EURASIP Journal on Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Virus propagation power of the dynamic network

Fu Cai[1], Huang Qingfeng[2*], Han LanSheng[1], Shen Li[1] and Liu Xiao-Yang[3]

Abstract

With the development of mobile networks, propagation characteristics and defense mechanism of the virus have attracted increasing research attention. But current researches are mainly concerned with static network topology or community structure and some studies focus on characteristics of virus and malware. Little attention is paid to the influence of dynamic changes of network topology to virus propagation. Meanwhile, many studies focus on the threshold rate of the infection (or the immunization rate) for virus outbreak. In this paper, we present a new way to assess and restrain virus propagation by proposing the concepts of propagation power and propagation structure. Three basic propagation structures are presented through which the infection risk can be quantified, and a framework is proposed to assess the impact of virus propagation in different network structures. The relationship between the speed of virus propagation and network structure is also explored. An algorithm is designed to assess the propagation power in a dynamic network with no need to redetect the community under the dynamic network which may significantly improve the efficiency of assessment in a large-scale dynamic network. This study offers a feasible approach for quantifying the risk of virus infection in the network community which is valuable for designing and optimizing the virus defense systems.

Keywords: Dynamic network; Propagation power; Virus restraint; Basic propagation structure

1. Introduction

With the increasing popularity of mobile devices, virus propagation in mobile networks has become a major security concern [1-4]. A virus brings not only a potential threat but also real problems such as leakage of personal data and remote access control [4,5]. A more dangerous threat is its large-scale outbreak which can paralyze the whole network.

The current research on network virus propagation can be divided into two categories: some researchers model the propagation of virus to find the threshold of a large-scale breakout, while others attempt to study the mechanism of a restraining virus.

Since network virus propagation in a mobile network is similar to that of biological virus propagation, four classic models (susceptible-infected-susceptible (SIS), susceptible-infected-recovered (SIR), susceptible-exposed-infected-recovered (SEIR), and susceptible-infected-directed-removed (SIDR)) used to describe biological virus propagation are used in describing network virus propagation. There are two major lines of research on mobile network adopting this approach.

One line of studies focuses on human epidemic through the mobile network. Using GPS and GSM, researchers can analyze people's social activities through collecting the position data of people carrying the devices. For example, researchers collect data from a high school in the USA by a set TelosB sensor node in 788 people in the school. After analyzing the collected data, it is concluded that biological virus can be immunized only in the area where vaccine injection rate is high [6]. Using the data collected from a mobile network, Fenichel et al. [7] observed contact between people and proposed a scheme to improve the resistance of biological virus infection by balancing the threat of virus and the benefit of social networking.

Another line of studies is concerned with virus in intelligent mobile devices. It is found that the virus will break out if the market share of intelligent mobile devices exceeds a certain threshold [8]. Some method has been put forward to address the problem of mobile phone virus [9-11]. But the mobility of smart devices can greatly

* Correspondence: qfhuang@mail.hust.edu.cn
[2]Network and Computer Center, Huazhong University of Science and Technology, Wuhan 430074, China
Full list of author information is available at the end of the article

increase the speed of virus propagation even if most nodes in the network remain still. Therefore, given the mobility of the nodes in the mobile network, the propagation of virus is likely to be exacerbated [12].

In the study of virus defense in the mobile network, it is found that virus propagation in a mobile network is difficult to restrain because the topology of the network is changing frequently due to the mobility of devices. There are two kinds of strategies restraining the virus in a mobile network: immunization strategy [13-22] and local strategy. In most cases, immunization strategies are implemented based on centralized distribution and static network [13-21]. But in a large-scale dynamic network, immunization strategy should cover 80% of nodes by stochastic immunization strategy or the whole topology should be known using a special immunization strategy [13,14]. Through local strategies, once a node finds itself infected, it sends an antivirus message to others immediately and the virus can be cleared accordingly [23-26].

Zyba et al. [23] presents an ideal scheme to restrain virus propagation through which a node can detect a virus locally. Once the virus has been found, the infected node immediately cuts off the communication or sends an antivirus message to adjacent nodes. However, this study is based on an idealistic but unrealistic assumption that the mobile devices possess superb computational power [27,28]. Hui et al. [24,25] propose a message dissemination mechanism through social community, but the mechanism requires that the virus is locally detected. Jackson and Creese [26] study virus propagation from the perspective of human behaviors.

In this paper, we propose a new way to research on virus propagation based on the network structure and its dynamic change. We propose the concepts of propagation power and propagation structure. It is assumed that the risk of virus propagation depends on the speed of virus propagation in the network. And the major factor influencing the speed of virus propagation is propagation power which can be quantified by analyzing the network structure based on three basic propagation structures. This approach has two distinct advantages. On the one hand, through the propagation power, we can assess the risk of the virus propagation quantitatively and take virus-restraining measures in advance. On the other hand, a guideline may be developed for network construction and optimization to restrain the virus.

The rest of the paper is organized as follows. Section 2 reviews the current researches and points out the major concern of our study. Section 3 offers definitions of various important concepts. Section 4 presents the calculation of fundamental propagation power in static network, whereas the propagation power in dynamic network is computed in Section 5. Section 6 presents the experiment and data analysis, and Section 7 offers several case studies. The paper is concluded in Section 8 along with some future works.

