A Sword with Two Edges: Propagation Studies on Both Positive and Negative Information in Online Social Networks

Sheng Wen, Mohammad Sayad Haghighi, Chao Chen, Yang Xiang, *Senior Member, IEEE*, Wanlei Zhou, *Senior Member, IEEE*, and Weijia Jia, *Senior Member, IEEE*

Abstract—Online social networks (OSN) have become one of the major platforms for people to exchange information. Both positive information (e.g., ideas, news and opinions) and negative information (e.g., rumors and gossips) spreading in social media can greatly influence our lives. Previously, researchers have proposed models to understand their propagation dynamics. However, those were merely simulations in nature and only focused on the spread of one type of information. Due to the human-related factors involved, simultaneous spread of negative and positive information cannot be thought of the superposition of two independent propagations. In order to fix these deficiencies, we propose an analytical model which is built stochastically from a node level up. It can present the temporal dynamics of spread such as the time people check newly arrived messages or forward them. Moreover, it is capable of capturing people's behavioral differences in preferring what to believe or disbelieve. We studied the social parameters impact on propagation using this model. We found that some factors such as people's preference and the injection time of the opposing information are critical to the propagation but some others such as the hearsay forwarding intention have little impact on it. The extensive simulations conducted on the real topologies confirm the high accuracy of our model.

Index Terms—Social network, modeling, propagation analysis

1 Introduction

The popularity of online social networks (OSN) such as Facebook [1], Google Plus [2] and Twitter [3] has greatly increased in recent years. OSNs have become an important platform for the dissemination of news, ideas, opinions, etc. Unfortunately, OSN is a double-edged sword. The openness of OSN platforms also enables rumors, gossips and other forms of disinformation to spread all around the Internet. To be generic, we name the authentic information as the positive information and the conflicting fake news (e.g., rumors) as the negative information.

1.1 How Information Spreads in OSNs

Both positive and negative information can spread in OSNs by posting on the wall or directly sending message to social neighbors. The propagation process continues when neighboring users believe the information and forward it to their social neighbors. When a user receives contradicting pieces of information (i.e., both positive and negative), he or she makes a choice. The user might go for the positive or negative, or even refute both. In the following, we list three real cases according to our investigation on the history of OSNs:

1) "*Two explosions in White House and Barack Obama is injured* (April 23, 2013)": Syrian hackers broke into the twitter account of Associated Press (AP) and spread news that explosions at White House have injured Obama [4]. The White House and AP assured the public minutes later that the report was not true but word did not appear to come fast enough to those frantically watching and responding on Wall Street. Both the DOW Jones industrial average and Standard & Poor's 500 Index plunged about 1 percent before regaining their losses.

People might have been misled by critical rumors they received, but once the White House (the positive information source) clarified the rumor to the public, they definitely believed the White House regardless of the rumor. In this case, we say people making ***optimistic*** choices upon their receiving. On the contrary, people can also make ***pessimistic*** choices if they absolutely believe negative information.

Technically, if OSN users receive both kinds of information and make optimistic choices, they will believe the positive information regardless of the negative one. Given the probabilities of people believing positive information (a) and negation information (b), we have $0 < a < 1, b = 0$. We let $a + b < 1$ since they can contradict both kinds of information like "there was an explosion in White House but Obama was not injured". Similarly, we have $a = 0, 0 < b < 1, a + b < 1$ for people making pessimistic choices.

2) "*Coup d'etat in Tunisia* (January 11, 2011)": The Arab world experienced a series of revolutions and power changes over the last few years. It started from Tunisia and OSNs like Facebook and Twitter had played an important

- *S. Wen, M. Haghighi, C. Chen, Y. Xiang and W.L. Zhou are with the School of Information Technology, Deakin University, 221 Burwood Hwy, Melbourne, VIC 3125, Australia. E-mail: {wsheng, m.sayadhaghighi, zvm, yang, wanlei}@deakin.edu.au.*
- *W. Jia is with Shanghai Jiaotong University, Shanghai, P.R. China. E-mail: weijiaj@gmail.com.*

Manuscript received 15 July 2013; revised 21 Nov. 2013; accepted 4 Dec. 2013. Date of publication 15 Jan. 2014; date of current version 11 Feb. 2015. Recommended for acceptance by I. Stojmenovic.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TC.2013.2295802

role in it. For example, a rumor from tweeters went round that the army has seized power and ousted the Tunisia president. The rumor was swiftly 'retweeted' by people but the coup story was later suggested to be untrue by Egyptian Chronicles since there had been no confirmation from the government [5].

People preferred to believe the wrong news with the expectation of this rumor to be true [6]. They were happy if their president was out of power. In this case, we say people making ***preferable*** choices on both kinds of competitive information.

Technically, people make choices according to their preference. If people prefer positive information, we have $0 < b < a < 1, a + b < 1$. On the contrary, if people prefer negative information, we have $0 < a < b < 1, a + b < 1$.

3) "*R.I.P Jackie Chan Dead* (June 19 2013)": Most recently, the action star Jackie Chan was reported to be dead in Facebook sending thousands of his devout fans into shock. The rumor spread even when some said he is still alive. The hoax finally stopped when Jackie Chan posted to Facebook a photo of himself with a newspaper [7]. In this case, we say people making ***alternative*** choices before Jackie Chan himself dispelled the hoax.

Technically, people making alternative choices is people answering "Yes-or-No" questions. People must take one side. They cannot say jackie Chan is neither dead nor alive. If people believe Jackie Chan has died (negative news) with probability a, there must be a probability b that people believe he is still alive (positive news) and we have $0 < a, b < 1, a + b = 1$. This is different from previous two cases where people may possibly contradict both kinds of information.

1.2 Motivation

A realistic propagation model for social networks shall take both the social and the digital aspects of these media into account. For example, hearing a rumor, some might believe it but some might not. In addition, due to the behavioral differences, some might keep silent, but some others might actively contribute to its spread. Also, there are certain delays in checking new digital messages and forwarding them which is specific to OSNs.

In previous works, the independent cascade model (ICM) [8], [9], [10], [11], [12], [13], [14], [15] and the linear threshold model (LTM) [14], [15], [16], [17] are two primary models for the propagation of both positive and negative information in OSNs. ICM is basically a simulation model. LTM provides deterministic spread process, but each node in LTM is either absolutely 'active' or 'inactive'. Thus, LTM is also more close to a simulation model rather than an analytical one. In simulation, it is possible to find the probability of being in a state by averaging over many runs, but this does not express the reasons why an initial set of parameters result in such results. Moreover, people in ICM and LTM are limited to two basic states of believing either positive or negative information. This is far from being enough to represent social behavioral differences to which we have referred before. Additionally, ICM and LTM family of models do not take temporal dynamics into account. These include the frequency people check social news with and the time they take for them to forward the information. Thus, their results may largely deviate from the real spreading dynamics in OSNs.

There are some other models discussing the propagation of single-type information [18], [19], [20], [21], [22], [23], [24]. However, those works are incapable of capturing the phenomena happening in the presence of contradictory information. This is because the model needs to present the process of people making choices if they receive both kinds of information.

In fact, propagation studies, such as modeling and parameter analysis, are fundamental to the research in this field. It is mandatory to provide an accurate analytical model before we convincingly investigate the way to control the spread of both positive and negative information. As far as we know, the work in this paper is the first to propose an analytical model and analysis discussing about the propagation of both positive and negative information in OSNs.

1.3 Contributions

The primary contributions of this paper are summarized as follows:

- We proposed an analytical model on the propagation of positive and negative information. This model presents both the propagation dynamics and the behaviors of people making choices when they receive both kinds of information.
- We carried out a series of experiments to evaluate the accuracy of our proposed model. The experiments were based on two real OSNs: Facebook and Google Plus. The results showed that our proposed analytical model are quite accurate compared with simulations.
- On the basis of the analytical model, we further studied the parameter impacts on the spreading dynamics. This part of analysis well supports the tactics of restraining negative information by spreading positive information.

The rest of this paper is organized as follows: Section 2 introduces the skeleton of the analytical model. In Section 3, we explain and model the processes of people making choices when they receive both kinds of information. Section 4 is the accuracy evaluation followed by comparisons with previous models in Section 5. We present the studies of parameter impacts in Section 6. Related works and conclusion will be presented finally in Sections 7 and 8 respectively.

2 Propagation Modeling Primer

2.1 Modeling Nodes, Topology and Social Events

Nodes and topology properties are basic elements for the propagation of OSN information. Given an OSN, we derive the topology of it. A node in the topology denotes a user in the OSN. In the real world, people may believe positive information, negative information or have not heard of the information yet. Let random variable $X_i(t)$ represent the state of node i at discrete time t. We borrow the concepts from epidemics and derive the values of $X_i(t)$ as follows:

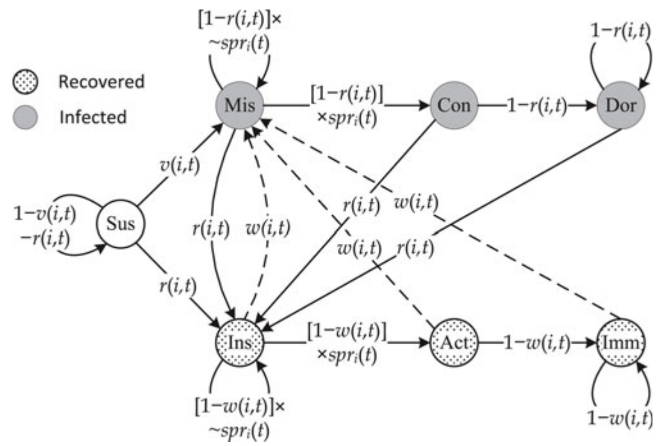


Fig. 1. State transition graph of a node in the topology.

$$X_i(t) = \begin{cases} \textbf{\textit{Sus.}}, & \textit{susceptible}, \\ \textbf{\textit{Rec.}}, & \textit{recovered}\begin{cases} \textbf{\textit{Ins.}}, & \textit{insider}, \\ \textbf{\textit{Act.}}, & \textit{active}, \\ \textbf{\textit{Imm.}}, & \textit{immunized}, \end{cases} \\ \textbf{\textit{Inf.}}, & \textit{infected}\begin{cases} \textbf{\textit{Mis.}}, & \textit{misled}, \\ \textbf{\textit{Con.}}, & \textit{contagious}, \\ \textbf{\textit{Dor.}}, & \textit{dormant}. \end{cases} \end{cases} \tag{1}$$

Every user is initially susceptible ($X_i(0) = Sus.$). During the spread, user i believes the positive information if $X_i(t) = Rec.$ and the negative information if $X_i(t) = Inf.$ We will further explain the child states $\{Ins., Mis., Act., Con., Imm., Dor.\}$ in Section 2.2. We have introduced more states to the model compared with previous works [25], [26].

