How the Anti-Rumor Kills the Rumor: Conflicting Information Propagation in Networks

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*—Online Social Networks (OSNs) is taking over television and newspapers, to be the dominant information dissemination option. The growing involvement of individuals create the situation that colliding, even contradicting information coexist and propagate in the same network, which gives rise to an interesting question: *how will the conflicting information propagate?* To answer this question, the propagation process is described to be an *Susceptible-Infected-Cured* (SIC) epidemic, and we propose an inference algorithm to study the transient behavior of the competing propagation processes in connected networks. Moreover, we provide an analytic method to derive the conditional infection count distribution for networks with special topologies, as a step further to understand the evolution. A trace collected from the Internet is analyzed to validate our model and methods.**

I. Introduction

As the advance of the networking technology, every individual is somehow connected with others via the Internet. The enhanced connectivity among individuals enables the rapid spread of information across the network. In the mean time, the growing involvement of individuals in the information production, exchange and consumption processes, contributes to the co-existence of colliding, conflicting, even contradicting information in OSNs, for example, posts with different opinions about an incident, and news versus articles about "the other side of the story". One interesting phenomenon worth noticing is, the latter injection of conflicting information into the same network *affects* the spread of the prior information.

Such phenomenon can be seen in various OSNs. On June 17th, 2015, in Wechat, a Chinese OSN formed mainly by close friends and family members, a post ("negative" information) advocating death sentence to anyone involved in child trafficking caused an epidemic spreading (540,000 re-posts by the 18th [1]). On the next day, a more rational post holding the opposite point of view ("positive" and conflicting) started to gain velocity (10,000 re-posts for the first day), while the spreading of the prior extreme post decreased significantly. Later news agency Sina revealed the advertisement nature of this incitement: a link is hidden in the first post to help a match-making website gain publicity. Another example is in Reddit and Twitter, the Boston bombing suspect hunt versus police clarification incident in April 2013 [2], in which the later police release eliminated the former viral posts accusing an innocent man based on a low-definition photo [3].

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

In both cases, the later injected "positive" information significantly affected the spreading of the former epidemic "negative" information, resulting in the extinction of a potential or on-going epidemic, which resembles distributing a replicative antidote to control an infectious virus. A natural question we will ask is *how the latter "anti-rumor" kills the former "rumor"*, i.e. we want to study *the evolution of the virus propagation under the influence of the antidote propagation*.

The exponential growth of re-posts in those examples reminds us of the epidemic propagation models, which are used to study disease [4], virus [5] failure [6] and information propagation [7] in different networks. The problem of conflicting information propagation, for instance, virus v.s. fixing patches, used to be modeled as a single epidemic problem, while the dissemination effect of the conflicting information is neglected. In this sense, two major categories of measures have been studied. The *preventing methods* [8], [9] focus on optimizing the network structure to *prevent* the epidemic from breaking out, without prior knowledge of the epidemic, while the *controlling methods* [10]–[12] are deployed to *control* the epidemic spreading when the network is under attack. On the other hand, the multiple epidemics competing problem has been introduced from the rivalry products advertising or belief propagation problems. Different propagation models have been proposed to study the competition and co-existence of multiple epidemics, in terms of time-invariant state of the network, for instances, the final infected fraction, the epidemic threshold of a large breakout, *etc*. To this end, a number of propagation models are introduced to describe the competition of epidemics, including the generalized linear threshold model [13], the Susceptible-Infected-Recovered (SIR) epidemic model,the Susceptible-Infected-Susceptible (SIS) model, and its extension SI_1I_2S model [14]. All of the aforementioned research focus on the asymptomatic analysis of *population epidemics*, where individuals in the network are viewed as a whole and influence of the topology is neglected.

For the conflicting information propagation problem, *i.e. how virus and antidote epidemics behave*, specifically, *how the number of infected and cured individuals change*, we adopted the Susceptible-Infected-Cured (SIC) propagation model to study the *transient behavior* of propagation processes in connected networks with different topologies.

The rest of the paper is organized as follows. First we introduce the system model and definitions of infection/cured counts in Section II. Then in Section III, we provide an inference method to estimate the expected infection/cured count

for arbitrary connected networks, followed by the derivation of conditional infection count distribution for connected networks with two special topologies in Section IV. A trace from the *Memetracker* [15] is analyzed for validation purpose in Section V. Finally we conclude the paper in Section VI.

II. System Model

The dissemination of conflicting information in a network is modeled as the completion between an *infectious* virus x and a *replicative* antidote a_x, later referred to as the SIC propagation model. It was first introduced in [16], but necessary introduction is given here for the sake of completeness.

A. SIC Propagation Model

1) Network $\mathcal{G}(\mathcal{V},\mathcal{E})$: The *network* $\mathcal{G}(\mathcal{V},\mathcal{E})$ is time-invariant, undirected and connected. For any vertex $v \in \mathcal{V}$, let $\mathcal{N}(v) := \{u \in \mathcal{V} |\ (u,v) \in \mathcal{E}\}$ be the *neighborhood* of v, which consists all of the vertices that can exchange data (may contain virus or antidote) with v. The adjacency matrix of network $\mathcal{G}$ is $\mathbf{A}_{n\times n} = (a_{u,v})_{u,v\in\mathcal{V}}$, where $n = |\mathcal{V}|$ is the *order* (*size* is used interchangeably) of the network.