2. Problem statement

As mentioned above, several schemes have been proposed to model and restrain virus propagation, but these schemes may have some weaknesses with regard to the current prevalent virus. We analyze some famous advanced persistent threat (APT) viruses to illustrate our research focus.

First, there are several classic virus propagation models such as SIS, SIR, SEIR, and SIDR where the threshold propagation rates λ are defined. But it is also found that the propagation rate is highly related to the size of the networks [29]. To reduce the propagation rate, it is assumed that the network size is infinite. Since in reality the network is usually finite, how can we assess the risk of virus propagation? For example, Figure 1 shows two real microblog communities. If we can accurately assess the risk of the two communities, then we can take measures in advance.

Second, the current researches on the outbreak of virus or the risk of propagation focus on the threshold infection rate. When the infection rate reaches a certain value, the virus will break out. But historically, viruses often utilize one or more 0-day vulnerabilities. And it will take a long time for virus patch to be released. Before the release of the patch, the infection rate of virus is almost 1, as shown in the cases of shockwave, Blaster Worm, and the APT virus in 2012. Table 1 shows some famous APT viruses,

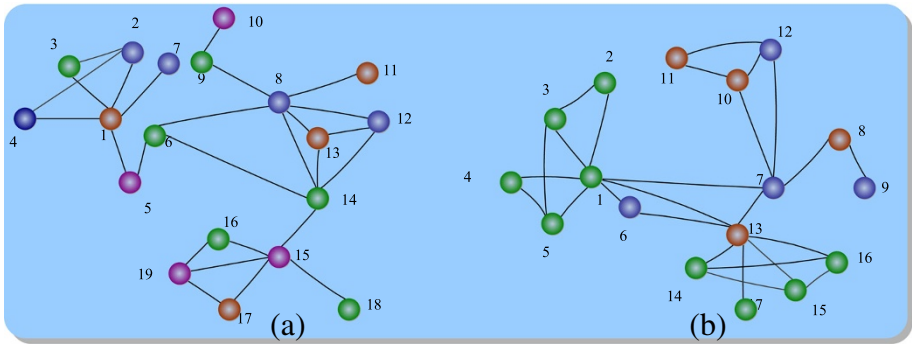

Figure 1 Two microblog communities (a, b).

Table 1 Famous virus utilizing 0-day vulnerability

Name	Date of discovery	0-day count
Aurora	January 2010	1
Stuxnet	June 2010	7
RSA	March 2011	1
Flame	May 2012	1

almost all of which utilize the 0-day vulnerability. The current research has uncovered 0-day malware in a mobile android device [30]. So what measures can be taken to restrain the propagation of such viruses?

Third, the current methods aim to remedy the situation after the discovery of the virus. What preventive measures can be taken in advance? More specifically, what can be done with respect to network design and optimization? For example, Figure 2 shows the topology of two independent networks and a number of possible ways to connect the two networks. Which is the best way to connect in order to restrain the virus propagation?

Fourth, due to the rapid development of mobile network and the mobility of mobile devices (smartphone, PDA), the network topology changes frequently. How can we assess the risk of the dynamic network efficiently? For example, Figure 3 shows three statuses of a microblog community. It is obviously inefficient to recalculate the propagation power of the entire community. Efficiency can be greatly improved if only the dynamic part of the community is calculated.

To address the issues above, it is insufficient to study the propagation model or virus defense. The network topology has to be taken into consideration as it is the basic environment for virus propagation. If the speed of virus propagation at the layer of network structure can be reduced, more time and space will be available to restrain the virus. Therefore, we propose the propagation power used to quantify the risk of virus propagation in the network. The network may be analyzed based on three basic propagation structures. We also propose a method to calculate the risk of virus propagation in the dynamic network. In short, we believe that our scheme is a new approach to restrain virus propagation taking into account the network structure which has been largely neglected in current researches.

3. Related definitions

In order to quantify the risk of virus propagation of a complex network structure we define three basic propagation structures of a network according to the speed of propagation. And then we define the propagation power to quantify the risk of virus propagation on the network. With the two definitions, we can better understand the relationship between the risk of network and the speed of virus propagation.

3.1. Definition of SBS and speed of virus propagation

Definition 1 Spreading network is defined as a connected graph $G = (V, E)$, where V is the set of vertices and E is the set of edges. For any e_{ij} that belongs to E, if $e_{ij} = 1$, v_i and v_j are connected.

Definition 2 The speed of virus propagation v_0 is defined as $v_0 = n/t$, where n is the number of infected nodes in the network and t is the time taken to infect the whole network.

