\mathcal{G}^*(t)$ and*

$k(t)$ denote the number of components at time t. The initial locking time τ_0 is defined as the first time that $\mathcal{G}^*(t)$ becomes disconnected, or equivalently

$$\tau_0 = \sup\{t > 0 \mid k(t) \geq 2\}.$$

Remark 1. *τ_0 marks a critical point of the virus epidemic, since at that instance, a topological property of the remaining graph $\mathcal{G}^*(t)$ has been changed, that is, the connectivity of $\mathcal{G}^*(t)$.*

Consequently in the dynamics, when $t < \tau_0$, $k(t) = 1$, the virus can potentially spread to every corner of the remaining graph $\mathcal{G}^*(t)$. As time t goes beyond τ_0, further fragmentation starts to happen in each $G_i \subset \mathcal{G}^*(t)$, and potential hazard of the virus can be treated as under control. Then we have the following theorem regarding the extinction time $\mathbb{E}(\tau_e)$.

THEOREM 1. *Let $C_0 = |\mathcal{C}(0)|$, $diam(G)$ and $\eta(G)$ denote the diameter and Cheeger constant of graph G, respectively. Then the expected extinction time can be upper bounded by*

$$\mathbb{E}(\tau_e) \leq \mathbb{E}(\tau_0) + \frac{1}{\gamma}[\frac{2}{(n - C_0)\eta(\mathcal{G})} + \max_{1\leq i\leq k(\tau_0)}\{diam(G_i)\}]. \tag{1}$$

Proof. Let r.v. Z_i and Y_i be defined as the following.

$$Z_i := \inf_{t>\tau_0}\{t - \tau_0 || V_i \cap \mathcal{I}(t)| = 1\}, \tag{2}$$

$$Y_i := \inf_{t>Z_i+\tau_0}\{t - (Z_i + \tau_0) || V_i \cap \mathcal{I}(t)| = 0\}. \tag{3}$$

Now we re-write $\mathbb{E}(\tau_e)$ with $\mathbb{E}(\tau_0)$,

$$\begin{aligned}\mathbb{E}(\tau_e) &\leq \mathbb{E}(\tau_0) + \mathbb{E}(\max_{1\leq i\leq k(\tau_0)}\{Z_i + Y_i\}) \\ &\overset{Jensen}{\leq} \mathbb{E}(\tau_0) + \max_{1\leq i\leq k(\tau_0)}\{\mathbb{E}(Z_i) + \mathbb{E}(Y_i)\} \\ &\leq \mathbb{E}(\tau_0) + \max_{1\leq i\leq k(\tau_0)}\{\mathbb{E}(Z_i)\} + \max_{1\leq i\leq k(\tau_0)}\{\mathbb{E}(Y_i)\}),\end{aligned} \tag{4}$$

Note that the first inequality of Eq. (4) follows from the fact that among all the components of $\mathcal{G}^*(\tau_0)$, some may not contain infected vertices. The physical meaning of Z_i is the time interval between the initial locking and the first vertex in G_i is cured, or equivalently, the minimum of $\delta(G_i)$ i.i.d exponential r.v.'s with parameter γ, where $\delta(G_i)$ denote the number of edges in edge cut $[V_i, \mathcal{V} \setminus V_i]$. Hence $Z_i \sim Exp(\delta(G_i)\gamma)$. Therefore

$$\begin{aligned}\max_{1\leq i\leq k(\tau_0)}\{\mathbb{E}(Z_i)\}) &= \frac{1}{\gamma}[\max_{1\leq i\leq k(\tau_0)}\{\frac{1}{|\delta(G_i)|}\} \\ &\leq \frac{1}{\gamma}[\max_{1\leq i\leq k(\tau_0)}\{\frac{1}{\eta(\mathcal{G})\min\{|V_i|, n - |V_i|\}}\} \\ &\leq \frac{1}{\gamma}\frac{2}{(n - C_0)\eta(\mathcal{G})}.\end{aligned} \tag{5}$$

