

Modeling and Analysis of Conflicting Information Propagation in a Finite Time Horizon

Jie Wang^{id}, *Student Member, IEEE*, Wenye Wang, *Fellow, IEEE*, and Cliff Wang *Fellow, IEEE*

Abstract—Emerging mobile applications enable people to connect with one another more easily than ever, which causes networked systems, *e.g.*, online social networks (OSN) and Internet-of-Things (IoT), to grow rapidly in size, and become more complex in structure. In these systems, different, even *conflicting information*, *e.g.*, rumor *v.s.* truth, and malware *v.s.* security patches, can compete with each other during their propagation over individual connections. For such information pairs, in which a desired information kills its undesired counterpart on contact, an interesting yet challenging question is *when and how fast the undesired information dies out*. To answer this question, we propose a *Susceptible-Infectious-Cured (SIC)* propagation model, which captures short-term competitions between the two pieces of information, and define *extinction time* and *half-life time*, as two pivots in time, to quantify the dying speed of the undesired information. Our analysis revealed the impact of network topology and initial conditions on the lifetime of the undesired information. In particular, we find that, the Cheeger constant that measures the edge expansion property of a network steers the scaling law of the lifetime with respect to the network size, and the vertex eccentricities that are easier to compute provide accurate estimation of the lifetime. Our analysis also sheds light on where to inject the desired information, such that its undesired counterpart can be eliminated faster.

Index Terms—Information propagation, conflicting information, epidemic models, network dynamics.

I. INTRODUCTION

DUE to proliferating mobile devices and emerging mobile applications, people are more connected with each other than ever. Consequently, networked systems, such as online social networks (OSN), institutional computer networks, and Internet-of-Things (IoT), are developing into much larger and more complex structures than before. For instance, the OSN giant Facebook has 1.6 billion daily active users, who generate more than 4 PetaByte new data every day [1], while the number of IoT devices is expected to exceed 500 billion by 2030 [2], imposing a tangible impact on mobile data traffic. As more and more individuals, *e.g.*, users in OSN, and devices

in IoT, join such systems, inconsistent, even *conflicting* information are injected into the network, leading to an interesting competition among different pieces of information.

By conflicting, we mean two pieces of information that can not be admitted by the same individual at the same time. For example, it is highly unlikely, if not impossible, for an OSN user to simultaneously admit the truth and a rumor that contradicts with the truth. Particularly for such pairs, in which one piece (the desired information, *e.g.*, truth) is apparently more credible than the other (the undesired information, *e.g.*, a rumor), an individual who has chosen to admit the desired information, will not be affected by its undesired counterpart. Therefore, the undesired information, spreading via individual connections itself, will be eliminated from the network (die out) by its desired counterpart, given sufficient time. In other words, the competition between the conflicting information pair takes place in *finite time*. Naturally, we ask: *when, and how fast will the undesired information die out?*

A. Motivating Examples

The phenomenon of conflicting information propagation is prevalent in OSNs, which has become the arena of clashing opinions, unverified reports, and publicity campaigns. For example, after the Boston bombing incident on April 15th, 2013, Reddit users started an online suspect hunt, which identified an innocent person as the bomber [3]. This rumor (undesired information in the form of image data) spread rapidly on both Reddit and Twitter, leading to serious cyber-harassment to the wrongly-accused. The number of mentions regarding this rumor quickly decreased after the police released the correct information (desired information) [4]. Before the rumor dies out completely, its impact, harassment to the wrongly-accused, as well as posts and clicks from users, will not disappear, so answer to the ‘when and how fast’ question directly characterizes the data traffic in the system over time.

On the other hand, such competing-while-spreading phenomenon can also be observed in engineered systems, such as computer networks for an institution, and IoT. For instance, to eliminate a computer malware (undesired information in the form of source code), such as SMS Trojans [5] that spreads over emails and messages, and Chameleon [6] that spreads over WiFi links, the system administrator can distribute self-replicable security patches (desired information) to ‘cure’ malware-infected devices, and ‘immune’ devices that have not been reached by the malware. In IoT, on the other hand, faults (undesired information) can quickly propagate due to the interdependence of data/power between devices, leading to a cascade-of-failures, while restoration operations (desired

Manuscript received October 11, 2017; revised August 19, 2018; accepted January 24, 2020; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor E. Yeh. This work was supported in part by the NSF under Grant CNS1423151 and Grant CNS1527696 and in part by the ARO under Grant W911NF-15-2-0102. This article was presented at the IEEE INFOCOM 2016. (*Corresponding author: Jie Wang.*)

Jie Wang and Wenye Wang are with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC 27606 USA (e-mail: jwang50@ncsu.edu; wwang@ncsu.edu).

Cliff Wang is with the Army Research Office, Research Triangle Park, Raleigh, NC 27709 USA (e-mail: cliff.wang@us.army.mil).

This article has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors.

Digital Object Identifier 10.1109/TNET.2020.2976972

information) on some devices, *e.g.*, load-shedding, will also trigger/enable the recovery of faulty devices that are connected to restored devices, gradually bringing the system back to normal operations. In these examples, desired information is proactively injected by the administrator as a countermeasure against malware or cascade-of-failures. In this sense, answer to the ‘when and how fast’ question also measures how effective a countermeasure is, in controlling the undesired information from epidemic spreading.

These conflicting information propagation processes have two defining characteristics: First, there are *two* pieces of information circulating the same network, both spread via contacts of individuals in an *epidemic* manner. Second, the later-injected desired information can convert victims of the undesired information back to normal states, just like a replicable *antidote* can cure/immune an individual from an infectious *virus*, as a result of which, the cured/immune individual will not be infected by the same virus again, but not vice versa. Our objective is to find out when and how fast the virus dies out in the network, after the injection of antidotes.

B. Related Work

Considering the resemblance between information spread over individual connections, and virus spread over individual contacts, information propagation has been extensively studied via epidemic models, in which information is modeled as an infectious virus. Based on the number of information pieces considered, existing literature can be broadly categorized into single-virus epidemics, and competing epidemics. However, none of existing approaches can describe the short-term competition between the *virus* and *antidote*.

Existing research on single-virus epidemics focus on propagation model [7], [8], epidemic threshold [9], [10], dependency on network topology [11]–[13] and epidemic control with non-self-replicable antidotes [12], [14]. However, these models only consider one piece of the information in the spreading process, and are hence not applicable to our problem. In fact, a single-virus epidemic is a special case of the conflicting information propagation process discussed in this paper, in the sense that the virus spreads to every corner of the network, and lives forever without any injection of antidote.

With respect to competing epidemics between conflicting information, existing literature can be further categorized into *population dynamics* and *network dynamics* [15], depending on whether network topology is taking into consideration. In population dynamics, participants of the propagation process are assumed to be a well-mixed population, *i.e.*, there is no notion of network, which does not apply to most propagation scenarios. In contrast, network dynamics view participants of the propagation process as heterogeneous, and model their connections with a graph structure. Among these, Lin *et al.* [16] utilized Mean-Field Approximation (MFA) to conduct asymptomatic and numerical analysis of the propagation process. Prakash *et al.* [17] proposed an SI_1I_2S model, and proved that a piece information with faster propagation speed will eliminate its slower counterpart as time approaches infinity. More recently, Dadlani *et al.* modeled competing memes as epidemic processes on multi-layered graphs, and

derived critical survival threshold of a meme [18] to be persistent. Newman [19] found the coexistence threshold of two competing epidemics on networks with known degree distributions, under Susceptible-Infected-Recovered (SIR) model. From the perspective of propagation model, both the linear threshold model and SI_1I_2S model allow any individual to switch back and forth between different information, in which context the asymptotic behavior (steady state of the system as time $t \rightarrow \infty$) of the dynamics is of more interest. In our case, however, the desired information is much more credible than its undesired counterpart, so competition between the two finishes in *finite time*. Consequently, the answer to our research question depends on *transient* changes of individual states, to which existing analysis on asymptotic behavior does not apply.

C. Summary of Contributions

Seemingly simple, the *when* and *how fast* question is actually challenging because of the short-term competition between conflicting information, and the possibly complex network topology of the system. First, answer to this question depends on occurrences of transient events, analysis of which involves time and is hence difficult. Second, the underlying network can be large, complex, or both, considering the broad applications of this problem, for which simple propagation rule at node level gives rise to complicated behavior of the system as a whole. Therefore, we need to identify useful topological properties of the network that can best answer the question, but are also accessible in practice for complex networks.

To address these challenges, contributions of this paper can be summarized as follows.

- 1) **Modeling:** We propose a novel SIC propagation model, to capture the competing and epidemic nature of a conflicting information pair in their propagation processes.
- 2) **Metrics:** We define new metrics, namely the *extinction time* τ_e and *half-life time* $\tau_{\frac{1}{2}}$, as two pivots in time, to quantify how fast the virus dies.
- 3) **Findings:** We derive bounds of the expected extinction time $\mathbb{E}(\tau_e)$ and half-life time $\mathbb{E}(\tau_{\frac{1}{2}})$ for network \mathcal{G} with an arbitrary topology, and find that:
 - Both $\mathbb{E}(\tau_e)$ and $\mathbb{E}(\tau_{\frac{1}{2}})$ are $O(\frac{\log n}{\eta(\mathcal{G})})$, where n is the network size, and $\eta(\mathcal{G})$ is the Cheeger constant measuring the level of bottleneck-ness of \mathcal{G} . It indicates that the undesired information dies out slower in larger networks with bottlenecks ($\eta(\mathcal{G}) \leq O(\log n)$), *e.g.*, networks with star topology. In contrast, the extinction time decreases with n when $\eta(\mathcal{G})$ is bounded below by $\log n$.
 - Considering $\eta(\mathcal{G})$ is difficult to obtain (NP-hard) for general networks, we show that the lifetime of the undesired information can also be upper bounded by functions of vertex eccentricities. These bounds enable us to estimate the extinction time in large complex networks, and also imply where to inject desired information such that its undesired counterpart can be eliminated faster.

The rest of the paper is organized as follows. First we introduce the SIC propagation model in Section II, followed by metrics definitions and problem formulation for a conflicting information propagation process. Then we study SIC epidemics on networks of two special topologies, *i.e.*, clique and star, in Sec. III. Based on the observation of these preliminary results, we identify two topological properties, namely, graph expansion and vertex eccentricity, by which we derive bounds for networks with arbitrary topology in Sec. IV, and provide practical guidelines on selecting antidote recipients to control the epidemic of undesired information. Finally, the paper is concluded in Sec. V.

II. THE SIC PROPAGATION MODEL

In this section, we introduce terminologies, assumptions and definitions of the *Susceptible-Infected-Cured* (SIC) propagation model, to capture the competition between a pair of *conflicting* information, referred to as a *virus* and an *antidote*, both of which propagate in an epidemic manner. This dynamic process takes place in a network, whose vertices represent individuals, *e.g.*, devices in IoT, and edges represent connections, *e.g.*, wireless links between devices. Based on the SIC model, we then formally define metrics to quantify the dying speed of the undesired information (virus), as the expected answers to the proposed *when* and *how fast* questions.

A. Conflicting Information Pair: Virus x and Antidote a_x

Conflicting information has been defined as “two pieces of text that are extremely unlikely to be considered true simultaneously” in relevant information collection studies [20]. Considering that information takes various forms other than text, such as source code, operating status of devices, and commands, we extend this definition to *mutually-exclusive* information that can not be possessed/admitted by the same individual at the same time. Particularly, we focus on a *virus-antidote* pair, in which the desired information (referred to as antidote a_x) is of dominant credibility/power over its undesired counterpart (referred to as virus x), such that it kills the virus if they are both present on the same vertex, but not vice versa.

