Modeling Propagation Dynamics of Social Network Worms

Sheng Wen, *Student Member*, *IEEE*, Wei Zhou, Jun Zhang, *Member*, *IEEE*, Yang Xiang, *Senior Member*, *IEEE*, Wanlei Zhou, *Senior Member*, *IEEE*, and Weijia Jia, *Senior Member*, *IEEE*

Abstract—Social network worms, such as email worms and facebook worms, pose a critical security threat to the Internet. Modeling their propagation dynamics is essential to predict their potential damages and develop countermeasures. Although several analytical models have been proposed for modeling propagation dynamics of social network worms, there are two critical problems unsolved: *temporal dynamics* and *spatial dependence*. First, previous models have not taken into account the different time periods of Internet users checking emails or social messages, namely, *temporal dynamics*. Second, the problem of *spatial dependence* results from the improper assumption that the states of neighboring nodes are independent. These two problems seriously affect the accuracy of the previous analytical models. To address these two problems, we propose a novel analytical model. This model implements a spatial-temporal synchronization process, which is able to capture the temporal dynamics. Additionally, we find the essence of spatial dependence is the spreading cycles. By eliminating the effect of these cycles, our model overcomes the computational challenge of spatial dependence and provides a stronger approximation to the propagation dynamics. To evaluate our susceptible-infectious-immunized (SII) model, we conduct both theoretical analysis and extensive simulations. Compared with previous epidemic models and the spatial-temporal model, the experimental results show our SII model achieves a greater accuracy. We also compare our model with the susceptible-infectious-susceptible and susceptible-infectious-recovered models. The results show that our model is more suitable for modeling the propagation of social network worms.

Index Terms—Security, social network worms, propagation dynamics, modeling

1 Introduction

The popularity of social network applications such as Facebook, Myspace, and Twitter has been boosted in recent years [1]. Unfortunately, criminal networks are growing at a rate equal to the social networking platforms upon which they parasitically feed. Representative examples are the worms like Koobface spreading in Facebook and the "Here you are" email worm emerging in 2010. According to the symantec Internet threats report [21], social network worms and resembling attacks account for 1/4 of the total threats in 2009 and nearly 1/5 of the total threats in 2010.

Social networks have become attractive targets for worm creators because of the following characteristics. First, they rely on the information like contact lists contained in a victim's machine to locate new targets. This intelligent mechanism allows far more *efficient* propagation than traditional scanning worms that make a large number of wild guesses for every successful infection [2]. Second, by using social engineering techniques exploiting trust in social networks, many users fail to recognize malicious codes that are sent by their friends and subsequently become infected. This results in a *wide* range that worms propagate to. Third, researchers have found that social networks exhibit both small world properties and scale-free behaviors [3], [4]. This means that the spreading of social network worms can be incredibly *fast* because the highly connected "hub" nodes of a scale-free network and the short paths in a strongly clustered small world will greatly facilitate the propagation of an infection over the whole network.

1.1 Motivation

To understand and allow for addressing defense strategies against social network worms, scientists have done much work in the last decade. For example, early works mainly refer to academic thoughts of epidemic propagation, including susceptible-infectious-susceptible (SIS) model [5] and susceptible-infectious-recovered (SIR) models [6], [7], [8]. Later, paper [13] presents a spatial-temporal model; papers [14], [15], [16], [32] rely on simulation models; paper[18] develops an SI model on a specific social network of smart phones. Moreover, on the basis of SIS models, papers [9], [10], [11], [12] focus on finding threshold conditions for fast extinction of worms.

Although a number of significant works have been done, there are two critical problems unsolved, which seriously affect the accuracy of modeling the propagation of social network worms. *First*, the spread of social network worms

- *S. Wen and W. Zhou are with the School of Information Science and Engineering, Central South University, Changsha, P.R. China 410083 and with the School of Information Technology, Deakin University, 221 Burwood Highway, Burwood, VIC 3125, Australia. E-mail: {wsheng, weiz}@deakin.edu.au.*
- *J. Zhang is with the School of Information Technology, Deakin University, 75 Pigdons Road, Waurn Ponds, VIC 3216, Australia. E-mail: jun.zhang@deakin.edu.au.*
- *Y. Xiang and W. Zhou are with the School of Information Technology, Deakin University, 221 Burwood Highway, Burwood, VIC 3125, Australia. E-mail: {yang, wanlei}@deakin.edu.au.*
- *W. Jia is with the Department of Computer Science, City University of Hong Kong, 83 Tat Chee Avenue, Kowloon, Hong Kong. E-mail: itjia@cityu.edu.hk.*

Manuscript received 20 Feb. 2012; revised 18 July 2012; accepted 17 Aug. 2012; published online 24 Aug. 2012.
Recommended for acceptance by S. Guo.
For information on obtaining reprints of this article, please send e-mail to: tpds@computer.org, and reference IEEECS Log Number TPDS-2012-02-0121.
Digital Object Identifier no. 10.1109/TPDS.2012.250.

depends on human involvement. That is Internet users periodically check their newly arrived emails or messages, and are lured to open those which are actually worm copies. The period for checking these emails or messages depends on users' own patterns. However, the *temporal dynamics*, which are to model the spread of worms with different checking time periods of users, have not been implemented in previous works. Most previous models, such as [9], [13], assume a user starts infecting others at the moment the user gets infected, which means the checking periods of users are identical and assumed to be one. Thus, previous models cannot present accurate and realistic spreading procedure. *Second*, the spreading of social network worms relies on the topology of social networks. In the modeling, the probability of a node being infected will increase when its neighbors have been infected. However, the increasing part in the probability cannot reversely increase the probabilities of its neighbors being infected. Otherwise, the modeling has redundant computation. This is the problem of *spatial dependence*. Previous analytical works, such as [10], [13], simplify this problem and assume nodes to be spatial independent. Our empirical study shows that their works greatly overestimate the number of infected users (see Section 4 of the supplementary file, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPDS.2012.250). To solve these two problems, we are motivated to propose a *susceptible-infectious-immunized* (SII) model which can *precisely* present the propagation dynamics of social network worms.

1.2 Contributions

The primary contributions of this work are summarized as follows:

1. We carry out extensive analyses on two important problems, *temporal dynamics* and *spatial dependence*, which crucially affect the accuracy of the existing analytical models. Our analysis shows that previous works cannot model the realistic propagation with different checking time periods of users. Moreover, the analysis shows the spreading cycles formed in the modeling lead to considerable errors in estimating the infection probabilities.
2. We propose a novel SII model. This model implements a spatial-temporal synchronization process, which helps us to solve the problem of modeling *temporal dynamics*. Furthermore, by eliminating the effect of those cycles, the SII model overcomes the computational challenge and provides a stronger approximation of *spatial dependence*.
3. We conduct a number of experiments to evaluate the proposed SII model. The results show that our SII model is superior to the state-of-the-art models for the spread of social network worms, which include the epidemic models, the spatial-temporal model, the SIS model, and the SIR model.

The remainder of this paper is organized as follows: Section 2 introduces the basis of social network worms. Section 3 states the problems in the modeling of social network worms. Section 4 explains the details of our SII model. Section 5 implements a series of experiments to evaluate the correctness of our model. Further discussion and related work are presented in Section 6 and 7, respectively. Finally, Section 8 concludes this paper.

2 The Spread of Social Network Worms

2.1 Social Network Worm Primer

Social network worms leverage social platforms to send worm emails or messages containing malicious links. Once the recipients are lured to read the infectious email attachments or visit the malicious web links, they will get infected and start spreading such worm copies or links to their friends found in the contact lists, such as email address lists. Different from scanning worms, the spread of social network worms mainly depends on the topology of social networks and need much human involvement.

In the early stage of the propagation, users have no knowledge of the newly emerging worm. After scientists detect it and encourage users to update their antivirus software, or when users become more skeptical of the emails or links which seem out-of-character, the infected computers will be cleaned and the momentum of the spreading will slow down until the worms become extinct in networks. We provide the state transition graph of a user in Fig. 1 of the supplementary file, which is available online.