For Y_i, it denotes the time interval between $\tau_0 + Z_i$ and the time that all the infected vertices in G_i are cured. Suppose the first cured vertex in G_i is $v_i \in V_i$, then Y_i is bounded above by the spreading time of the antidote along the shortest-path spanning tree of G_i rooted at v_i. Since $v_i \in {}_i$ can be any vertex that is directed connected to $\mathcal{C}(\tau_0)$. Then

$$Y_i \leq \max_{v_i\in E_i}\{\sum_{s=1}^{depth(G_i,v_i)}\Delta_C^s\} \leq \sum_{s=1}^{diam(G_i)}\Delta_C^s, \tag{6}$$

where $\Delta_C^s \sim Exp(\gamma)$ denote the time intervals between the $s-1$-th and the s-th curing, $depth(G_i, v_i)$ denotes the depth of the shortest-path spanning tree of G_i, and $diam(G_i)$ is the diameter (length of the longest shortest-path) of G_i. The second inequality of Eq. (6) follows from the fact that $depth(G_i, v_i) \leq diam(G_i)$, $\forall v_i \in E_i$, so $\sum_{s=1}^{depth(G_i,v_i)}\Delta_C^s \leq \sum_{s=1}^{diam(G_i)}\Delta_C^s$, $\forall v_i \in E_i$. So

$$\max_{1\leq i\leq k(\tau_0)}\{\mathbb{E}(Y_i)\}) \leq \frac{1}{\gamma}[\max_{1\leq i\leq k(\tau_0)}\{diam(G_i)\}] \tag{7}$$

Combining Eq. 4, (5) and (7) completes the proof. □

Theorem 1 implies the infection is defeated through a *divide and conquer* procedure. Specially, the extinction time can be mainly determined by $\mathbb{E}(\tau_0)$ and $max_{1\leq i\leq k(\tau_0)}\{diam(G_i)\}$, since $\eta(G) \geq \frac{n}{2}$ (the case that all vertices form a single line) indicates the second term in Eq. (1) is at most $O(1)$.

B. Ideal Antidote Distribution Policy

From a holistic view, there are two influential factors on the extinction time: the initial antidote distribution policy, and the network topology. The former includes the initial cured count C_0 and the recipient of the C_0 antidotes, *i.e.* assignment of $\mathcal{C}(0)$. Now consider the initial locking time τ_0. Apparently it is decreasing as the initial cured count C_0 increases, as well as $diam(G_i)$. However, it is not reasonable nor realistic to let C_0 approach n. We want to find the most effective way to distribute as less antidote as possible, under condition that the extinction time can be mostly shortened. So in this subsection, we assume that we have just enough antidote such that $\tau_0 = 0$, *i.e.* $\mathcal{C}(0)$ is a *vertex cut* of $\mathcal{G}$. Then we define the following metric to describe the impact of the graph topology on the extinction time with respect to an assignment of $\mathcal{C}(0)$.

DEFINITION 4. *The hazard index $\phi(\mathcal{C}(0))$ is defined as the maximum diameter of components of the initially disconnected graph $\mathcal{G}^*(t) = \cup_{1\leq i\leq k(t)}G_i$,*

$$\phi(\mathcal{C}(0)) = \max\{diam(G_i) \mid G_i \subset \mathcal{G}^*(\tau_0), 1 \leq i \leq k(\tau_0)\}.$$

Considering diameter is rather difficult to attain especially when the network is large, we provide the following upper bound of $\phi(\mathcal{C}(0))$ in terms of $|V_i|$, the number of vertices in each components of $\mathcal{G}^*(0)$.

THEOREM 2. *Let $d'(\mathcal{G})$ denote the minimum degree of graph $\mathcal{G}$, then*

$$\phi(\mathcal{C}(0)) \leq \frac{3}{d'(\mathcal{G}) + 1}\max_{1\leq i\leq k(\tau_0)}\{|V_i|\} - 1. \tag{8}$$

Proof. Let the minimum degree of each component G_i be denoted as d_i'. If it satisfies $d_i' \leq \frac{|V_i|-2}{2}$ ($d_i' > \frac{|V_i|-2}{2}$ is

highly unlikely, because it would suggest every component G_i is dense), then [14]

$$diam(G_i) \leq 3\lfloor \frac{|V_i|}{d'_i+1} \rfloor - \begin{cases} 3, & n \mod (d'_i+1) = 0; \\ 2, & n \mod (d'_i+1) = 1; \\ 1, & \text{otherwise.} \end{cases} \tag{9}$$