Second, we propose employing an $m \times m$ square matrix with elements η_{ij} to describe the topology of an OSN with m nodes, as in

$$\begin{pmatrix} \eta_{11} & \cdots & \eta_{1m} \\ \vdots & \eta_{ij} & \vdots \\ \eta_{m1} & \cdots & \eta_{mm} \end{pmatrix} \eta_{ij} \in [0, 1], \tag{2}$$

wherein η_{ij} represents the probability of information spreading from user i to user j, including the probability of user i forwarding information to user j (p_{ij}) and the probability of user j believing it (q_{ij}). Therefore, we generally have $\eta_{ij} = p_{ij} \times q_{ij}$.

Third, we introduce two indicators $open_i(t)$ and $spr_i(t)$ to represent the events of OSN users checking newly arrived information and forwarding it to their social neighbors if they are willing to do that. In the real world, most people may not stay online in OSN all over the day. They will not receive information and forward it to others instantly. Therefore, we let $open_i(t) = 1$ if users read new information at time t. Otherwise, we let $open_i(t) = 0$. Similarly, we have $spr_i(t) = 1$ when users spread the information but 0 if they decide not to do so. Note that $\sim open_i(t)$ and $\sim spr_i(t)$ are the negations of $open_i(t)$ and $spr_i(t)$.

2.2 Susceptible-X-X (SXX) and Susceptible-X-Recovered (SXR)

We derive the state transition graph of an arbitrary node in OSN. As shown in Fig. 1, a node enters the *misled* or *insider* state when the user checks receivings ($open_i(t) = 1$) and believes negative ($Mis.$) or positive information ($Ins.$). This node then becomes contagious or active if the user is willing to forward the information to social neighbors ($spr_i(t) = 1$). After that, user stays in the *dormant* or *immunized* state until being infected or recovered. We use $v(i, t)$ and $r(i, t)$ to denote the probability of user i being infected or recovered. Note that people spread information only when they are contagious or active.

In this paper, we propose using the *Susceptible-X-X* and the *Susceptible-X-Recovered* to describe the information propagation in OSNs. For SXX, people are originally susceptible to both kinds of information. They will then switch between the states of believing positive ($Rec.$) or negative information ($Inf.$). In contrast, people of SXR will not believe the negative information any more after they accept the positive information once. To facilitate the modeling, we introduce $w(i, t)$ as the probability of user i being infected from the *recovered* state. Then, depending on the propagation of SXX or SXR, we have

$$w(i, t) = \begin{cases} v(i, t), & \textit{for}\ \text{SXX}, \\ 0, & \textit{for}\ \text{SXR}. \end{cases} \tag{3}$$

2.3 Modeling Propagation Dynamics

Given a topology of an OSN with m nodes, we can estimate the number of susceptible, infected and recovered users at time t, $S(t)$, $I(t)$ and $R(t)$, as in

$$\begin{cases} S(t) = m - I(t) - R(t), \\ I(t) = \sum_{i=1}^{m} P(X_i(t) = Inf.), \\ R(t) = \sum_{i=1}^{m} P(X_i(t) = Rec.). \end{cases} \tag{4}$$

Proof. Take $I(t)$ for example. We use value 1 to substitute the *infected* state and value 0 as the states excluding the infected state.Then, we have

$$\begin{aligned} E[X_i(t)] &= P(X_i(t) = 1) \times 1 + P(X_i(t) = 0) \times 0 \\ &= P(X_i(t) = Inf.). \end{aligned} \tag{5}$$

In probability theory, we generally have the identity $E[\sum_{i=1}^{m} X_i(t)] = \sum_{i=1}^{m} E[X_i(t)]$. Thus, we can easily derive the following:

$$I(t) = E\left[\sum_{i=1}^{m} X_i(t)\right] = \sum_{i=1}^{m} P(X_i(t) = Inf.). \tag{6}$$

Similarly, we can derive the calculation of $R(t)$ and obtain $S(t) = m - I(t) - R(t)$. □

As shown in Fig. 1, a susceptible user may believe the negative information and the node enters the *infected* state. An infected node may also be recovered if this user accepts the positive information. Based on the state transition graph (Fig. 1), we can easily iterate the value of $P(X_i(t) = Inf.)$ using the discrete difference equation as in

TABLE 1
Major Notations Used in This Paper

Symbol	Explanation
$v(i,t)$	Probability of user i believing negative information.
$r(i,t)$	Probability of user i believing positive information.
$u(i,t)$	Probability of user i refuting both kinds of information.
$w(i,t)$	Probability of recovered user i being misled.
$X_i(t)$	The state of a network node i at time t.
p_{ij}	Probability of user i forwarding information to user j.
q_{ij}	Probability of user i believing information from user j.
η_{ij}	Probability of information spreading from user i to user j.
$open_i(t)$	The indicator of user i checking new information at time t.
$spr_i(t)$	The indicator of user i forwarding information at time t.
τ	The arbitrary time between user i last checking new information and the current time t (excluding t).
$I(t)$	The number of infected users in networks at time t.
$R(t)$	The number of recovered users in networks at time t.
$S(t)$	The number of susceptible users in networks at time t.
m	The size of users in the network.
$Pos(i,t)$	Probability of user i ***not*** affected by positive information.
$Neg(i,t)$	Probability of user i ***not*** affected by negative information.
$C^{pos}(i,t)$	The number of positive copies that user i believes.
$C^{neg}(i,t)$	The number of negative copies that user i believes.

$$P(X_i(t)=Inf.) = v(i,t)\cdot P(X_i(t-1)=Sus.) + w(i,t)\cdot P(X_i(t-1)=Rec.) + [1-r(i,t)]\cdot P(X_i(t-1)=Inf.). \tag{7}$$

A node will stay in the *susceptible* state if the user has not been infected or recovered. Thus, we can also iterate the value of $P(X_i(t)=Sus.)$ and further derive the value of $P(X_i(t)=Rec.)$ as in

$$P(X_i(t)=Sus.) = [1-v(i,t)-r(i,t)]\cdot P(X_i(t-1)=Sus.), \tag{8}$$

$$P(X_i(t)=Rec.) = 1 - P(X_i(t)=Sus.) - P(X_i(t)=Inf.). \tag{9}$$

We adopt discrete time to model the propagation dynamics. The length of each time tick relies on the real environment. It can be 1 minute, 1 hour or one day. For the convenience of readers, we list the major variables in Table 1.

3 User Making Choice on Information

Depending on the ways user believes OSN messages, we can derive different values of $r(i,t)$ and $v(i,t)$.

3.1 Optimistic or Pessimistic Choices

First, we consider the case that people are optimistic or pessimistic to the received information. If people are optimistic, OSN users will absolutely believe the positive information except they only receive negative ones. In contrast, people will absolutely believe the negative information if they are pessimistic.

In order to calculate the values of $r(i,t)$ and $v(i,t)$, a temporal variable τ is required to represent the arbitrary time after users last check new information. In the modeling, we need to estimate the number of unread information on each user at current time t. However, these new pieces of information may be forwarded to users at any time after users last log in OSN. As shown in Fig. 2, we let $\tau \in [t', t)$. This can help us accumulate the number of unread information (excluding the ones arrived at current time t).

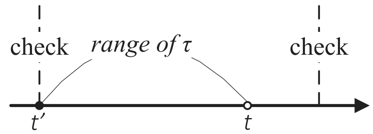


Fig. 2. The value range of the arbitrary time τ for each user. t' denotes the time of user last checking new information.

We introduce $Pos(i,t)$ and $Neg(i,t)$ to be the probability of user i *not* believing positive or negative information. We can derive $Pos(i,t)$ by assuming all social neighbors cannot convince user i of positive information. Then, according to the principle of multiplication, we have

$$Pos(i,t) = \prod_{j\in N_i}[1-\eta_{ji}\cdot P(X_j(\tau)=Act.|X_i(t)\neq Rec.)] \tag{10}$$

$$\overset{Markov}{\approx} \prod_{j\in N_i}\left[1-\eta_{ji}\cdot P^{\bar{i}}(X_j(\tau)=Act.)\right] \tag{11}$$

$$= (\sim open_i(t-1))\cdot Pos(i,t-1) \cdot \prod_{j\in N_i}\left[1-\eta_{ji}\cdot P^{\bar{i}}(X_j(t-1)=Act.)\right]. \tag{12}$$

We can also derive $Neg(i,t)$ by assuming user i refute all negative information from social neighbors

$$Neg(i,t) = \prod_{j\in N_i}[1-\eta_{ji}\cdot P(X_j(\tau)=Con.|X_i(t)\neq Inf.)] \tag{13}$$

$$\overset{Markov}{\approx} \prod_{j\in N_i}\left[1-\eta_{ji}\cdot P^{\bar{i}}(X_j(\tau)=Con.)\right] \tag{14}$$

$$= (\sim open_i(t-1))\cdot Neg(i,t-1) \cdot \prod_{j\in N_i}\left[1-\eta_{ji}\cdot P^{\bar{i}}(X_j(t-1)=Con.)\right]. \tag{15}$$

Take equations (10)-(12) for example to explain the above derivations. *First*, user j will spread positive information to user i if user i is susceptible or has believed the negative one. This is presented by the conditional probability $P(X_j(\tau)=Act.|X_i(t)\neq Rec.)$. However, its value is computationally too expensive to obtain, especially when the size of neighborhood is large [25], [26]. For example, if node i has k neighbors, the total number of states needed to calculate the probability is $O(2kt)$. Chen's work [25] suggests that the value using Markov approximation will be accurate enough if η_{ij} is large. Since we mainly focus on the propagation of critical information ($\eta_{ij} > 0.5$), we adopt Markov approximation in equations (11) and (14). In this paper, we use $P^{\bar{i}}(X_j(t-1)=Act.)$ and $P^{\bar{i}}(X_j(t-1)=Con.)$ to denote the approximate values. Readers could refer to Section 3.4 for detailed calculation. *Second*, τ is a temporary temporal variable which help us accumulate the number of unread pieces of information received by each user. We then relax this temporary variable by iteration of equations (11)-(12) and (14)-(15). Readers could refer to our previous work ([26, Section 4.2]) for details.