2) Propagation Model: With respect to a virus and antidote pair (x, a_x), each vertex v is associated with a *state*, r.v. $X_v(t) : \Omega \to \{0, 1, -1\}$. The three possible states, *susceptible* to x, *infected* by x and *cured* by a_x, correspond to value $0, 1, -1$ respectively. Actions that can change the state of a vertex are *infect* by an infected neighbor and *cure* by a cured neighbor. Based on states, $\mathcal{V}$ can be decomposed into three disjoint time-varying subsets: the susceptible set $\mathcal{S}(t) = \{v \in \mathcal{V} :\ X_v(t) = 0\}$, the infected set $\mathcal{I}(t) = \{v \in \mathcal{V} :\ X_v(t) = 1\}$ and the cured set $\mathcal{C}(t) = \{v \in \mathcal{V} :\ X_v(t) = -1\}$. When infected, vertex v will pass copies of virus x to its susceptible neighbors $\mathcal{N}_S(t, v) = \mathcal{N}(v) \cap \mathcal{S}(t)$. When cured, v will pass copies of a_x to its non-cured neighbors $\mathcal{N}_{NC}(t, v) = (\mathcal{N}(v) \cap \mathcal{I}(t)) \cup (\mathcal{N}(v) \cap \mathcal{S}(t))$.

Considering the curing effect of antidote a_x on virus x, which is motivated by the observation that the "positive" anti-rumor information has a dominant credibility over the "negative" rumor information, we make the following *asymmetrical immunity* assumption.

Assumption 1: Antidote a_x can cure infected vertices and *immune* susceptible vertices (we don't differentiate these two actions). Virus x can infect any susceptible vertex in the network, but can not infect a vertex that has already been exposed to a_x (*i.e.* in cured state). Thus $\lim_{t\to\infty} X_i(t) = -1,\ \forall i \in \mathcal{V}$.

A simple example of an SIC epidemic evolution in a connected network of eight vertices is shown in Fig. 1. At t_0, the infected set $\mathcal{I}(t_0) = \{a, g\}$, while $\mathcal{C}(t_0) = \{d\}$. During t_0 to t_0+1, a tries to infect susceptible vertices b and h, while d cures (or rather, immunes) susceptible vertex e. As a result, at t_0+1, $\mathcal{I}(t_0+1) = \{a, g, b, h\}$ and $\mathcal{C}(t_0+1) = \{d, e\}$. Based on Assumption 1, at $t_0 + 2$, infected vertex $h \in \mathcal{I}(t_0 + 1)$ is cured, since d tries to pass an antidote to it before $t_0 + 2$. Also, we can infer that after some time, cured vertex f won't be infected by g even if g tries to pass the virus to it.

As can be seen in the previous example, the propagation of virus x and antidote a_x drives the change in sizes of

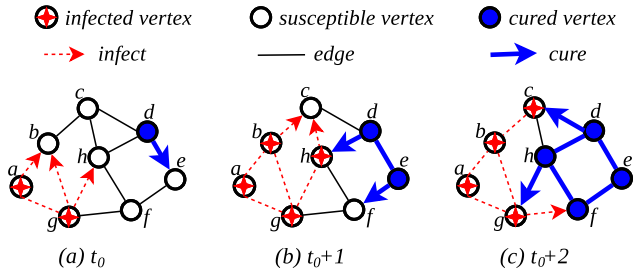

Fig. 1: An SIC Epimdemic in a Connected Network.

the infected set $\mathcal{I}(t)$ and the cured set $\mathcal{S}(t)$. To capture the behavior of x and a_x, it is important to characterize the time intervals between infections and curings. Suppose vertex u is infected at t, let $t+s_u(v)$ denote the time that infected vertex u infects one of its susceptible neighbors $v \in \mathcal{N}_S(t, u)$. Suppose u is cured at t, let $t+a_u(v)$ denote the time that vertex u cures one of its non-cured neighbors $v \in \mathcal{N}_{NC}(t, u)$. We make the following assumption about random intervals $s_u(v)$ and $a_u(v)$.

Assumption 2: Random intervals $\{s_u(v)\}_{v\in\mathcal{N}_S(t,u)}$ and $\{a_u(v)\}_{v\in\mathcal{N}_{NC}(t,u)}$ are two groups of r.v.'s each satisfying (1) pairwise independent; (2) exponentially distributed with time-invariant parameters $\beta_{u,v}$ and $\gamma_{u,v}$ respectively.

$\beta_{u,v}$ is called the *virulence* (or *infection rate*) of the virus x on edge $e(u, v)$, which determines how frequently vertex u (if in infected state) distributes a copy of virus x to its susceptible neighbor v along edge $e(u, v)$. Note that $\beta_{u,v}$ is in fact the derivative of probability $\mathbb{P}(s_u(v) \le t)$ over time t. Similarly, $\gamma_{u,v}$ is the *curing rate* of the antidote a_x on edge $e(u, v)$.

B. The Infection Count and Cured Count

From the network-wise point of view, the evolution of SIC epidemics can be described by the change of sizes of the infected set and cured set over time. So we define the *infection count* and *cured count* to study this transient behavior.