Then

$$\begin{aligned} \phi(\mathcal{C}(0)) &\leq \max_{1\leq i\leq k(\tau_0)} \{3\lfloor \frac{|V_i|}{d'_i+1} \rfloor - 1\} \\ &\leq \max_{1\leq i\leq k(\tau_0)} \{3(\frac{|V_i|}{d'(\mathcal{G}^*(\tau_0))+1}) - 1\} \\ &\overset{\text{Note}}{\leq} \max_{1\leq i\leq k(\tau_0)} \{3(\frac{|V_i|}{d'(\mathcal{G})+1}) - 1\} \\ &\leq \frac{3}{d'(\mathcal{G})+1} \max_{1\leq i\leq k(\tau_0)} \{|V_i|\} - 1. \end{aligned} \tag{10}$$

Note that the third inequality in Eq. (10) is not strict, but still reasonable. Since $d'(\mathcal{G}^*(\tau_0)) \geq d'(\mathcal{G})$ holds except the case that the vertext with the minimal degree (denote as h) is directly adjacent to a vertex in $\mathcal{C}(0)$. If the degree of h is decreased much due to the removal of $\mathcal{C}(0)$, then it would be more convenient to include h in $\mathcal{C}(0)$ at the initial antidote distribution, and in this case, the degree of the remaining vertices is decreased at most 1, due to the fact that $\mathcal{G}$ is simple. □

Theorem 2 provides an upper bound of $\phi(\mathcal{C}(0))$ in terms of the size of the giant component $\max_{1\leq i\leq k(\tau_0)}(|V_i|)$ in $\mathcal{G}(\tau_0)$, where $\tau_0 = 0$ in this case. In discrete mathematics, this is equivalent to the *most balanced cut* problem, that is, finding such vertex-cut constrained on different balance requirements, which in our case, is the number of vertices in each component G_i (order of G_i).

On the other hand, clearly from Eq. (1), our goal of assigning $\mathcal{C}(0)$ is to minimize the hazard index, *i.e.* $\min_{\mathcal{C}(0)}\{\phi(\mathcal{C}(0))\}$. From the Moore Bound [15] Inequality, we know that the diameter of each subgraph G_i can be lower bounded by[1]

$$diam(G_i) \geq \begin{cases} \frac{|V_i|-2}{2}, & d_i = 2, \\ \log_{d_i-1}[(|V_i|-1)\frac{d_i-2}{d_i}) + 1], & d_i > 2, \end{cases}$$

where d_i is the maximum degree and $|V_i|$ is the number of vertices of graph G_i respectively. Let $f(d_i, |V_i|) := \log_{d_i-1}[(|V_i|-1)\frac{d_i-2}{d_i}) + 1]$. We can show that $f(d_i, |V_i|)$ is decreasing in d_i, but increasing in $|V_i|$. In the mean time, $\frac{|V_i|-2}{2}$ is also increasing in $|V_i|$, irrelevant to d_i. So

$$f(d_i, |V_i|) \geq f(d(\mathcal{G}), |V_i|),$$

where $d(\mathcal{G})$ denotes the maximum degree of the network $\mathcal{G}$.

Now minimizing $\max\{diam(G_i)\}$ becomes minimizing $\max\{|V_i|\}$, that is, we want to find a minimum vertex cut $\mathcal{C}(0) \subset \mathcal{V}$, such that in the induced subgraph $\mathcal{G}^*(\tau_0) = \mathcal{G} \setminus \mathcal{C}(0)$, $\min_{\mathcal{C}(0)}\{\max_i\{|V_i|\}\}$ can be achieved. Again, the upper bound of $\phi(\mathcal{C}(0))$, (and hence that of $\mathbb{E}(\tau_e)$) is related to the maximum number of vertices in each component of $\mathcal{G}^*(\tau_0)$, which also leads to the *minimum most balanced vertex-cut* problem. By assigning $\mathcal{C}(0)$ to the minimum most balanced cut, the upper bound of $\mathbb{E}(\tau_e)$ in Eq. (1) can be tightened because $\max_i\{diam(G_i)\}$ is tightened.