Note that virus x can not re-infect an individual, who has admitted antidote a_x , which is different to the symmetric setting in existing models [17], [21], where virus x can re-infect an individual who already has a_x , so a_x is treated as another virus, instead of an antidote to x . The rationale behind the asymmetry in our model comes from observations in real-world examples and concerns on modeling accuracy. First, for the case of security patch v.s. computer malware, and restoration commands v.s. (cascading) faulty status, the desired information (*e.g.*, security patch) is injected purposely by the system to eliminate the undesired information (*e.g.*, malware), and it is only reasonable that a malware can not leverage a fixed bug to attack the system. Second, for the case of clashing opinions and adoption of different products, once an individual is convinced (infected) by the newer and better product a_x , it is unlikely for him/her to switch back to the older product x , unless the older product has an upgrade (to the newest version) b_x . This new injection of b_x is considered as the

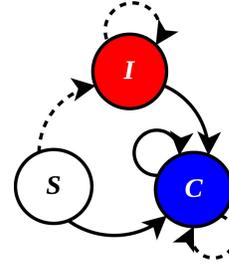


Fig. 1. State transitions of the SIC epidemic model.

beginning of a new epidemic process with b_x as the antidote to virus a_x in our model. In the $SI_1 \cdots I_k S$ model, b_x and x are considered as the same virus competing with a_x , such that re-infection (or switching-back from a_x to x) is allowed, because of its focus on long-term (time $t \rightarrow \infty$) behaviors. A drawback of the existing approach, *e.g.*, [17], [21], is that b_x and x are automatically associated with the same propagation speed. Our approach, on the contrary, allows x , a_x , b_x to have different propagation speeds, capturing competitions of both a_x v.s. x , and b_x v.s. a_x , and is hence more accurate.

B. Network Model $\mathcal{G}(\mathcal{V}, \mathcal{E})$

The *network* is described as a graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$, where vertex set \mathcal{V} corresponds to the set of individuals in the system, and edge set \mathcal{E} corresponds to the set of connections between any of two individuals. An edge $e(i, j)$ exists when vertex i and j can directly exchange information. For any vertex $v \in \mathcal{V}$, its *neighborhood* $\mathcal{N}(v) := \{u \in \mathcal{V} \mid (u, v) \in \mathcal{E}\}$ is defined as the set of vertices that is directly connected to vertex v . We make the following assumptions of the network \mathcal{G} : i) It is undirected, that is, edge $e(i, j) = e(j, i)$ identifies mutual connections between vertices i and j . ii) It is connected, such that information (x and a_x) can spread to every vertex in \mathcal{V} . iii) It is static, that is, both size $n := |\mathcal{V}|$ and topology of the network remain the same during the epidemic evolution.

C. Epidemic Propagation Process

Both virus x and antidote a_x spread in an *epidemic* manner. To describe this epidemic process, each vertex is associated with a *state* that can change over time.

1) *State Transitions*: Let r.v. $X_v^x(t) : \Omega \rightarrow \Lambda = \{0, 1, -1\}$ denote the state of vertex $v \in \mathcal{V}$ at time t . Values of X_v^x correspond to different nodal states, as shown in Fig. 1.

Susceptible at time t : The default state $X_v^x(t) = 0$ (white circle with letter S in Fig. 1 indicates that neither the virus x nor the antidote a_x has reached vertex v by time t , so it is possible for v to be infected by x , or cured/immunized by a_x in the future, if any of them propagates to v via contacts.

Infected at time t_i : If a copy of virus x reaches susceptible vertex v at t_i , v becomes infected at t_i , which means $X_v^x(t_i) = 1$ and $\lim_{t \rightarrow t_i^-} X_v^x(t) = 0$. This *infect* action is shown by the dashed arrow from susceptible state to the *infected* state (red circle with letter I) in Fig. 1. At this state, vertex v will try to *infect*, *i.e.*, pass copies of virus x to any of its susceptible neighbors, that is, $u \in \mathcal{N}_S^x(t_i, v) = \{u \in \mathcal{N}(v) \mid X_u^x(t_i) = 0\}$,

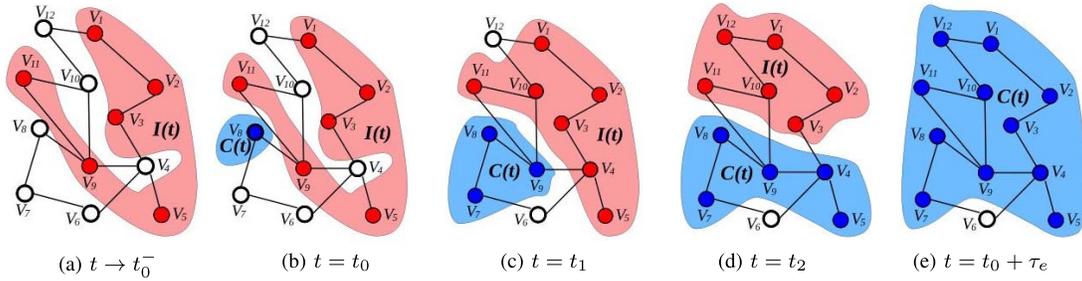
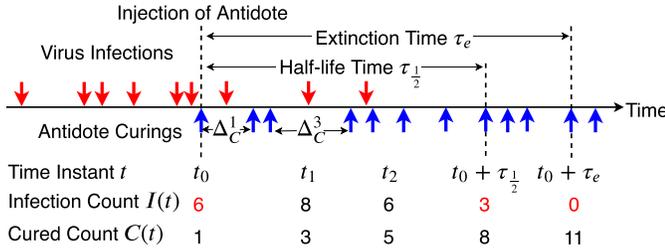


Fig. 2. An example of an SIC epidemics on a network of 12 vertices.

Fig. 3. Illustration of the extinction time τ_e and half-life time $\tau_{\frac{1}{2}}$ of the virus for the example shown in Fig. 2: At time t_0 , $C_0 = 1$ copy of antidote is injected into the network.

after a random period of time $s_v^x(u)$. Vertex v will stay in infected state until it receives a copy of antidote a_x .

Cured at time t_c : If at t_c , a copy of antidote a_x reaches vertex v (solid arrows in Fig. 1) for the first time, i.e., $\lim_{t \rightarrow t_c^-} X_v^x(t) \geq 0$, the state of v changes to *cured* at t_c , that is, $X_v^x(t) = -1$, as shown as the blue circle with letter C in Fig. 1. At this state, vertex v will pass copies of antidote to any u of its neighbors $\mathcal{N}_{NC}^x(t, v) = \{u \in \mathcal{N}(v) | X_u^x(t_i) \geq 0\}$ after a random period of time $s_v^{a_x}(u)$. Vertex v will stay cured for the rest of the time, i.e., $X_v^x(t) = -1$ for any $t > t_c$.

Fig. 2 shows an example of a simple network with 12 vertices under an SIC dynamics. As shown in Fig. 2(a), before antidote a_x is injected into the network at time t_0 , vertices in $\{v_1, v_2, v_3, v_5, v_9, v_{11}\}$ are infected (colored in red). Then at t_0 , one unit of antidote is given to vertex v_8 , and cures it immediately (indicated as blue), as shown in Fig. 3(b). As time t proceeds, states of the 12 vertices change in Fig. 3(c-e). Eventually, the virus dies out at time $t = t_0 + \tau_e$.

2) **Propagation Rules**: Individual state changes are driven by the propagation of virus x and antidote a_x , whose speeds are controlled by intervals $\{s_u^x(v)\}_{v \in \mathcal{N}_S(t, u)}$ and $\{s_u^{a_x}(v)\}_{v \in \mathcal{N}_{NC}(t, u)}$. To make this problem tractable, we follow the convention in [13], [17] and assume time homogeneity for the propagation process: For any vertex u , random intervals $\{s_u^x(v)\}_{v \in \mathcal{N}_S(t, u)}$ and $\{s_u^{a_x}(v)\}_{v \in \mathcal{N}_{NC}(t, u)}$ are two groups of r.v.'s satisfying i) pairwise independent; and ii) exponentially distributed with parameters $\beta_{u,v}^x$ and $\gamma_{u,v}^x$, respectively.

From the perspective of time, $\beta_{u,v}^x = \beta_{v,u}^x$ is known as the *virulence* (or *infection rate*) of virus x , while $\gamma_{u,v}^x = \gamma_{v,u}^x$ as the *curing rate* of antidote a_x , representing how frequently a copy of virus x and antidote a_x is exchanged via edge $e(u, v)$, respectively. We give their formal definitions as follows.

Definition 1: For an infected vertex u , and a susceptible vertex $v \in \mathcal{N}_S^x(t, v)$, the **virulence** of virus x on edge $e(u, v)$ is defined as

$$\beta_{u,v}^x := \lim_{t \rightarrow 0^+} \frac{\mathbb{P}(s_u^x(v) \leq t)}{t}, \quad (1)$$

where $s_u^x(v)$ is the time period between the infection of u , and the time when u passes a copy of virus x to v via edge $e(u, v)$.

Definition 2: For a vertex u in cured state, and its neighbor $v \in \mathcal{N}_{NC}^x(t, v)$, the **curing rate** of antidote a_x on edge $e(u, v)$ is defined as

$$\gamma_{u,v}^x := \lim_{t \rightarrow 0^+} \frac{\mathbb{P}(s_u^{a_x}(v) \leq t)}{t}, \quad (2)$$

where $s_u^{a_x}(v)$ is the time period between the curing of u , and the time when u passes a copy of antidote to v via edge $e(u, v)$.

From the perspective of probability, $\beta_{u,v}^x$ is also known as the infection probability over unit time, which can be explained by considering a simple network composed of two connected vertices, i.e., $\mathcal{V} = \{u, v\}$. At time t , given $X_v^x(t) = 0$, the probability that v gets infected by u in Δt is

$$\begin{aligned} \mathbb{P}(X_v^x(t + \Delta t) = 1 | X_v^x(t) = 0) \\ = \Delta t \cdot \beta_{u,v}^x \cdot \mathbb{1}_{\{X_u^x(t)=1\}} + o(\Delta t), \end{aligned} \quad (3)$$

where r.v. $s_u^x(v) \sim \text{Exp}(\beta_{u,v}^x)$, with mean $\mathbb{E}(s_u^x(v)) = \frac{1}{\beta_{u,v}^x}$. Similar results also apply to r.v. $s_u^{a_x}(v)$ and curing rate $\gamma_{u,v}^x$. Therefore, when the time interval is of unit length, i.e., $\Delta t = 1$, the state transition probability $\mathbb{P}(X_v^x(t + \Delta t) = 1 | X_v^x(t) = 0) = \mathbb{P}(s_u^x(v) \leq t) = \beta_{u,v}^x$ equals to the infection rate $\beta_{u,v}^x$ in value, which bridges the gap between continuous-time modeling/analysis and discrete-time simulations.

At any time t , vertices set $\mathcal{V} = \mathcal{S}^x(t) \cup \mathcal{I}^x(t) \cup \mathcal{C}^x(t)$, where $\mathcal{S}^x(t)$, $\mathcal{I}^x(t)$, $\mathcal{C}^x(t)$ are mutually disjoint sets: the susceptible set $\mathcal{S}^x(t) := \{v \in \mathcal{V} : X_v^x(t) := 0\}$, the infected set $\mathcal{I}^x(t) := \{v \in \mathcal{V} : X_v^x(t) := 1\}$, and the cured set $\mathcal{C}^x(t) = \{v \in \mathcal{V} : X_v^x(t) = -1\}$. Evolution of an SIC epidemic process can be captured by the time-varying *infection count*, $I^x(t) := |\mathcal{I}^x(t)|$, and the *cured count*, $C^x(t) := |\mathcal{C}^x(t)|$. For the ease of notation, we suppress x in $X_v^x(t)$, $\mathcal{S}^x(t)$, $\mathcal{I}^x(t)$, $\mathcal{C}^x(t)$, and write $X_v(t)$, $\mathcal{S}(t)$, $\mathcal{I}(t)$, $\mathcal{C}(t)$ instead.

D. When and How Fast Will the Virus Die Out?

From the state transition diagram (Fig. 2) of the SIC epidemics and the fact that network \mathcal{G} is connected, it is clear

that eventually the virus will be eliminated from \mathcal{G} , *i.e.*, as time $t \rightarrow \infty$, state $X_v(t) = -1$ for every vertex v . The question is *when* and *how fast* the virus dies out. In other words, we need to examine short-term transitions of individual states, which all take place in finite time. To answer this question, we define two pivots in the lifetime of virus x to quantify its dying speed.

Definition 3: For an SIC dynamic of virus x and antidote a_x , the **extinction time** of virus x , denoted as τ_e , is defined as the length of the time interval between t_0 and the first time that the infected set $\mathcal{I}(t)$ becomes empty, that is,

$$\tau_e := \inf\{t > t_0 : \mathcal{I}(t) = \emptyset\} - t_0, \quad (4)$$

where t_0 is the time instant when C_0 copies of antidote a_x are injected into the network \mathcal{G} .