If people are optimistic, we have the probability of being infected $v(i,t)$ or recovered $r(i,t)$ as

$$\begin{cases} v(i,t) = Pos(i,t) \cdot [1 - Neg(i,t)] \cdot open_i(t), \\ r(i,t) = [1 - Pos(i,t)] \cdot open_i(t). \end{cases} \tag{16}$$

Similarly, when people are pessimistic, we can also derive the following as

$$\begin{cases} v(i,t) = [1 - Neg(i,t)] \cdot open_i(t), \\ r(i,t) = Neg(i,t) \cdot [1 - Pos(i,t)] \cdot open_i(t). \end{cases} \tag{17}$$

In this case, when $Pos(i,t) = Neg(i,t) = 1$, we have $v(i,t) = r(i,t) = 0$. When $Pos(i,t) \times Neg(i,t) \neq 1$, user i receives information from social neighbors. We can distribute the probabilities that people choose either positive or negative information according to the ratio of $C^{pos}(i,t)$ and $C^{neg}(i,t)$ as in

$$r(i,t) = \frac{(1 - Pos(i,t) \cdot Neg(i,t)) \cdot C^{pos}(i,t)}{C^{pos}(i,t) + C^{neg}(i,t)} \cdot open_i(t), \tag{24}$$

$$v(i,t) = \frac{(1 - Pos(i,t) \cdot Neg(i,t)) \cdot C^{neg}(i,t)}{C^{pos}(i,t) + C^{neg}(i,t)} \cdot open_i(t). \tag{25}$$

3.2 Preferable Choices

When people receive both kinds of information, they will believe either positive or negative information according to how much either can be trusted. We introduce α as the preference of people and η_{ij}^{pre} as the biased probability of user i receiving and believing preferable information from user j. Then, we have

$$\eta_{ij}^{pre} = p_{ij} \cdot [q_{ij} + \alpha \cdot (1 - q_{ij})]. \tag{18}$$

In accordance, we use β as the resistence of people and η_{ij}^{dis} as the biased probability of user i receiving and believing disliked information from user j

$$\eta_{ij}^{dis} = p_{ij} \cdot q_{ij} \cdot (1 - \beta). \tag{19}$$

We assume α ($1 > \alpha > 0$) and β ($1 > \beta > 0$) are independent to each other. Given a preference (α) and an resistence (β), we have $\eta_{ij}^{pre} > \eta_{ij} > \eta_{ij}^{dis}$.

Taking people preferring positive information for example, people have no bias on making choices if $\alpha, \beta = 0$. When $\alpha, \beta > 0$, we can compute $Pos(i,t)$ and $Neg(i,t)$ by replacing η_{ij} with η_{ij}^{pre} and η_{ij}^{dis} in equations (12) and (15). We can also estimate the number of positive and negative information copies that user i believes ($C^{pos}(i,t)$ and $C^{neg}(i,t)$) as in

$$C^{pos}(i,t) = \sum_{j \in N_i} \left[\eta_{ji}^{pre} \cdot P(X_j(\tau) = Act. | X_i(t) \neq Rec.)\right] \tag{20}$$

$$\begin{aligned} &\approx (\sim open_i(t-1)) \cdot C^{pos}(i,t-1) \\ &\quad + \sum_{j \in N_i} \left[\eta_{ji}^{pre} \cdot P^{\bar{i}}(X_j(t-1) = Act.)\right], \end{aligned} \tag{21}$$

$$C^{neg}(i,t) = \sum_{j \in N_i} \left[\eta_{ji}^{dis} \cdot P(X_j(\tau) = Con. | X_i(t) \neq Inf.)\right] \tag{22}$$

$$\begin{aligned} &\approx (\sim open_i(t-1)) \cdot C^{neg}(i,t-1) \\ &\quad + \sum_{j \in N_i} \left[\eta_{ji}^{dis} \cdot P^{\bar{i}}(X_j(t-1) = Con.)\right]. \end{aligned} \tag{23}$$

The derivation of $C^{pos}(i,t)$ and $C^{neg}(i,t)$ is similar to equations (10)-(12) and (13)-(15) in Section 3.1.

3.3 Alternative Choices

When people are making alternative choices, people are actually doing "Yes-or-No" questions. They have to accept either positive or negative information, but cannot refute both of them. We say an arbitrary user i believing positive information if this user accepts positive information or refuses to accept negative information. We can also say user i believing negative information if this user accepts negative information or reject positive information. Thus, we can estimate the values of $C^{pos}(i,t)$ and $C^{neg}(i,t)$ as in

$$\begin{aligned} C^{pos}(i,t) = &(\sim open_i(t-1)) \cdot C^{pos}(i,t-1) \\ &+ \sum_{j \in N_i} \left[p_{ji} \cdot q_{ji} \cdot P^{\bar{i}}(X_j(t-1) = Act.)\right] \\ &+ \sum_{j \in N_i} \left[p_{ji} \cdot (1 - q_{ji}) \cdot P^{\bar{i}}(X_j(t-1) = Con.)\right], \end{aligned} \tag{26}$$

$$\begin{aligned} C^{neg}(i,t) = &(\sim open_i(t-1)) \cdot C^{neg}(i,t-1) \\ &+ \sum_{j \in N_i} \left[p_{ji} \cdot q_{ji} \cdot P^{\bar{i}}(X_j(t-1) = Con.)\right] \\ &+ \sum_{j \in N_i} \left[p_{ji} \cdot (1 - q_{ji}) \cdot P^{\bar{i}}(X_j(t-1) = Act.)\right]. \end{aligned} \tag{27}$$

Similar to $Pos(i,t)$ and $Neg(i,t)$, we compute the probability of user i *not* receiving any positive or negative information from social neighbors at time t ($\Upsilon^{pos}(i,t)$ and $\Upsilon^{neg}(i,t)$) as in

$$\begin{aligned} \Upsilon^{pos}(i,t) = &(\sim open_i(t-1)) \cdot \Upsilon^{pos}(i,t-1) \\ &\cdot \prod_{j \in N_i} \left[1 - p_{ji} \cdot P^{\bar{i}}(X_j(t-1) = Act.)\right], \end{aligned} \tag{28}$$

$$\begin{aligned} \Upsilon^{neg}(i,t) = &(\sim open_i(t-1)) \cdot \Upsilon^{neg}(i,t-1) \\ &\cdot \prod_{j \in N_i} \left[1 - p_{ji} \cdot P^{\bar{i}}(X_j(t-1) = Con.)\right]. \end{aligned} \tag{29}$$

The value of $1 - \Upsilon^{pos}(i,t) \cdot \Upsilon^{neg}(i,t)$ is the probability that user i has received positive, negative or both kinds of

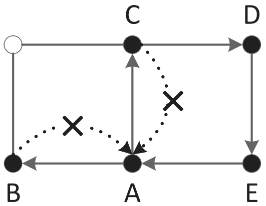


Fig. 3. The illustration of Markov approximation.

information. When $\Upsilon^{pos}(i,t) \cdot \Upsilon^{neg}(i,t) = 1$, user i does not receive any information from social neighbors. According to the ratio of $C^{pos}(i,t)$ and $C^{neg}(i,t)$, we then compute $r(i,t)$ and $v(i,t)$ as in

$$r(i,t) = \frac{(1-\Upsilon^{pos}(i,t) \cdot \Upsilon^{neg}(i,t)) \cdot C^{pos}(i,t)}{C^{pos}(i,t) + C^{neg}(i,t)} \cdot open_i(t), \quad (30)$$

$$v(i,t) = \frac{(1-\Upsilon^{pos}(i,t) \cdot \Upsilon^{neg}(i,t)) \cdot C^{neg}(i,t)}{C^{pos}(i,t) + C^{neg}(i,t)} \cdot open_i(t). \quad (31)$$

3.4 Markov Approximation

In our model, we use Markov approximation in the derivations (equations (11), (14), (21), (23), (26), (27), (28) and (29)). Given a simple example in Fig. 3, node A spreads information to node B and C. Node C further affects D, E and back to A. In Markov approximation, the modeling does not allow node B and C to spread information to node A reversely, but it admits the overestimation from D, E back to A.

First, we introduce the probability of user i not believing the positive information from social neighbors except the neighbor x, $Pos^{\bar{x}}(i,t)$. Similarly, we introduce $Neg^{\bar{x}}(i,t)$, and then we have

$$Pos^{\bar{x}}(i,t) = \prod_{\substack{x,j \in N_i \\ x \neq j}} [1 - \eta_{ji} \cdot P^{\bar{i}}(X_j(\tau) = Act.)], \quad (32)$$

$$Neg^{\bar{x}}(i,t) = \prod_{\substack{x,j \in N_i \\ x \neq j}} [1 - \eta_{ji} \cdot P^{\bar{i}}(X_j(\tau) = Con.)]. \quad (33)$$

Using equations (16), (17), (24) and (25), we can easily obtain the probability of user i being infected or recovered by social neighbors except the neighbor x ($v^{\bar{x}}(i,t)$, $r^{\bar{x}}(i,t)$ and $w^{\bar{x}}(i,t)$). Based on the state transition graph (Fig. 1), we can derive the followings:

$$P^{\bar{x}}(X_i(t) = Act.) = spr_i(t) \cdot \left[1 - w^{\bar{x}}(i,t)\right] \cdot P^{\bar{x}}(X_i(t-1) = Ins.), \quad (34)$$

$$P^{\bar{x}}(X_i(t) = Con.) = spr_i(t) \cdot \left[1 - r^{\bar{x}}(i,t)\right] \cdot P^{\bar{x}}(X_i(t-1) = Mis.). \quad (35)$$

Since the value of $w^{\bar{x}}(i,t)$ is equal to $v^{\bar{x}}(i,t)$ for SXX and 0 for SXR (refer to equation (3)), we can compute the value of $P^{\bar{x}}(X_i(t) = Mis.)$ and $P^{\bar{x}}(X_i(t) = Ins.)$ for SXX as in

TABLE 2
Basic Properties of the Network Topologies

	Facebook [1]	Google Plus [2]
Number of nodes	45814	264004
Number of links	4693129	47130325
Average degree	5.76	10.04
Max outdegree	199	5739
Max indegree	157	3063

$$\begin{aligned} P^{\bar{x}}(X_i(t) = Mis.) = {} & \left[1 - P(X_i(t-1) = Inf.)\right] \cdot v^{\bar{x}}(i,t) \\ & + \left[1 - r^{\bar{x}}(i,t)\right] \cdot [1 - spr_i(t)] \\ & \cdot P^{\bar{x}}(X_i(t-1) = Mis.), \end{aligned} \quad (36)$$

$$\begin{aligned} P^{\bar{x}}(X_i(t) = Ins.) = {} & \left[1 - P(X_i(t-1) = Rec.)\right] \cdot r^{\bar{x}}(i,t) \\ & + \left[1 - v^{\bar{x}}(i,t)\right] \cdot [1 - spr_i(t)] \\ & \cdot P^{\bar{x}}(X_i(t-1) = Ins.). \end{aligned} \quad (37)$$

Similarly for SXR, we can also derive the followings:

$$\begin{aligned} P^{\bar{x}}(X_i(t) = Mis.) = {} & P(X_i(t-1) = Sus.) \cdot v^{\bar{x}}(i,t) \\ & + \left[1 - r^{\bar{x}}(i,t)\right] \cdot [1 - spr_i(t)] \\ & \cdot P^{\bar{x}}(X_i(t-1) = Mis.), \end{aligned} \quad (38)$$

$$\begin{aligned} P^{\bar{x}}(X_i(t) = Ins.) = {} & \left[1 - P(X_i(t-1) = Rec.)\right] \\ & \cdot r^{\bar{x}}(i,t) + [1 - spr_i(t)] \\ & \cdot P^{\bar{x}}(X_i(t-1) = Ins.). \end{aligned} \quad (39)$$

Equations (34)-(39) have provided an iteration mechanism to compute the values of $P^{\bar{x}}(X_i(t) = Mis.)$ and $P^{\bar{x}}(X_i(t) = Ins.)$. Give an arbitrary user i with k neighbors, we can see that the complexity has largely decreased by only keeping $2 \times k$ states.