Therefore, to maximize the effect of antidote, as well as to minimize the extinction time, the target of an ideal initial antidote distribution strategy is to assign $\mathcal{C}(0)$ to the minimum most balanced vertext-cut of the network $\mathcal{G}$.

C. Realistic Approaches

However, in some networks, it is not possible to find a "small" vertex-cut, let alone a minimum most balanced vertex-cut. For example, in the complete graph K_n, the minimum vertex-cut contains $n-1$ vertices. It translates to either a long τ_0, or a large enough C_0, but $\mathbb{E}(Y_i) \simeq \frac{1}{\gamma(n-1)}$, in which case the upper bound in [12] Section IV.A is more applicable. In addition, searching for the minimum most balanced cut or directly searching for vertices whose removal will result in a smaller $\phi(\mathcal{C}(0))$ is difficult. Especially, finding the minimum most balanced cut for general graphs is NP-Hard [16]. So we introduce the following operable approaches under the guideline of minimizing $\phi(\mathcal{C}(0))$, or minimizing the size of the giant component in $\mathcal{G}^*(\tau_0)$.

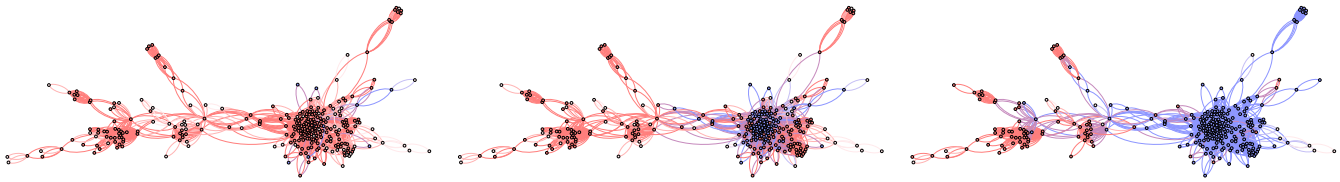
1) betcen-based approach: Betweenness centrality of a vertex v is defined as $g(v) = \sum_{s\neq v\neq t} \frac{\delta_{st}(v)}{\delta_{st}}$, where δ_{st} denotes the number of shortest paths between vertex s and v, while $\delta_{st}(v)$ are the number of those paths that passes through vertex v. $g(v)$ indicates how likely vertex v sits in other vertices' shortest paths. However, it requires global knowledge and takes $O(|\mathcal{V}||\mathcal{E}|)$ [17] time to calculate. Removal of a vertex with high g value will likely break more shortest paths, rendering a disconnected network.

2) ccfs-based approach: Clustering Coefficient of a vertex v is defined as $C(v) = \frac{2|\{e_{st}|e_{st},e_{vs},e_{vt}\in\mathbb{E}\}|}{d(v)[d(v)-1]}$, which shows how densely connected is v's neighborhood. To calculate $C(v)$, knowledge of vertices within the distance of two to v is required. The reason of using C value as an indicator is that when $C(v)$ is a small but strictly positive value, it implies that v's neighbor is not well-connected, and relies v to function as a bridge between its neighbors. This is especially true when the graph does not have many "long edges".

3) degree-based approach: Degree of vertex v, $d(v)$ is easy to attain because it only requires knowledge of one-hop neighbors of a vertex. Higher degree indicates a vertex has a higher chance of being a hub. Hence the removal of such a vertex will result in the removal of a lot of edges.

In the implementation of these approaches in Section IV, we first calculate the g, C and d values for each vertex, and then sort them to find the best candidates. Considering the sorting process require global knowledge of the graph, in real world implementation, the sorting process can be substituted with a cut-off mechanism with predetermined threshold values.

[1]The Moore Bound is fairly difficult to attain, but it is a valid lower bound with respect to the number of vertices of each component.