The extinction time τ_e is a finite¹ r.v. on measurable space $(\Omega^n, 2^{\Omega^n}, \mathbb{P})$. It answers *when the virus dies out*, because time instant $t_0 + \tau_e$ marks the *end point* of the virus's life in \mathcal{G} . In other words, the infection count decreases from $I(t_0)$ to 0 during a time interval of length τ_e , so we know the virus (undesired information) dies out at an average speed of $\frac{I(t_0)}{\tau_e}$. But it is still not clear how such speed changes during the lifetime of the virus, *i.e.*, *how fast* the virus dies out, which answers a lot of realistic questions, *e.g.*, when will the majority of individuals be free of the undesired information? To answer this question, we identify another pivot in time.

Definition 4: For an SIC dynamic of virus x and antidote a_x , the **half-life time** of the virus epidemic, denoted as $\tau_{\frac{1}{2}}$, is defined as the length of the time interval between t_0 and the last time that event $\{I(t) \geq \frac{1}{2}I(t_0)\}$ happens after t_0 , that is,

$$\tau_{\frac{1}{2}} := \sup\{t \in [t_0, t_0 + \tau_e] : I(t) \geq \frac{1}{2}I(t_0)\} - t_0, \quad (5)$$

where $I(t_0) > 0$ is the initial infection count at t_0 .

The Half-life time $\tau_{\frac{1}{2}} : \Omega^n \rightarrow [t_0, t_0 + \tau_e]$ is also a finite r.v. on the same measurable space as r.v. τ_e . The term *half-life* is originally from Chemical Kinetics, which describes the decay of discrete entities. But, unlike in Chemical Kinetics, where half-life is the mean, we define half-life as the actual time interval until event $\{I(t) \geq \frac{1}{2}I(t_0)\}$ happens for the *last* time. The physical meaning of $\tau_{\frac{1}{2}}$ can be explain as follows: the competition between the undesired information and its desired counterpart mainly takes place before pivot time $t_0 + \tau_{\frac{1}{2}}$, after which the undesired information can be viewed as controlled, because the number of its victim will never exceed the threshold $I(t_0)/2$.

In this sense, the two lifetime metrics, extinction time τ_e and half-life time $\tau_{\frac{1}{2}}$, illustrate *how fast* the virus dies, because for a fixed initial condition ($I(t_0)$ and $C(t_0)$), the larger the gap $\tau_e - \tau_{\frac{1}{2}}$, the faster the undesired information dies during its most hazardous phase $[t_0, t_0 + \tau_{\frac{1}{2}}]$. Consequently, if desired information is purposely injected by the system as a countermeasure against the undesired information, its effectiveness can be reflected by these lifetime metrics.

¹To be more accurate, r.v. τ_e is almost surely (a.s.) finite, that is, $\mathbb{P}(\tau_e < \infty) = 1$, when the network \mathcal{G} is connected, because it can be written as the summation of a finite number of exponential r.v.'s, each of which is a.s. finite.

Fig. 3 illustrates the extinction time and the half-life time for the example of the SIC dynamics shown in Fig. 2, where a red arrow corresponds to an infection, and a blue one represents a curing event. At $t_0 + \tau_{\frac{1}{2}}$, the infection count of the system drops to 3 ($= \frac{1}{2}I_0$), and never exceeds 3 again, which implies that the virus epidemic has been restricted to a limited area, or equivalently, under control. At $t_0 + \tau_e$, the virus dies out.

Without loss of generality, let $t_0 = 0$, and denote $I(0)$ as I_0 (and $C(0) = C_0$) for the ease of notation. All the events we discuss hereafter take place in the observation window $[0, \tau_e]$. We further assume that both the infection rate β and curing rate γ are constant on every edge of the network, which is commonly adopted in information propagation studies.

Under the proposed SIC model, we restate our research question as follows: Consider a pair of conflicting information (x , a_x), in which x is the virus with virulence β , and a_x is the antidote with curing rate γ . At time $t = 0$, C_0 copies of antidote are distributed in network $\mathcal{G}(\mathcal{V}, \mathcal{E})$, when the infection count equals to I_0 . What is the expected extinction time $\mathbb{E}(\tau_e)$ and half-life time $\mathbb{E}(\tau_{\frac{1}{2}})$ of virus x (undesired information)?

III. CONFLICTING INFORMATION PROPAGATION IN COMPLETE NETWORKS AND STAR NETWORKS

There are three sets of defining factors, which control the spreading behavior and hence the lifetime of the undesired information in an SIC dynamics: propagation parameters (β , γ), initial condition (I_0 , C_0), and topological properties of network \mathcal{G} . Among these, the key challenge arises from the underlying network \mathcal{G} , a structure with numerous topological properties, some of which are of particular importance to the epidemic spreading of information, as evidenced in [11], [13], [19]. To answer the *when* and *how fast* question for *conflicting* information propagation in an unspecified, and possibly complex, network \mathcal{G} , we start from simple cases, in which \mathcal{G} has special topologies, in order to clarify which topological properties to consider for more general cases. Specifically, we examine complete graph K_n and star graph S_n , which are not only common network topology themselves, but also essential components of complex networked systems.

A. SIC Epidemics on Complete Networks K_n

A complete graph K_n , is a fully connected graph with n vertices, in which for any pair of vertices $v_i, v_j \in V(K_n), i \neq j$, there exists an edge $e(i, j) \in E(K_n)$ between them, and hence $|E(K_n)| = n(n-1)/2$. This topology is frequently seen in networks that require high reliability, *e.g.*, network of AS routers, or networks that are densely connected everywhere, *e.g.*, a household or a community, where every individual is familiar with one another. It has two key characteristics: i) it is the most densely connected simple graph, because it has the maximum number of edges; ii) it is regular, because every vertex is inter-changable with another, which means they have the same vertex metrics, such as degree and

centrality. In such networks, the expected lifetime metrics of the undesired information are upper bounded by the following theorem.

Theorem 1: Consider an SIC dynamics in action on a complete network K_n , with curing rate γ , and initial condition (I_0, C_0) . The expected extinction time $\mathbb{E}(\tau_e)$ and half-life time $\mathbb{E}(\tau_{\frac{1}{2}})$ of the virus can be bounded above as

$$\mathbb{E}(\tau_e) < \frac{1}{\gamma n} \left[2 + \ln \frac{(n-1)(n-C_0)}{C_0} \right], \quad (6)$$

$$\mathbb{E}(\tau_{\frac{1}{2}}) < \frac{1}{\gamma n} \left[2 + \ln \frac{(n-1)(n-1-\lceil I_0/2 \rceil)}{\lceil I_0/2 \rceil + 2} \right]; \quad (7)$$

when $C_0 \geq 2$, we also have

$$\mathbb{E}(\tau_e) < \frac{2}{\gamma(n-1+C_0)} \ln(n-C_0), \quad (8)$$

$$\mathbb{E}(\tau_{\frac{1}{2}}) \leq \frac{2}{\gamma(n-1-\lceil I_0/2 \rceil + C_0)} \left(1 + \ln \frac{n-C_0}{\lceil I_0/2 \rceil + 1} \right), \quad (9)$$

where $n = |\mathcal{V}|$ is the size of the complete network K_n .

Proof of Theorem 1 can be found in Appendix A.

Theorem 1 implies that both the expected extinction time and half-life time decreases with size n of the clique as $O(\frac{\log n}{n})$, which means the larger the network, the quicker the undesired information dies, given the same initial condition of the dynamics. The reason behind this is that, dense connections among vertices make it difficult for the virus to dodge contact with antidotes. In other words, even though both virus and antidote can spread faster due to the large number of edges, dense connections work in favor of the antidote propagation. On the other hand, with respect to the severeness of virus infection upon antidote injection, we can see the half-life time decreases with the initial infection count I_0 as $O(\log \frac{A}{I_0})$, where A is a function of n and C_0 that does not vary with I_0 . It shows that the severer the infection, the less time it takes to get the virus epidemics under control. In this case (larger I_0), the larger gap $\tau_e - \tau_{\frac{1}{2}}$ indicates that competition between the antidote and the virus is only fierce for a short period of time, and the larger portion of extinction time is spent on extinguishing cornered virus infections.

From the perspective of connectivity, any other simple network topology of the same size n has strictly less edges than the complete graph K_n , as a result of which the undesired information (virus) will die slower with high probability in these networks, even with the same initial condition (I_0 and C_0). Bases on this observation, we also have the following lower bounds as a corollary of Theorem 1.

Corollary 1: For an SIC epidemic with curing rate γ on an arbitrary network \mathcal{G} of size n , the expected extinction time and half-life time of the virus can be bounded below as

$$\mathbb{E}(\tau_e) \geq \frac{1}{\gamma n} [\mathcal{H}_{C_0+I_0-1} - \mathcal{H}_{C_0-1} + \mathcal{H}_{n-C_0} - \mathcal{H}_{n-C_0-I_0}], \quad (10)$$

$$\mathbb{E}(\tau_{\frac{1}{2}}) \geq \frac{1}{\gamma n} [\mathcal{H}_{C_0+\lceil I_0/2 \rceil-1} - \mathcal{H}_{C_0-1} + \mathcal{H}_{n-C_0} - \mathcal{H}_{n-C_0-\lceil I_0/2 \rceil}], \quad (11)$$

where $\mathcal{H}_k = \sum_{j=1}^k \frac{1}{j}$ is the k -th *Harmonic Number*.

Proof of Corollary 1 can be found in Appendix B.

B. SIC Epidemics on Star Networks S_n

As opposed to the most densely connected complete network K_n , a star network S_n is composed of a hub v_1 and $n-1$ leaves (peripheral vertices), such that every piece of information from a leaf vertex has to go through the hub to reach another vertex. In other words, edges only exists between the hub and leaves, creating a huge bottleneck in the middle, which can be viewed as a high-degree of heterogeneity among vertices. The star topology is also of great importance, due to its natural link to artificial structures, *e.g.*, a WiFi access network within the coverage of an access point, and the ego-network of a high-degree node in OSNs.

Consider the same SIC dynamic with infection rate β and curing rate γ on a star S_n . In an extreme case, if a copy of antidote is first given to the hub, then infection count $I(t)$ will be monotonically decreasing with time t , which is less interesting, because it is impossible for the virus to claim new victims, and hence no competition. Therefore, we consider the case that antidotes can only be distributed to peripheral vertices at time 0, and present the following upper bounds.

Theorem 2: For an SIC dynamics on a start network in which the initial infection count satisfies $I_0 \geq 2$, the expected extinction time $\mathbb{E}(\tau_e)$ and half-life $\mathbb{E}(\tau_{\frac{1}{2}})$ follow upper bounds:

$$\mathbb{E}(\tau_e) < \frac{1}{\gamma} \left[\frac{1}{C_0} + 1 + \ln(I_0 - 1) + \frac{\beta}{\gamma C_0 (I_0 - 1)} \right], \quad (12)$$

$$\mathbb{E}(\tau_{\frac{1}{2}}) < \frac{1}{\gamma} \left[\frac{1}{C_0} + 1 + \ln\left(\frac{I_0 - 1}{\lceil I_0/2 \rceil}\right) + \frac{\beta}{\gamma C_0 (I_0 - 1)} \right]. \quad (13)$$

Proof of Theorem 2 can be found in Appendix C.

It is intuitive that the undesired information (virus) will die slower in a star network than in a complete network of the same size, due to the much sparser connections. However, it is rather interesting to observe from Theorem 2 that, both the expected extinction time and half-life time increases with the network size n in the star topology. This clear contrast between the two topologies leads to a conjecture: existence of bottlenecks changes how the network size impacts when and how fast the virus dies out. In addition to this difference, we also observe a similarity with SIC epidemics on the complete network, that is, the half-life time of the virus in a star network also decreases with the initial infection count I_0 , but with a milder decreasing slope, as $O(\frac{1}{I_0})$.