4 Correctness Investigation

In this field, there are no real traces of both positive and negative information spreading in popular OSNs. All the existing research, such as [8], [9], [11], [25], [26], adopts simulation to evaluate analytical models. In order to evaluate the accuracy of our proposed model, we run the modeling and simulations on two real OSNs: Facebook [1] and Google Plus [2], [27]. We mainly focus on the critical information in our modeling. Since the real critical information, such as widespread rumors and official announcements, generally spreads from popular or highly authorized sources [4], we start the modeling and simulations from two highly-connected nodes in the networks. The spread of the two kinds of information will start at different time. We introduce t_{inject} to denote the delay of the second kind of information.

All the experiments were conducted on a server running Microsoft Windows Server 2008 with eight CPUs and 32 G memory. The implementation was done in C++ and Matlab2012. The random numbers are produced by the C++ TR1 library extensions. The simulation results are averaged over 100 runs. The number of 100 comes

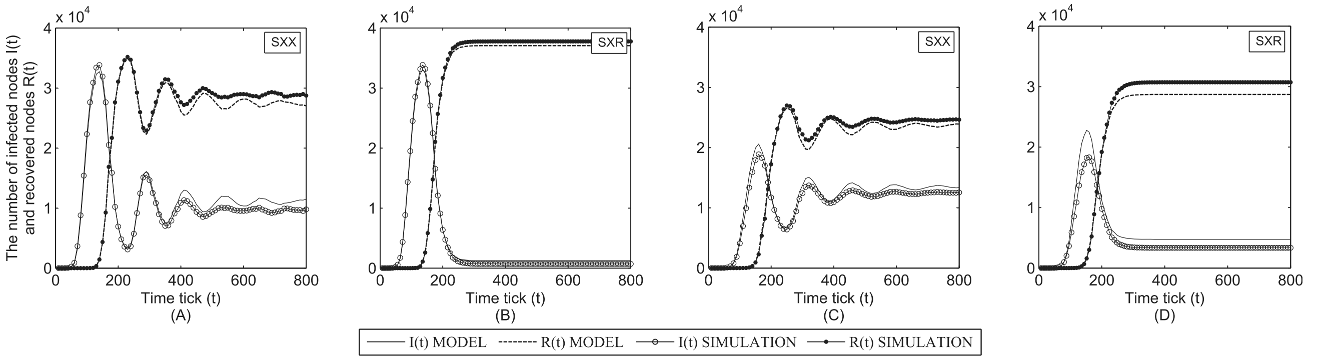


Fig. 4. Empirical proofs of the modeling accuracy. General settings: 1) Facebook; 2) Optimistic choices; 3) $t_{inject} = 100$. Specific settings: (A) $E(\eta_{ij}) = 0.9$, SXX; (B) $E(\eta_{ij}) = 0.9$, SXR; (C) $E(\eta_{ij}) = 0.6$, SXX; (D) $E(\eta_{ij}) = 0.6$, SXR.

from the discussion in [28]. We choose typical parameters to validate the accuracy but leave the analysis of parameter impact to Section 6. The basic properties of the two tested topologies are listed in Table 2.

4.1 Evaluate Optimistic or Pessimistic Choices

First, we evaluate people making optimistic or pessimistic choices. Due to symmetry of the model, we only take optimistic case as the example (refer to Section 6.1 for details).

1) *Simulation*. Given an infected user i, information is forwarded to neighbors by comparing a random number with η_{ij}. Recall that $\eta_{ij} = p_{ij} \cdot q_{ij}$. Thus, once the delivery succeeds, user j will receive and believe this piece of information. A user will not move into the *inf.* state if he obtains at least one piece of positive information. $I(t)$ and $R(t)$ are obtained by counting the infected and recovered users in the network.

2) *Settings*. We assume user i check new information every T_i time ticks and forward messages every F_i time ticks ($T_i, F_i \sim N(20, 10)$, refer to Section 6.5 for details). The positive information will be injected into the networks at time 100 ($T_{inject} = 100$, refer to Section 6.2 for details). The accuracy of SXX and SXR will be examined by setting $E(\eta_{ij}) = 0.9$ or 0.6.

3) *Results*. The Facebook results are shown in Fig. 4. We can see that our modeling results are quite close to the simulations. The error in Fig. 4D is a bit large, but we still have ($error < 10\% \times I(t)$). We then examine the accuracy in the Google Plus network. As shown in Fig. 5, our modeling results are also very accurate.

4) *Analysis*. For SXX model, people can change their original state by believing the opposite kind of information. Thus, users may sway between two kinds of information. That is the reason why we can see many oscillations in the SXX results. For SXR model, people recover and will not believe the negative information again. Thus, $I(t)$ decreases fast in the SXR results.

4.2 Evaluate Preferable Choices

Second, we evaluate the case of people making preferable choices. We choose the typical values of α and β for the evaluation. For the impact of these two variables, please refer to Section 6.3 for details.

1) *Simulation*. Given an recovered user i, positive information is forwarded to neighbors according to the value of η_{ij}^{pre}. A random number is compared with η_{ij}^{pre} to see if the delivery succeeds or not. The same happens on negative information with η_{ij}^{dis}. Once the delivery succeeds, user j believes this piece of information. The final decision depends on the ratio of positive and negative information copies.

2) *Settings*. We still assume $T_i, F_i \sim N(20, 10)$ and $T_{inject} = 100$. To be generic, we set $\alpha, \beta = 10\%$ and $\eta_{ij} = 0.75$. Thus, we have $E(\eta_{ij}^{pre}) = 0.825$ and $E(\eta_{ij}^{dis}) = 0.675$.

3) *Results*. The Facebook results are shown in Figs. 6A and 6B. The Google Plus results are shown in Figs. 6C and 6D. All modeling results are very close to the simulations.

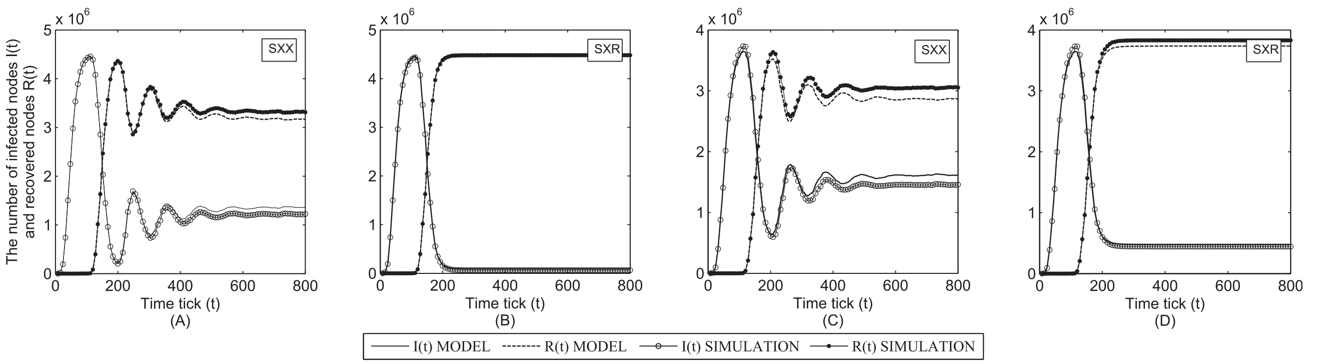


Fig. 5. Empirical proofs of the modeling accuracy. General settings: 1) Google+; 2) Optimistic choices; 3) $t_{inject} = 100$. Specific settings: (A) $E(\eta_{ij}) = 0.9$, SXX; (B) $E(\eta_{ij}) = 0.9$, SXR; (C) $E(\eta_{ij}) = 0.6$, SXX; (D) $E(\eta_{ij}) = 0.6$, SXR.

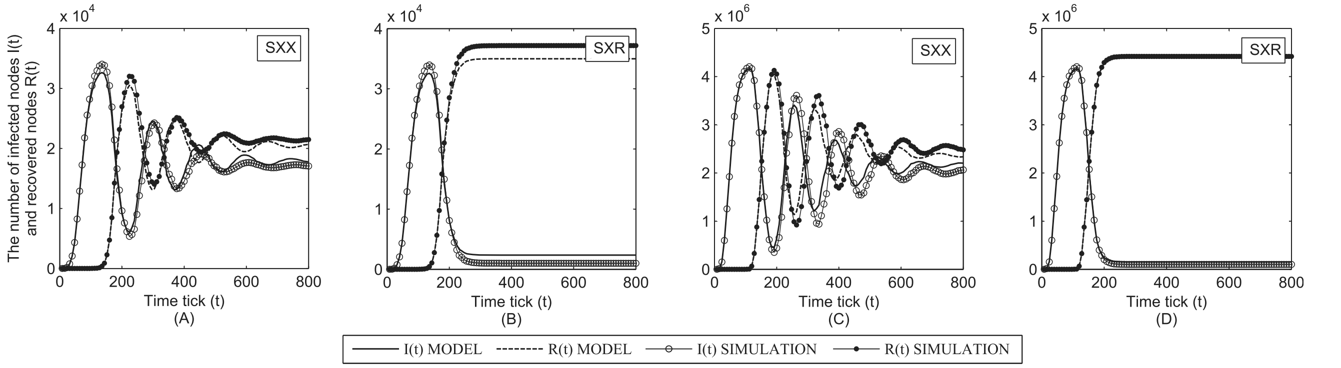


Fig. 6. Empirical proofs of the modeling accuracy. General settings: 1) $E(\eta_{ij}^{pre}) = 0.825$, $E(\eta_{ij}^{dis}) = 0.675$; 2) Preferable choices; 3) $t_{inject} = 100$. Specific settings: (A) Facebook, SXX; (B) Facebook, SXR; (C) Google+, SXX; (D) Google+, SXR.

The errors in Fig. 6B are a bit large, but it is still acceptable ($error < 10\% \times I(t)$).