2.2 Social Network Topology Characters

The social network topologies have the following characters:

1. They can be thought of as a "semidirected network," a graph in which some edges are directed and others are undirected [4], [22]. In this paper, we use *reciprocity rate* to depict the fraction of edges which point back to the sources.
2. The indegree of nodes tends to match the outdegree [4], [22]. We simply let outdegree be equal to indegree and both follow the power-law distribution [3], [4], [5], [19].
3. Users who have large groups of friends in social networks tend to appear in the contact lists of many others [22]. Thus, any edge from one node to another node is chosen by the *large-degree-preferential* principle.
4. The weight of each edge denotes the probability of an unwary user reading malicious messages from their friends. This probability is determined by human factor. Similar to [14], we let this parameter follow Gaussian distribution.

Existing works [8], [13], [14] have shown topological factors strongly affect the spreading speed and scale. In this paper, we derive the structure of the network topology for simulations on the basis of previous statistical analysis in real social networks [3], [4], [15], [19], [22]. The topologies are generated by using a 2K-series method which was proposed and verified by Mahadevan et al. [29]. The detailed algorithm is shown in Algorithm 1 of the supplementary file, which is available online.

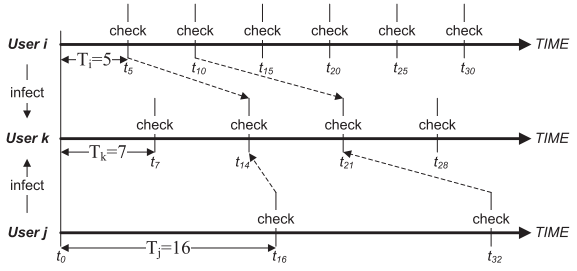


Fig. 1. An example to explain why previous models have not presented accurate temporal dynamics on users in the propagation.

3 Problem Statement

For the convenience of description, in the following, we use neighbors N_i instead of friends of user i and messages instead of those malicious attachments or links, if not otherwise stated. We also let T_i denote the message checking period of user i.

3.1 The Problem of Temporal Dynamics

Previous models [6], [7], [8], [10], [13], [18] assume users check newly arrived messages at every moment and the message checking time T_i is identical to each user (usually one time interval). That means, any malicious messages, which arrive at node i at time t, will be definitely opened and possibly forwarded to their neighbors N_i at time t. These neighbors repeat this process at time $t+1$ and the neighbors of N_i do so at time $t+2$, and so forth. This process is equivalent to the k-hop modeling where k denotes the number of edges from the initial infectious node to the current node i.

In the real world, users have different message checking time. We take Fig. 1 as an example: user k is a neighbor of both user i and user j in a social network. In the previous models, if node i and node j are compromised at their first checking time (t_5 and t_{16}), they will send worm copies to user k and possibly infect user k at the second checking time of user k (t_{14}). Considering different message checking time of users, user i is compromised at t_5 (the first checking time of user i), the malicious message from user i will only be read by user k at t_7 (the first checking time of user k) rather than t_{14} (the second checking time of user k). Besides, user j is infected at t_{16} (the first checking time of user j), the malicious message from user j will be read by user k at t_{21} (the third checking time of user k) rather than t_{14}. Therefore, previous k-hops models suffer from the problem of temporal dynamics. To tackle this problem, we derive a concept as follows.

Definition 1. *If the state of node i being infected is timekeeping with its compromised neighbors sending malicious messages (spatial factor) and the message checking time for possibly visiting those messages (temporal factor), we call this process "the spatial-temporal synchronization process."*

This process means the modeling needs to *record and accumulate* every newly arrived malicious message in an arbitrary T_i, and calculate the joint infection probability of those messages when the user checks them. By realizing this process in the modeling, we can solve the problem of temporal dynamics.

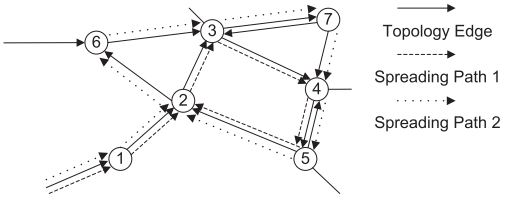


Fig. 2. Worms spread in a small episode of a social network. Paths 1 and 2 are two examples for the propagation in the topology.

To demonstrate the importance of this process, we further compare the previous analytical models with our simulation. The results show that the problem of temporal dynamics is critical in the previous models. The details are in Section 3 of the supplementary file, which is available online.

3.2 The Problem of Spatial Dependence

We use a small episode of a social network to depict the spatial dependence. As shown in Fig. 2, there are two spreading paths: 1→2→3→4→5→2 and 1→2→6→3→7→4→5→2. In the analytical modeling, the event of a node being infected is presented by a probability at time t. Thus, the probability of node 5 being infected will increase due to node 2 having been infected before. However, the increasing part in the probability of node 5 being infected cannot further affect the probability of node 2 being infected. Otherwise, the probability of node 2 having been infected before will be implicitly *recounted*, which causes overestimation of the number of infected nodes. More importantly, we find the essence of spatial dependence is the cycles formed in the spreading paths [20]. Based on the number of intermediate nodes in a cycle, we derive a concept as follows.

Definition 2. *If a spreading path originates from an arbitrary node i, passes through k intermediate nodes, and then return to node i, we call this spreading path a "k-order cycle."*

In Fig. 2, there are a 3-order cycle (2→3→4→5→2) and a 5-order cycle (2→6→3→7→4→5→2) on node 2. To realize the spatial dependence, we have to trace each spreading cycle and remove their effect to avoid overestimation.

Through empirical analysis on the social topologies, we find the cycles from 1-order to 5-order have significant effect on the propagation. Thus, we need to remove the effect of 1-order to 5-order cycles in the modeling. Previous Markov approximation that focuses on removing only 1-order cycles is not enough. The detailed analysis is provided in Section 4 of the supplementary file, which is available online.

4 SII Model

4.1 Modeling Nodes, Topology, and Events of Users

Nodes and topology information are basic elements for the propagation of social network worms. Given a social network, we derive the topology of it. A node in the topology presents a user in the social network. Let random variable $X_i(t)$ denote the state of a node i at discrete time t. Then, we have

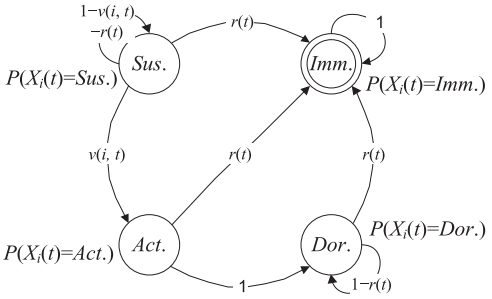


Fig. 3. State transition graph of a node in the topology.

$$X_i(t) = \begin{cases} \boldsymbol{Hea.}, \text{healthy} \begin{cases} \boldsymbol{Sus.}, \text{susceptible}, \\ \boldsymbol{Imm.}, \text{immunizied}, \end{cases} \\ \boldsymbol{Inf.}, \text{infected} \begin{cases} \boldsymbol{Act.}, \text{active}, \\ \boldsymbol{Dor.}, \text{ dormant}. \end{cases} \end{cases} \quad (1)$$

We derive the state transition graph of an arbitrary node i in a social network. As shown in Fig. 3, node i transits to the $infected$ state if it is at the $susceptible$ state. The infection probability is presented by $v(i,t)$. A user is infectious at the $active$ state. Then, the node of this user will stay at the *dormant* state until this node is immunized. Moreover, no matter which state the node is at, it may transit to the *immunized* state. The recovery probability is presented by $r(t)$. For the convenience of readers, we list major notations of this paper in Table 1.