C. Numerical Simulation and Discussion

To validate upper bounds (dashed lines) in Theorem 1 and Theorem 2, numerical results (solid lines with markers) with respect to network size n and the initial infection count I_0 are shown in Fig. 4 and Fig. 5-6 respectively. As can be seen from all three figures, trends of both the expected extinction time and half-life time are well captured by the derived bounds. Comparing the bounds and simulation results for the two simple topologies, we highlight the following observations.

1) *Propagation Parameters β and γ* : The infection/curing rates have the same effect on both temporal metrics regardless of network topology, which can be observed by comparing the sub-figures in Fig. 5 and Fig. 6. To be more specific, they only lengthen or shorten time intervals between events, but does not change the order of event occurrences, resulting in a homogeneous scaling effect on both lifetime metrics.

2) *Network Size n* : There is a clear dichotomy in the dying speed of the virus with respect to the network size n , that is, virus dies faster in a larger complete network, but dies slower in a larger star network, as shown in the semi-log plots of Fig. 4. Both $\mathbb{E}(\tau_e)$ and $\mathbb{E}(\tau_{\frac{1}{2}})$ are $O(\frac{\log n}{n})$ for complete networks (decreasing blue lines), while both are $O(\log n)$ for star networks (increasing red lines), as indicated by Theorem 1 and Theorem 2, respectively. This implies an interesting change in the conflicting information propagation process, when a hub (bottleneck) emerges in the network.

3) *Initial Infection Count I_0* : Impact of I_0 on the lifetime of the virus is shown in Fig. 5 and 6, for complete network K_{100} and star network S_{100} , respectively. The mild zig-zag pattern of the half-life time (left end of all blue solid lines with markers) is caused by rounding-off the threshold $\lfloor \frac{1}{I_0} \rfloor$ when I_0 is small. Though the general trend is the same in both types of networks, *i.e.*, extinction time increases with I_0 , while half-life time decreases with I_0 , impact of topology is still apparent, as indicated by the different scaling behavior over I_0 (black texts in Fig. 5 and 6).

As two extremes of network topology, the complete graph is regular (every vertex has the same degree) and most densely connected ($|\mathcal{E}| = \frac{n(n-1)}{2}$), while the star graph is highly irregular (due to the existence of the central hub) and sparse ($|\mathcal{E}| = n - 1$). Comparing the two, we observe when edges concentrate and form a bottleneck in the network, the impact of network size n and initial condition I_0 changes drastically. Though results in these two simple networks do not directly apply to general cases, they shed lights on studying conflicting information propagation in general networks, by recognizing the importance of bottlenecks in networks.

IV. CONFLICTING INFORMATION PROPAGATION IN NETWORKS WITH ARBITRARY TOPOLOGY

Considering the broad application scenarios of conflicting information propagation, it is possible, and even more likely, that the underlying network \mathcal{G} does not have nice topological properties like regularity, and is hence a *complex network* with unique topologies. To study *when* and *how fast* the undesired information dies out in such networks, we examine graph metrics, namely, the Cheeger constant $\eta(\mathcal{G})$ and vertex eccentricities $\{\epsilon(v)\}_{v \in \mathcal{V}}$, to quantitatively link topological properties to the lifetime metrics ($\mathbb{E}(\tau_e)$ and $\mathbb{E}(\tau_{\frac{1}{2}})$) of the virus in an SIC epidemics through upper bounds.

A. Bounds by Considering the Edge-Expansion Property

Recall in the comparison between complete networks and star networks, we find that the key to propagation behavior change is the central hub of a star, which forms a bottleneck in the network. Therefore, we first consider the metric that

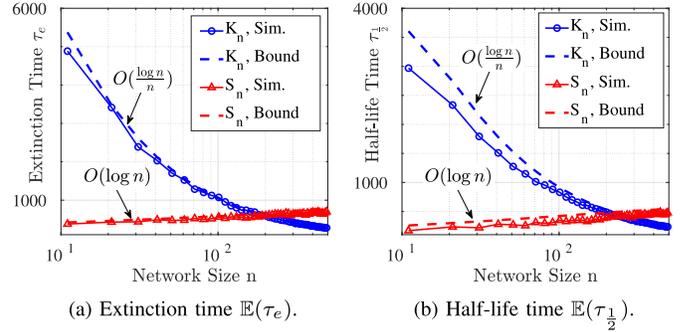


Fig. 4. Expected extinction time $\mathbb{E}(\tau_e)$ and half-life time $\mathbb{E}(\tau_{\frac{1}{2}})$ of an SIC epidemics ($C_0 = 1$, $I_0 = 10$), with respect to the network size n . In this simulation, we set the infection(β) and curing (γ) rates as $\beta = \gamma = 0.01$ in the star network S_n , and $\beta = \gamma = 0.0001$ in the complete network K_n , for a clearer comparison between the two topologies.

quantitatively measures the level of ‘bottleneckness’ of a network \mathcal{G} , the Cheeger constant $\eta(\mathcal{G})$, which is defined as $\eta(\mathcal{G}) := \inf_{S \subset \mathcal{V}, |S| \leq n/2} \frac{|\delta(S)|}{|S|}$, where $\delta(S) := [S, \mathcal{V} \setminus S]$ is the edge-cut of vertex set S , that is, the set of boundary edges between set S and its complement set $\mathcal{V} \setminus S$.

As a graph expansion property [22], $\eta(\mathcal{G})$ identifies the ‘narrowest’ part of network \mathcal{G} , *i.e.*, the minimum boundary edges of as many as vertices. Intuitively, the larger the boundary set, the more difficult it is to break the network into isolated components by disconnecting edges, so $\eta(\mathcal{G})$ and other expansion properties are viewed as indicators of robustness, and hence studied in many applications [22]. Due to this property, it has been shown that $\eta(\mathcal{G})$ is of key importance to the spreading behavior of an epidemic process, under different propagation models, including SI [13], [23], SIS [13], and SIR [11] models. For the proposed SIC propagation model, in which *two* epidemic processes compete in finite time, we show in the following theorem that both the expected extinction time and half-life time are $O(\frac{\log n}{\gamma \eta(\mathcal{G})})$.

Theorem 3: For a network \mathcal{G} with Cheeger constant $\eta(\mathcal{G})$, given the initial cured count C_0 , the extinction time of the virus in an SIC epidemic with curing rate γ can be bounded above by

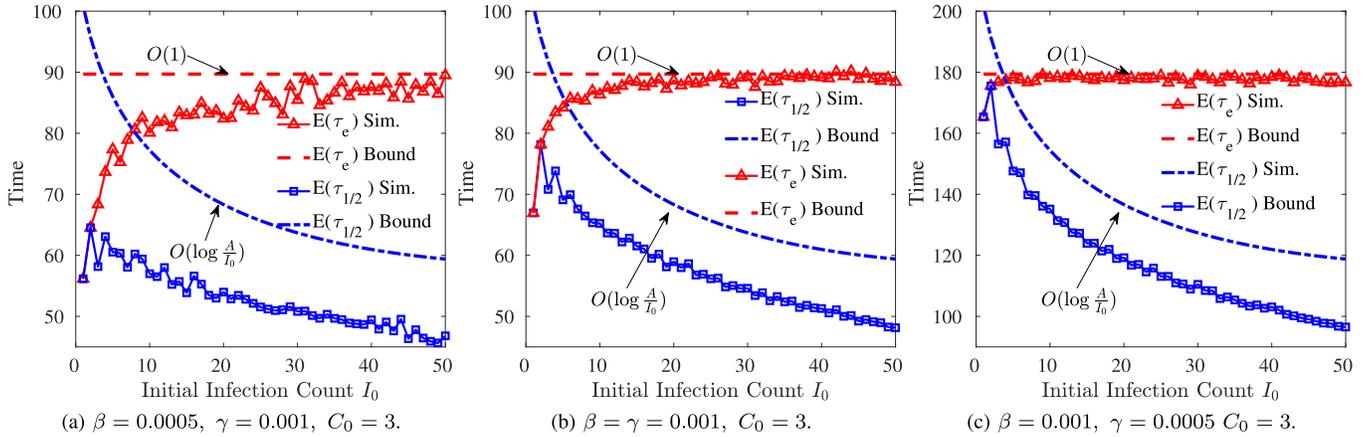
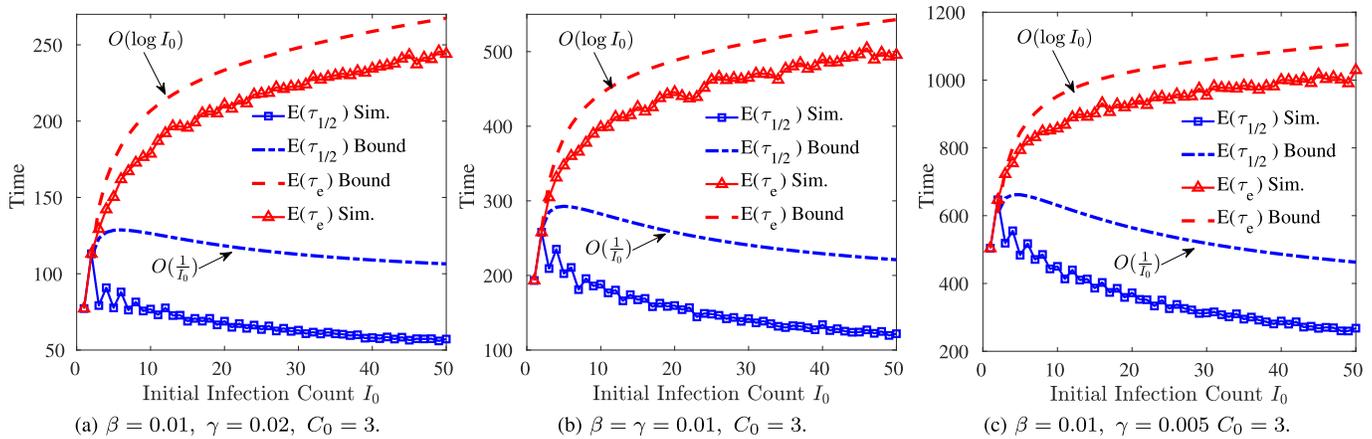
$$\mathbb{E}(\tau_e) \leq \frac{1}{\gamma \eta(\mathcal{G})} [a \ln(n+b) + c], \quad (14)$$

- 1) If $1 \leq C_0 < n/2$, $b = 0$, and when $C_0 = 1$, $a = \frac{2}{\ln 2}$, $c = 2\gamma_E$, otherwise $a = \frac{2}{\ln 4(C_0-1)}$, $c = -\left[\frac{4C_0-5}{8(C_0-1)^2} + \gamma_E\right]$;
 - 2) If $C_0 \geq n/2$, $a = 1$, $b = -C_0 + 1$, $c = \frac{1}{2(n-C_0+1)} + \gamma_E$,
- where $\gamma_E \approx 0.577$ is the Euler-Mascheroni constant.

When the initial infection count I_0 is given, the expected half-life time can be upper bounded² when $2 \leq C_0 < n/2$,

$$\mathbb{E}(\tau_{\frac{1}{2}}) \leq \frac{1}{\gamma \eta(\mathcal{G})} \left[\ln \frac{n^2}{2(C_0-1)I_0} - \frac{4C_0-5}{8(C_0-1)^2} + \frac{2I_0+1}{I_0^2} \right], \quad (15)$$

²For the case of $C_0 > n/2$, it is with high probability that $\tau_{\frac{1}{2}} = 0$, which is a less interesting case and hence not discussed.


 Fig. 5. Extinction time $\mathbb{E}(\tau_e)$ and Half-life time $\mathbb{E}(\tau_{\frac{1}{2}})$ in the complete graph K_{100} , over the initial infection count I_0 .

 Fig. 6. Extinction time $\mathbb{E}(\tau_e)$ and Half-life time $\mathbb{E}(\tau_{\frac{1}{2}})$ in the star network S_{100} , over the initial infection count I_0 .

and when $C_0 = 1$, $\mathbb{E}(\tau_{\frac{1}{2}}) \leq \frac{1}{\gamma\eta(\mathcal{G})} \left[\ln \frac{n^2}{2I_0} + \frac{2I_0+1}{I_0^2} \right]$.