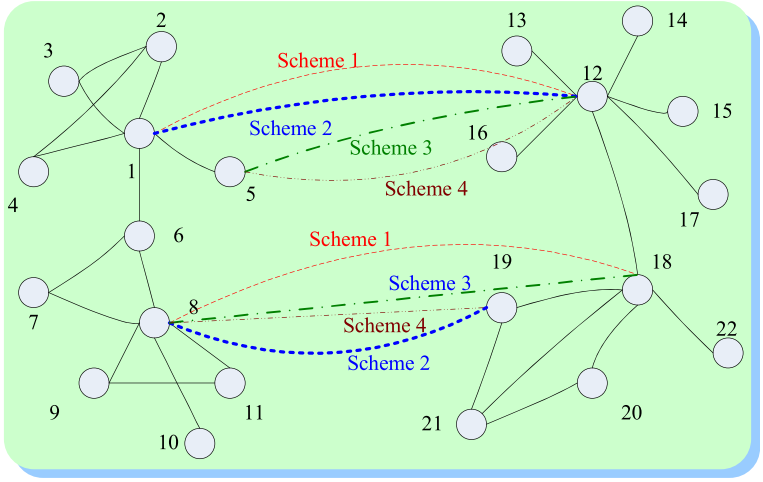

Figure 2 Optimization of network based on virus defense.

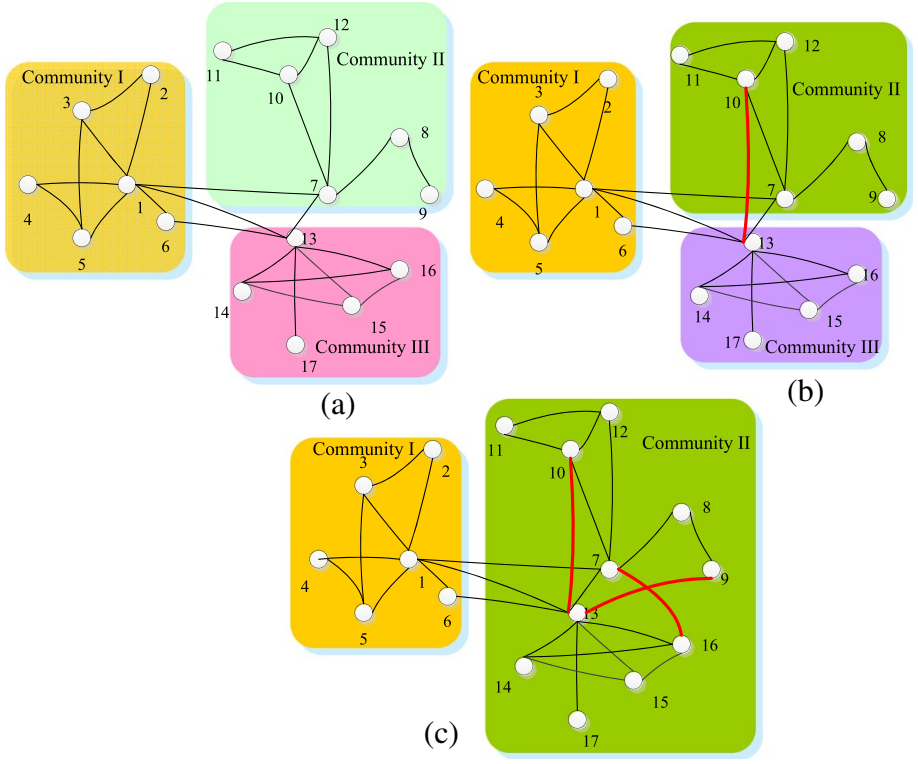

Figure 3 The changes of risk of virus propagation in a microblog community (a, b, c).

Definition 3 The speed of virus propagation of the network v is defined as $v = \left(\frac{1}{n}\right)\sum_{i=1}^{n}\frac{n}{a_i t_0}$, where n is the number of the nodes in the network, α_i is the number of infected in the whole network by the node i, and t_0 is the time required for a successful infection.

Definition 4 The three basic propagation structures include linear propagation structure (LPS), ring propagation structure (RPS), and star propagation structure (SPS).

Definition 5 LPS is a connected graph, which satisfies $1 \le d_i \le 2$, where d_i is the degree of any node i in the LPS. There must exist two nodes whose degrees are equal to 1.

Definition 6 RPS is a connected graph, where the degrees of all nodes are equal to 2 and the number of the nodes is greater than 2.

Definition 7 SPS is a connected graph, where the degree of one node is greater than or equal to 3 and the degree of other nodes is equal to 1 and the number of the nodes in the structure is greater than 3. All basic structures are shown in Figure 4.

3.2. Definition of propagation power

Speeds of virus propagation vary in different network structures. In this section, the propagation power is defined to quantify the effect of different structures on the speed of virus propagation. Intuitively, virus propagation is faster in SPS than in the other two structures, whereas RPS is faster than LPS. Under the same situation, the speed of virus propagation is mainly determined by the structures.