In the SII model, we propose employing an M by M square matrix with elements p_{ij} to describe a topology consisting of M nodes, as in

$$\begin{pmatrix} p_{11} & \cdots & p_{1M} \\ \vdots & p_{ij} & \vdots \\ p_{M1} & \cdots & p_{MM} \end{pmatrix} \quad p_{ij} \varepsilon [0,1], \quad (2)$$

wherein p_{ij} represents the probability of user j reading malicious messages from user i. $p_{ij} = 0$ means no contact between node i and j. Therefore, the matrix reflects the topology of a social network.

The infection of social network worms relies on unwary users checking malicious messages. To properly model this process, we introduce two new variables. *First*, we define a random variable $open_i(t)$. We have $open_i(t) = 1$ if the user is checking new messages at time t, otherwise $open_i(t) = 0$. The probability of a user checking messages is determined by the value of T_i. Thus, we have

$$P(open_i(t) = 1) = \begin{cases} 0, & \text{otherwise}, \\ 1, & t \bmod T_i = 0. \end{cases} \quad (3)$$

Second, it is significant to obtain the number of unread messages of each user at current time t. These messages may arrive at the users at an arbitrary time between the users last checking new messages and the current time t (excluding t). Thus, we introduce a variable t' to indicate the arbitrary time. As shown in Fig. 4, the value of t' has two forms depending on whether user checks messages at current time t or not. Then, we have

$$\begin{cases} t - T_i \le t' < t, & if\ open_i(t) = 1, \\ t - (t \bmod T_i) \le t' < t, & \text{otherwise}. \end{cases} \quad (4)$$

TABLE 1
Major Notations Used in This Paper

Symbol	Explanation
T_i	Message checking time of user i.
$r(t)$	The recovery function of users, which provides the probability for any user to be immunized at time t.
$X_i(t)$	The state of a network node i at time t: "*Sus.*" susceptible, "*Imm.*" immunized, "*Act.*" active and "*Dor.*" dormant.
p_{ij}	The probability of user j reading malicious messages from i.
$open_j(t)$	The event of user j checking newly arrived messages at time t.
t'	The arbitrary time between the time of user i last checking messages and the current time t (excluding t).
M	The size of the social network.
N_i	The set of neighboring nodes of node i.
$n(t)$	The number of infected nodes in the social network at time t.
$v(i,t)$	The infection probability of node i at time t.
$s(i,t)$	The probability of user i reading malicious messages from neighboring nodes at time t.
C_{ij}	An arbitrary spreading path from node i to node j.
$\Theta(t)$	The probability for node j being infected by node i through spreading path C_{ij} at time t.
$\delta(C_{ij})$	The probabilistic effect of C_{ij} to the propagation modeling.
$\tau(t,C_{ij})$	The beginning time of the spreading path C_{ij}.
te_i	The earliest time of messages arriving at node i.
tl_i	The latest time of messages arriving at node i.

A compromised user can only spread worm copies to the neighboring users in social networks. Thus, for each user in networks, we record and accumulate every newly arrived malicious message from neighboring users at each t', then we can finally obtain the joint infected probability of each user who checks those messages.

4.2 Modeling Propagation Dynamics

We use the values 0 and 1 to substitute the *healthy* state and the *infected* state, respectively. Given a topology of a social network with M nodes, the expected number of infected users at time t, $n(t)$, can be computed as in

$$\begin{aligned} n(t) &= E\left[\sum_{i=1}^{M} X_i(t)\right] = \sum_{i=1}^{M} E[X_i(t)] = \sum_{i=1}^{M} P(X_i(t) = 1) \\ &= \sum_{i=1}^{M} P(X_i(t) = Inf.). \end{aligned} \quad (5)$$

The expected number of infected nodes, $n(t)$, is ascribed to the sum of the probabilities of each node being infected at time t, $P(X_i(t) = Inf.)$. As shown in Fig. 3, a susceptible node can be compromised and be at *infected* state, and an infected node can be recovered and be at *immunized* state. Based on the state transition graph, we derive the computation of $P(X_i(t) = Inf.)$ by difference equations as follows:

$$\begin{aligned} P(X_i(t) = Inf.) = (1 - r(t)) * P(X_i(t-1) = \inf.) \\ + v(i,t) * P(X_i(t-1) = Sus.). \end{aligned} \quad (6)$$

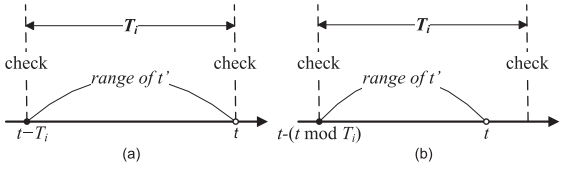


Fig. 4. Different cases of variable t'. (a) User checks messages at current time t. (b) User does not check messages at current time t.

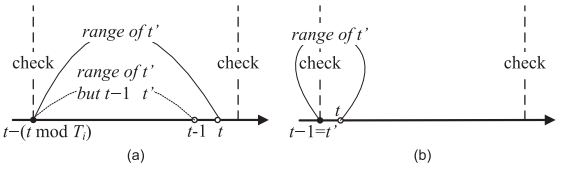


Fig. 5. (a) and (b) Different cases for the computation of $s(i,t)$. (b) is different from (a), because user checks messages at time $t-1$ in (b).

For the computation of $P(X_i(t) = Sus.)$, we have

$$P(X_i(t) = Sus.) = 1 - P(X_i(t) = Inf.) - P(X_i(t) = Imm.). \quad (7)$$

Moreover, for the computation of $P(X_i(t) = Imm.)$, we have

$$P(X_i(t) = Imm.) = P(X_i(t-1) = Imm.) + r(t) * [1 - P(X_i(t-1) = Imm.)]. \quad (8)$$

Once we obtain the values of $v(i,t)$ and $r(t)$, the value of $P(X_i(t) = Inf.)$ can be computed by the iteration of the above (6), (7), and (8). We provide the algorithm of the iteration process in Section 5 of the supplementary file, which is available online.

In fact, there are three preconditions for each infection of social network worms: 1) user has not been immunized; 2) user checks new messages; and 3) susceptible user reads those malicious messages. When the first and the second preconditions are satisfied, we use $s(i,t)$ to represent the probability of a susceptible user i reading malicious messages from their neighboring nodes at time t. Then, the infection probability $v(i,t)$ can be derived as in

$$v(i,t) = s(i,t) * P(open_i(t) = 1) * [1 - r(t)]. \quad (9)$$

For each pair of neighboring users, when the second precondition is satisfied, the probability of a susceptible user i reading malicious messages from user j at time t is $p_{ji} * P(X_j(t') = Act | X_i(t-1) = Sus.)$. Here, the event $X_j(t') = Act.$ means node j is infected and sents out malicious message copies to neighboring nodes at time t'. Since N_i denotes the set of neighboring nodes of node i, we can compute $s(i,t)$ as in

$$s(i,t) = 1 - \prod_{j \in N_i} [1 - p_{ji} * P(X_j(t') = Act.|X_i(t-1) = Sus.)]. \quad (10)$$

Considering different values that the variable t' may take, we disassemble (10) by excluding $t-1$ from the range of value t'. There are two cases. First, as shown in Fig. 5a, user does not check new messages at time $t-1$. Thus, we have

$$\begin{aligned} &\prod_{j \in N_i} [1 - p_{ji} * P(X_j(t') = Act.|X_i(t-1) = Sus.)] \\ &= \prod_{\substack{j \in N_i \\ t' \neq t-1}} [1 - p_{ji} * P(X_j(t') = Act.|X_i(t-1) = Sus.)] \\ &\quad * \prod_{j \in N_i} [1 - p_{ji} * P(X_j(t-1) = Act.|X_i(t-1) = Sus.)] \end{aligned} \quad (11)$$