Proof: Let $\mathcal{V}_k := \{V | V \subset \mathcal{V}, |V| = k\}$ be the set of all vertex sets containing k vertices from the network $\mathcal{G}(\mathcal{V}, \mathcal{E})$. Let $\mathcal{C}_k \in \mathcal{V}_k$ be the set of cured k vertices when $C(t) = k$. Let $T_k := \inf\{t | C(t) = k\}$ be the time that the cured count $C(t)$ reaches k . Note that $C(t)$ is monotonically increasing, and $C(0) = C_0$, $T_{C_0} = 0$. For any $k \geq C_0$,

$$\begin{aligned} \mathbb{E}(T_{k+1} - T_k) &= \sum_{A \subset \mathcal{V}_k} \mathbb{E}(T_{k+1} - T_k | C(t) = A) \mathbb{P}(C(t) = A) \\ &= \mathbb{E} \left(\frac{1}{\gamma |\delta(\mathcal{C}_k)|} \right), \end{aligned} \quad (16)$$

where $\delta(\mathcal{C}_k) := [C_k, \mathcal{V} \setminus C_k]$ is the edge cut of set \mathcal{C}_k , and

$$|\delta(\mathcal{C}_k)| \geq \begin{cases} \eta(\mathcal{G})k, & k < \frac{n}{2}, \\ \eta(\mathcal{G})(n-k), & k \geq \frac{n}{2}. \end{cases}$$

Also, we have $\mathbb{E}(\tau_e) \leq \sum_{k=C_0}^{n-1} \mathbb{E}(T_{k+1} - T_k)$ from Lemma 3 in the Appendix. So when $2 \leq C_0 < \frac{n}{2}$, the expected

extinction time can be upper bounded by

$$\begin{aligned} \mathbb{E}(\tau_e) &\leq \sum_{k=C_0}^{n/2} \frac{1}{\gamma\eta(\mathcal{G})k} + \sum_{k=\frac{n}{2}+1}^{n-1} \frac{1}{\gamma\eta(\mathcal{G})(n-k)} \\ &= \frac{1}{\gamma\eta(\mathcal{G})} \left(2\mathcal{H}_{\frac{n}{2}} - \frac{2}{n} - \mathcal{H}_{C_0-1} \right) \\ &\stackrel{\text{Franel}}{<} \frac{1}{\gamma\eta(\mathcal{G})} \left[\ln \frac{n^2}{4(C_0-1)} - \frac{4C_0-5}{8(C_0-1)^2} - \gamma_E \right], \end{aligned}$$

where the last line follows from Franel's Inequality [24]. Similarly, the expected half-life time can be bounded by

$$\mathbb{E}(\tau_{\frac{1}{2}}) \leq \sum_{k=C_0}^{n/2} \frac{1}{\gamma\eta(\mathcal{G})k} + \sum_{k=\frac{n}{2}+1}^{n-1-\lceil I_0/2 \rceil} \frac{1}{\gamma\eta(\mathcal{G})(n-k)}.$$

As for the case of $C_0 \geq \frac{n}{2}$, the upper bound can be obtained by considering $\mathbb{E}(\tau_e) \leq \sum_{k=C_0}^{n-1} \frac{1}{\gamma\eta(\mathcal{G})(n-k)}$. ■

Though not a sharp bound, Theorem 3 confirms the importance of the edge-expansion property, *i.e.*, the Cheeger constant $\eta(\mathcal{G})$, in epidemic spreading of information, particularly two pieces of conflicting information in this paper. As can be seen, the more 'expandable' (strong connectivity with less vertices, towards the clique topology), the less time it takes to eliminate the virus, which explains the shorter rumor

circulation time [25] than before, as OSNs (and Internet at large) become more connected nowadays. Particularly for the two simple network topologies, clique and star, studied in the previous section, Theorem 3 clearly explains the dichotomy of the scaling laws over the network size n : i) $\eta(K_n) = n/2 = O(n)$, which implies the $O(\frac{\log n}{n})$ scaling for the complete network K_n ; while ii) $\eta(S_n) = O(1)$ for the star network S_n , and hence the $O(\log n)$ increasing over network size n .

For some large networks with special topological properties [22], existing results on its edge expansion properties allow us to estimate the lifetime of the undesired information in such networks as the network grow in size. For instance, Krishnasamy et.al. showed that for random graphs (Erdős-Rényi graphs) $G(n, p)$ with $p > \frac{32 \log n}{4np}$, there is a high probability (over $1 - \frac{1}{n^2}$) that $\eta(G) \geq \frac{np}{4}$ [13, Corollary 2]. Applying Theorem 3, we know that the undesired information dies out in $O(\frac{\log n}{\gamma n})$ time with high probability.

For a complex network with arbitrary topology, however, obtaining the Cheeger constant is a well-known hard problem (NP-hard [26]), especially when the network is large. In this case, $\eta(\mathcal{G})$ can be bounded by the *Cheeger Inequality* [22], [27], $\frac{\lambda_1}{2} \leq \eta(\mathcal{G}) \leq 2\sqrt{\lambda_1}$, where λ_1 is the second smallest eigenvalue of the graph Laplacian³ of network \mathcal{G} . As we will show in the simulation later, $\frac{\lambda_1}{2}$ does not always lead to a tight bound of the extinction time, so next we consider properties that are more accessible for general large networks.

B. Bounds by Considering Vertex Eccentricity

Recall in Theorem 2, after the hub of a star network is cured, the extinction time of the virus is bounded by the last curing event of an infected peripheral vertex, which depends on the longest time it takes a copy of antidote to pass through multiple hub-peripheral paths of length 1. In a general network \mathcal{G} , if we view every initially cured vertex $v \in \mathcal{C}(0)$ as a ‘hub’, the extinction time also depends on such hub-peripheral paths, whose lengths are measured in hop-count distances $\text{dist}(\cdot, \cdot)$. In practice, distance between two vertices is easy to obtain, and is therefore widely used in various applications, such as routing and influential node detection. Based on this graph metric, the *eccentricity* of a vertex v in network \mathcal{G} is defined as the longest distance of between v and any other vertex in \mathcal{V} , that is, $\epsilon_{\mathcal{G}}(v) := \max_{u \in \mathcal{V}} \text{dist}_{\mathcal{G}}(u, v)$, and the *diameter* of network \mathcal{G} is defined as the largest eccentricity, i.e., $\text{diam}(\mathcal{G}) := \max_{v \in \mathcal{V}} \epsilon_{\mathcal{G}}(v)$.

As the first step, consider the case that a single copy of antidote is distributed to vertex $c \in \mathcal{V}$ at time 0, i.e. $\mathcal{C}_0 = \{c\}$. Then propagation time of the antidote to any specific vertex i can be bounded by the following lemma.

³Let \mathbf{A} denote the adjacency matrix of \mathcal{G} . The graph Laplacian of \mathcal{G} is defined as $\mathcal{L}(\mathcal{G}) = \text{diag}(\bar{D}) - \mathbf{A}$, where \bar{D} is the degree sequence of \mathcal{G} , and $\text{diag}(\bar{D})$ is the diagonal matrix with \bar{D} as its main diagonal. Since \mathcal{G} is undirected and connected, its Laplacian $\mathcal{L}(\mathcal{G})$ is symmetric and positive-semidefinite, and has n non-negative real eigenvalues. Among these, the second smallest eigenvalue λ_1 , referred to as the *algebraic connectivity*, measures the expanding property of \mathcal{G} . Particularly, the lower bound of $\eta(\mathcal{G})$ is referred to as the Buser Inequality.

Lemma 1: Let $T_{c,i}$ denote the time that vertex i gets a copy of the antidote originated from vertex $c \in \mathcal{C}_0$. We have

$$\mathbb{E}(T_{c,i}) \leq \frac{\text{dist}(c, i)}{\gamma}, \quad (17)$$

Proof: Let $\{P_k\}_{k=1}^K$ denote the set of paths between vertex c and i , such that their lengths are in an ascending order, i.e., $l_k \leq l_{k+1}$. For any $1 \leq k \leq K$, we have

$$\text{dist}_{\mathcal{G}}(c, i) \leq l_1 \leq l_k \leq l_K. \quad (18)$$

Consequently, the curing time $T_{c,i}$ of vertex i (by the antidote originated from vertex c) can be re-written as

$$T_{c,i} = \min_{1 \leq k \leq K} T_{c,i}^k \leq T_{c,i}^1, \quad (19)$$

where $T_{c,i}^k$ is the attempted curing time of vertex i , by the antidote copy originated from c , and transmitted along path P_k . For every k , time $T_{c,i}^k$ is the sum of l_k i.i.d. Exponential r.v.’s with mean $\frac{1}{\gamma}$, as a result of which r.v. $T_{c,i}^k$ satisfies Gamma distribution, i.e., $T_{c,i}^k \sim \Gamma(l_k, \gamma)$, with mean $\mu_k = \frac{l_k}{\gamma}$ and variance $\sigma_k^2 = \frac{l_k}{\gamma^2}$. Then the upper bound in (17) can be obtained through (18) because

$$\mathbb{E}(T_{c,i}) = \mathbb{E}(\min_{1 \leq k \leq K} T_{c,i}^k) \leq \min_{1 \leq k \leq K} \mathbb{E}(T_{c,i}^k) = \frac{l_1}{\gamma}, \quad (20)$$

where $l_1 = \text{dist}_{\mathcal{G}}(c, i)$ is the length of the shortest path between vertex c and vertex i in graph \mathcal{G} . ■

Lemma 1 provides an upper bound of the expected antidote dissemination time from a designated vertex c to vertex i , which is also an upper bound of the curing time of vertex i , if $c \in \mathcal{C}(0)$. To better facilitate the analysis of the extinction time, we first provide a technical lemma on the maximum and minimum of multiple Gamma r.v.’s.

Lemma 2: Let $X_n = \min_{1 \leq i \leq n} X_i$ be the minimum of n r.v.’s, and $\bar{X}_n = \max_{1 \leq i \leq n} X_i$ be the maximum of them, where each $X_i \sim \Gamma(k_i, \theta_i)$, where θ_i ’s are the rate parameters. Then $\mathbb{E}(X_n)$ and $\mathbb{E}(\bar{X}_n)$ are bounded:

$$\mathbb{E}(\bar{X}_n) \leq \max_i \left\{ \frac{k_i}{\theta_i} \right\} + \left(\frac{n-1}{n} \sum_{i=1}^n \frac{k_i}{\theta_i^2} \right)^{1/2}, \quad (21)$$

$$\mathbb{E}(X_n) \geq \min_i \left\{ \frac{k_i}{\theta_i} \right\} - \left(\frac{n-1}{n} \sum_{i=1}^n \frac{k_i}{\theta_i^2} \right)^{1/2}. \quad (22)$$

When $k_i = k$ and $\theta_i = \theta$ for every $1 \leq i \leq n$ such that $\{X_i\}_i$ are a set of i.i.d Gamma r.v.’s, tighter bounds exist:

$$\mathbb{E}(\bar{X}_n) \leq \frac{2 \ln n}{\theta(1 - n^{-1/k})}, \quad (23)$$

$$\mathbb{E}(X_n) \geq \frac{k \ln 2 - \ln n}{\theta}. \quad (24)$$

Proof: Equation (21) and (22) follow from [28, Theorem 2.1] and [28, Corollary 2.1] by substituting the mean $\mu_i = \mathbb{E}(X_i) = \frac{k_i}{\theta_i}$ and variance $\sigma_i^2 = \frac{k_i}{\theta_i^2}$ of Gamma distributions.

For the i.i.d case, (23) is proved in [29, Eq. (7)]. Now we prove the lower bound in (24). Let r.v. $Y_i = -X_i$, then $X_n = \min_i X_i = -\max_i Y_i \triangleq Y_n$. The moment generating function (MGF) of r.v. Y_i can be derived as

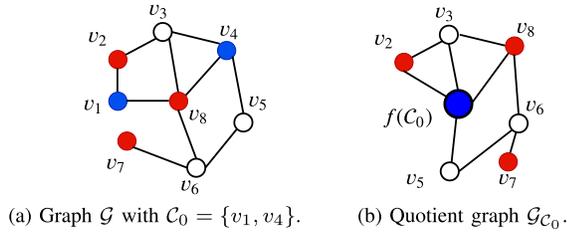


Fig. 7. An example of contracting cured set \mathcal{C}_0 into vertex $f(\mathcal{C}_0)$. In the quotient graph $\mathcal{G}_{\mathcal{C}_0}$, there is $K_{\mathcal{G}/\mathcal{C}_0}^* = 1$ path of length $\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0)) = 3$ from $f(\mathcal{C}_0)$ to vertex v_7 , and $K' = 1$.