To quantify the propagation power, we should consider $E_n = \frac{1}{n}\sum_{i=1}^{n} p_i$ to assess the risk of virus of a structure, where probability $p_i = p(x_1 = x_2 = \ldots = x_n = 1 | x_i = 1)$ and n is the number of nodes in network. Quantification is

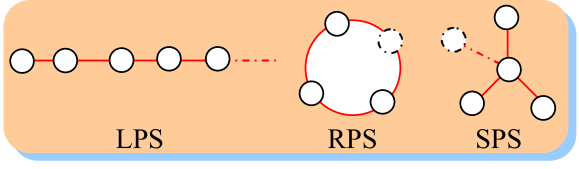

Figure 4 Basic propagation structures.

reasonable in terms of theory for the reason that the larger the value of E_n is, the higher the risk of virus propagation. However, it is impracticable to work out the probability p_i. Thus, we chose another way which investigated the number of hops a virus needed from an original node spreading to all nodes of the network. To simplify the calculation, it is assumed that the probability of virus spreading from one to another is 1. The propagation power was defined as follows:

Definition 8 Propagation power of a network is $F = \left(\frac{t_0}{n}\sum_{i=1}^{n}\text{disp}_i\right)^{-1}$, where disp_i is the number of infection caused by the ith node, n is the number of nodes in the network, and t_0 is the time required for a successful infection.

With the definition of propagation power, the speed of virus propagation in a network can be evaluated. If the propagation power is greater, the time for virus propagation would be smaller and vice versa.

4. Calculation of fundamental propagation power

4.1. Calculation of propagation power in a large network

To calculate the propagation power, we present the algorithm according to the definition of propagation power to be applied in a large-scale network as follows.

The number of hops required for a virus to spread from an original node to all nodes of the network can be considered as the maximum value of the shortest path length which is the shortest length one needs to travel from the original node to another. Thus, the problem comes down to a task of obtaining the shortest path.

The main idea of calculating the propagation power is that a shortest path tree for each node in the network is worked out by the Dijkstra algorithm first, and then the height of the tree which represents the number of hops is obtained by calculating F. The algorithm is described as follows:

(1) Input graph G and set propagation power $F = 0$;
(2) Calculate the shortest path tree SPT_i $(1 \le i \le n)$ by Dijkstra (G, v), the root of which is v;
(3) Get the height of tree SPT_i, $\text{disp}_i = \text{TreeHeight}(v)$;
(4) Return F, $F = \left(\frac{t_0}{n}\sum_{i=1}^{n}\text{disp}_i\right)^{-1}$.

As the time complexity of the optimized Dijkstra algorithm is $O(|E| \log |V|)$ [31], since every vertex must be computed, the complexity of the algorithm above is $O(|V| \times |E| \log |V|)$. Figure 5 shows the algorithm.

As shown in Figure 6, we present an example to show how the algorithm works:

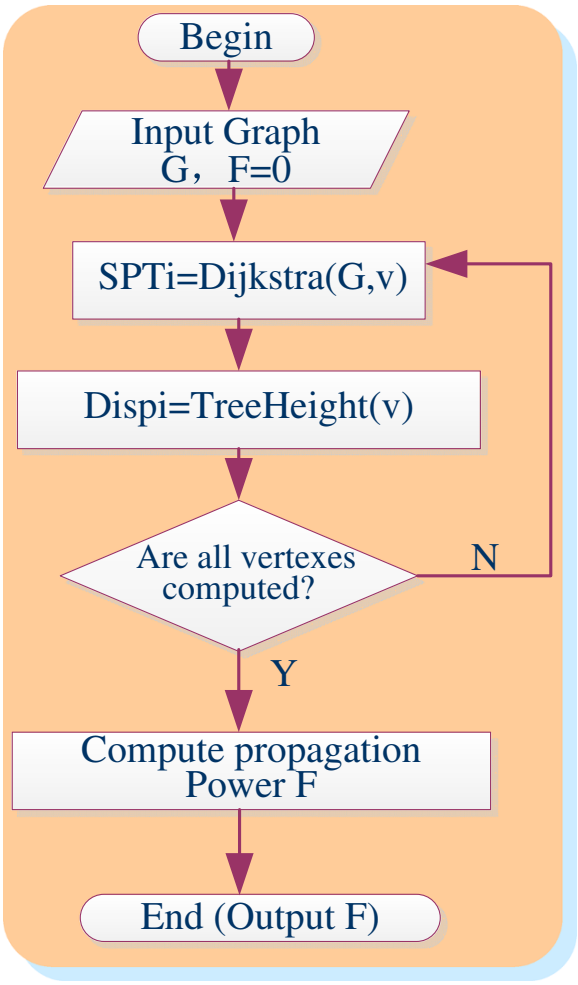

Figure 5 The propagation power calculation algorithm.

1. Calculate the propagation power of Figure 6a:

$$F = \left[\frac{t_0}{5} \times (3+2+3+2+3)\right]^{-1} = \frac{0.38}{t_0}$$

2. Calculate the propagation power of Figure 6b:

$$F = \left[\frac{t_0}{5} \times (2+2+2+2+2)\right]^{-1} = \frac{0.5}{t_0}$$

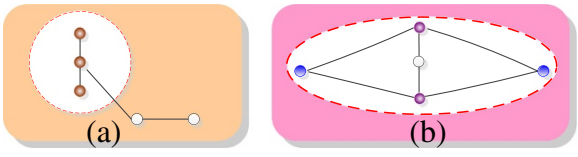

Figure 6 Example of calculation (a, b).

In Figure 6, although the vertex numbers are the same, the propagation powers are different, which suggests that the virus risk is different.

4.2. Relation between propagation power and infection time

In this section, we present some theorems about the propagation power F and the infection time n/v.