Definition 1: Infection count, $I(t) : \Omega^n \to \mathbb{R}$, is the size of the infected set $\mathcal{I}(t)$ at time t. *Cured count* $C(t) : \Omega^n \to \mathbb{R}$, is the size of the cured set $\mathcal{C}(t)$ at time t.

$$I(t) = \sum_{v\in\mathcal{V}} \mathbb{1}_{\mathcal{I}(t)} = \sum_{v\in\mathcal{V}} \mathbb{1}_{\{1\}}(X_v(t)).$$
$$C(t) = \sum_{v\in\mathcal{V}} \mathbb{1}_{\mathcal{C}(t)} = \sum_{v\in\mathcal{V}} \mathbb{1}_{\{-1\}}(X_v(t)).$$

The SIC dynamic in a network drives the evolution of random processes $\{I(t)\}_{t\in\Gamma}$ and $\{C(t)\}_{t\in\Gamma}$. By the asymmetric immunization assumption, the network will stabilize at all-cured state, that is, $\lim_{t\to\infty} C(t) = n$. In addition, $C(t)$ is a counting process. To study $\{I(t)\}_{t\in\Gamma}$ and $\{C(t)\}_{t\in\Gamma}$, we adopt different approaches for networks with different topologies. First for any connected network $\mathcal{G}$, we develop an inference method to derive the expected infection count as a function of time, then for networks with special topologies (fully-connected and star), based on the fact that $C(t)$ is a counting process, we define the *conditional infection count distribution* to give more details about the evolving SIC dynamics.

III. SIC Epidemic in Connected Networks with Arbitrary Topologies

Given an arbitrary connected network $\mathcal{G}(\mathcal{V},\mathcal{E})$, let the time be discrete and a time step is small enough, such that within a time step the state of a vertex $i\in\mathcal{V}$, $X_i(t)$ depends solely on the last state of its neighbors $\{X_j(t-1)|j\in\mathcal{N}(i)\}$ and itself $X_i(t-1)$, which resembles the *Local Markov property* of Markov Random Fields (MRF), but with a correlated temporal and spatial dependent relationship. Inspired by Chen and Ji's idea [17] of separating spatial dependence and temporal dependence, we derive a time-recursive expression of $\mathbb{P}(X_i(t)=1)$.

With states of i's neighbors at time t, the probability that a susceptible vertex i remains susceptible in the next time step can be written as

$$\begin{aligned}&\mathbb{P}(X_i(t+1)=0|X_i(t)=0,X_{\mathcal{N}(i)}(t)=x_{\mathcal{N}(i)}(t))\\&=\prod_{j\in\mathcal{N}_I(t,i)}(1-\beta_{j,i})\cdot\prod_{k\in\mathcal{N}_C(t,i)}(1-\gamma_{k,i})\\&=\prod_{j\in\mathcal{N}(i)}(1-\beta_{j,i})^{\frac{1}{2}(x_j^2(t)+x_j(t))}\cdot(1-\gamma_{j,i})^{\frac{1}{2}(x_j^2(t)-x_j(t))}.\end{aligned}\tag{1}$$

Let $\mathbb{P}_{uv}(t)=\mathbb{P}(X_i(t+1)=v|X_i(t)=u)$ denote the transition probability from state u to v during one time step at time t. Note that only three equations are needed because $\sum_{v\in\Lambda}\mathbb{P}_{u,v}=1$, $\mathbb{P}_{10}=0$ and $\mathbb{P}_{-1-1}=1$. Let $I_i(t)=1-\prod_{j\in\mathcal{N}(i)}(1-\beta_{j,i})^{\frac{1}{2}(x_j^2(t)+x_j(t))}$, $C_i(t)=1-\prod_{k\in\mathcal{N}(i)}(1-\gamma_{k,i})^{\frac{1}{2}(x_k^2(t)-x_k(t))}$, then the evolution can be described by the following one-step transition probabilities.

$$\begin{aligned}\mathbb{P}_{00}(i,t)=&\sum_{x_{\mathcal{N}(i)}(t)}\mathbb{P}(X_{\mathcal{N}(i)}(t)=x_{\mathcal{N}(i)}(t)|X_i(t)=0)\\&\cdot(1-I_i(t))(1-C_i(t)).\end{aligned}\tag{2}$$

$$\begin{aligned}\mathbb{P}_{01}(i,t)=&\sum_{x_{\mathcal{N}(i)}(t)}\mathbb{P}(X_{\mathcal{N}(i)}(t)=x_{\mathcal{N}(i)}(t)|X_i(t)=0)\\&\cdot I_i(t)(1-C_i(t)).\end{aligned}\tag{3}$$

$$\mathbb{P}_{11}(i,t)=\sum_{x_{\mathcal{N}(i)}(t)}\mathbb{P}(X_{\mathcal{N}(i)}(t)=x_{\mathcal{N}(i)}(t)|X_i(t)=1)C_i(t).\tag{4}$$

The temporal dependence is included in $I_i(t)$ and $C_i(t)$, while the spatial dependence is captured by the joint conditional probability $\mathbb{P}(X_{\mathcal{N}(i)}(t)=x_{\mathcal{N}(i)}(t)|X_i(t)=x_i(t))$, which is the reason that calculating the exact marginal probability distribution is expensive. A common assumption is that during one time step, states of different vertices are mutually independent, *i.e.* $\mathbb{P}(X_{\mathcal{N}(i)}(t)=x_{\mathcal{N}(i)}(t)|X_i(t)=x_i(t))=\prod_{j\in\mathcal{N}(i)}\mathbb{P}(X_j(t)=x_j(t))$. However, this assumption will result in a noticeable deviation [17]. Based on the observation,

$$\begin{aligned}&\mathbb{P}(X_{\mathcal{N}(i)}(t)=x_{\mathcal{N}(i)}(t)|X_i(t)=x_i(t))\\&\qquad=\prod_{j\in\mathcal{N}(i)}\mathbb{P}(X_j(t)=x_j(t)|X_i(t)=x_i(t)).\end{aligned}\tag{5}$$