$M_Y(t) = M_X(-t) = (1 + \frac{t}{\theta})^{-k}$, where $t < \theta$ as required by MGF $M_X(t)$ of r.v. X_i .

Let set $D(M) = \{t \geq 0 | M_Y(t) \geq 1\}$. Then with the technique from [29, Eq. (6)], we have

$$\begin{aligned} \mathbb{E}(Y_n) &\leq \inf_{t \in D(M)} \frac{1}{t} [\ln n + \ln M_Y(t)] \\ &\leq \frac{1}{t} \left[\ln n - k \ln \left(1 + \frac{t}{\theta}\right) \right] \triangleq g(t), \end{aligned}$$

which is true for every $t \in [0, \theta)$. Notice that $g(t)$ is monotonically decreasing in $[0, \theta)$, so a tighter bound of $\mathbb{E}(Y_n)$ can be upper derived as $g(\theta)$. Then (22) follows from the fact that $E(X_n) = -E(Y_n) \geq -g(\theta)$. ■

Based on Lemma 1 and Lemma 2, the extinction time of the undesired information can be analyzed with graph augmentation, as illustrated in the following theorem.

Theorem 4: Let $\epsilon_{\mathcal{G}}(v)$ denote the eccentricity of vertex v in graph \mathcal{G} . Let $\mathcal{G}_V(\mathcal{V}/V, \mathcal{E}')$ denote the resulting graph induced by contracting vertices in set $V \subset \mathcal{V}$ to a single vertex $f(V)$, and removing all multiple edges.⁴ Given that antidotes are distributed to the set \mathcal{C}_0 at time $t = 0$, The expected extinction time τ_e of the virus can be upper bounded by

$$\mathbb{E}(\tau_e) \leq \frac{1}{\gamma} \begin{cases} \frac{2 \ln K_{\mathcal{G}/\mathcal{C}_0}^*}{1 - (K_{\mathcal{G}/\mathcal{C}_0}^*)^{-\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0))}}, & \text{if } \epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0)) \leq 10, \\ \epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0)) + \sqrt{(K^s - 1)\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0))}, & \text{in general,} \end{cases} \quad (25)$$

where $K_{\mathcal{G}/\mathcal{C}_0}^*$ is the number of longest shortest paths starting from $f(\mathcal{C}_0)$ in the quotient graph $\mathcal{G}_V(\mathcal{V}/V, \mathcal{E}')$, and $K^s \geq K_{\mathcal{G}/\mathcal{C}_0}^* + 1$ is the number of shortest paths that start from $f(\mathcal{C}_0)$, and are longer than $\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0)) - s$.

Proof: For an SIC dynamics on graph \mathcal{G} , the extinction time of the virus can be bounded by

$$\max_{i \in \mathcal{I}_0} \left\{ \min_{c \in \mathcal{C}_0} T_{c,i} \right\} \leq \tau_e \leq \max_{i \in \mathcal{V} \setminus \mathcal{C}_0} \left\{ \min_{c \in \mathcal{C}_0} T_{c,i} \right\}. \quad (26)$$

First we show when $\mathcal{C}_0 \geq 2$, the extinction time τ_e on the original graph \mathcal{G} is upper bounded by the extinction time $\hat{\tau}_e$ of the SIC epidemic on the quotient graph $\mathcal{G}_{\mathcal{C}_0}$, in which one unit of antidote is distributed to vertex $f(\mathcal{C}_0)$ at time $t = 0$.

⁴Graph \mathcal{G}_V is the quotient graph of \mathcal{G} through an equivalence relationship induced by partition $\{V, \{v_1\}, \{v_2\}, \dots\}$ where $v_i \in \mathcal{V} \setminus V$, so that its vertex set $\mathcal{V}/V = (\mathcal{V} \setminus V) \cup f(V)$ has $|\mathcal{V}| - |V| + 1$ vertices.

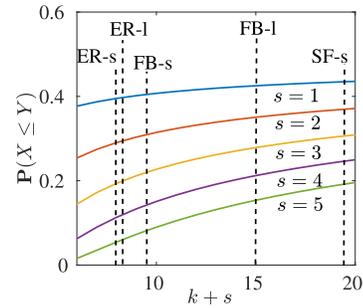


Fig. 8. Probability $\mathbb{P}(X \leq Y)$ for $k + s$ (\leq graph diameter $\text{diam}(\mathcal{G})$) ranging from 5 to 20.

An example of the graph contraction procedure is shown in Fig. 7. Consider vertex $v_1, v_4 \in \mathcal{C}_0$. Evolution of the SIC epidemic (spread of virus and antidote) will not be affected by adding an edge (v_1, v_4) to \mathcal{G} , no matter such edge (v_1, v_4) exists in \mathcal{G} or not. This is because: i) transmission of antidotes between any vertex pair is independent with transmission actions between others; and ii) transmission of antidote between u and v will not result in any state change. Since v_1 and v_4 will stay in *cured* state forever, we can combine and contract them into one vertex $f(\{v_1, v_4\})$ and keep all their edges in \mathcal{G} , without affecting the evolution process. The contraction process may result in multiple edges, by removing which the extinction time may increase. Then by induction on the contraction of \mathcal{C}_0 , the extinction time $\tau_e \leq \hat{\tau}_e$, which is the extinction time of the virus on quotient graph $\mathcal{G}_{\mathcal{C}_0}$.

Therefore, it is sufficient to consider the SIC epidemic on graph $\mathcal{G}_{\mathcal{C}_0}$, with one unit of antidote distributed to $f(\mathcal{C}_0)$ at time $t = 0$. Then we have the following inequality,

$$\tau_e \leq \hat{\tau}_e := \max_{i \in \mathcal{V} \setminus \mathcal{C}_0} T_{f(\mathcal{C}_0), i} \leq \max_{i \in \mathcal{V} \setminus \mathcal{C}_0} T_{f(\mathcal{C}_0), i}^1, \quad (27)$$

where $T_{f(\mathcal{C}_0), i}^1 \sim \Gamma(\text{dist}_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0), i), \gamma)$. The last inequality of (27) follows from (19) in Lemma 1. Next we discuss $\max_{i \in \mathcal{V} \setminus \mathcal{C}_0} T_{f(\mathcal{C}_0), i}^1$ and the eccentricity $\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0))$ of the central cured hub $f(\mathcal{C}_0)$.

By definition, distance between any vertex i and the cured hub $f(\mathcal{C}_0)$ satisfies

$$\text{dist}_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0), i) \leq \epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0)) \leq \text{diam}(\mathcal{G}_{\mathcal{C}_0}) \leq \text{diam}(\mathcal{G}). \quad (28)$$

Without loss of generality, suppose $\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0)) = k + s$ is achieved by the path from $f(\mathcal{C}_0)$ to vertex u in the quotient graph $\mathcal{G}_{\mathcal{C}_0}$. Consider infected vertex i that is $\text{dist}_{\mathcal{G}_{\mathcal{C}_0}}(i, f(\mathcal{C}_0)) = k$ hops away from the cured hub $f(\mathcal{C}_0)$, such that $s \geq 1$. Let r.v. X and Y denote the time it takes a copy of antidote to reach vertex u and v from the hub $f(\mathcal{C}_0)$, respectively. We first show that when $k + s$ is small, the probability that r.v. $X \leq Y$ is small.

Clearly, $X \sim \Gamma(k + s, \theta)$ and $Y \sim \Gamma(k, \theta)$, where θ is the rate parameter, and equals to the curing rate γ in our model. As a result, r.v. $\frac{X}{X+Y}$ satisfies Beta distribution with

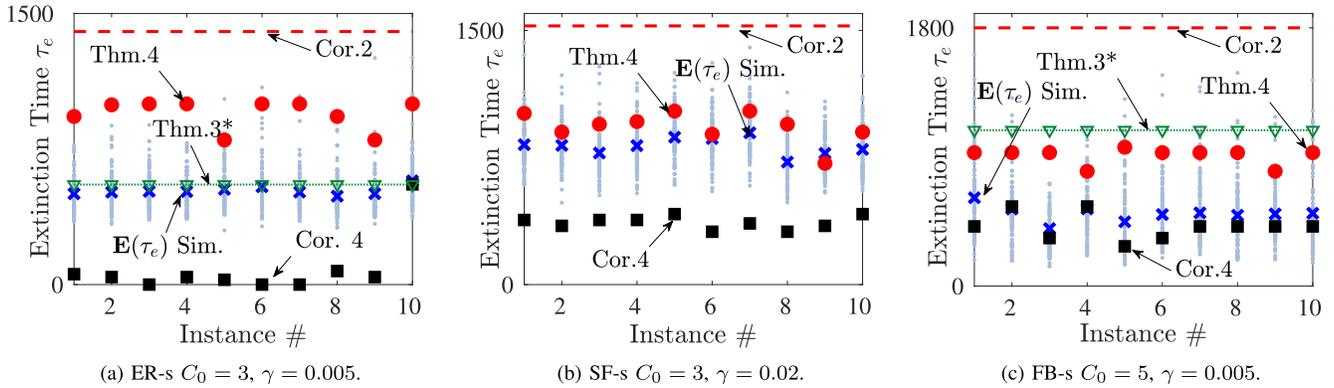


Fig. 9. Bounds and simulation of the extinction time $\mathbb{E}(\tau_e)$ in the general network group of small sizes ($n \simeq 1000$). The curing rate γ in the SF-s scenario is increased to avoid a lengthy simulation in a sparse graph (with less edges) like SF-s.

TABLE I
STATISTICS OF NETWORKS USED IN SIMULATION

Network	ER-s	SF-s	FB-s	ER-l	FB-l
Size n	1000	1000	1033	22470	22232
Avg. degree \bar{d}	7.362	1.998	51.785	15.183	11.164
Diameter $\text{diam}(\mathcal{G})$	7	19	9	6	15
Radius $\text{radi}(\mathcal{G})$	5	10	5	5	8
Algebraic conn. λ_1	0.762	0.002	0.133	1.820	0.071

parameters $k + s$ and k , and

$$\begin{aligned} \mathbb{P}(X \leq Y) &= \mathbb{P}\left(\frac{X}{X+Y} \leq 0.5\right) \\ &= \frac{\Gamma(2k+s)}{\Gamma(k+s)\Gamma(k)} \int_0^{0.5} t^{k+s-1}(1-t)^{k-1} dt, \end{aligned} \quad (29)$$

where $\Gamma(k) = \int_0^\infty x^{k-1}e^{-x}dx$ is the Gamma function. It is difficult to bound (29) due to the integral, so we plot this probability for k and s values in Fig. 8, such that $k+s$ reflects the range of diameter for frequently⁵ seen networks.

In Fig. 8, the dashed vertical lines identify the diameter $\text{diam}(\mathcal{G})$ of five networks used for simulation validation, whose statistics are shown in Table I. When $k+s \leq \text{diam}(\mathcal{G}) \leq 10$, the probability that the antidote will reach u sooner than i , who is s -hops closer than u from $f(\mathcal{C}_0)$ is small (< 0.2) when $s \geq 3$. In addition, the rate parameter θ of the Gamma distribution equals to the curing rate γ , so when γ is also small, the gap between the two time intervals Y and X is therefore small as well, such that we only need to consider vertices who are located the furthest from vertex $f(\mathcal{C}_0)$ in graph $\mathcal{G}_{\mathcal{C}_0}$, *i.e.*, the $K_{\mathcal{G}/\mathcal{C}_0}^*$ distinct vertices, such as u , satisfying $\text{dist}_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0), i) = \epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0))$. Then the first line on the right-hand side of (25) follows from the tighter upper bound (23) in Lemma 2. Note that when $K_{\mathcal{G}/\mathcal{C}_0}^* = 1$, the first term is not well-defined, because the denominator equals to 0. In this case, we take $\mathbb{E}(\tau_e) \leq \frac{\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0))}{\gamma}$ instead.

⁵In most networks we consider (see Table I for detailed statistics), the network diameter $\text{diam}(\mathcal{G}) = O(\log|\mathcal{V}|)$ is small due to the *small-world* effect, so probability $\mathbb{P}(X \leq Y)$ in (29) is also small.