First, we show the mathematical expressions of the propagation power of SPS and the propagation speed of SPS.

Relationships between the propagation power F of LPS and number of nodes n in the network and relationships between the speed v of LPS and number of nodes n in the network can be expressed as follows:

$$v = \frac{1}{n}\sum_{1}^{n}\frac{n}{\alpha_i t_0} = 2\times\sum_{i=\frac{n}{2}}^{n-1}\frac{1}{it_0} + k \tag{1}$$

$$\text{where } k = \begin{cases} 0 & (\text{while } n \text{ is an even number}) \\ \dfrac{1}{\left\lfloor\frac{n}{2}\right\rfloor t_0} & (\text{while } n \text{ is an odd number}) \end{cases}$$

$$F = \begin{cases} \left(\dfrac{2t_0}{n}\sum\limits_{i=\frac{n}{2}}^{n-1} i\right)^{-1} = \left[t_0\left(n-1+\dfrac{n}{2}\right)\times\dfrac{n}{2}\times\dfrac{1}{2}\times\dfrac{2}{n}\right]^{-1} \\ \qquad = \left(\dfrac{3n-2}{4}t_0\right)^{-1} \ (\text{while } n \text{ is an even number}) \\ \left(\dfrac{2t_0}{n}\sum\limits_{i=\frac{(n-1)}{2}}^{n-1} i + \dfrac{n-1}{2\times n}\right)^{-1} = \left[\dfrac{(3n-1)(n-1)t_0}{4n}\right]^{-1} \\ \qquad (\text{while } n \text{ is an odd number}) \end{cases} \tag{2}$$

Relationships between propagation power F of RPS and number of nodes n in the network and relation between speed v of RPS and number of nodes n in the network can be expressed as follows:

$$v = \frac{1}{n}\sum_{1}^{n}\frac{n}{\alpha_i t_0} = \frac{n}{\lfloor n/2\rfloor t_0} = 2/t_0 \tag{3}$$

$$F = \left(\frac{t_0}{n}\sum_{i=1}^{n} n/2\right)^{-1} = (t_0\lfloor n/2\rfloor)^{-1} \tag{4}$$

Relationship between propagation power F of SPS and number of nodes n in the network and relationship between speed v of SPS and number of nodes n in the network can be expressed as follows:

$$v = \frac{1}{n}\sum_{1}^{n}\frac{n}{\alpha_i t_0} = \frac{n-1}{2t_0} + \frac{1}{t_0} = \frac{n+1}{2t_0} \tag{5}$$

$$F = \left(\frac{t_0}{n}\sum_{n-1} 2 + \frac{1}{n}\right)^{-1} = \left(\frac{2n-1}{n}t_0\right)^{-1} \tag{6}$$

According to the above formulae, it may be concluded that v is the average speed of the nodes infecting the whole network, and propagation power expresses the risk of virus propagation in the network. Assuming the two variables are related, three theorems are presented as follows:

Theorem 1 *When n approaches infinity in LPS,* $F^{-1} = \alpha(1/v)$, *where* $3/4 \le \alpha \le 3/2$; *In RPS,* $F^{-1} = \alpha(1/v)$, *where* $\alpha = 1$; *when n approaches infinity in SPS,* $F^{-1} = \alpha(1/v)$, *where* $\alpha = 1$.

Proof In LPS:

$$n/v = n/\left(\frac{1}{n}\sum_{i=\frac{n}{2}}^{[n-1]}\frac{n}{it_0} + k\right)$$

$$n/\left(\frac{1}{n}\sum_{i=\frac{n}{2}}^{[n-1]}\frac{n}{it_0} + k\right) \le n/v \le n/\frac{1}{n}\sum_{i=\frac{n}{2}}^{[n-1]}\frac{n}{it_0}$$

$$n/\left(\frac{1}{n}\sum_{i=\frac{n}{2}}^{[n-1]}\frac{n}{n/2t_0} + k\right) \le n/v \le n/\left(\frac{1}{n}\sum_{i=\frac{n}{2}}^{[n-1]}\frac{1}{t_0}\right)$$

$$t_0/2 \le 1/v \le t_0$$

When n approaches infinity, substitute $F^{-1} = (3/4)t_0$ into the above inequalities. Hence, $3/4 \le \alpha \le 3/2$.

We can prove the relationships in RPS and SPS in the same way.

Theorem 2 *The ratio of the reciprocal of propagation energy of any network structure to time* (n/v) *has upper and lower limits. The upper limit function of the two variables is* $f(n, \alpha) = \alpha \times n$, *and the lower limit function is* $f(n, \alpha) = a/n$, *where* $\alpha = {t_0}^2$.

Proof Since $F = \left(t_0\frac{1}{n}\sum \mathrm{disp}_i\right)^{-1}$

and $1 \le \mathrm{disp}_i \le n$.