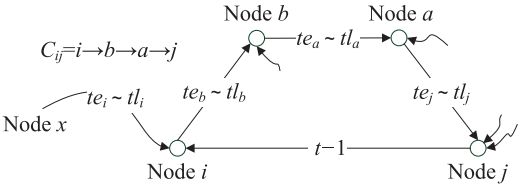


Fig. 6. An example of the propagation cycle on node i.

$$\begin{aligned} &= [1 - s(i,t-1)] * \prod_{j \in N_i} [1 - p_{ji} * P(X_j(t-1) \\ &= Act.|X_i(t-1) = Sus.)]. \end{aligned} \quad (12)$$

Second, as shown in Fig. 5b, user checks new messages at time $t-1$. Thus, the malicious messages received at time t are those sent at time $t-1$ by the infected neighboring users. The variable t' only takes the value $t-1$. In this case, we have

$$\begin{aligned} &\prod_{j \in N_i} [1 - p_{ji} * P(X_j(t') = Act.|X_i(t-1) = Sus.)] \\ &= \prod_{j \in N_i} [1 - p_{ji} * P(X_j(t-1) = Act.|X_i(t-1) = Sus.)]. \end{aligned} \quad (13)$$

Actually, the difference of (12) and (13) is caused by user checking newly arrived messages at time $t-1$. Thus, we can unify the computation of $s(i,t)$ as in

$$\begin{aligned} s(i,t) = 1 - \{1 - s(i,t-1) * [1 - P(open_i(t-1) = Sus)]\} \\ * \prod_{j \in N_i} [1 - p_{ji} * P(X_j(t-1) = Act.|X_i(t-1) = Sus.)]. \end{aligned} \quad (14)$$

Through (9) and (14), we implement the spatial-temporal synchronization process in the SII model. Once the $P(X_j(t-1) = Act.|X_i(t-1) = Sus.)$ and $r(t)$ are obtained, the expected number of infected nodes $n(t)$ can be calculated through iterations of previous difference equations.

4.3 Relaxing of Spatial Dependence Condition

The conditional probability $P(X_j(t-1) = Act.|X_i(t-1) = Sus.)$ in (14) represents the spatial dependence between node i and node j. In Section 3.2, we have explained the spatial dependence is essentially the spreading cycles formed in the propagation procedure. Thus, to obtain the value of $P(X_j(t-1) = Act.|X_i(t-1) = Sus.)$, we need to: 1) find out those cycles which start at node i, pass by node j and return back to node i; and 2) remove their effects in the modeling.

Assume node i is a direct neighbor of node j. As shown in Fig. 6, if node i has a k-hop spreading path to node j, we then have a k-order cycle which starts at node i, passes by node j and returns back to node i. Actually, once a network topology is created and presented in a matrix, we can find out and record those k-hop spreading paths by the iteration of this topology matrix. We introduce a variable C_{ij} to denote an arbitrary spreading path from node i to node j, and $\delta(C_{ij})$ as its probabilistic effect to the spreading. We also introduce $\tau(t, C_{ij})$ as the beginning time of the spreading path C_{ij} which ends at time t. Then, we are able to compute

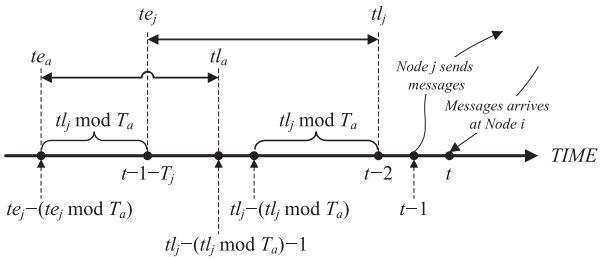


Fig. 7. An example to explain how the values of te and tl are computed in the modeling procedure.

the probability for node j being infected by node i through path C_{ij} at time $t, \Theta(t)$, as in

$$\theta(t) = \delta(C_{ij}) \cdot P(X_i(\tau(t, C_{ij})) = Act.). \quad (15)$$

The probability for node j being infected by its neighbors at time t is $v(j,t)$. Assuming there are totally $H(k)$ k-hop spreading paths which start at node i, and end at node j, we can then remove the probabilistic effect of those spreading paths, and obtain the value of the spatial dependence condition as in

$$P(X_j(t-1) = Act.|X_i(t-1) = Sus.) = \left\{1 - \frac{1 - v(j, t-1)}{\prod_{k=1}^{K}\prod_{h=1}^{H(k)}[1 - \theta_h(t-1)]}\right\} \cdot P(X_j(t-2) = Sus.), \quad (16)$$

wherein K is the maximal length of the spreading paths. If the values $\delta(C_{ij})$ and $\tau(t, C_{ij})$ are obtained, (16) can be iterated with (9) and (15) in the modeling.

How to compute the beginning time of the spreading path $C_{ij}, \tau(t, C_{ij})$? As shown in Fig. 7, if node j is infected and sends messages to node i at time $t-1$, we can derive the *earliest time* (te_j) and the *latest time* (tl_j) of messages arriving at node j as $t-1$-T_j and t-2, respectively; Additionally, node a is a neighbor of node j in Fig. 6. Because the latest time for node a sending malicious messages to node j is at tl_j-(tl_j mod T_a), the value tl_a becomes tl_j-(tl_j mod T_a)-1. Besides, since the messages arriving at node a after the time te_j-(te_j mod T_a) will not be forwarded to node j before te_j, the value te_a can be computed as te_j-(te_j mod T_a). Similarly, we can also derive the earliest and latest time of messages arriving at node b and node i in Fig. 6. Actually, the beginning time $\tau(t, C_{ij})$ ranges from te_i to tl_i. If node i sends malicious messages to neighbors N_i during the period of te_i to tl_i, node i will possibly receive the messages originated from itself.

For the convenience of description, we label the nodes in an arbitrary spreading path C_{ij} as the number 0 to k from i to j. Then, we have the earliest (te) and latest (tl) time of messages arriving at each node in C_{ij} as

$$\begin{cases} te_k = t - T_k - 1, \\ tl_k = t - 2, \end{cases} \quad (17)$$

$$\begin{cases} te_{k-x} = te_{k-x+1} - (te_{k-x+1} \bmod T_{k-x}), \\ tl_{k-x} = tl_{k-x+1} - (tl_{k-x+1} \bmod T_{k-x}), \end{cases} \quad 0 < x \le k, \quad (18)$$

when $x = k, te_{k-x}$, and tl_{k-x} are actually the earliest and latest time of messages arriving at node i, which means $\tau(t, C_{ij})$ falls in the time range $[te_0, tl_0]$. Assume E is the set of edges in a C_{ij} and U is the set of its nodes excluding node i and node j. Then, we can also compute the probabilistic effect of this path $\delta(C_{ij})$ by multiplying the propagation probability of each edge in E and the probability of each node being susceptible in U. The time for each node being infected in C_{ij}, (th), can be obtained by looking back upon the computation of $\tau(t, C_{ij})$. That is

$$\begin{cases} th_1 = \tau(t, C_{ij}) - [\tau(t, C_{ij}) \bmod T_1] + T_k \\ th_x = th_{x-1} - (th_{x-1} \bmod T_{x-1}) + T_{x-1} \end{cases} \quad 1 < x \le k. \quad (19)$$

Then, the nodes in C_{ij} should be susceptible in $th-1$. The computation of the values $\delta(C_{ij})$ and $\tau(t, C_{ij})$ is shown in Algorithm 3 of the supplementary file, which is available online. Currently, we are able to compute the value of $P(X_j(t-1) = Act.|X_i(t-1) = Sus.)$ by removing the effect of the spreading cycles by using (16) between te_0 to tl_0.