If this condition is not satisfied, *e.g.*, when $\text{diam}(\mathcal{G})$ is large, such that $\mathbb{P}(X \leq Y)$ can not be omitted, we always have the option to consider more paths, which are possibly shorter, to upper bound the extinction time. Then the upper bound (second line in (25)) follows from the upper bound for general distributions, as described in (21) of Lemma 2. ■

Note that the eccentricity $\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0))$ is a property of the augmented quotient graph $\mathcal{G}_{\mathcal{C}_0}$. To apply Theorem 4, we need to know the exact locations (vertices) where copies of antidotes are disseminated, *i.e.*, set \mathcal{C}_0 . If such knowledge is not readily available, the extinction time can still be bounded by the diameter (or more generally, distribution of vertex eccentricities) of graph \mathcal{G} , as given in the following two corollaries. Proofs of Corollary 2 and 4 are simple as they directly follow from the fact that $\epsilon_{\mathcal{G}_{\mathcal{C}_0}}(f(\mathcal{C}_0)) \leq \min_{c \in \mathcal{C}_0} \epsilon_{\mathcal{G}}(c) \leq \text{diam}(\mathcal{G})$.

Corollary 2: Particularly when $\mathcal{C}_0 = 1$, for any initial antidote recipient $c \in \mathcal{V}$,

$$\mathbb{E}(\tau_e) \leq \frac{1}{\gamma} \left[\text{diam}(\mathcal{G}) + \sqrt{\text{diam}(\mathcal{G})(|\text{Peri}(\mathcal{G})|/2 - 1)} \right], \quad (30)$$

where $\text{diam}(\mathcal{G})$ is the diameter of \mathcal{G} , and $\text{Peri}(\mathcal{G}) = \{v \in \mathcal{V} \mid \epsilon(v) = \text{diam}(\mathcal{G})\}$ is the set of peripheral vertices in \mathcal{G} .

Corollary 2 is also an upper bound of the expected extinction time $\mathbb{E}(\tau_e)$ for the case of $\mathcal{C}_0 > 1$, because the more copies of antidote distributed initially at $t = 0$, the less time it takes to fully remove virus from the network, *i.e.*, a shorter extinction time. A direct implication of Corollary 2 is that for small-world graphs, which naturally emerges in various contexts, such as social networks, the extinction time is $O(\frac{\log n}{\gamma})$, due to the $O(\log n)$ scaling of the network diameter. Though not a tight bound compared to Theorem 3, this corollary is much more accessible, especially for large networks. Obtaining the eccentricity distribution of a graph, including diameter and peripheral size, is at most $O(n|\mathcal{E}|)$ in time complexity with further speedups [30], which is much faster than obtaining the Cheeger constant $\eta(\mathcal{G})$. By the eccentricity distribution and the initial cured count \mathcal{C}_0 , a more accurate upper bound can be derived as follows.

Corollary 3: Suppose vertices in \mathcal{C}_0 are chosen uniformly at random from \mathcal{V} , and CDF of the eccentricity distribution

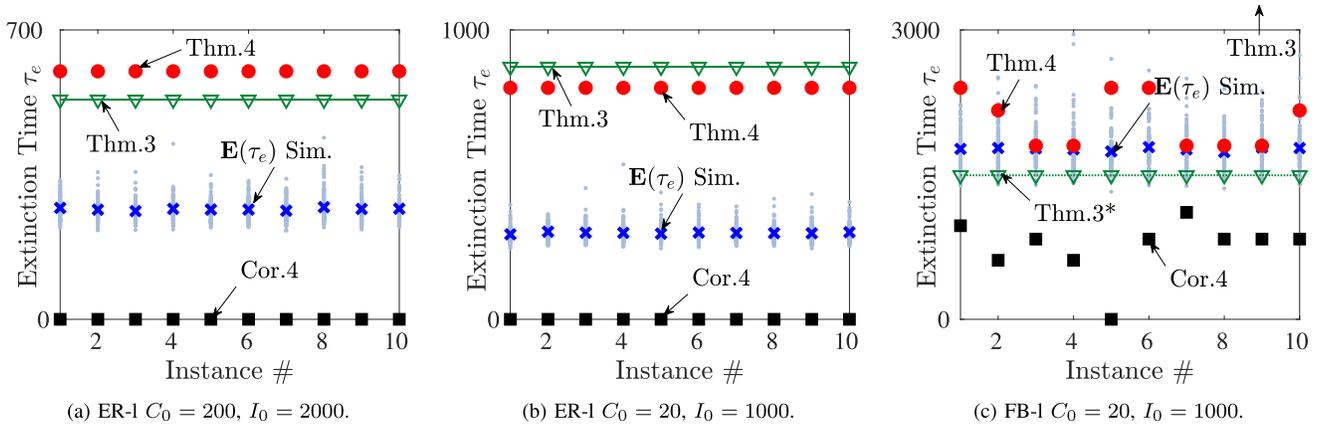


Fig. 10. Bounds and simulation of the extinction time $\mathbb{E}(\tau_e)$ in the general network group of large sizes ($n \simeq 22000$).

of graph \mathcal{G} is $F(x)$ ($\text{radi}(\mathcal{G}) \leq x \leq \text{diam}(\mathcal{G})$). The expected extinction time can be upper bounded by

$$\mathbb{E}(\tau_e) \leq \frac{1}{\gamma} \left[\mu_{C_0} + \sqrt{\mu_{C_0} \left(\frac{n - C_0}{\text{diam}(\mathcal{G})} - 1 \right)} \right], \quad (31)$$

where $\mu_{C_0} = \sum_{k=\text{radi}(\mathcal{G})}^{\text{diam}(\mathcal{G})} k[1 - F(k)]^{C_0}$ is the expected maximum eccentricity of vertices in the cured set C_0 .

Corollary 3 also stems from Theorem 4, by considering the expected maximum eccentricity of the C_0 antidote recipients in C_0 , if antidotes are randomly distributed at time 0. Specifically, for any vertex $v \in V \setminus C_0$, the distance between i and vertex $f(C_0)$ in the quotient graph \mathcal{G}_{C_0} satisfies $\text{dist}_{\mathcal{G}_{C_0}}(f(C_0), v) \leq \epsilon_{\mathcal{G}_{C_0}}(f(C_0)) \leq \min_{c \in C_0} \epsilon_{\mathcal{G}}(c) \leq \text{diam}(\mathcal{G})$, which is the diameter of the original graph \mathcal{G} .

Changing perspective to the initial infected set \mathcal{I}_0 , which contains all infected vertices at time 0, the expected extinction time $\mathbb{E}(\tau_e)$ can be bounded below by considering the shortest paths between vertices in set \mathcal{I}_0 and set C_0 .

Corollary 4: Given the initial infected set $\mathcal{I}_0 = \mathcal{I}(t = 0)$, the expected extinction time can be lower bounded by

$$\mathbb{E}(\tau_e) \geq \frac{1}{\gamma} [\text{dist}(C_0, i^*) \ln 2 - \ln K_{i^*}], \quad (32)$$

where $\text{dist}(C_0, i) = \min_{c \in C_0} \{\text{dist}(c, i)\}$ denotes the shortest distance between set C_0 and vertex i , $i^* = \arg \max_{i \in \mathcal{I}_0} \{\text{dist}(C_0, i)\}$, and K_{i^*} is the number of shortest paths⁶ between set C_0 and vertex i^* .

Proof: From (26), we have

$$\tau_e \geq \max_{i \in \mathcal{I}_0} \left\{ \min_{c \in C_0} T_{c,i} \right\} \geq \min_{c \in C_0} T_{c,i^*}, \quad (33)$$

where $T_{c,i^*} = \min_{1 \leq k \leq K_{i^*}} T_{c,i^*}^k$ is the minimum of K_{i^*} i.i.d. Gamma r.v.'s, and each $T_{c,i^*}^k \sim \Gamma(\text{dist}(C_0, i), \frac{1}{\gamma})$, which comes from Lemma 1. Then the lower bound in (32) follows immediately from (24) in Lemma 2. ■

⁶In practice, obtaining quantity K_{i^*} requires executing path searches repetitively. To avoid high computation load in the simulation, we use the upper bound of K_{i^*} , that is, $|\{i \in \mathcal{I}_0 | \text{dist}(C_0, i) = \text{dist}(C_0, i^*)\}|$ when evaluating the lower bound in Corollary 4.

C. Validation in Synthetic and Real-World Networks

To validate the derived bounds, especially the extinction time of the undesired information, we test the derived bounds in five networks of three different topologies: the Erdős-Rényi (random graph) topology, including ER-s (edge connecting probability $p = 7.50 \times 10^{-3}$) and ER-1 ($p = 6.76 \times 10^{-4}$); the scale-free topology, including SF-s (generated with the Barabási-Albert model); and real world fractions of Facebook, FB-s (Dataset No. 107 from [31]) and FB-1⁷ (webgraph from [32]). Their statistics are presented in Table I, in which the networks are grouped into two sets, as indicated by the ‘-s’ (size $n \simeq 1000$) and ‘-l’ (size $n \simeq 22000$) suffixes.

In Fig. 9 and Fig. 10, the simulated extinction time are shown in grey dots, whose mean $\mathbb{E}(\tau_e)$ is identified by blue ‘×’ markers. Each instance along the x-axis corresponds to the same initial cured set C_0 and infected set \mathcal{I}_0 , which is realized 100 times to obtain $\mathbb{E}(\tau_e)$. The red round markers and black squares markers correspond to the upper bounds in Theorem 4 and Corollary 4, respectively, both of which are determined by vertex eccentricities. The red dashed line and green lines with triangle markers correspond to the upper bounds presented in Corollary 2 that relies on network diameter $\text{diam}(\mathcal{G})$, and Theorem 3 that relies on the Cheeger constant,⁸ respectively.

We highlight the following observations:

1) *Topological Properties:* The small diameter $\text{diam}(\mathcal{G})$ and the relatively large algebraic connectivity λ_1 of the random graph topology (ER-s, ER-1) indicate that they are much more ‘regular’ than SF-s, FB-s, and FB-1, in the sense that vertices differ less in degree, centrality, etc. (indicating different importance/influential in status of individuals). As a result, the expected extinction time $\mathbb{E}(\tau_e)$ varies less violently in Fig. 9(a), Fig. 10(a), and Fig. 10(b).

⁷The actual FB-1 network used in simulation has 238 less vertices than the original dataset, as can be seen from Table I, because isolated vertices are removed to avoid the case of infinite extinction time ($\tau_e = \infty$).

⁸Approximating the Cheeger constant $\eta(\mathcal{G})$ with the lower bound $\frac{\lambda_1}{2}$ results in a very loose bound of $\mathbb{E}(\tau_e)$ in smaller networks and FB-1, so the upper bound (solid green line) is not shown in Fig. 9 or Fig.10(c). The dashed green lines correspond to the upper bound in Cheeger’s Inequality $\eta(\mathcal{G}) \leq 2\sqrt{\lambda_1}$, which are presented to make a comparison with bounds based on eccentricities, hence the star sign in the legend.

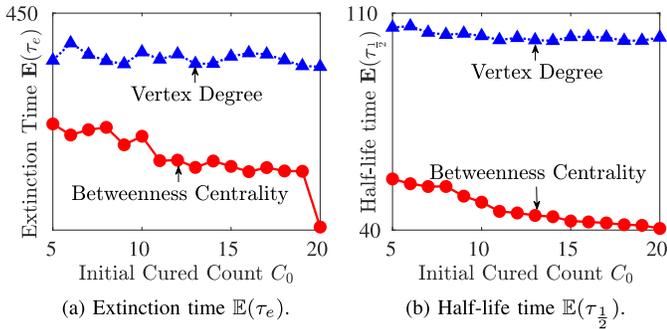


Fig. 11. Effectiveness of antidote distribution strategies based on betweenness centrality and vertex degree ($I_0 = 150$).

Despite differences in average degree \bar{d} and algebraic connectivity λ_1 , the similarity in vertex eccentricity properties (diameter and radius) between ER-s and FB-s leads to the same range of extinction time in Fig. 9(a) and Fig. 9(c), which indicates the usefulness of Corollary 2 and Corollary 4.