Let $N_0^s(i,j,t)=\mathbb{P}(X_j(t)=s|X_i(t)=0)$,
$N_1^s(i,j,t)=\mathbb{P}(X_j(t)=s|X_i(t)=1)$, $s\in\{0,-1,1\}$.
Then Eq. (2),(3),(4) can be simplified to

$$\begin{aligned}\mathbb{P}_{00}(i,t)=&\sum_{x_{\mathcal{N}(i)}(t)}\prod_{j\in\mathcal{N}(i)}\big[N_0^{x_j(t)}(i,j,t)\cdot(1-\beta_{j,i})^{\frac{1}{2}(x_j^2(t)+x_j(t))}\\&\cdot(1-\gamma_{j,i})^{\frac{1}{2}(x_j^2(t)-x_j(t))}N_0^{x_j(t)}(i,j,t)\big]\\=&\prod_{j\in\mathcal{N}(i)}\big[1-N_0^1(i,j,t)\beta_{j,i}\\&-\big(1-N_0^0(i,j,t)-N_0^1(i,j,t)\big)\gamma_{j,i}\big]\end{aligned}\tag{6}$$

$$\begin{aligned}\mathbb{P}_{01}(i,t)=&\sum_{x_{\mathcal{N}(i)}(t)}\big[1-\prod_{j\in\mathcal{N}(i)}(1-\beta_{j,i})^{\frac{1}{2}(x_j^2(t)+x_j(t))}\big]\\&\cdot\prod_{j\in\mathcal{N}(i)}(1-\gamma_{j,i})^{\frac{1}{2}(x_j^2(t)-x_j(t))}N_0^{x_j(t)}(i,j,t)\\=&\prod_{j\in\mathcal{N}(i)}\big[1-\big(1-N_0^0(i,j,t)-N_0^1(i,j,t)\big)\gamma_{j,i}\big]\\&-\mathbb{P}_{00}(i,t)\end{aligned}\tag{7}$$

$$\begin{aligned}\mathbb{P}_{11}(i,t)=1-&\sum_{x_{\mathcal{N}(i)}(t)}\big[1-\prod_{j\in\mathcal{N}(i)}(1-\gamma_{j,i})^{\frac{1}{2}(x_j^2(t)-x_j(t))}\big]\\&\cdot\prod_{j\in\mathcal{N}(i)}N_0^{x_j(t)}(i,j,t)\\=&\prod_{j\in\mathcal{N}(i)}\big[1-\big(1-N_1^0(i,j,t)-N_1^1(i,j,t)\big)\gamma_{j,i}\big]\end{aligned}\tag{8}$$

With Eq. (5)-(7), the recursive master equations of the system over time can be written as

$$\mathbb{P}(X_i(t+1)=0)=\mathbb{P}(X_i(t)=0)\cdot\mathbb{P}_{00}(i,t),\tag{9}$$

$$\begin{aligned}\mathbb{P}(X_i(t+1)=1)=&\mathbb{P}(X_i(t)=0)\cdot\mathbb{P}_{01}(t)\\&+\mathbb{P}(X_i(t)=1)\cdot\mathbb{P}_{11}(i,t).\end{aligned}\tag{10}$$

Clearly, $N_0^0(i,j,t), N_0^1(i,j,t), N_1^0(i,j,t)$ and $N_1^1(i,j,t)$ are necessary to solve these equations. Though it is hard to give closed form equations of these quantities, it is possible to derive a recursive relationship with respect to time.

$$\begin{aligned}N_0^0(i,j,t)=&\frac{\mathbb{P}(X_j(t)=0,X_i(t)=0)}{\mathbb{P}(X_i(t)=0)}\\=&\frac{\mathbb{P}(X_i(t-1)=0)}{\mathbb{P}(X_i(t)=0)}\cdot N_0^0(i,j,t-1)\\&\cdot\mathbb{P}_{00}(i,j,t-1)\cdot\mathbb{P}_{00}(j,i,t-1).\end{aligned}\tag{11}$$

$$\begin{aligned}N_0^1(i,j,t)=&\frac{\mathbb{P}(X_j(t)=1,X_i(t)=0)}{\mathbb{P}(X_i(t)=0)}\\=&\frac{\mathbb{P}(X_i(t-1)=0)}{\mathbb{P}(X_i(t)=0)}\cdot P_{00}(i,j,t-1)\\&\cdot\big[N_0^0(i,j,t-1)\cdot\mathbb{P}_{01}(j,i,t-1)\\&+(1-\beta_{j,i})\cdot N_0^1(i,j,t-1)\cdot\mathbb{P}_{11}(j,i,t-1)\big].\end{aligned}\tag{12}$$

$$N_1^0(i,j,t) = \frac{\mathbb{P}(X_j(t)=0, X_i(t)=1)}{\mathbb{P}(X_i(t)=1)}$$
$$= \frac{\mathbb{P}_{00}(j,i,t-1)}{\mathbb{P}(X_i(t)=1)} \cdot \big[\mathbb{P}(X_i(t-1)=0)\cdot N_0^0(i,j,t-1)$$
$$\cdot \mathbb{P}_{01}(i,j,t-1) + \mathbb{P}(X_i(t-1)=1)\cdot(1-\beta_{i,j})$$
$$\cdot N_1^0(i,j,t-1)\cdot \mathbb{P}_{11}(i,j,t-1)\big]. \quad (13)$$