(a) $t = 0$ (b) $t = 150$ (c) $t = 400$

Fig. 3. A realization of an SIC information dynamic evolution after an initial distribution with ccfs-based strategy. The topology of the network is based on dataset0, initial cured count $|\mathcal{C}(0)| = C_0$ equals to 40, while the initial infection count $|\mathcal{I}(0)| = I_0$ equals to 200. Propagation parameters: $\beta = \gamma = 0.003$.

IV. Numerical Results and Discussion

To validate and compare the proposed approaches, we first analyze an extreme case scenario, and then present simulation results in network portions acquired from the real world OSN, Facebook. Fig. 3 shows a set of snapshots of an SIC information dynamics evolution in dataset0, where color red, white and blue indicates infected, susceptible and cured respectively.

A. A Special Case: the Star Network

Consider a special case when $\mathcal{G} = S_n$, i.e. the star network with one hub and $n-1$ peripherals. Based on all three approaches, the hub will be the first one selected in $\mathcal{C}(0)$, due to its high g and d value, as well as its low C value. When the hub is cured, the SIC dynamic is in a *locked* condition, that is, the SIC dynamic fragmented the remaining of the star into disconnected vertices, leaving no further expansion space for the virus. The extinction time τ_e in S_n will be the maximum of I_0 i.i.d. r.v.'s, each with distribution $Exp(\gamma)$, and $\mathbb{E}(\tau_e) = \frac{\mathcal{H}_{I_0}}{\gamma}$, where $\mathcal{H}_k$ is the k-th Harmonic number.

B. SIC Information Dynamics in Real Networks

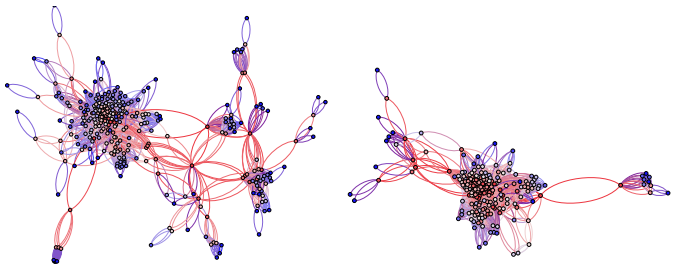
To examine the effectiveness and efficiency of the proposed antidote dissemination policy, we conducted simulation of SIC dynamics on two connected network, both fractions from Facebook [18], dataset0[2] and dataset348. Statistics of the two networks fraction are shown in Table I. As can be seen from the average degree, dataset348 is much denser than dataset0. In addition, from the average ccfs (clustering coefficient), dataset348 is more clustered than dataset0.

TABLE I
Statistics of the two underlying graphs.

Statistics	dataset0	dataset348
order (number of vertices)	324	224
size (number of edges)	5028	6384
average degree	31.037	57
diameter	11	9
average ccfs	0.522	0.544
average path length	3.573	3.042

Fig. 4 shows topology of the two networks, where the betweenness centrality value is indicated by color. Candidates for $\mathcal{C}(0)$ in the betcen-based approach are the vertices in red.

[2]Originally, dataset0 contains 342 vertices and is disconnected, so we select the giant component to be dataset0 during the simulation.



(a) dataset0 (b) dataset348

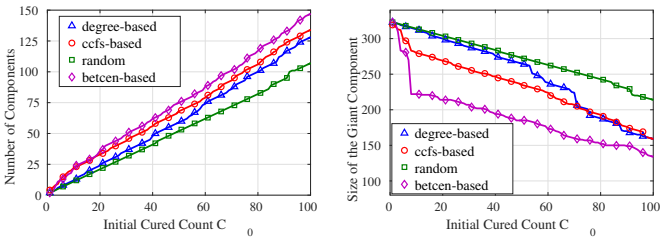
Fig. 4. Topology of the two networks. Red vertices have high betweenness centrality (betcen, g) values, while blue ones have low g values.

Intuitively from the figure, we can tell that dataset0 is more "scattered" than dataset348, which implies that it will be easier (with a smaller $C_0 = |\mathcal{C}(0)|$) to achieve the locking condition, *i.e.* $\mathcal{G}^*(0)$ is disconnected.