Therefore, $\frac{1}{t_0 n}\sum_i 1 \le F^{-1} \le \frac{1}{t_0 n}\sum_i n$

$$\frac{1}{nt_0} \le F^{-1} \le 1/t_0$$

$$v = (1/n)\sum_n \frac{n}{\alpha_i t_0}$$

And because $1 \le \alpha_i \le n$,

$$\frac{1}{n}\sum_{n}\frac{1}{t_0} \le v \le \frac{1}{n}\sum_{n}\frac{n}{t_0}$$

$$1/t_0 \le v \le n/t_0$$

$$t_0 \le \frac{n}{v} \le nt_0$$

From above, it can be concluded that $\sup \frac{n}{v} = F^{-1} \times f(n,\alpha)$, where $f(n, \alpha) = \alpha n$ $\inf \frac{n}{v} = F^{-1} \times f(n,\alpha)$, where $f(n, \alpha) = a/n$.

Theorem 3 *The speeds of virus propagation in three SBSs are different when the numbers of nodes are the same. The speed in LPS v_{line}, the speed in RPS v_{ring}, and the speed in SPS v_{star} satisfy the inequality $v_{line} < v_{ring} < v_{star}$.*

Proof Since

$$v_{ring} = \frac{1}{n}\sum_{1}^{n}\frac{n}{\alpha_i t_0} = 2 \times \sum_{n-1}^{\lfloor n/2 \rfloor}\frac{n}{it_0} + k$$

$$2 \times \sum_{n-1}^{\lfloor n/2 \rfloor}\frac{n}{it_0} < \frac{1}{n}\sum_{1}^{n}\frac{n}{\alpha_i t_0}$$

According to the Equation 3, $v_{\text{line}} < v_{\text{ring}}$

$$\frac{n}{\lfloor n/2 \rfloor t_0} < \frac{n+1}{2t_0}$$

According to Equations 3 and 5, $v_{\text{ring}} < v_{\text{star}}$.

Therefore, $v_{\text{line}} < v_{\text{ring}} < v_{\text{star}}$.

From formulae 1, 3, and 5, the following conclusions can be made:

1. The speed of virus propagation in LPS decreases when the number of nodes increases.
2. The speed of virus propagation in RPS fluctuates around $2\,t_0$.
3. The speed of virus propagation in SPS increases when the number of nodes increases. The relationship between the speed and the number of nodes is linear.

Through the above analysis, we can see that the structure of the network is the main factor in virus propagation and the most dangerous structure is the SPS. When a network has more SPS, the risk of propagation is higher.

According to Equations 1, 3, and 5, the speed of virus propagation is not only related to the scale of the network, but it also largely depends on the structure of the network. The star structure has played the most significant role in increasing the speed of virus propagation. In any given network, the more star structures there are and the bigger the propagation power F is, the faster the virus propagation would be. The influence of the star structure also increases with the increase of the number of nodes. The influence of the ring structure is smaller and that of the linear structure is the smallest. Overall, the propagation is slower in a network with more linear structures and ring structures than that in one with more star structures.

5. Calculating propagation power in dynamic network

It is complex to evaluate the risk of virus propagation in a large-scale dynamic network due to the following reasons: first, when the network scale is large, it would be too complex to compute the propagation power of the entire network; second, since the risk of virus propagation changes frequently due to the change of the network structure, a large amount of computation is required to reevaluate the propagation risk after each change. In this section, we propose a feasible algorithm to compute the propagation power of virus in large-scale dynamic networks based on dynamic community mining algorithm in [26], which can efficiently evaluate the propagation risk of virus in a community network.

Four types of changes can be observed in a dynamic network, namely adding edges, reducing edges, increasing nodes, and reducing nodes, which lead to different derived networks and the need to recalculate propagation power in the network. For each type of change, we will specifically describe how the algorithm calculates the propagation power in the dynamic network.

5.1. Add a new node

Adding a new node in the network can change the community structure in the following two ways for which the propagation power of the community has to be recalculated.

First, when a new node not connected to any other nodes in the network is added, a new community is added to the entire network. Since there is no change in

propagation power in the community, the value of propagation power remains unchanged.

Second, when a new node connected with other nodes in the network is added, this situation is more complicated since the structure of the network community may change. When the node is added, since other nodes connected to this node may belong to one or more of the communities, only the propagation power of the community where the node is added needs to be recalculated. The propagation power of the entire community has to be updated accordingly. Algorithm 1 is described in detail below:

Algorithm 1 AddNode

Input: F_t represented value of propagation energy of the community in time t, and community structure and new node u

Output: F_{t+1} represented the updated value of propagation energy of the community in time $t+1$ and the new community structure

If(d_u ==0) then

Remain F_t unchanged

Update $C^{t+1} = C^t \cup \{u\}$

Else

Recalculate F for the community which u join in

Update $F_{t+1} = F_t + \Delta F$

Update $C^{t+1} = (C^t \setminus C) \cup (C \cup \{u\})$

End if

5.2. Add a new edge

Adding an edge may lead to a variety of changes in structure. We recalculate the propagation power of community after each change, which may be one of the two following cases.

When the two nodes of a new edge belong to the same community, the new edge is added to the same community. We simply need to recalculate the propagation power after adding the new edge and update the value of propagation power of the entire community.