$$N_1^1(i,j,t) = \frac{\mathbb{P}(X_j(t)=1, X_i(t)=1)}{\mathbb{P}(X_i(t)=1)}$$
$$= \frac{\mathbb{P}(X_i(t-1)=0)}{\mathbb{P}(X_i(t)=1)} \cdot \Big[N_0^0(i,j,t-1)\cdot \mathbb{P}_{01}(i,j,t-1)$$
$$\cdot \mathbb{P}_{01}(j,i,t-1) + N_0^1(i,j,t-1)\cdot \mathbb{P}_{11}(j,i,t-1)$$
$$\cdot [\beta_{j,i} + \mathbb{P}_{01}(i,j,t-1) - \beta_{j,i}\mathbb{P}_{01}(i,j,t-1)]\Big]$$
$$+ \frac{\mathbb{P}(X_i(t-1)=1)}{\mathbb{P}(X_i(t)=1)} \cdot \mathbb{P}_{11}(j,i,t-1)$$
$$\cdot \Big[[\beta_{i,j} + \mathbb{P}_{01}(j,i,t-1) - \beta_{i,j}\mathbb{P}_{01}(j,i,t-1)]$$
$$\cdot N_1^0(i,j,t-1) + N_1^1(i,j,t-1)\cdot \mathbb{P}_{11}(i,j,t-1)\Big]. \quad (14)$$

where

$$\mathbb{P}_{00}(i,j,t) = \prod_{k\in\mathcal{N}(i)\setminus\{j\}} \big[1 - N_0^1(i,k,t)\beta_{k,i}$$
$$- \big(1 - N_0^0(i,k,t) - N_0^1(i,k,t)\big)\gamma_{k,i}\big] \quad (15)$$
$$\mathbb{P}_{01}(i,j,t) = \prod_{k\in\mathcal{N}(i)\setminus\{j\}} \big[1 - \big(1 - N_0^0(i,k,t) - N_0^1(i,k,t)\big)\gamma_{k,i}\big]$$
$$- P_{00}(i,j,t) \quad (16)$$
$$\mathbb{P}_{11}(i,j,t) = \prod_{k\in\mathcal{N}(i)\setminus\{j\}} \big[1 - \big(1 - N_1^0(i,k,t) - N_1^1(i,k,t)\big)\gamma_{k,i}\big]. \quad (17)$$

denote the probability that without considering vertex $j \in \mathcal{N}(i)$, vertex i remains susceptible, becomes infected, and remains infected during time step t respectively.

Note that Eq. (11)-(14) are time-recursive, allowing us to estimate the state evolution as long as the initial state of the network is known (equivalent to the case that the initial distribution of state vector $\mathbf{X}(0)$ is a δ-distribution at time $t=0$). It is clear that $\mathbb{P}(X_v(0)=1)=1\ \forall v\in\mathcal{I}(0)$, $\mathbb{P}(X_u(0)=-1)=1\ \forall u\in\mathcal{C}(0)$ and $\mathbb{P}(X_w(0)=0)=1\ \forall w\in\mathcal{S}(0)$. Therefore, $\mathbb{P}(\mathbf{X}(0)=\mathbf{x}(0)) = 1 = \prod_{v\in\mathcal{V}}\mathbb{P}(X_v(0)=x_v(0))$, where $\mathbf{x}(0)=\{x_v(0)\}_{v\in\mathcal{V}}$ is the state vector of the network at time 0, which indicates r.v.'s $\{X_v(0)\}_{v\in\mathcal{V}}$ are mutually independent. Hence $N_r^s(j,0)=\mathbb{P}(X_j(0)=s|X_i(0)=r)=\mathbb{P}(X_j(0)=s)$ is determined for any $r,s\in\{0,1,-1\}$. Then $\mathbb{P}(X_i(t)=1)$ can be solved iteratively, and the expected infection count at time t can be calculated by

$$\mathbb{E}(I(t)) = \mathbb{E}[\sum_{i\in\mathcal{V}} \mathbb{1}_{\{1\}}(X_i(t))] = \sum_{i\in\mathcal{V}} \mathbb{P}(X_i(t)=1), \quad (18)$$
$$\mathbb{E}(C(t)) = \sum_{i\in\mathcal{V}} \big[1 - \mathbb{P}(X_i(t)=0) - \mathbb{P}(X_i(t)=1)\big]. \quad (19)$$

The manipulations described above can be summarized into the following iterative algorithm. Note that $\vec{P}_0(t) = \{\mathcal{P}(X_i(t)=0)\}_{i\in\mathcal{V}}$ and $\vec{P}_1(t) = \{\mathcal{P}(X_i(t)=1)\}_{i\in\mathcal{V}}$ are two $n\times 1$ vectors. $\vec{RS}(t) = \{\mathbb{P}_{00}(i,j,t)\}_{i,j\in\mathcal{V}}$, $\vec{NC}(t) = \{\mathbb{P}_{11}(i,j,t)\}_{i,j\in\mathcal{V}}$, $\vec{N}_0^0(t) = \{N_0^0(i,j,t)\}_{i,j\in\mathcal{V}}$, $\vec{N}_1^0(t) = \{N_0^1(i,j,t)\}_{i,j\in\mathcal{V}}$, $\vec{N}_0^1(t) = \{N_1^0(i,j,t)\}_{i,j\in\mathcal{V}}$, $\vec{N}_1^1(t) = \{N_1^1(i,j,t)\}_{i,j\in\mathcal{V}}$ are $n\times n$ matrices.