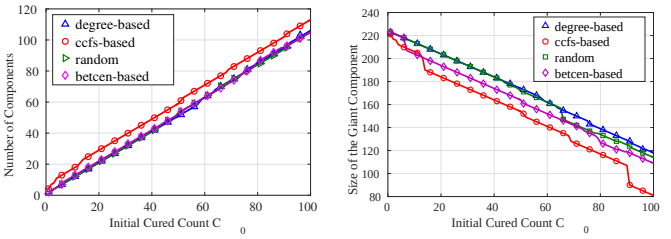
Fig. 5 illustrates topological changes of dataset0 and dataset348, induced by different initial antidote distribution strategies, *i.e.* different assignments of $\mathcal{C}(0)$. These results can be used to predict the effectiveness of those three approaches, plus a random distribution, in terms of $\mathbb{E}(\tau_e)$ and $\mathbb{E}(\tau_{\frac{1}{2}})$. As discussed in Section III, the approach that can minimize the size of the giant component will most effectively shorten the extinction time. From the topological characteristics shown in Fig. 5, it is interesting to see that betcen-based approach will be outperformed by ccfs-based approach in dataset348, while in dataset0 it is the opposite. The possible reason is that dataset348 is much denser (avg. degree 57) and more clustered (avg. ccfs 0.544) than dataset0, which is also manifested in Fig. 4. This implies the betcen-based approach will disconnect dataset0 more easily, while leaving dataset348 still connected during the initial distribution.

Fig. 6 illustrate the mean extinction time and half-life time, each over 1000 simulation runs. The propagation parameters of the SIC epidemic are: infection rate $\beta = 0.01$, the curing rate $\gamma = 0.01$ and initial infection count $I_0 = |\mathcal{I}(0)| \simeq \frac{1}{2}|\mathcal{V}|$, *i.e.* 150 for dataset0 and 110 for dataset348.

Simulation results in Fig. 6 echoes with the prediction we had from Fig. 5, in which performance of degree-based approach is poor, and the best approach is either betcen-based or ccfs-based. In terms of extinction time, the degree-based approach is even worse than random distribution, because high degree does not imply high importance. Due to the

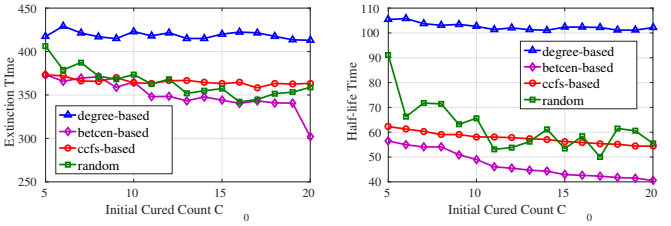


(a) No. of components in dataset0 (b) Giant component size in dataset0

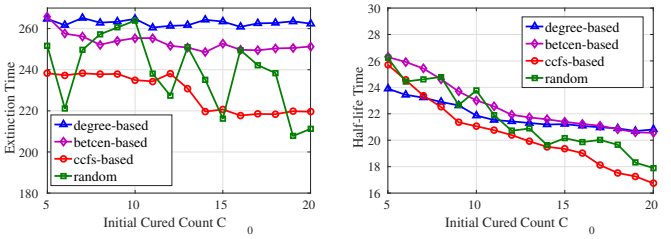


(c) No. of components in dataset348 (d) Giant component size in dataset348

Fig. 5. Topological change of $\mathcal{G}^*(0)$ with different $\mathcal{C}(0)$.



(a) $\mathbb{E}(\tau_e)$ on dataset0 (b) $\mathbb{E}(\tau_{\frac{1}{2}})$ on dataset0



(c) $\mathbb{E}(\tau_e)$ on dataset348 (d) $\mathbb{E}(\tau_{\frac{1}{2}})$ on dataset348

Fig. 6. Extinction time $\mathbb{E}(\tau_e)$ and Half-life time $\mathbb{E}(\tau_{\frac{1}{2}})$ under different initial distibution strategies on network dataset0 and dataset 348.

clustering effect of human social interactions, high degree vertices is often located in a densely connected core, where the removal of such a vertex can be compensated by its neighbors. However, the zig-zag pattern of the green line, *i.e.* the random distribution, indicates the instability of this approach. What's more, considering the ccfs-based approach only requires knowledge of two-hop neighbors, rather than global knowledge in the betcen-based approach, it is an optimal choice, especially when the network is denser. As for the half-life time, which indicates the effectiveness in alleviating heavy infection conditions, simulation suggests similar conclusion as the extinction time, except that degree-based distribution performs better when C_0 is small in dataset348, where the network is more clustered.