When the two nodes of a new edge belong to different communities, a new external edge is added which may cause two different circumstances: first, the original community structure remains unchanged, and we keep the value of propagation power of the whole community unchanged; second, when the addition of the two nodes to the community increases the value of the function of the entire community ratings, we will add the two nodes to that community and calculate the value of propagation power of the transformed community, and then update the value of propagation power of entire community network. The details of the Algorithm 2 are described as follows:

Algorithm 2 AddEdge

Input: F_t represented value of propagation energy of the community intime t, and community structure and new edge u

Output: F_{t+1} represented the updated value of propagation energy of the community in time $t+1$ and the new community structure

If($C(i) \neq C(j)$) then

Recalculate F for the community which i and j in

Update $F_{t+1} = F_t + \Delta F$

If($C(i) \neq C(j)$) then

If (i join the community which j in)

Recalculate F for the communities which changes

Update $F_{t+1} = F_t + \Delta F$

End if

Endif

5.3. Delete an edge

Compared with the above cases, deleting an edge is less frequent. When an edge is deleted from the network, we recalculate the propagation power based on a different case.

When the two nodes of the edge belong to different communities, the overall value of propagation power remains unchanged.

When the two nodes of the edge belong to the same community and the degree of one of the nodes is 1, this node will become an isolated node, forming a new and separate community that contains only one node. The structure of the rest of the community remains unchanged. In particular, when the degrees of both nodes are 1, the deletion of the edge will cause both nodes to become isolated nodes, and two separate communities would form. We need to

recalculate the F value of the two one-node communities.

When the two nodes of the edge belong to the same community, but the degrees of the nodes are different from the above situations, the value of the function of the community ratings will decrease. The community may remain unchanged or split into two communities after an internal edge is deleted. In the latter case, the F value of the divided communities has to be recalculated. A detailed Algorithm 3 is presented as follows:

Algorithm 3 DeleteEdge

Input: F_t represented value of propagation energy of the community in time t, and community structure and edge u to be removed

Output: F_{t+1} represented the updated value of propagation energy of thecommunity in time $t+1$ and the new community structure

Remain F_t unchanged 1

$C^{t+1} = (C^t \setminus \{i, j\}) \cup \{i\} \cup \{j\}$

Else if (either i (or j) is one degree) then

$C^{t+1} = (C^t \setminus C(i)) \cup \{i\} \cup (C(i) \setminus \{i\})$

Recalculate the F for the two communities

Update $F_{t+1} = F_t + \Delta F$

Else if(i and j are not belong the same community) then

$C^{t+1} = C^t$

Else

Int value = Get the betweenness of edge (i, j);

If (value is top n)

Let the rest nodes consider their best community;

Recalculate the F for the communities of the new separate

End if

Update $F_{t+1} = F_t + \Delta F$

End if

Update C^{t+1}

5.4. Delete a node

When a node is deleted from the network, recalculating the F value is a complex process. When the degree of the deleted node is 1, only one edge is deleted. Thus, the rest of the community structure and the F value of the original community remain unchanged. When the degree of the deleted node is more than 1, it will appreciably impact the original community and loosen the structure of the original community.

When we consider other nodes adjacent to the deleted node, the adjacent nodes connected to different communities should be added to the appropriate community so that the function value of the community ratings is higher. F values of the changed communities have to be recalculated. A detailed Algorithm 4 is described as follows:

Algorithm 4 DeleteNode

Input: F_t represented value of propagation energy of the community in time t, and community structure and node u to be removed

Output: F_{t+1} represented the updated value of propagation energy of the community in time $t+1$ and the new community structure

k=1;

If (node I's degree is 1)

$C^{t+1} = C^{'} \setminus C(i) \cup (C(i) \setminus i) \cup \{i\}$

Else

While ($N(i)$ is not empty)do

$S(k)$ = {nodes have connection with nodes which don't in $C(i)$ };

k=k+1;

End while

End if

Let every $S(i)$ consider its best communities;

Recalculate F for all the communities changes

Update $F_{t+1} = F_t + \Delta F$

Update C^{t+1}

6. Experiments and analysis

6.1. Experimental environment and simulation process

In this study, the algorithms are implemented in C++ or VS2008 (Microsoft, Albuquerque, NM, USA) and run on Intel (R) Celeron (R) (Santa Clara, CA, USA) 2.2 GHz, 2G memory in Win7.

In general, three methods are used to simulate the propagation of the virus [22]: log files, infection model, and synthetic model. Log files can reflect the real users' behaviors; with incomplete geographical coverage, it cannot represent all behaviors. The virus infection model can be computed efficiently, but many details are neglected. The synthetic model is very flexible and can cover the entire region. In this paper, we adopt the third method for model simulation.

Figure 7 shows the whole algorithm process, where G represents the network in which the virus will spread and *time[i]* represents the time taken to infect the entire network, with i representing the ith node that causes infection. *Infected[i]* denotes whether the ith node is infected. *Allnodestart* denotes the number of nodes that causes infection in graph G.

6.2. Relationship between propagation power and virus infection

Section 4 shows how to calculate the propagation speed of virus. We analyze the relationship between propagation speed v and propagation power F in basic structure. We firstly use three basic structures as input of graph G to obtain the corresponding propagation speed v and then calculate the corresponding propagation power F. Here, to simplify the simulation, the time taken for a node to infect another node is assumed to be 1 (this assumption does not affect the results).