5 Model Validation

In this field, all existing research adopts simulation to evaluate analytical models, such as [10], [13]. To validate the correctness of the SII model, we draw an SII-compatible propagation simulator from existing simulation models [14], [15], [16], [32]. The implementation is in C++ and Matlab7. The random numbers in experiments are produced by the C++ TR1 library extensions. The topologies adopted for evaluation are medium topologies (10,000 nodes). The simulation results are averaged by 100 runs. The number of 100 comes from the discussion "how many simulation runs are needed before we obtain a steady curve? [14]." Each run takes 800 time intervals. Each run of the spread has two infected nodes at the beginning, which are randomly chosen from the network. Moreover, we set the two nodes have a distance of 6 (the number of edges between them) in the topology.

5.1 Test the Performance of Modeling of Temporal Dynamics and Spatial Dependence

To evaluate the accuracy of our model, we conduct experiments with different parameter settings. First, to exclude the impact of different recovery processes, the experiments are carried out without the recovery process ($r(t) = 0$) in this section. The topology has a power-law exponent $\alpha = 2.5$, an average degree $E(D) = 5.5$ and a reciprocity rate $\lambda = 0.23$. The infection probability p_{ij} follows Gaussian distribution $N(0.5, 0.2)$. The checking time T_i follows Exponential distribution $Exp(40)$. These parameter settings come from previous works [14], [22].

Fig. 8 shows the modeling results of adopting different approximations to spatial dependence. We can see that the approximation by removing 1- to 5-order cycles is quite close to the simulation result. Since removing more cycles will lead to fewer conflicts in the spread of worms, the result of removing first to 5-order cycles are better than removing only 1-order cycles or never removing any cycles. Actually, the differences come from the computation of (16). We assume $\varphi(\infty) = P(X_j(t-1) = Act.|X_i(t-1) = Sus.)$ and use $\varphi(K)$ to present the asymptotic value of $\varphi(\infty)$ by removing 1 to K-order cycles. Then, we have

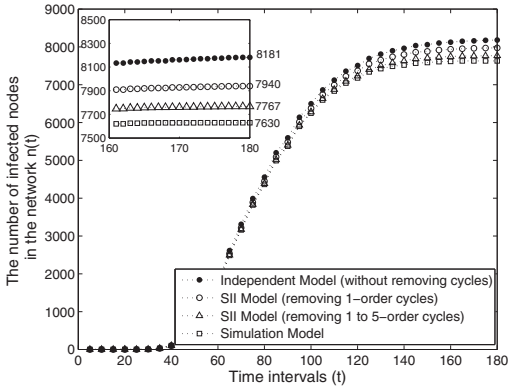


Fig. 8. The spread of social network worms modeled by SII models (with or without removing cycles). Topology: $\alpha = 2.5,\ E(D) = 5.5, \lambda = 0.23,$ $p_{ij} \sim N(0.5, 0.2)$; Message checking time: $T_i \sim Exp(40)$.

$$\begin{aligned}\Delta(K) &= \varphi(K-1) - \varphi(K) \\ &= \frac{(1 - v(j,t-1)) \cdot \left[1 - \prod_{h=1}^{H(K)} (1 - \Theta_h(t-1))\right]}{\gamma} \\ &\cdot P(X_j(t-2) = Sus.),\end{aligned}$$

wherein $0 < \gamma = \prod_{k=1}^{K} \prod_{h=1}^{H(k)} (1 - \Theta_h(t-1)) < 1$. Because we have $0 < v(j, t-1) < 1$ and $0 < \Theta(t-1) < 1$, we can prove that $\Delta(K) > 0$, which leads to

$$\varphi(0) > \varphi(1) > \cdots > \varphi(K) > \varphi(\infty).$$

This means the result of the SII model becomes more accurate when more cycles are removed. Additionally, when $K \to \infty$, we have

$$\lim_{K \to \infty} \gamma = 0 \quad \text{and} \quad \lim_{K \to \infty} \Theta(t-1) = 0,$$

which lead to

$$\lim_{K \to \infty} \frac{\Delta(K)}{K} = 0.$$

This means $\Delta(K)$ is high-order infinitesimal to the value $K(\Delta(K) = o(K))$ and the approximating benefit by removing K-order cycles decreases rapidly when the order K increases. Therefore, it is not necessary to remove all the k-hop spreading paths in the topology to approximate the spatial dependence. We have investigated the effect of spreading cycles in Section 3 of the supplementary file, which is available online. Because the 1- to 5-order cycles have noticeable effects on the modeling, we mainly focus on the k-hops cycles ($k \leq 5$). This saves much computation.

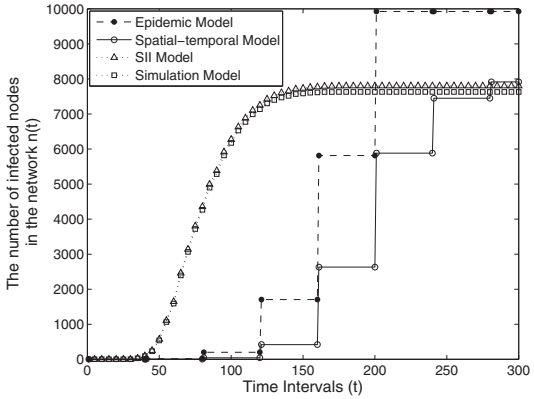


Fig. 9. The comparison of different models with message checking time: $T_i \sim Exp(40)$; topology: $\alpha = 2.5,\ E(D) = 5.5, \lambda = 0.23, p_{ij} \sim N(0.5, 0.2)$.

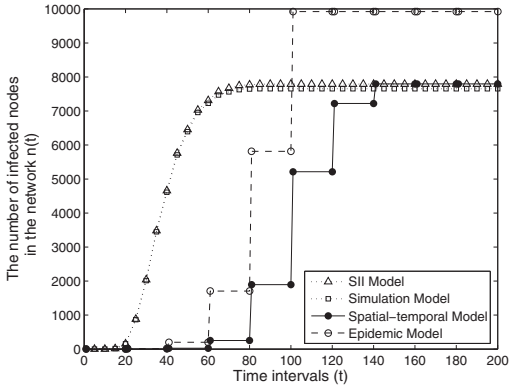


Fig. 10. The comparison of different models with message checking time: $T_i \sim Exp(20)$; topology: $\alpha = 2.5, E(D) = 5.5, \lambda = 0.23, p_{ij} \sim N(0.5, 0.2)$.

We also compare the SII model (removing 1- to 5-order cycles) with other models. As shown in Fig. 9, because the epidemic model [8] and the spatial-temporal model [13] do not implement message *checking time*, the spread in their modeling is actually determined by hops. When $T_i = 40$, the spreading of 1 hop will cost 40 time intervals. As a result, their curves show shapes of stairways, which largely deviate from the simulation result.

SII models with different T_i and p_{ij} values are also investigated. We have *Case 1* (what if users check messages more frequently): $T_i \sim Exp(20)$, $p_{ij} \sim N(0.5, 0.2)$; *Case 2* (what if users are more vigilant against this worm): $T_i \sim Exp(40)$, $p_{ij} \sim N(0.25, 0.1)$. As shown in Fig. 9, the SII model, which has removed 1- to 5-order cycles, achieves a better accuracy. We also compare the results of these two cases with other models under the same parameter settings. As shown in Figs. 10 and 11, our SII model (removing 1- to 5-order cycles) is much better than previous epidemic model and spatial-temporal model.

Moreover, we investigate the accuracy of the SII model in different topologies. Compared with the topology settings used in Fig. 8 ($\alpha = 2.5,\ E(D) = 5.5, \lambda = 0.23$), we have *Case 1*: (what if users averagely have more friends in the network, which means α becomes smaller and $E(D)$ becomes larger): $\alpha = 2.0, E(D = 10),\ \lambda = 0.23$; *Case 2* (what

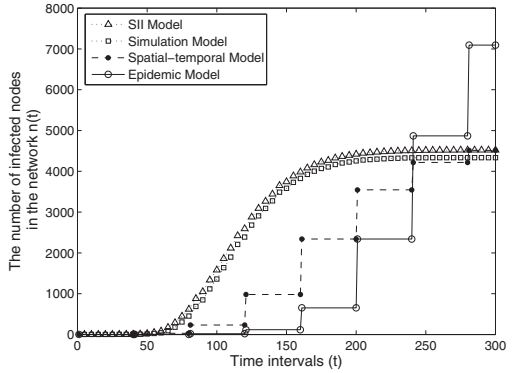


Fig. 11. The comparison of different models with message checking time: $T_i \sim Exp(40)$; topology: $\alpha = 2.5,\ E(D) = 5.5,\ \lambda = 0.23,\ p_{ij} \sim N(0.25, 0.1)$.