Due to the relatively large diameter $\text{diam}(\mathcal{G})$ of SF-s and FB-I, the probability in (29) can not be neglected, so the first condition of (25) in Theorem 4 does not hold. However, the second condition still holds, and leads to tight bounds (red round markers) in Fig. 9(b) and Fig. 10(c).

2) *Tightness of the Bounds*: Corollary 2 (dashed red lines) that relies only on network diameter $\text{diam}(\mathcal{G})$ is an easy-to-obtain, but not tight upper bound. Theorem 4 (round red markers) and Corollary 4 (black square markers) are tighter in the less ‘regular’ SF-s, FB-s, and FB-I, due to the existence of few ‘dominant’ longest shortest paths, *i.e.*, K^* is small.

In contrast, for the ER-s and ER-I network, vertex degrees and pair-wise distances are highly homogeneous (small variance in degree distribution), resulting in a large K^* that affects the tightness of both the upper and lower bounds by eccentricities. This is especially true when the initial cured count C_0 is large, *e.g.*, in Fig. 10(a). In this case, further upper bounding Theorem 3 by setting $\eta(\mathcal{G}) = \frac{\lambda_1}{2}$ (solid green lines with triangle markers) proves to be a tighter bound.

However, Theorem 3 that relies on the Cheeger constant $\eta(\mathcal{G})$, or more accurately, the algebraic connectivity λ_1 , becomes less effective than Theorem 4, when the initial cured count C_0 decreases, as shown in Fig. 10(b). This effect is more obvious in smaller networks (Fig. 9), where even the maximum possible value $\eta(\mathcal{G}) = 2\sqrt{\lambda_1}$ of the Cheeger Constant (dotted green lines) are not tight, especially in SF-s and FB-s.

D. Implication on Antidote Distribution

From the proof of Theorem 4, another implication is that, the extinction time of the undesired information can be bounded by the number and the length of shortest paths, which go through vertices of the initial cured set C_0 . This observation sheds light on antidote distribution (injection of desired information), when used as a counter-measure against the epidemic spreading of undesired information. A general guideline on selecting vertices for set C_0 is to choose vertices that sit on the most number of shortest paths, *i.e.*, vertices with high *betweenness centrality*. We test this

strategy on a portion of Facebook with 324 vertices, as illustrated in Fig. 11. Compare it with the distribution strategy based on degree (dotted blue line with triangle markers), adopting the betweenness centrality metric (solid red line with round markers) as guideline can drastically reduce both the extinction time and the half-life time, which indicates that the undesired information is effectively controlled (shorter half-life) and quickly exterminated (shorter extinction time).

V. CONCLUSION

In this paper, we studied conflicting information propagation in networks. Motivated by examples in social networks, computer networks, and IoT, we introduced a Susceptible-Infected-Cured epidemic model, to capture the transient competition between the desired and undesired information. Our model allowed us to examine the impact of network topology, propagation intensity, and the initial condition on the propagation processes of both pieces of information, such that short-term behaviors of the dynamics, namely, the lifetime of the undesired information can be characterized.

Our analysis, validated by simulation in network of various topologies, provided insights on the transient evolution of such network dynamics. For instance, the derived bounds revealed the level of ‘bottleneckness’ in a network may lead to different, even contrasting scaling behavior of information’s lifetime, with respect to the size of a network. Our findings on the impact of network topology and initial condition permitted the estimation of information lifetime before the dynamics unfolds, and can guide information distributor to efficiently allocate the limited resource to minimize the the impact of the undesired information on individuals. For example, a computer network operator who aims to effectively exterminate a spreading virus by distributing security patches, may first give patches to computers that locate on the most shortest paths, instead of computers with the most connections.

There are several directions that can be explored to further our understanding on the network dynamics of conflicting information propagation: The virus-antidote model can be generalized to allow semi-conflicting information pairs, such that the more complex competition between half-rumor and half-truth can be examined. The analysis can be extended to include multiple epochs of short-term competition, such that the best time to inject conflicting information can be studied. In addition, the decay of propagation intensity (popularity of information) over time can also be introduced to more accurately investigate this transient dynamics in finer temporal grains. We hope that our model and findings contribute to the knowledge of information propagation, and they serve as the foundation and motivation for future research in this direction.

REFERENCES

- [1] J. Wiener and N. Bronson. (Oct. 2014). *Facebook’s Top Open Data Problems*. [Online]. Available: <https://research.fb.com/facebook-s-top-open-data-problems/>
- [2] *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2017–2022*. Cisco, San Jose, CA, USA, Feb. 2019.
- [3] H. Alexander. (2013). *Boston Bomb: Reddit Apologises for Identifying Wrong Suspects*. [Online]. Available: <http://www.telegraph.co.uk/news/worldnews/northamerica/usa/10012382/>

- [4] D. Lee. (2013). *Boston Bombing: How Internet Detectives Got it Very Wrong*. [Online]. Available: <http://www.bbc.com/news/technology-22214511>
- [5] M. Garnaeva, V. Chebyshev, D. Makrushin, R. Unuchek, and A. Ivanov. (2014). *Kaspersky Security Bulletin 2014. Overall Statistics for 2014*. [Online]. Available: <http://securelist.com/analysis/kaspersky-security-bulletin/68010/kaspersky-security-bulletin-2014-overall-statistics-for-2014/>
- [6] J. Milliken, V. Selis, and A. Marshall, "Detection and analysis of the chameleon WiFi access point virus," *EURASIP J. Inf. Secur.*, vol. 2013, no. 1, p. 2, Oct. 2013.
- [7] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Phys. Rev. Lett.*, vol. 86, no. 14, pp. 3200–3203, Apr. 2001.
- [8] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the US power grid," *Saf. Sci.*, vol. 47, no. 10, pp. 1332–1336, Dec. 2009.
- [9] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Trans. Inf. Syst. Secur.*, vol. 10, no. 4, pp. 1:1–1:26, Jan. 2008.
- [10] B. A. Prakash, D. Chakrabarti, N. C. Valler, M. Faloutsos, and C. Faloutsos, "Threshold conditions for arbitrary cascade models on arbitrary networks," *Knowl. Inf. Syst.*, vol. 33, no. 3, pp. 549–575, Jul. 2012.
- [11] A. Ganesh, L. Massoulié, and D. Towsley, "The effect of network topology on the spread of epidemics," in *Proc. 24th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 2, Mar. 2005, pp. 1455–1466.
- [12] M. Lelarge, "Efficient control of epidemics over random networks," in *Proc. 11th Int. Joint Conf. Meas. Modeling Comput. Syst. (SIGMETRICS)*, New York, NY, USA, 2009, pp. 1–12.
- [13] S. Krishnasamy, S. Banerjee, and S. Shakkottai, "The behavior of epidemics under bounded susceptibility," in *Proc. ACM Int. Conf. Meas. Modeling Comput. Syst. (SIGMETRICS)*, New York, NY, USA, 2014, pp. 263–275.
- [14] B. A. Prakash, H. Tong, N. Valler, M. Faloutsos, and C. Faloutsos, "Virus propagation on time-varying networks: Theory and immunization algorithms," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases (ECML PKDD)*, Berlin, Germany: Springer-Verlag, 2010, pp. 99–114.
- [15] C. Nowzari, V. M. Preciado, and G. J. Pappas, "Analysis and control of epidemics: A survey of spreading processes on complex networks," 2015, *arXiv:1505.00768*. [Online]. Available: <http://arxiv.org/abs/1505.00768>
- [16] Y. Lin, J. C. S. Lui, K. Jung, and S. Lim, "Modeling multi-state diffusion process in complex networks: Theory and applications," in *Proc. Int. Conf. Signal-Image Technol. Internet-Based Syst.*, Dec. 2013, pp. 506–513.
- [17] B. A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos, "Winner takes all: Competing viruses or ideas on fair-play networks," in *Proc. 21st Int. Conf. World Wide Web (WWW)*, New York, NY, USA, 2012, pp. 1037–1046.
- [18] A. Dadlani, M. S. Kumar, M. G. Maddi, and K. Kim, "Mean-field dynamics of inter-switching memes competing over multiplex social networks," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 967–970, May 2017.
- [19] M. E. J. Newman, "Threshold effects for two pathogens spreading on a network," *Phys. Rev. Lett.*, vol. 95, no. 10, Sep. 2005, Art. no. 108701.
- [20] M.-C. De Marneffe, A. N. Rafferty, and C. D. Manning, "Finding contradictions in text," in *Proc. ACL*, vol. 8, 2008, pp. 1039–1047.
- [21] A. Beutel, B. A. Prakash, R. Rosenfeld, and C. Faloutsos, "Interacting viruses in networks: Can both survive?" in *Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, New York, NY, USA, 2012, pp. 426–434.
- [22] S. Hoory, N. Linial, and A. Wigderson, "Expander graphs and their applications," *Bull. Amer. Math. Soc.*, vol. 43, no. 4, pp. 439–562, Aug. 2006.
- [23] H. Zhang, Z. Zhang, and H. Dai, "Gossip-based information spreading in mobile networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5918–5928, Nov. 2013.
- [24] F. Qi and B.-N. Guo, "Sharp bounds for harmonic numbers," 2010, *arXiv:1002.3856*. [Online]. Available: <http://arxiv.org/abs/1002.3856>
- [25] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," in *Proc. 4th ACM Int. Conf. Web Search Data Mining (WSDM)*, New York, NY, USA, 2011, pp. 177–186.
- [26] A. Louis, "The complexity of expansion problems," Ph.D. dissertation, Georgia Inst. Technol., Atlanta, GA, USA, 2014.
- [27] F. R. Chung, "Laplacians of graphs and Cheeger's inequalities," *Combinatorics, Paul Erdos Eighty*, vol. 2, nos. 157–172, pp. 2–13, 1996.
- [28] T. Aven, "Upper (lower) bounds on the mean of the maximum (minimum) of a number of random variables," *J. Appl. Probab.*, vol. 22, no. 3, pp. 723–728, 1985.
- [29] G. Dasarathy. (2011). *A Simple Probability Trick for Bounding the Expected Maximum of N Random Variables*. [Online]. Available: <https://www.ece.rice.edu/~gd14/files/maxGaussians.pdf>
- [30] F. Takes and W. Kusters, "Computing the eccentricity distribution of large graphs," *Algorithms*, vol. 6, no. 1, pp. 100–118, Feb. 2013.
- [31] J. Leskovec and J. J. McAuley, "Learning to discover social circles in ego networks," in *Advances in Neural Information Processing Systems*, P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. 2012, pp. 548–556.
- [32] B. Rozemberczki, C. Allen, and R. Sarkar, "Multi-scale attributed node embedding," 2019. [Online]. Available: <https://snap.stanford.edu/data/facebook-large-page-page-network.html>



Jie Wang (Student Member, IEEE) received the Ph.D. degree in computer engineering from North Carolina State University, Raleigh, NC, USA, in 2019. She is currently a Post-Doctoral Research Scholar with the Department of Electrical and Computer Engineering, North Carolina State University. Her research interests lie in the area of networking, including modeling and performance evaluation of networked systems, such as the Internet-of-Things, cyber-physical systems, and social networks, dynamic spectrum access, mobile/edge computing, and mobile big data.



Wenyue Wang (Fellow, IEEE) received the M.S.E.E. and Ph.D. degrees in computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 1999 and 2002, respectively. She is currently a Professor with the Department of Electrical and Computer Engineering, North Carolina State University, Raleigh, NC, USA. Her research interests include mobile and secure computing, modeling and analysis of wireless networks, network topology, and architecture design. She has been a member of the Association for Computing Machinery since 1998 and a member of the Eta Kappa Nu and Gamma Beta Phi honorary societies since 2001. She was the recipient of the NSF CAREER Award 2006. She was a corecipient of the 2006 IEEE GLOBECOM Best Student Paper Award—Communication Networks—and the 2004 IEEE Conference on Computer Communications and Networks Best Student Paper Award.



Cliff Wang (Fellow, IEEE) received the Ph.D. degree in computer engineering from North Carolina State University in 1996. He currently serves as the Computing Sciences Division Chief of the Army Research Office. He is also appointed as an Adjunct Faculty Member of computer science with the College of Engineering, North Carolina State University. He has been carrying out research in the areas of computer vision, medical imaging, high speed networks, and most recently information security.