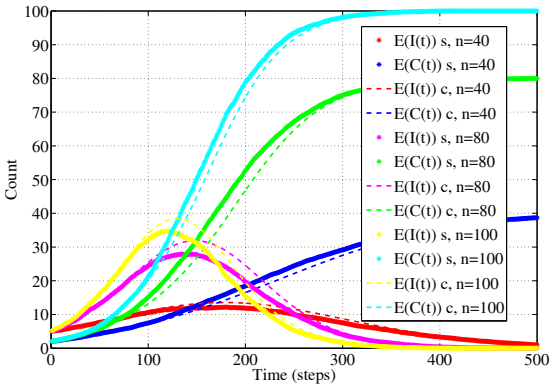

Fig. 2: Simulated *v.s.* Calculated Infection/Cured Count in Networks of Size 40,80 and 100, with Arbitrary Topologies.

Algorithm 1 Iterative Inference Method.

Input: $\mathcal{I}(0)$, $\mathcal{C}(0)$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\mathcal{V}$
Output: $\mathbb{E}(I(t))$, $\mathbb{E}(C(t))$
Initialize: $n=|\mathcal{V}|$, $I(0)=|\mathcal{I}(0)|$, $C(0)=n-I(0)-|\mathcal{C}(0)|$
$\vec{P}_0(0)=\vec{N}_0^0(0)=\vec{N}_1^0(0)=\mathbb{1}_{\mathcal{V}\setminus(\mathcal{I}(0)\cup\mathcal{C}(0))}$,
$\vec{P}_1(0)=\vec{N}_0^1(0)=\vec{N}_1^1(0)=\mathbb{1}_{\mathcal{I}(0)}$
Calculate $\vec{P}_0(1)$ with Eq. (4) and (8)
Calculate $\vec{P}_1(1)$ with Eq. (6), (7) and (10)
Output $\mathbb{E}(I(1)) = sum(\vec{P}_1(1))$,
$\mathbb{E}(C(1)) = n - \mathbb{E}(I(1)) - sum(\vec{P}_0(1))$
Calculate $\vec{RS}(0)$ and $\vec{NC}(0)$ with Eq. (15) and (16)
$t=1$
while $\mathbb{E}(C(t)) < n$ **do**
  Calculate $\vec{N}_0^0(t)$, $\vec{N}_0^1(t)$, $\vec{N}_1^0(t)$, $\vec{N}_1^1(t)$ with Eq. (11)-(15)
  Calculate $\vec{P}_0(t+1)$ with Eq. (4) and (8)
  Calculate $\vec{P}_1(t+1)$ with Eq. (6), (7) and (10)
  Output $\mathbb{E}(I(t+1)) = sum(\vec{P}_1(t+1))$,
  $\mathbb{E}(C(1)) = n - \mathbb{E}(I(t+1)) - sum(\vec{P}_0(t+1))$
  Calculate $\vec{RS}(t)$ and $\vec{NC}(t)$ with Eq. (15) and (16)
  $t=t+1$
end while

Fig. 2 shows the simulation (dashed lines with s in the legend) v.s. calculation (solid lines with c in the legend) of expected value of infection and cured counts, in arbitrary connected networks with various size. As shown, the infection/cured counts can be well captured by the algorithm.

IV. SIC Epidemic in Networks with Special Topologies

When the underlying graph of the network is of special topologies, *i.e.* a complete graph (K_n) or a star with one hub(S_n) , more information can be achieved. We can derive a "conditional" *distribution* of the infection count given the cured count if the following additional assumption is given.

Assumption 3: For any edge in the network, the infection rate and the curing rate are constant, *i.e.*

$$\beta_{i,j} = \beta, \gamma_{i,j} = \gamma, \ \forall i,j \in \mathcal{V} \ s.t. \ e(i,j) \in \mathcal{E}.$$

The state of the network can be described by tuple $(C(t), I(t))$ for any given time t, which takes value in a *discrete* state space $[0,n]\times[0,n]$. Since we are more interested in *how* the epidemic evolves, rather than *when*, we index $(C(t), I(t))$ with $k \in \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, to record a sequence of infection and curing events ordered by their occurrences. Thus we have a discrete time discrete state space Markov Chain $\{(C_k, I_k)\}_{k\in\mathbf{N}_0}$ on $M = [0,n] \times [0,n] \subset \mathbb{R}^2$. Due to the assumption of asymmetrical immunity, the evolution of the SIC epidemic is a path connecting the initial state (C_0, I_0) and the steady state $(n, 0)$ on state space M.

At any state (c,i), there are three possible events that can result a state change: an infection of a susceptible vertex with probability $p_{c,i} = \mathbb{P}((c,i) \to (c,i+1))$, a curing of a susceptible vertex with probability $r_{c,i} = \mathbb{P}((c,i) \to (c+1,i))$, and a curing of an infected vertex with probability $q_{c,i} = \mathbb{P}((c,i) \to (c+1, i-1))$. Note that $p_{0,i} = 1$, $q_{c,n-c} = 1$, and $p_{c,i} + r_{c,i} + q_{c,i} = 1$. We refer to $P = (p_{c,i})_{c,i\in M}$, $R = (r_{c,i})_{c,i\in M}$ and $Q = (q_{c,i})_{c,i\in M}$ as the *partial transition (probability) matrix*, since all together they uniquely determine the transition probability of the system.