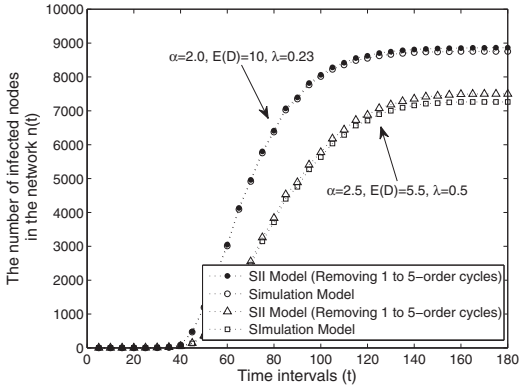


Fig. 12. The accuracy of the SII model in different topologies. Case 1: $\alpha = 2.0$, $E(D) = 10$, $\lambda = 0.23$; Case 2: $\alpha = 2.5$, $E(D) = 5.5$, $\lambda = 0.5$.

if users are more likely to add each other into their contacts, which means λ becomes larger and the topology has more cycles): $\alpha = 2.5, E(D = 5.5)$, $\lambda = 0.5$. As shown in Fig. 12, the results of our SII model still fit the simulation result by removing 1- to 5-order spreading cycles.

5.2 Test the Performance of SII Model with the Recovery Function $r(t)$

In the following experiments, we use the SII model which has removed the effect of 1- to 5-order spreading cycles. First, we evaluate the performance of the SII model with different recovery functions $r(t)$: "Qualys," "Constant," and "Ratio" (see details in Section 7 of the supplementary file, which is available online). As shown in Fig. 13, the curves of our SII models match the simulations well, even if the recovery functions are different.

Considering the recovery function "*Qualys*," we also evaluate our SII model with different settings of the *first discovery period* (d) and the *time for 50 percent decreasing* (D). This concerns two cases: *Case 1* (what if the worms are easily to be cleaned from infected users, which means the number of infected users decreases quickly): $d = 40$ and the variable D equals to 40, 80, and 120 intervals; *Case 2* (what if security experts detect new worm incidents quickly): the variable d equals to 40, 80, and 120 intervals and $D = 40$. As shown in Fig. 14, our SII model presents a good approximation to the simulation result. When variable D is smaller, the speed of worms being cleaned is quicker. For case 2, as shown in Fig. 15, the curves of our SII models are close to the

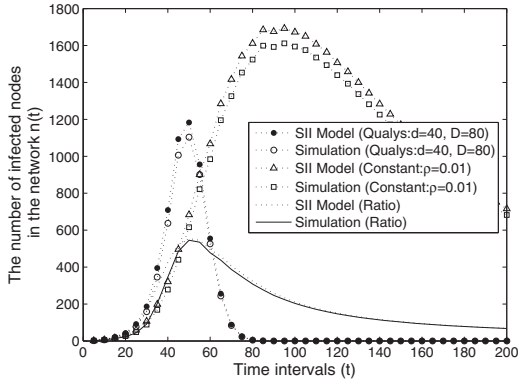


Fig. 13. The accuracy of the SII model affected by different recovery functions: Qualys, Constant, and Ratio.

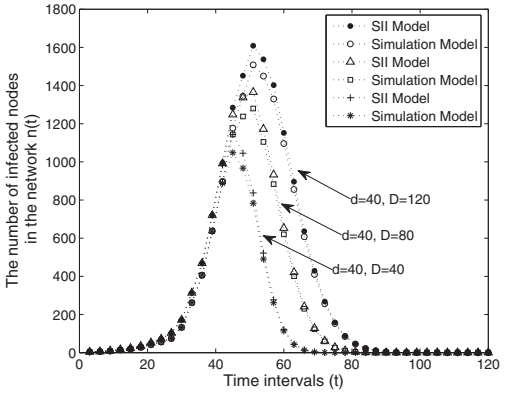


Fig. 14. The accuracy of recovery processes (Case 1: $d = 40$). Topology: $\alpha = 2.5$, $E(D) = 5.5$, $\lambda = 0.23$; message checking time: $T_i \sim Exp(40)$; propagation probability: $p_{ij} \sim N(0.25, 0.1)$.

simulation results, which indicate our SII model is accurate. We also find the *first discovery period d* has a critical effect on the spreading scale of social network worms. When d is small, which means security experts detect this worm quickly, the spreading scale will be largely constrained.

We also compare the SII model with those SIS models [5], [10], [13] and SIR models [6], [7], [8]. The difference of these models is caused by different considerations on the state transition of nodes. To exclude the impact of other factors, we derive the SIS and SIR model on the basis of the SII model. A susceptible user can be immunized in the SII model, but in SIR model this user cannot. Thus, we obtain an SIR model by revising (8) as $P(X_i(t) = Imm.) = P(X_i(t-1) = Imm.) + r(t) \times P(X_i(t-1) = Inf.)$.

Moreover, there is no *immunized* state in SIS model. An infected user is recovered and becomes susceptible again. Thus, we obtain an SIS model by setting $P(X_i(t) = Imm = 0.)$. As shown in Fig. 16, the result of the SII model decreases more rapidly than the SIS and SIR models, and is closer to the simulation.

6 Further Discussion

There is still much work to do on modeling the propagation of social network worms. First, the SII model proposed in this paper mainly focuses on nonreinfection worms [14],

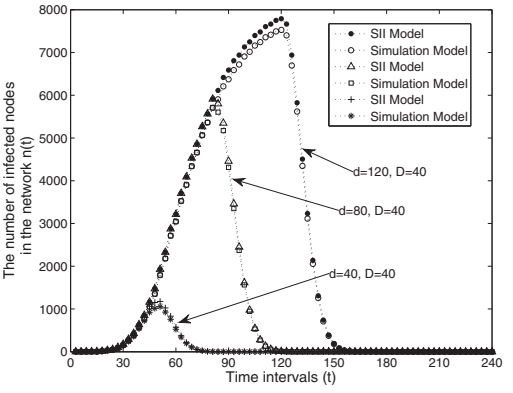


Fig. 15. The accuracy of recovery processes (Case 2: $D = 40$). Topology: $\alpha = 2.5, E(D = 5.5)$, $\lambda = 0.23$; message checking time: $T_i \sim Exp(40)$; propagation probability: $p_{ij} \sim N(0.25, 0.1)$.

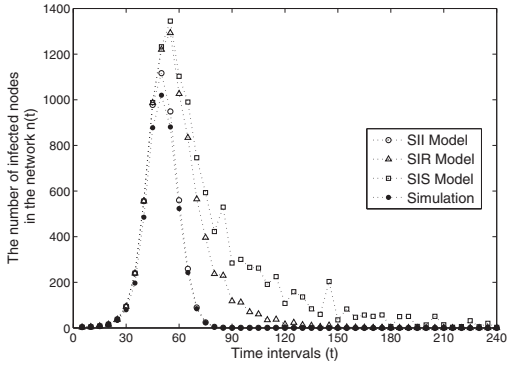


Fig. 16. The comparison of the SII model, SIR model, and SIS model with message checking time: $T_i \sim Exp(40)$; topology: $\alpha = 2.5$, $E(D = 5.5)$, $\lambda = 0.23, p_{ij} \sim N(0.5, 0.2)$; Qualys recovery function: $d = 40, D = 80$.

such as Melissa, Love Letter spreading in email network and Koobface spreading in Facebook. However, some other popular social network worms spread in a more aggressive way called reinfection [33]. Reinfection allows users to send out worm copies whenever they open the malicious messages, regardless of whether this user has been previously infected or not. Infected users also send out worm copies when certain events are triggered, such as restart of computers. This kind of worms can spread faster than nonreinfection worms in social networks. We leave the modeling of reinfection worms to the future work.