4) *Analysis*. For people making preferable choices, users finally choose an information by the ratio of different information copies. They will not absolutely believe positive or negative information even when they have some preference (α) or resistence (β). Thus, we can see many strong oscillations and curve crosses in the SXX results. For SXR model, similar to the previous case, $I(t)$ drops quickly in the SXR results after we inject the positive information.

4.3 Evaluate Alternative Choices

Finally, we evaluate the accuracy when people make alternative choices. The impact of p_{ij} and q_{ij} will be discussed in Section 6.4.

1) *Simulation*. Given an infected user i, information is forwarded to neighbors according to the values of p_{ij}. Random numbers will be compared with q_{ij} to check if user j believes the positive information or the opposite negative one. The final decision depends on the ratio of believed copies of positive and negative information.

2) *Settings*. We still assume $T_i, F_i \sim N(20, 10)$ and $T_{inject} = 100$. To be generic, we set $p_{ij} = 0.75$ and $q_{ij} = 0.75$. Thus, we have $E(\eta_{ij}) = 0.56$.

3) *Results*. The Facebook results are shown in Figs. 7A and 7B. The Google Plus results are shown in Figs. 7C and 7D. All modeling results are quite accurate. Another fact in Fig. 7 is that the number of $I(t)$ and $R(t)$ in the SXX modeling are very close to each other. For the convenience of readers, we have zoomed the results in the inset figures.

4) *Analysis*. For people making alternative choices, a user refuting negative information means this user believes the opposite positive one. Thus, we can see $R(t)$ goes up with $I(t)$ before we inject positive information into the network. Users have to choose one kind of information (either positive or negative) with the probabilities q_{ij} or $1 - q_{ij}$. Thus, $I(t)$ and $R(t)$ are very close to each other in the SXX modeling. Similar to the previous cases, the number of $I(t)$ drops quickly in the SXR results.

5 Comparison with Previous Models

5.1 ICM and LTM

For the propagation of competitive information, the most basic and well-studied models are the independent cascade model and the linear threshold model. In this field, we find many deviations of these two models [8], [9], [10], [12], [13], [15], [16], [17] but the following ICM and LTM lie at the cores of most model variants.

ICM. ICM starts with an initial set of active nodes. The process unfolds in discrete steps according to the randomized rules: when an arbitrary node i first becomes active in step t, it is given a single chance to activate each of the currently inactive neighbors; it succeeds with the probability x_{ij} ($j \in N_i$); if user i has multiple newly activated neighbors, their attempts are sequenced in an arbitrary order. Once a node becomes active, it will remain active forever.

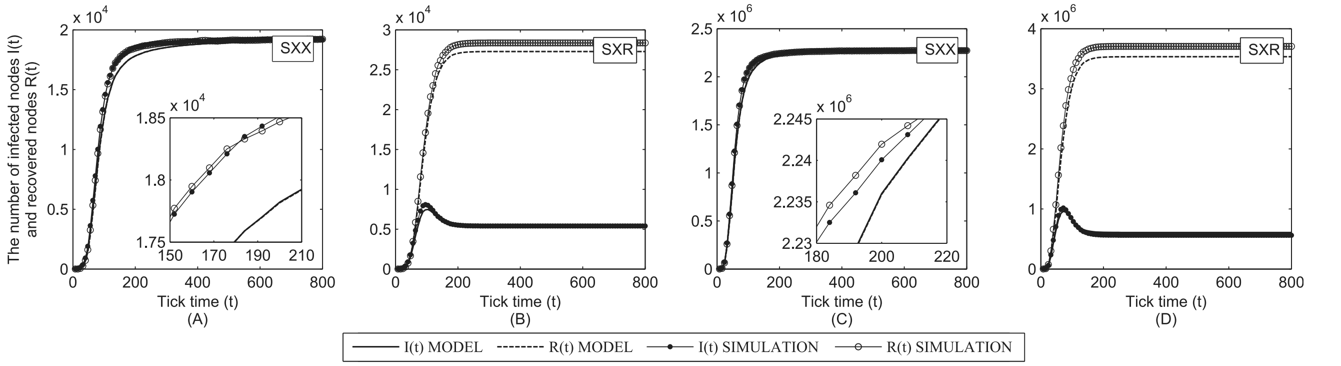


Fig. 7. Empirical proofs of the modeling accuracy. General settings: 1) $E(p_{ij}) = 0.75$, $E(q_{ij}) = 0.75$; 2) Alternative choices; 3) $t_{inject} = 100$. Specific settings: (A) Facebook, SXX; (B) Facebook, SXR; (C) Google+, SXX; (D) Google+, SXR.

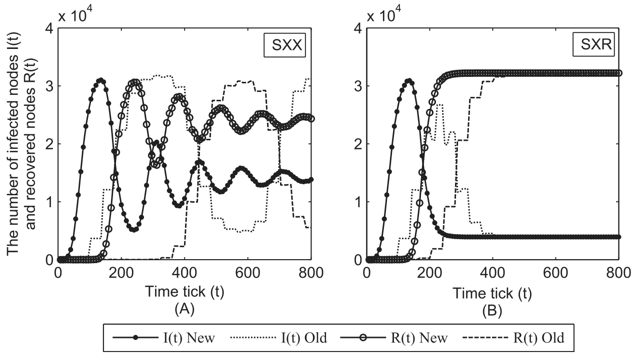


Fig. 8. Compare differences in temporal spread dynamics. "New": Our model; "Old": ICM and LTM. Settings: 1) Facebook; 2) Optimistic choices; 3) $E(\eta_{ij}) = 0.75$; 4) $t_{inject} = 100$.

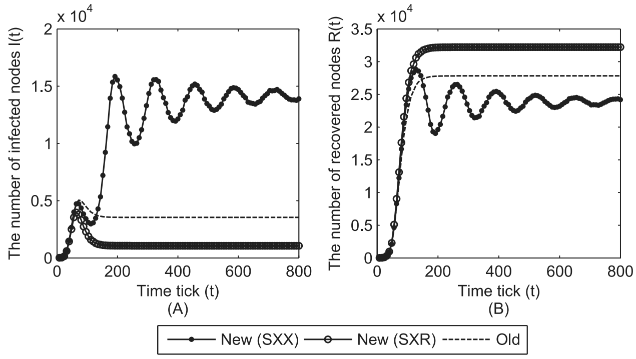


Fig. 9. Compare differences in state transition schema. "New": Our model; "Old": ICM and LTM. Settings: 1) Facebook; 2) Optimistic choices; 3)$E(\eta_{ij}) = 0.75$; 4) $t_{inject} = 0$.

LTM. Node i is influenced by neighbors according to the weight x_{ij}, ($\sum_{j \in N_i} x_{ij} \leqslant 1$). Given a threshold θ_i and an initial set of active nodes, the diffusion process unfolds deterministically in discrete steps. Node i is activated if the following two conditions are satisfied: 1) $\sum_{j \in N_i} x_{ij} \geqslant \theta_i$, 2) j is active. Once a node becomes active, similar to ICM, it remains active forever.

When a user receives two kinds of competitive information in ICM and LTM, the strategy adopted to make final decisions varies according to different environments. Some chose "optimistic or pessimistic" [8], [17]. Some chose "alternative" [10]. We can also see some adopted "game theory" [12], [15] and "first come first win" [13] to find out optimized strategies.

5.2 Superiority Analysis

Compared to ICM and LTM, our model provides an analytical way to present the propagation. We summarize the major differences as follows:

First, ICM is a simulation model. LTM provides deterministic spread process, but each node in LTM is absolutely active or inactive. Thus, LTM is more close to a simulation model rather than an analytical one. Researchers can derive the probability of being in either state for each node by averaging over many runs of simulation, but simulation models cannot quantify the reasons why initial parameters result in such probabilities and further disclose the essence.

Second, ICM and LTM are very basic models. We separately analyze two of the differences from our model. The experiments are carried out taking Facebook and people making optimistic choices as an example. First, when our model can present the processes of people checking and forwarding information randomly (e.g., $T_i, F_i \sim N(20, 10)$), ICM and LTM can only use equivalent constants for T_i and F_i (e.g., $T_i, F_i = 20$). As shown in Fig. 8, their results consequently show stair-like behaviors which are obviously not realistic in the real world. Second, we investigate the state transition processes. To avoid the impacts from temporal factors, both positive and negative information are injected at the beginning ($T_{inject} = 0$). We can see from Fig. 9 that their results largely deviate from our SXX and SXR results. In ICM and LTM, a node will remain infected or recovered till the spread ends. This assumption does not suit for the spread cases in real OSNs.

Third, we cannot quantify the superiority of our model in presenting people making choices because it is highly social environment related. Previous ICM and LTM were originally devoted to the marketing area and particle systems [14]. On the contrary, we derive the strategies based on the real information propagation in OSNs (refer to Section 1.1). Thus, our approach is more suitable to model the propagation of OSN competitive information.

6 Parameter Studies

Based on the analytical model, we further explore the impacts of different parameters to the propagation dynamics, including 1) optimistic and pessimistic; 2) T_{inject}; 3) α, β; 4) q_{ij}; 5) $open_i(t), spr_i(t)$.

6.1 Optimistic and Pessimistic

We investigate the differences of people making optimistic or pessimistic choices. Technically, if people receive both kinds of information and make optimistic choices, they will absolutely believe the positive information regardless the negative one. On the contrary, they will absolutely believe the negative information. In the experiments, to avoid the influence from other parameters, we set $T_{inject} = 0$, $T_i, F_i \sim N(20, 10)$ and $E(\eta_{ij}) = 0.75$. We run experiments on both Facebook and Google Plus using SXX and SXR mechanisms.