V. Conclusion

In this paper, we examined the influence of topology on the expected extinction time of an SIC information dynamic through theoretic analysis, and leveraged such influence in designing effective antidote distribution strategies. Real world network portions were included in the simulation validation. We hope our work would contribute to the knowledge of information dynamic in networks.

References

[1] N. B. Janet Wiener, "Facebook's top open data problems." https://research.facebook.com/blog/facebook-s-top-open-data-problems/, 2014.

[2] D. Lee, "Boston bombing: How internet detectives got it very wrong." http://www.bbc.com/news/technology-22214511, 2013.

[3] Y.-C. Chen, H.-S. Ma, and J.-W. Huang, "Multi-state open opinion model based on positive and negative social influences," in *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, ASONAM '15, (New York, NY, USA), pp. 170–177, ACM, 2015.

[4] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the us power grid," *Safety Science*, vol. 47, no. 10, pp. 1332–1336, 2009.

[5] B. A. Prakash, H. Tong, N. Valler, M. Faloutsos, and C. Faloutsos, "Virus propagation on time-varying networks: Theory and immunization algorithms," in *Proceedings of the 2010 European Conference on Machine Learning and Knowledge Discovery in Databases: Part III*, ECML PKDD'10, (Berlin, Heidelberg), pp. 99–114, Springer-Verlag, 2010.

[6] S. Krishnasamy, S. Banerjee, and S. Shakkottai, "The behavior of epidemics under bounded susceptibility," in *The 2014 ACM International Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '14, (New York, NY, USA), pp. 263–275, ACM, 2014.

[7] V. Preciado, M. Zargham, and D. Sun, "A convex framework to control spreading processes in directed networks," in *Information Sciences and Systems (CISS), 2014 48th Annual Conference on*, pp. 1–6, March 2014.

[8] C. Borgs, J. Chayes, A. Ganesh, and A. Saberi, "How to distribute antidote to control epidemics," *Random Structures & Algorithms*, vol. 37, no. 2, pp. 204–222, 2010.

[9] M. Khouzani, S. Sarkar, and E. Altman, "Optimal control of epidemic evolution," in *INFOCOM, 2011 Proceedings IEEE*, pp. 1683–1691, April 2011.

[10] P. Y. Chen and K. C. Chen, "Optimal control of epidemic information dissemination in mobile ad hoc networks," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pp. 1–5, Dec 2011.

[11] M. Lelarge, "Efficient control of epidemics over random networks," in *Proceedings of the Eleventh International Joint Conference on Measurement and Modeling of Computer Systems*, SIGMETRICS '09, (New York, NY, USA), pp. 1–12, ACM, 2009.

[12] J. Wang and W. Wang, "To live or to die: Encountering conflict information propagation over simple networks," in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications (INFOCOM 2016)*, (San Francisco, USA), Apr. 2016.

[13] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, (New York, NY, USA), pp. 497–506, ACM, 2009.

[14] L. Caccetta and W. Smyth, "Graphs of maximum diameter," *Discrete Mathematics*, vol. 102, no. 2, pp. 121 – 141, 1992.

[15] M. Miller and J. Siran, "Moore graphs and beyond: A survey of the degree/diameter problem," *Electronic Journal of Combinatorics (Dynamic Surveys)*, pp. 1–92, May 2013.

[16] M. Kim and K. S. Candan, "Sbv-cut: Vertex-cut based graph partitioning using structural balance vertices," *Data & Knowledge Engineering*, vol. 72, pp. 285 – 303, 2012.

[17] U. Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Sociology*, vol. 25, pp. 163–177, 2001.

[18] J. Leskovec and J. J. Mcauley, "Learning to discover social circles in ego networks," in *Advances in Neural Information Processing Systems 25* (P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), pp. 548–556, 2012.