Second, we use simulations to evaluate the proposed SII model in this paper. In real-world scenarios, the spread of most social network worms is typically impossible to track given the directed, topological manner in which they spread. Some email worms, like Nyxem [34], once compromising a computer, will automatically generate a single http request for the URL of an online statistics page. However, as the report [34] said, the statistics of Nyxem also cannot present a precise investigation on the propagation of email worms due to the legitimate access, repeated probes and DDoS attacks to the web page. It should be pointed out that there is no real data set available for the evaluation of analytical models in this field.

Third, by implementing different recovery functions $r(t)$, our SII model can be used to derive best defense strategies, such as *where*, *when* and *how many* users are needed to deploy defense so that we can prevent social network worms from spreading. For example, a common view for the preferable positions of defense is at the highly connected users [14], [17] or those with most active neighbors [15]. However, according to our latest investigation [23], the users with higher random walk betweenness [24] or lower closeness values may be more suitable for slowing down the spread of social network worms.

Finally, we have a simple assumption in the modeling. That is users check social messages periodically. However, in the real world, the message checking time is more varied than assumed here. A possible solution is to presume the nonperiodic checking time for each user in the implementation of modeling. We leave this part of discussion to the future work.

7 Related Work

There have been substantial efforts in modeling the propagation dynamics of Internet worms in the last decade. Researchers mainly focus on two different spreading mechanisms [2]: the scanning-based spread and the topology-based spread. The former relies on infectious nodes searching victims which contain vulnerabilities in the whole or a specific IP space. The later depends on infectious nodes sending malicious messages to their topological neighbors, luring them to open those messages and then get infected. Readers could find scanning-based models in two surveys [27], [28]. The propagation of social network worms belongs to the topology-based spread. In this section, we mainly discuss this kind of models.

First, based on whether infected users return to susceptible state after being recovered, previous models can be classified into SIS [5], [9], [10], [11], [12] and SIR models [6], [7], [8]. Considering the spread character of social network worms, however, after users clean their infected computers or become more vigilant against a worm, they are unlikely to be infected any more. Thus, SIS models [5], [9], [10], [11], [12] are not appropriate for modeling the spread of social network worms. SIR models [6], [7], [8] may suit for the propagation of social network worms, but a more realistic scenario is that a susceptible user can be immunized directly without being infected at first, like the consideration in modeling scanning worms [25], [26]. Moreover, as discussed in [14], early epidemic models [5], [6], [7], [8] greatly overestimate the spreading speed due to their implicit "*homogeneous mixing*" assumption.

Second, the works [14], [15], [32], [16] rely on simulations to model the propagation of social network worms. Their simulation models avoid the problem of "*homogeneous mixing*" assumption but cannot provide analytical study on the propagation.

Third, the spatial-temporal model in [13] captures accurate topological information, and thus, can provide a better modeling than previous epidemic models. However, as discussed in Section 3.1, this model cannot presents the *temporal dynamics* and does not provide a strong approximation to *spatial dependence*.

Finally, Ganesh et al. [11] study the effect of topology on the spread of worms, but their discussion is based on the SIS model. The work [18] develops an SI model on a specific social network formed by smart phones, which also has the problem of "*homogeneous mixing*" assumption.

There are also some works which characterize the propagation dynamics of isomorphic worms, such as P2P worms [9], [10], [12], [17] and mobile worms [30], [31]. The works [9], [10], [12] focus on finding threshold conditions for fast extinction of worms. The work in [17] use a 0-1 matrix to model the propagation of P2P worms, but 0-1 matrix is not suitable to be used on social networks because the weight of any edge should be a probability rather than a logic value (0 or 1). In [30] and [31] Yan et al. explore the propagation dynamics of Bluetooth worms. Their models satisfy the mobile property of nodes by assuming devices homogeneously mixed, and thus, they are unlikely to work in real mobile social networks.

8 Conclusion

In this paper, we propose a novel SII model for the propagation of social network worms. This model is able to address two critical problems unsolved in the previous analytical models: temporal dynamics and spatial dependence. For the problem of temporal dynamics, we introduce a spatial-temporal synchronization process. By recording and accumulating each malicious message from neighbors in the last checking time period, the infection probability of each user can be estimated by computing joint probabilities of this user opening arrived messages. For the problem of spatial dependence, we find the essence of spatial dependence is the spreading cycles formed in the propagation procedure. By removing the effect of these spreading cycles, we are able to obtain a stronger approximation to the spatial dependence.

We conduct a number of experiments to evaluate the correctness of the SII model. The experiments show that the result of the SII model is close to the simulation, which means our SII model is accurate for modeling the spread of social network worms. We also compare the SII model with epidemic models, a spatial-temporal model, the SIS model and the SIR model. The evaluation results indicate that our SII model is more suitable for modeling the propagation of social network worms. We believe our work presented in this paper is of great significance to both academic aims and practical usage.

Acknowledgments

This work is supported in part by grants from the General Research Fund of the Hong Kong SAR, China No. (CityU 114609 and CityU 114012) and CityU Applied R&D Funding (ARD) No. 9681001 and CityU Strategic Research Grant No. 7008110; CityU Applied Research Grant (ARG) No. 9667052; ShenZhen-HK Innovation Cycle Grant No. ZYB200907080078A and NSF (China) No. 61070222.

References

[1] M.E.J. Newman, *Networks: An Introduction,* pp. 36-39, Oxford, 2010.

[2] N. Weaver, V. Paxson, S. Staniford, and R. Cunningham, "A Taxonomy of Computer Worms," *Proc. First ACM Workshop Rapid Malcode (WORM '03),* Oct. 2003.

[3] H. Ebel, L. Mielsch, and S. Bornholdt, "Scale-Free Topology of E-Mail Networks," *Physical Rev. E,* vol. 66, 2002.

[4] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," *Proc. ACM SIGCOMM,* 2007.

[5] R. Pastor-Satorras and A. Vespignani, "Epidemic Spreading in Scale-Free Networks," *Physical Rev. Letters,* vol. 86, pp. 3200-3203, 2001.

[6] Y. Moreno, J.B. Gomez, and A.F. Pacheco, "Epidemic Incidence in Correlated Complex Networks," *Physical Rev. E,* vol. 68, 2003.

[7] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, "Epidemic Outbreaks in Complex Heterogeneous Networks," *European Physical J. B,* vol. 26, pp. 521-529, 2002.

[8] M. Boguna, R. Pastor-Satorras, and A. Vespignani, "Epidemic Spreading in Complex Networks with Degree Correlations," *Statistical Mechanics of Complex Networks,* vol. 625, pp. 127-147, 2003.

[9] R. Thommes and M. Coates, "Epidemiological Modeling of Peer-to-Peer Viruses and Pollution," *Proc. IEEE INFOCOM,* pp. 1-12, 2006.

[10] D. Chakrabarti, J. Leskovec, C. Faloutsos, S. Madden, C. Guestrin, and M. Faloutsos, "Information Survival Threshold in Sensor and p2p Networks," *Proc. IEEE INFOCOM,* pp. 1316-1324, 2007.

[11] A. Ganesh, L. Massoulie, and D. Towsley, "The Effect of Network Topology on the Spread of Epidemics," *Proc. IEEE INFOCOM,* pp. 1455-1466, 2005.

[12] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, "Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint," *Proc. IEEE Symp. Reliable Distributed Systems,* pp. 25-34, 2003.

[13] 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.