Considering $\{C_k\}_{k\in\mathbb{N}_0}$ is a counting process, and antidote is easier to monitor since it is distributed by the network administrator, we define the *conditional infection count distribution* to study the evolution of the virus epidemic.

Definition 2: Let $\mathbb{P}_0(\cdot) = \mathbb{P}(\cdot \mid (C_0, I_0) = (c_0, i_0))$ denote the probability that the chain visits a state on condition that the initial state is (c_0, i_0). The *conditional infection count distribution* $\pi_c(i)$ is defined as the probability that the infection count is i on condition that the cured count is c, *i.e.*

$$\pi_c(i) = \sum_{k\in\mathbb{N}_0} \frac{\mathbb{P}_0((C_k, I_k) = (c,i))}{\sum_{j=0}^{n-1} \mathbb{P}_0((C_k, I_k) = (c,j))}.$$

Note that we require at least one cured vertex and one infected vertex at $t = 0$, or else the SIC epidemic will degenerate to an SI epidemic. $\mathbb{P}_0((C_k, I_k) = (c,j))$ can be calculated with matrices P, R and Q, which we omit here due to the limitation of space. If the network is of a special topology, the P, R and Q matrix can be obtained with the graphical characteristics of the network.

1) Complete Graph K_n:

Theorem 1: For an SIC epidemic with propagation parameters β, γ and initial state (c_0, i_0), if $\mathcal{G} = K_n$, then

$$p_{c,i} = \frac{i(n-i-c)}{\rho c(n-c) + i(n-c-i)}, \tag{20}$$

$$r_{c,i} = \frac{\rho c(n-i-c)}{\rho c(n-c) + i(n-c-i)}, \tag{21}$$

where $\rho = \gamma/\beta$.

Proof: Since $I(t) + C(t) + S(t) = n$ in the SIC dynamic, and the network is fully connected, the master equations of the system are

$$C'(t) = \gamma C(t)\,(n - C(t)) \tag{22}$$

$$I'(t) = \beta I(t) S(t) - \gamma C(t) I(t) \tag{23}$$

$$S'(t) = -\beta I(t) S(t) - \gamma C(t) S(t). \tag{24}$$

By relating the rate to probability, it is clear to see that $p_{c,i}$ equals to the probability that an infection happens before any curing when the chain is at state (c,i). Therefore

$$\begin{aligned} p_{c,i} &= \mathbb{P}((C_{k+1}, I_{k+1}) = (c, i+1) | (C_k, I_k) = (c,i)) \\ &= \frac{\beta i(n-i-c)}{\gamma c(n-c) + \beta i(n-i-c)}. \end{aligned} \tag{25}$$

Similarly, $r_{c,i}$ is the probability that a curing of a susceptible vertex happens before any infection when the chain is at state (c,i) and $q_{c,i}$ corresponds to the probability that a curing of an infected vertex happens before any infection when the chain is at state (c,i). ■

2) Star Graph S_n:

Theorem 2: For an SIC epidemic with parameters β, γ and initial infection count $i_0 \geq 2$ (i.e. the hub is infected at $t_0 = 0$), if $\mathcal{G} = S_n$ with one hub and $n-1$ leaves, then for $c \geq C_0 + 1$,

$$p_{c,i} = 0, \tag{26}$$

$$r_{c,i} = \frac{n-c-i}{n-c}, \tag{27}$$

where $\rho = \gamma/\beta$.

Proof: Note that $p_{c_0,i} = \frac{(n-C_0-i)}{(n-C_0-i)+C_0\rho}$ and $r_{c_0,i} = 0$. The curing of the hub is the turning point of the evolution. First, since the antidote are distributed to susceptible vertices at $t_0 = 0$, the dissemination of the antidote can not reach other peripheral vertices without passing through the bub. Hence $r_{C_0,i} = 0$. In addition, due to similar reasons, once the antidote is passed to the hub, virus can't infect the hub again, therefore after $c = C_0$, no peripheral susceptible vertices will suffer from infection again, i.e. $p_{C_0,i} = 0$. ■

V. Trace Analysis

In this section, we present a trace of a propaganda incident captured in the Internet, analyze the data with our SIC propagation model, and compare the simulation results with the captured data to show how our SIC model describes the evolution of information propagation in networks.

On March 20th 2009, President Obama joked about his bowling skills, saying "It was like a Special Olympics, or something" on 'The Tonight Show with Jay Leno'. Realizing this joke could be inappropriate, he then apologized to Special Olympics chairman Tim Shriver, before the program was aired. With information about both the show and the clarification released around the same time, the meme tracker [18] recorded the evolution of this incident in terms of popularity, as shown in Figure 3. In this case, the clarification afterwords functions as the antidote to the viral offensive joke. The vertical axis, that is, the number of mentions, can be treated as a quantification of popularity, or equivalently, the infection count $I(t)$. Due to the injection of "positive" information, the spread of the viral information was limited, and the whole incident died down within 65 hours after its initiation.