As shown in Figs. 10A and 10C, the SXX results of both strategies are symmetric. In the results of Figs. 10B and 10D, we find $I(t)$ of SXR drops faster if people make pessimistic choices. Both the optimistic and pessimistic strategies have so far behaved as one would expect it to. We further investigate the estimated number of contagious nodes ($\sum_i P(X_i(t) = Con.)$) and active nodes ($\sum_i P(X_i(t) = Act.)$). As shown in Fig. 11, $I(t)$ and $R(t)$ fluctuate in the propagation dynamics. In Figs. 11A and 11C, we introduce two ellipses. We find active nodes are more than contagious nodes if people make optimistic choices. Otherwise, contagious nodes will be more. However, when the propagation continues to the outside of the ellipses, contagious node and active nodes are comparable. We have similar results in Figs. 11B and 11D, but there will be no contagious nodes and active nodes after 200 time ticks more or less. From the results of Figs. 10 and 11, *we find that the propagation is mainly*

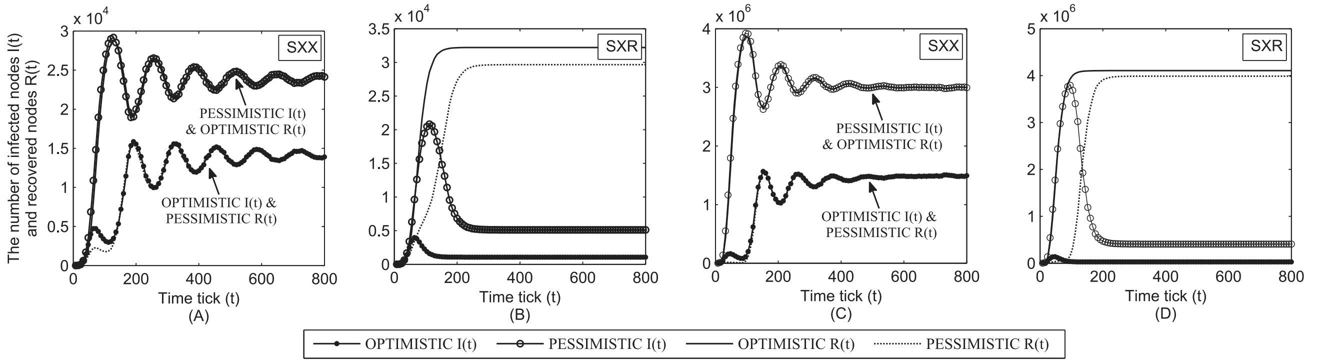


Fig. 10. The differences of people making optimistic and pessimistic choices. (A) Facebook; (B) Facebook; (C) Google+; (D) Google+. Settings: 1) $E(\eta_{ij}) = 0.75$; 2) $t_{inject} = 0$.

decided by the early spreading dynamics if people make optimistic or pessimistic choices on their receiving.

6.2 Impact of T_{inject}

In Section 4, we assume the positive information is injected into the network at time 100. However, the value of T_{inject} may considerably affect the spreading dynamics. To exclusively investigate the impact of T_{inject}, we set $T_i, F_i \sim N(20, 10)$, $E(\eta_{ij}) = 0.75$ and people making preferable choices ($\alpha = \beta = 0$). At this moment, the group of people who believe positive information is fair to the group of people who believe negative information. The experiments are run on both Facebook and Google Plus using SXX and SXR. We test T_{inject} at the values 0, 100 and 200.

We see two features according to the results in Fig. 12. First, the propagation under different settings will finally become steady even though there are oscillations in Figs. 12A and 12B. *The final results $(T(t), R(t))$ will be approximately equal to a constant*. Second, we can observe in Figs. 12A and 12C that *the spread of negative information reaches the largest scale at the early time stage*. This feature is the same to the results in Section 6.1.

The results inspire us something in the real world. We informally take the rumor "Barack Obama was born in Kenya" for example. During the campaign for president, Obama was questioned to be a native-born citizen. If not, under Article 2 of the US Constitution, he was ineligible to be President of the United States. In response to the rumor, Obama posted an image of his birth certificate. Based on our analysis, the time for the conspirator surfacing the rumor is an important issue. First, the number of people believing the rumor would be identical in the long term. Second, Obama needed time to collect evidences to clarify the rumor. If the conspirator spread the rumor a short time before the poll, Obama's opponents might possibly win more votes in the campaign.

6.3 Impact of α,β

The values of α and β decide the preference and the resistence of people when they make preferable choices. In order to exclusively investigate the impacts of α and β, we set $T_{inject} = 0$, $T_i, F_i \sim N(20, 10)$ and $p_{ij} = q_{ij} = 0.75$. We test α and β at values 25 and 50 percent. We can compute $\eta_{ij}^{pre} = 0.8325$ and $\eta_{ij}^{dis} = 0.6075$ for $\alpha = \beta = 25\%$. We also have $\eta_{ij}^{pre} = 0.855$ and $\eta_{ij}^{dis} = 0.405$ for $\alpha = \beta = 50\%$. Experiments are run in Facebook and Google Plus platforms with SXX and SXR spreading mechanisms.

As shown in Fig. 13, *the values of α and β have considerable impacts on the propagation dynamics*. Particularly for SXX, when α and β increase from 25 to 50 percent, the results of $I(t)$ and $R(t)$ deviate more from each other. For SXR, the impacts are a bit less. This is because the probability for people believing negative information largely decreases when people have resistence on them. As a result, the number of

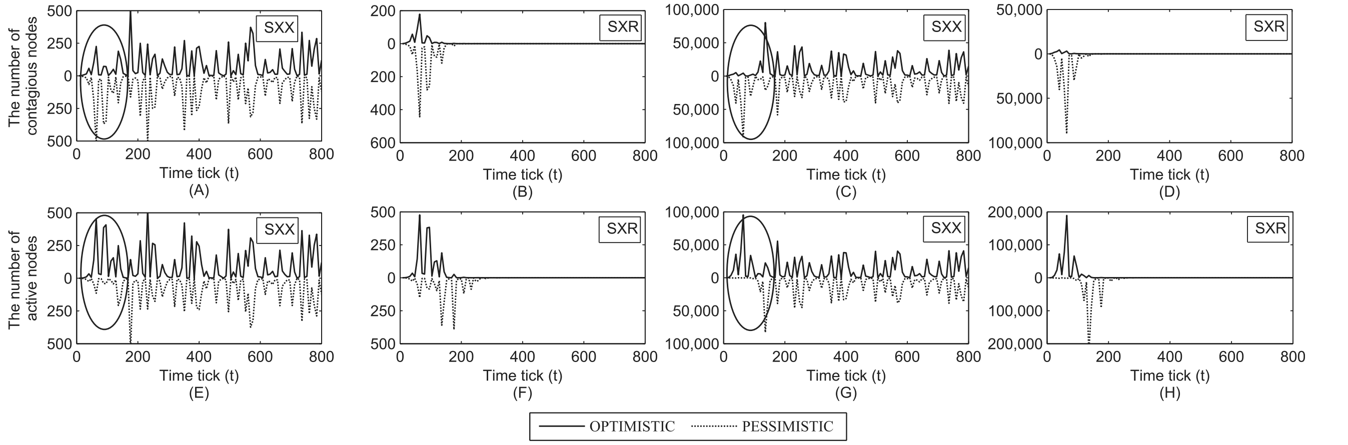


Fig. 11. The number of contagious and active nodes. (A) Facebook; (B) Facebook; (C) Google+; (D) Google+. Settings: 1)$E(\eta_{ij}) = 0.75$; 2) $t_{inject} = 0$.

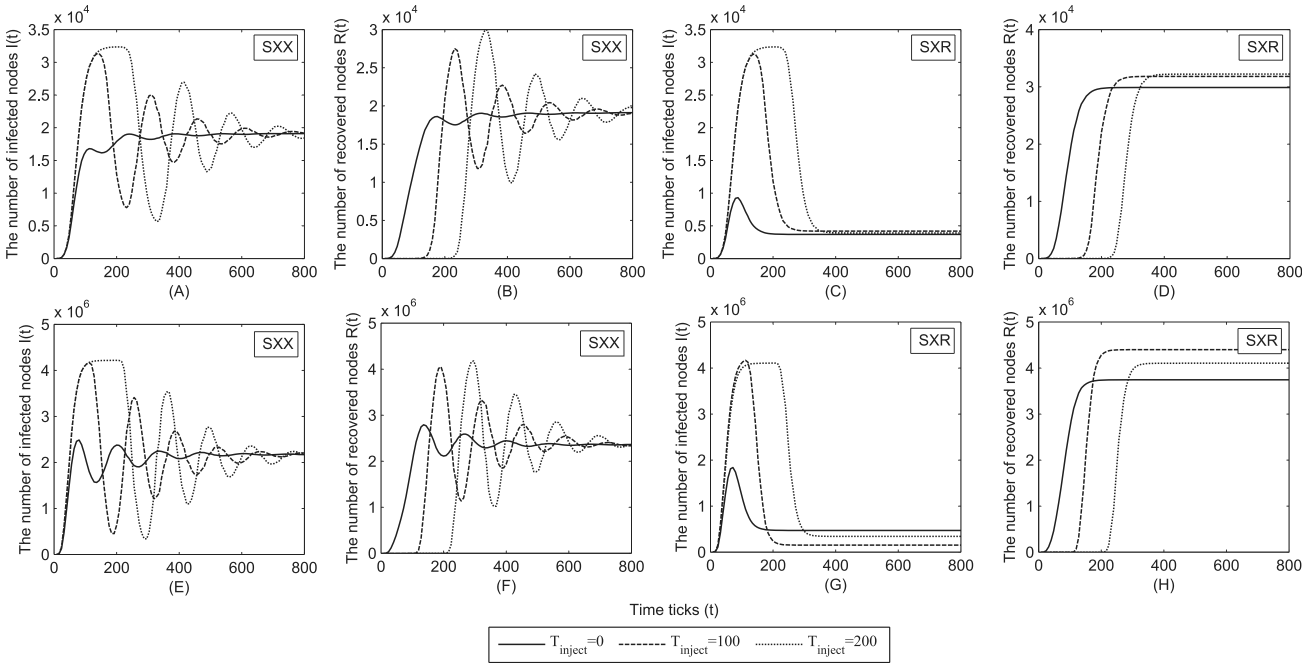


Fig. 12. Positive information is injected at different time. (A)-(D) Facebook; (E)-(H) Google+. Settings: 1) $E(\eta_{ij}) = 0.75$; 2) Preferable; 3) $\alpha = 0, \beta = 0$.

people believing negative information $R(t)$ drops at the very beginning stage of the propagation.

6.4 Impact of q_{ij} for Alternative Choices

When people make alternative choices, if they believe positive information with probability q_{ij}, they will believe negative information with probability $1 - q_{ij}$ (refer to Section 3.3). In order to exclusively investigate the impact of q_{ij}, we set $T_{inject} = 0, T_i, F_i \sim N(20, 10)$ and $p_{ij} = 1$. We test q_{ij} at values 0.6, 0.75 and 0.9.

The results are shown in Fig. 14. We can see that *the values of q_{ij} almost have no impact on the propagation when people make alternative choices.* Particularly for SXX spreading mechanism, all results coincide with each other. For SXR case, when q_{ij} increases, $I(t)$ will increase and $R(t)$ will decrease. But, the differences are not significant.

In the real world, people may have preference on making alternative choices. For example, they prefer positive information more than negative information. Under this condition, the results will not coincide with each other. However, the differences come from the preference but not the value of q_{ij}. Readers could refer to Section 6.3 for the impacts of people's preference.