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

[15] 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 Comm. and Security (ASIACCS '11),* 2011.

[16] W. Fan and K. Yeung, "Online Social Networks Paradise of Computer Viruses," *Physics A: Statistical Mechanics and its Applications,* vol. 390, pp. 189-197, 2011.

[17] X. Fan and Y. Xiang, "Modeling the Propagation of Peer-to-Peer Worms," *Future Generation Computer Systems,* vol. 26, pp. 1433-1443, 2010.

[18] S.M. Cheng, W. 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.

[19] Y.Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, "Analysis of Topological Characteristics of Huge Online Social Networking Services," *Proc. ACM Int'l Conf. World Wide Web (WWW '07),* 2007.

[20] Y. Wang, S. Wen, S. Cesare, W. Zhou, and Y. Xiang, "Eliminating Errors in Worm Propagation Models," *IEEE Comm. Letters,* vol. 15, no. 9, pp. 1022-1024, Sept. 2011.

[21] M. Fossi and J. Blackbird, "Symantec Internet Security Threat Report 2010," technical report, Symantec Corp., Mar. 2011.

[22] M.E.J Newman, S. Forrest, and J. Balthrop, "Email Networks and the Spread of Computer Viruses," *Physical Rev. E,* vol. 66, 2002.

[23] S. Wen, W. Zhou, Y. Wang, W.L. 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.

[24] M. Newman, "A Measure of Betweenness Centrality Based on Random Walks," *Social Networks,* vol. 27, pp. 39-54, 2005.

[25] Z. Chen, L. Gao, and K. Kwiat, "Modelling the Spread of Active Worms," *Proc. IEEE INFOCOM,* pp. 1890-1900, 2003.

[26] C.C. Zou, W. Gong, and D. Towsley, "Code Red Worm Propagation Modelling and Analysis," *Proc. Ninth ACM Computer and Comm. Security Conf. (CCS '02),* pp. 138-147, 2002.

[27] Y. Xiang, X. Fan, and W. Zhu, "Propagation of Active Worms: A Survey," *Int'l J. Computer Systems Science and Eng.,* vol. 24, pp. 157-172, 2009.

[28] C.C. Zou, D. Towsley, and W. Gong, "On the Performance of Internet Worm Scanning Strategies," *Performance Evaluation,* vol. 63, pp. 700-723, 2006.

[29] P. Mahadevan, D. Krioukov, K. Fall, and A. Vahdat, "Systematic Topology Analysis and Generation Using Degree Correlations," *Proc. ACM SIGCOMM,* 2006.

[30] G. Yan, H.D. Flores, L. Cuellar, N. Hengartner, S. Eidenbenz, and V. Vu, "Bluetooth Worm Propagation: Mobility Pattern Matters," *Proc. Second ACM Symp. Information Computer and Comm. Security (ASIACCS '07),* 2007.

[31] G. Yan and S. Eidenbenz, "Modeling Propagation Dynamics of Bluetooth Worms (Extended Version)," *IEEE Trans. Mobile Computing,* vol. 8, no. 3, pp. 353-367, Mar. 2009.

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

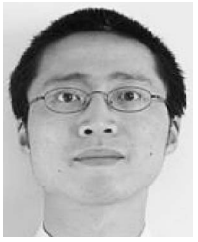
[33] M.C. Calzarossa and E. Gelenbe, *Performance Tools and Applications to Networked Systems: Revised Tutorial Lectures.* Springer-Verlag, 2004.

[34] Moore and C. Shannon, "The Nyxem Email Virus: Analysis and Inferences," technical report, CAIDA, Feb. 2006.

Sheng Wen received the graduate degree in computer science and technology from Lanzhou Jiaotong University, Gansu, China, in 2003. He is currently working toward the PhD degree at the Central South University, Changsha, China, under the supervision of Prof. Weijia Jia. Since 2011, he has been working with the school of information technology, Deakin University, Victoria, Australia, under the supervision of Prof. Wanlei Zhou and Dr. Yang Xiang. His research interests include modeling of virus spread, defence strategies of the Internet threats, and locating the authors of computer viruses. He is a student member of the IEEE and the IEEE Computer Society.

Wei Zhou received the BEng and MEng degrees from Central South University, Changsha, China, in 2005 and 2008, respectively, both in computer science. She is currently working toward the PhD degree in the School of Information Science and Engineering, Central South University and a joint training PhD student in the School of Information Technology, Deakin University, Victoria, Australia. Her research interests include distributed systems, computer networks, and network security.

Jun Zhang received the PhD degree from the University of Wollongong, Australia, in 2011. He is currently with the School of Information Technology, Deakin University, Victoria, Australia. His research interests include network and system security, pattern recognition, and multimedia retrieval. He has published more than 30 research papers in the reputed journals and conferences, such as *IEEE Transactions on Image Processing*, *IEEE Transactions on Parallel and Distributed Systems*, and IEEE International Conference on Image Processing. He received 2009 Chinese government award for outstanding self-financed student abroad. He is a member of the IEEE and the IEEE Computer Society.



Yang Xiang received the PhD degree in computer science from Deakin University, Victoria, Australia. He is currently with the School of Information Technology, Deakin University. His research interests include network and system security, distributed systems, and networking. In particular, he is currently leading the Network Security and Computing Lab developing active defense systems against large-scale distributed network attacks. He is the chief investigator of several projects in network and system security, funded by the Australian Research Council. He has published more than 100 research papers in many international journals and conferences, such as *IEEE Transactions on Computers*, *IEEE Transactions on Parallel and Distributed Systems*, *IEEE Transactions on Information Security and Forensics*, and *IEEE Journal on Selected Areas in Communications*. One of his papers was selected as the featured article in the April 2009 issue of *IEEE Transactions on Parallel and Distributed Systems*. He has published two books: *Software Similarity and Classification* (Springer) and *Dynamic and Advanced Data Mining for Progressing Technological Development* (IGI-Global). He has served as the program/general chair for many international conferences such as ICA3PP 12/11, IEEE/IFIP EUC 11, IEEE TrustCom 11, IEEE HPCC 10/09, IEEE ICPADS 08, NSS 11/10/09/08/07. He has been the PC member for more than 50 international conferences in distributed systems, networking, and security. He serves as the associate editor of *IEEE Transactions on Parallel and Distributed Systems* and the editor of Journal o*f Network and Computer Applications*. He is a senior member of the IEEE and the IEEE Computer Society.

Wanlei Zhou received the BEng and MEng degrees from Harbin Institute of Technology, China, in 1982 and 1984, respectively, the PhD degree from the Australian National University, Canberra, in 1991, and the DSc degree from Deakin University, Victoria, Australia, in 2002. He is currently the chair professor of information technology and the head of School of Information Technology, Faculty of Science and Technology, Deakin University, Melbourne, Australia. His research interests include distributed and parallel systems, network security, mobile computing, bioinformatics, and e-learning. He has published more than 200 papers in refereed international journals and refereed international conferences proceedings. Since 1997, he has been involved in more than 50 international conferences as general chair, steering committee chair, PC chair, session chair, publication chair, and PC member. He is a senior member of the IEEE and the IEEE Computer Society.

Weijia Jia received the BSc and MSc degrees from Central South University, Changsha, China, in 1982 and 1984, respectively, and the master of applied science and PhD degrees from the Polytechnic Faculty of Mons, Mons, Belgium, in 1992 and 1993, respectively, all in computer science. He is currently a full professor with the Department of Computer Science and the director of Future Networking Center, ShenZhen Research Institute, City University of Hong Kong (CityU). He joined the German National Research Center for Information Science, Bonn (St. Augustine), Germany, from 1993 to 1995 as a research fellow. In 1995, he joined the Department of Computer Science, CityU, as an assistant professor. His research interests include next-generation wireless communication, protocols, and heterogeneous networks, distributed systems, and multicast and anycast QoS routing protocols. He is a senior member of the IEEE and the IEEE Computer Society.