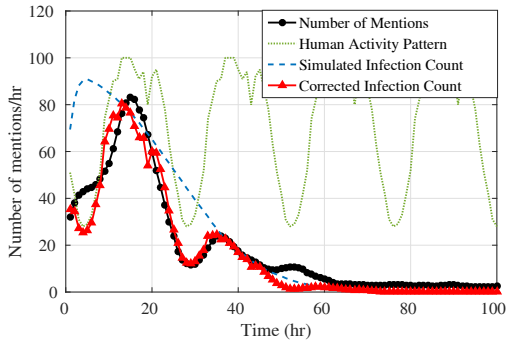

Fig. 3: President Obama's '*Special Olympics*' Joke Incident: Real Number of Mentions v.s. SIC Model Simulation with parameters: $\beta = 0.011,\ \gamma = 0.0012,\ C_0 = 5,\ I_0 = 51$.

Data points ($I(t)$) are shown as the black solid line with round markers in Fig.3. We then simulate the evolution of an SIC epidemic on K_{100} with propagation parameters $\beta = 0.011,\ \gamma = 0.0012$, and initial condition $C_0 = 5,\ I_0 = 51$[1], as the corrected infection counts $\hat{I}(t)$ (the red solid line with triangle markers). By "corrected", we mean the human online activity pattern [19] is taken into consideration, to calibrate the simulation, since the spread of this incident lasts more than 24 hours. The trend of the dynamics evolution is well captured by the SIC model, especially at the following turning points: (i) at 12:00 am March 20th ($t = 0$) when the incident started spreading, $I(0) = 31.95 \simeq \hat{I}(0)$; (ii) both peaks occurred around $t = 14$, and the peak value $I(t = 14)$ is 83.14[2]; (iii) the extinction of the virus happens at $t = 65$. However, there are two noticeable gaps at the initiation phase and the extinction phase, i.e. $\hat{I}(t) < I(t)$ for $t \in [1, 8]$ and $t \in [49, 57]$. The possible reason are: first, the human activity pattern is an average over a large sample space, hence is less accurate when applied to small group of people, that is, at the beginning and the extinction phases; second, information propagation rate decays as time goes on, so in the initiation phase where the virus dominants the propagation, β is actually larger, while in the extinction phase where the antidote dominants the propagation, γ is smaller than the middle part.

VI. CONCLUSION

In this paper, we adopted an SIC information propagation model to study the the conflicting information propagation problem. For connected networks with arbitrary topology and varying propagation parameters, an inference method is proposed to study the mean infection count and cured count as a function of time, while for networks with special topologies, the transition probability matrix and conditional infection count are analyzed. The proposed methods are validated through a trace analysis. We hope this work will be useful in studying information dissemination problem in networks.

[1]Complete graph is chosen as the topology due to the universal availability of the content. Parameters are set to best fit the shape of this meme.

[2]This value can be considered as normalized quantity out of 100.

REFERENCES

[1] M. Danning, "Debate erupts on social media over dath for traffickers." www.chinadaily.com.cn/china/2015-06/18/content_21042257.htm, 2015.

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

[3] H. Alexander, "Boston bomb: Reddit apologises for identifying wrong suspects." http://www.telegraph.co.uk/news/worldnews/northamerica/usa/10012382/, 2013.

[4] R. Kao, L. Danon, D. Green, and I. Kiss, "Demographic structure and pathogen dynamics on the network of livestock movements in great britain," *Proc. of the Royal Society of London B: Biological Sciences*, vol. 273, no. 1597, pp. 1999–2007, 2006.

[5] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical review letters*, vol. 86, no. 14, p. 3200, 2001.

[6] 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.

[7] A. Ganesh, L. Massoulie, and D. Towsley, "The effect of network topology on the spread of epidemics," in *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, vol. 2, pp. 1455–1466 vol. 2, March 2005.

[8] 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.

[9] A. J. O'Donnell and H. Sethu, "Software diversity as a defense against viral propagation: Models and simulations," in *Proceedings of the 19th Workshop on Principles of Advanced and Distributed Simulation*, PADS '05, pp. 247–253, IEEE Computer Society, 2005.

[10] 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.

[11] 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.

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

[13] N. Pathak, A. Banerjee, and J. Srivastava, "A generalized linear threshold model for multiple cascades," in *Proceedings of the 2010 IEEE International Conference on Data Mining*, ICDM '10, (Washington, DC, USA), pp. 965–970, IEEE Computer Society, 2010.

[14] B. A. Prakash, A. Beutel, R. Rosenfeld, and C. Faloutsos, "Winner takes all: Competing viruses or ideas on fair-play networks," in *Proceedings of the 21st International Conference on World Wide Web*, WWW '12, (New York, NY, USA), pp. 1037–1046, ACM, 2012.

[15] Y. Lin, J. Lui, K. Jung, and S. Lim, "Modeling multi-state diffusion process in complex networks: Theory and applications," in *Signal-Image Technology Internet-Based Systems (SITIS), 2013 International Conference on*, pp. 506–513, Dec 2013.

[16] 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.

[17] Z. Chen and C. Ji, "Spatial-temporal modeling of malware propagation in networks," *Neural Networks, IEEE Transactions on*, vol. 16, pp. 1291–1303, Sept 2005.

[18] J. Yang and J. Leskovec, "Patterns of temporal variation in online media," in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, (New York, NY, USA), pp. 177–186, ACM, 2011.

[19] Chitika, "Hour-by-hour examination: Smartphone, tablet, and desktop usage rates." https://chitika.com/browsing-activity-by-hour, Mar. 2013.