6.5 Impact of $open_i(t)$, $spr_i(t)$

The temporal propagation dynamics are mainly presented by the flags $open_i(t)$ and $spr_i(t)$ in our model. Following the considerations in [25], [26], [28], we assume user i check newly arrived information every T_i time ticks and forward information every F_i time ticks. Thus, we have $open_i(t) = 1$ if $(t \bmod T_i = 0)$ and $spr_i(t) = 1$ if $(t \bmod F_i = 0)$. The values of T_i and F_i are generated by Gaussian distribution. We set $\eta_{ij}^{pos} = \eta_{ij}^{neg} = 0.75$, $T_{inject} = 0$ and people making preferable choices. We let T_i and F_i follow $N(20, 10)$ and $N(40, 20)$.

As shown in Fig. 15, *the values of $open_i(t)$ and $spr_i(t)$ have some impacts on the propagation dynamics*. We summarize two features from the results of Fig. 15: 1) the final results of the propagation will stay the same when the values of $open_i(t)$ and $spr_i(t)$ change; 2) the spreading speed will decrease if

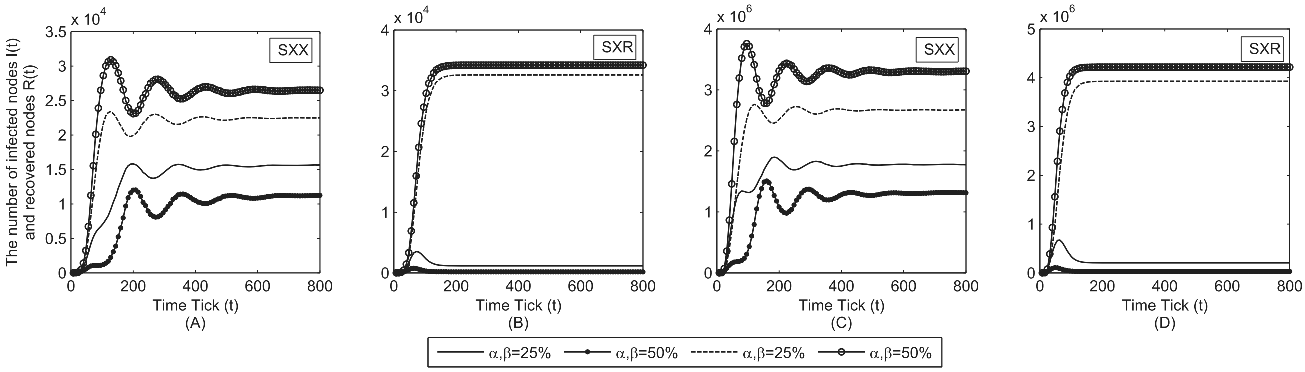


Fig. 13. Impact of α and β for people making alternative choices. (A) Facebook; (B) Facebook; (C) Google+; (D) Google+. Settings: 1) $T_{inject} = 0$; 2) $T_i, F_i \sim N(20, 10)$; 3) $p_{ij} = q_{ij} = 0.75$.

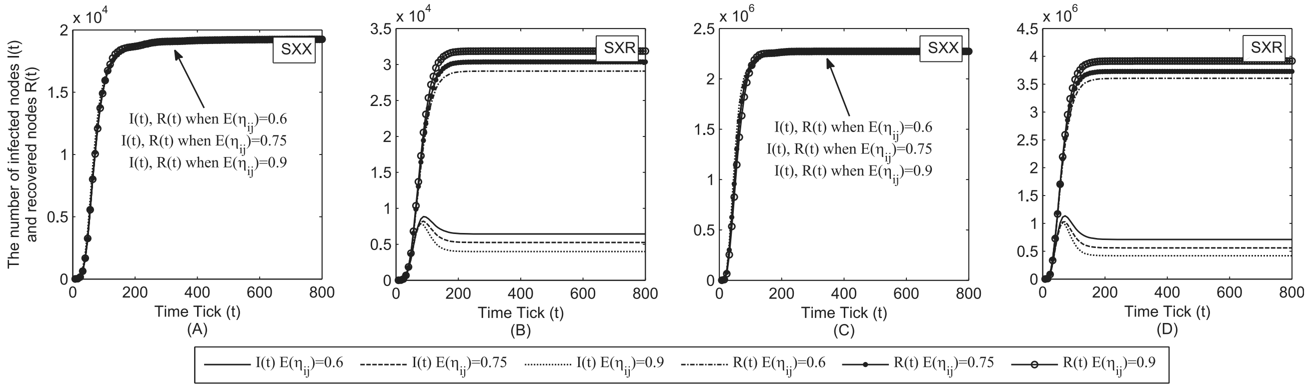


Fig. 14. Impact of q_{ij} for people making alternative choices. (A) Facebook; (B) Facebook; (C) Google+; (D) Google+. Settings: 1) $T_{inject} = 0$; 2) $T_i, F_i \sim N(20, 10)$; 3) $p_{ij} = 1$.

the values of $open_i(t)$ and $spr_i(t)$ increase. These two features line with our expectation on the impacts of $open_i(t)$ and $spr_i(t)$.

7 Related Work

7.1 Propagation Modeling Techniques

There have been substantial efforts in modeling the propagation of information in the last decade. For competitive information, most researchers borrowed basic and well-studied ICM and LTM [8], [9], [10], [12], [13], [15], [16], [17] from marketing area and particle systems. We have compared our model with ICM and LTM in Section 5. In the following, we mainly focus on the propagation models of single information.

Most widely adopted propagation models of single information [18], [19], [20], [21], [22], [23], [24] came from epidemiology since the epidemic spreads are similar to the processes of information dissemination. Epidemic models use differential equations to calculate the number of infected nodes in networks without considering the probabilities of each node being infected or not. Thus, this kind of models are weak to investigate where, when and how many nodes are needed to control the information dissemination [29]. Moreover, as early discussed in [28], the epidemic models [18], [19], [20], [21], [22], [23], [24] may greatly overestimate the spreading speed due to their implicit 'homogeneous mixing' assumption.

The works [28], [30], [31], [32], [33] relied on simulations to model the propagation of malicious information, such as Internet worms. Their simulation models avoid the problem of 'homogeneous mixing' assumption but cannot provide analytical study on the propagation.

There are some other propagation models such as [25], [26], [34], [35] adopting difference equations to present the propagation dynamics of information. Our proposed model is close to their works but our work collaborates the spread processes and presents the interaction of two kinds of information.

7.2 Propagation Control Techniques

On the basis of propagation models, researchers have studied the way to control (restrain or accelerate) the propagation of information. In fact, the problem of selecting most influential nodes is NP-hard [14], [16]. Thus, to maximize the influence of information, some researchers [8], [10], [11], [12], [14] adopted heuristic algorithms to approximate the optimal solution. There are also some works [9], [12], [13], [15] using game theory to find the optimal strategies.

In another side, to restrain the propagation of information. He et al. [17] adopted a greedy algorithm to search the most controllable nodes. Wang et al. [36] studied the propagation of mobile viruses. Their results explained the lack of a major mobile virus breakout so far. The works [33], [37], [38] explored the counter-intuitive fact that the most

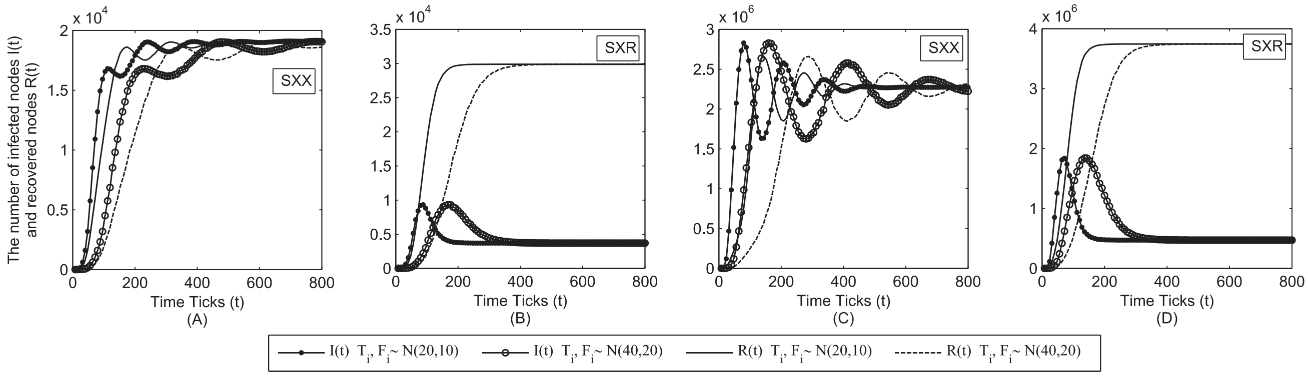


Fig. 15. Impact of $open_i(t), spr_i(t)$. (A) Facebook; (B) Facebook; (C) Google+; (D) Google+. Settings: 1) $T_{inject} = 0$; 2) Preferable; 2) $\eta_{ij}^{pos} = \eta_{ij}^{neg} = 0.75$.

influential nodes in OSNs may not be the most highly-connected nodes. Moreover, the works [39], [40] examined the most influential edges in networks.

Compared with this part of work, our paper provides an accurate propagation model. This model can serve as a fundamental work to support the research of propagation control techniques.

8 Conclusion

In this paper, we studied the propagation of both positive and negative information in OSNs. We proposed an analytical model using difference equations and considering people making different choices. The experiment results showed the accuracy of our model. On the basis of it, we further examined the impacts of parameters impact to the propagation.

In the future, the propagation of multiple kinds of information will be modeled. The information can be supportive or competitive. We will optimize the controllability of the propagation on the basis of our proposed analytical model. Another important work is to use our model to explain or predict the real information propagation. We believe our work presented in this paper is of great significance to both academic aims and practical usage.

Acknowledgments

Weijia Jia's work was supported in part by grants from the General Research Fund of the Hong Kong SAR, China No. (CityU 114012 and 114513); NSF (China) No. 61070222 and Shenzhen (China) Basic Research Project No. JCYJ20120618115257259.

References

[1] B. Viswanath, A. Mislove, M. Cha, and K.P. Gummadi, "On the Evolution of User Interaction in Facebook," *Proc. Second ACM Workshop Online Social Networks (WOSN '09)*, pp. 37-42, 2009.

[2] N.Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, and D. Song, "Evolution of Social-Attribute Networks: Measurements, Modeling, and Implications Using Google+," *Proc. ACM Conf. Internet Measurement Conf. (IMC '12)*, pp. 131-144, 2012.

[3] H. Kwak, C. Lee, H. Park, and S. Moon, "What Is Twitter, a Social Network or a News Media?" *Proc. 19th Int'l Conf. World Wide Web (WWW '10)*, pp. 591-600, 2010.

[4] F. Peter, "'bogus' ap Tweet about Explosion at the White House Wipes Billions Off US Markets," The Telegraph, Finance/Market, Apr. 2013.

[5] Zeinobia, "Unconfirmed News: Coup d'État in Tunisia 'Updated'," Egyptian Chronicles, Jan. 2012.

[6] Al Jazeera, "Tunisia's Ben Ali Flees Amid Unrest," Jan. 2011.

[7] J. Campbell, "Jackie Chan Dead 2013: Death Hoax Rumors Dispelled by Facebook Post," The Christian Post, June 2013.

[8] C. Budak, D. Agrawal, and A. El Abbadi, "Limiting the Spread of Misinformation in Social Networks," *Proc. 20th Int'l Conf. World Wide Web (WWW '11)*, pp. 665-674, 2011.

[9] S. Shirazipourazad, B. Bogard, H. Vachhani, A. Sen, and P. Horn, "Influence Propagation in Adversarial Setting: How to Defeat Competition with Least Amount of Investment," *Proc. 21st ACM Int'l Conf. Information and Knowledge Management (CIKM '12)*, pp. 585-594, 2012.

[10] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence Maximization in Social Networks When Negative Opinions May Emerge and Propagate," *Proc. SIAM Int'l Conf. Data Mining (SDM '11)*, 2011.

[11] W. Chen, W. Lu, and N. Zhang, "Time-Critical Influence Maximization in Social Networks with Time-Delayed Diffusion Process," Arxiv preprint arXiv:1204.3074, 2012.

[12] J. Kostka, Y. Oswald, and R. Wattenhofer, "Word of Mouth: Rumor Dissemination in Social Networks," *Structural Information and Comm. Complexity*, vol. 5058, pp. 185-196, 2008.

[13] S. Bharathi, D. Kempe, and M. Salek, "Competitive Influence Maximization in Social Networks," *Proc. Third Int'l Conf. Internet and Network Economics (WINE '07)*, pp. 306-311, 2007.

[14] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the Spread of Influence through a Social Network," *Proc. Ninth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '03)*, pp. 137-146, 2003.

[15] A. Clark and R. Poovendran, "Maximizing Influence in Competitive Environments: A Game-Theoretic Approach," *Proc. Second Int'l Conf. Decision and Game Theory for Security (GameSec'11)*, pp. 151-162, 2011.

[16] A. Borodin, Y. Filmus, and J. Oren, "Threshold Models for Competitive Influence in Social Networks," *Proc. Sixth Int'l Conf. Internet and Network Economics (WINE '10)*, pp. 539-550, 2010.

[17] X. He, G. Song, W. Chen, and Q. Jiang, "Influence Blocking Maximization in Social Networks under the Competitive Linear Threshold Model," *Proc. SIAM Int'l Conf. Data Mining (SDM '12)*, pp. 463-474, 2012.

[18] W. Yu, X. Wang, P. Calyam, D. Xuan, and W. Zhao, "Modeling and Detection of Camouflaging Worm," *IEEE Trans. Dependable and Secure Computing*, vol. 8, no. 3, pp. 377-390, May/June 2011.

[19] C. Gao and J. Liu, "Modeling and Restraining Mobile Virus Propagation," *IEEE Trans. Mobile Computing*, vol. 12, no. 3, pp. 529-541, Mar. 2013.

[20] S.M. Cheng, W. C. Ao, P. Y. Chen, and K. C. Chen, "On Modeling Malware Propagation in Generalized Social Networks," *IEEE Comm. Letters*, vol. 15, no. 1, pp. 25-27, Jan. 2011.

[21] Y.M. Ko and N. Gautam, "Epidemic-Based Information Dissemination in Wireless Mobile Sensor Networks," *IEEE/ACM Trans. Networking*, vol. 18, no. 6, pp. 1738-1751, Dec. 2010.

[22] M. Vojnovic, V. Gupta, T. Karagiannis, and C. Gkantsidis, "Sampling Strategies for Epidemic-Style Information Dissemination," *IEEE/ACM Trans. Networking*, vol. 18, no. 4, pp. 1013-1025, Aug. 2010.

[23] P. Eugster, R. Guerraoui, A.M. Kermarrec, and L. Massoulie, "Epidemic Information Dissemination in Distributed Systems," *IEEE Computer Magazine*, vol. 37, no. 5, pp. 60-67, May 2004.

[24] G. Yan and S. Eidenbenz, "Modeling Propagation Dynamics of Bluetooth Worms (Extended Version)," *IEEE Trans. Mobile Computing*, vol. 8, no. 3, pp. 353-368, Mar. 2009.

[25] Z. Chen and C. Ji, "Spatial-Temporal Modeling of Malware Propagation in Networks," *IEEE Trans. Neural Networks*, vol. 16, no. 5, pp. 1291-1303, Sept. 2005.

[26] S. Wen, W. Zhou, J. Zhang, Y. Xiang, W. Zhou, and W. Jia, "Modeling Propagation Dynamics of Social Network Worms," *IEEE Trans. Parallel and Distributed Systems*, vol. 24, no. 8, pp. 1633-1643, Aug. 2013.

[27] N.Z. Gong, A. Talwalkar, L. Mackey, L. Huang, E.C.R. Shin, E. Stefanov, E. Shi, and D. Song, "Joint Link Prediction and Attribute Inference using a Social-Attribute Network," *ACM Trans. Intelligent Systems and Technology*, accepted for publication, 2013.

[28] C.C. Zou, D. Towsley, and W. Gong, "Modeling and Simulation Study of the Propagation and Defense of Internet Email Worms," *IEEE Trans. Dependable and Secure Computing*, vol. 4, no. 2, pp. 105-118, Apr. 2007.

[29] Y. Wang, S. Wen, S. Cesare, W. Zhou, and Y. Xiang, "The Microcosmic Model of Worm Propagation," *Computer J.*, vol. 54, no. 10, pp. 1700-1720, Oct. 2011.

[30] W. Fan and K.H. Yeung, "Online Social Networks-Paradise of Computer Viruses," *Physica A: Statistical Mechanics and Its Applications*, vol. 390, no. 2, pp. 189-197, 2011.

[31] W. Yu, N. Zhang, X. Fu, and W. Zhao, "Self-Disciplinary Worms and Countermeasures: Modeling and Analysis," *IEEE Trans. Parallel and Distributed Systems*, vol. 21, no. 10, pp. 1501-1514, Oct. 2010.

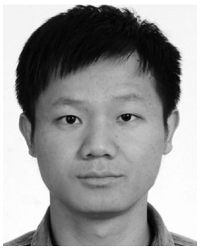
[32] G. Yan, G. Chen, S. Eidenbenz, and N. Li, "Malware Propagation in Online Social Networks: Nature, Dynamics, and Defense Implications," *Proc. Sixth ACM Symp. Information, Computer and Comm. Security (ASIACCS '11)*, pp. 196-206, 2011.

[33] C. Gao, J. Liu, and N. Zhong, "Network Immunization and Virus Propagation in Email Networks: Experimental Evaluation and Analysis," *Knowledge and Information Systems*, vol. 27, pp. 253-279, 2011.

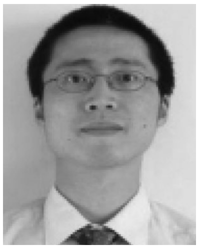
[34] D. Chakrabarti, J. Leskovec, C. Faloutsos, S. Madden, C. Guestrin, and M. Faloutsos, "Information Survival Threshold in Sensor and P2P Networks," *Proc. IEEE INFOCOMM*, pp. 1316-1324, 2007.
[35] C. Anagnostopoulos, S. Hadjiefthymiades, and E. Zervas, "Information Dissemination between Mobile Nodes for Collaborative Context Awareness," *IEEE Trans. Mobile Computing*, vol. 10, no. 12, pp. 1710-1725, Dec. 2011.
[36] P. Wang, M.C. Gonzalez, C.A. hidalgo, and A.L. Barabasi, "Understanding the Spreading Patterns of Mobile Phone Viruses," *Science*, vol. 324, pp. 1071-1076, 2009.
[37] Y.Y. Liu, J.J. Slotine, and A.-L. Barabasi, "Controllability of Complex Networks," *Nature*, vol. 473, pp. 167-173, 2011.
[38] S. Wen, W. Zhou, Y. Wang, W. Zhou, and Y. Xiang, "Locating Defense Positions for Thwarting the Propagation of Topological Worms," *IEEE Comm. Letters*, vol. 16, no. 4, pp. 560-563, Apr. 2012.
[39] T. Nepusz and T. Vicsek, "Controlling Edge Dynamics in Complex Networks," *Nature*, vol. 8, pp. 568-573, 2012.
[40] Y. Li, W. Chen, Y. Wang, and Z.L. Zhang, "Influence Diffusion Dynamics and Influence Maximization in Social Networks with Friend and Foe Relationships," *Proc. Sixth ACM Int'l Conf. Web Search and Data Mining (WSDM '13)*, pp. 657-666, 2013.

Sheng Wen graduated with a degree in computer science from Central South University of China in 2012. He is currently working toward the PhD degree in the School of Information Technology, Deakin University, Melbourne, Australia, under the supervision of Prof. Wanlei Zhou and Yang Xiang. His focus is on modeling of virus spread, information dissemination, and defence strategies of Internet threats.

Mohammad Sayad Haghighi received the BSc, MSc, and PhD degrees in electrical engineering, telecommunications from the University of Tehran, Iran, Sharif University of Technology, Iran, and K.N. Toosi University of Technology, Iran, in 2003, 2005, and 2011, respectively. From 2009 to 2013, he held research positions with Macquarie University and Deakin University. He is currently an assistant professor at the Iran Telecom Research Center. His interests include both distributed networks and network security.



Chao Chen received the bachelor of information technology degree with first class honors from Deakin University, Australia, in 2012. He is currently working toward the PhD degree with the School of Information Technology, Deakin University. His research interests include network security and social network security.



Yang Xiang received the PhD degree in computer science from Deakin University, Australia. He is currently a full professor with the School of Information Technology, Deakin University. His research interests include network and system security, distributed systems, and networking. He has published more than 100 research papers in many international journals and conferences. He is a senior member of the IEEE.

Wanlei Zhou received the PhD degree from the Australian National University, Canberra, in 1991. He is currently the chair professor of information technology and the head of the School of Information Technology, Faculty of Science and Technology, Deakin University, Melbourne, Australia. His research interests include distributed and parallel systems and network security. He has published more than 200 papers in refereed international journals and refereed international conferences proceedings. He is a senior member of the IEEE.

Weijia Jia received the PhD degree from the Polytechnic Faculty of Mons, Mons, Belgium, in 1993. He is currently a full professor with the Department of Computer Science, City University of Hong Kong (CityU), Hong Kong. His research interests include next-generation wireless communication, QoS routing protocols, and multicast. He has published more than 200 papers in refereed international journals and refereed international conference proceedings. He is a senior member of the IEEE.