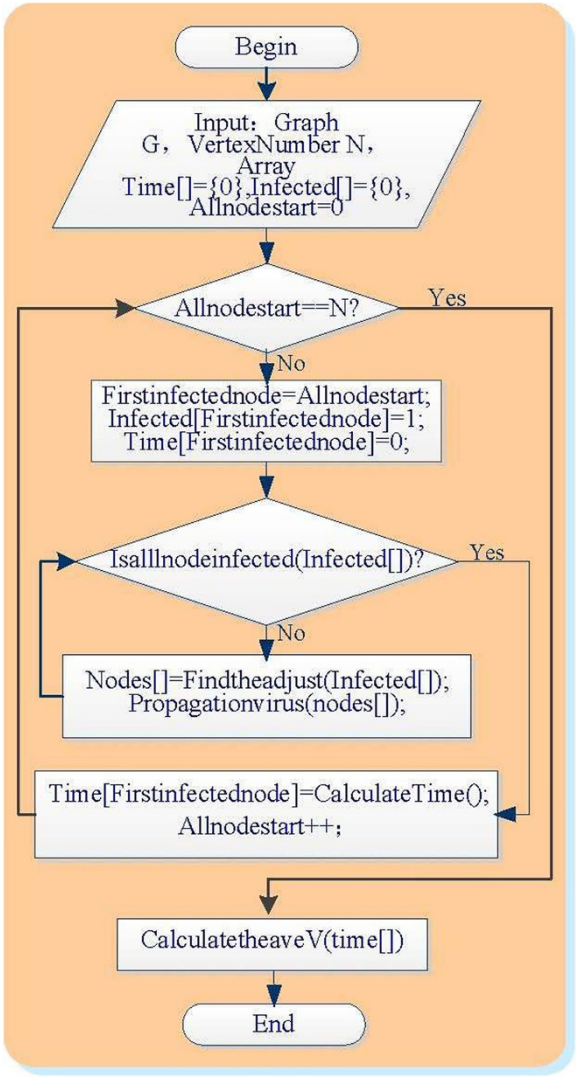

Figure 7 Simulation algorithm flowchart.

6.2.1. Relationship between propagation power and virus infection in LPS

We select 100%, 80%, 60%, and 50% of the nodes from the network respectively to constitute a linear structure. The remaining nodes are randomly connected to the linear networks. The relationships between propagation speed v and the number of nodes in the linear structure under different proportions are observed by experiment as shown in Figure 8a.

As can be seen in Figure 8a, when a network is closer to a linear structure, the propagation speed is lower. Thus, we conclude that the linear structure may help restrain the spread of the virus. Figure 8b shows that the value of propagation power F is not only related to the propagation time; it is also influenced by the number of the nodes in the linear structure.

When the number of nodes increases, the reciprocal value F^{-1} of the propagation power will also increase and the value of propagation power F will decrease.

The propagation time n/v (the time t_0 taken for a node to infect another node is assumed to be 1) and reciprocal value F^{-1} of the propagation power are almost linearly related in the linear structure, which is consistent with the theoretical analysis in Section 4.

6.2.2. Relationship between propagation power and viral infection in RPS

We select 100%, 80%, 60%, and 50% of the nodes from the network respectively to constitute a ring structure. The remaining nodes are randomly connected to the network. The relationships between propagation speed v and the number of nodes n in the ring structure under different proportions are observed by experiment as shown in Figure 9a.

As shown in Figure 9a, when a network is closer to a ring structure, the propagation speed is lower. Thus, we conclude that the ring structure can also help restrain the spread of the virus. Figure 9b shows that the value of propagation power F is not only related to the propagation time; it is also influenced by the number of nodes in the ring structure.

When the number of nodes increases, the reciprocal value F^{-1} of propagation power will also increase and the value of propagation power F will decrease.

The propagation time n/v (the time t_0 taken by a node to infect another node is assumed to be 1) and

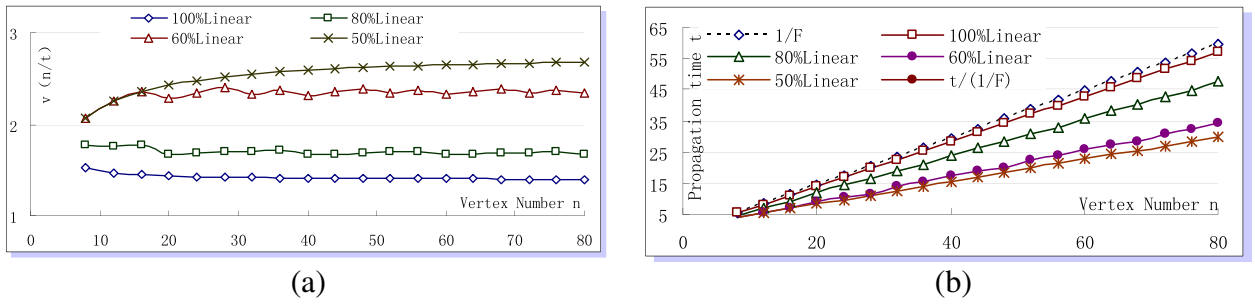

Figure 8 Simulation of virus propagation in LPS (a, b).

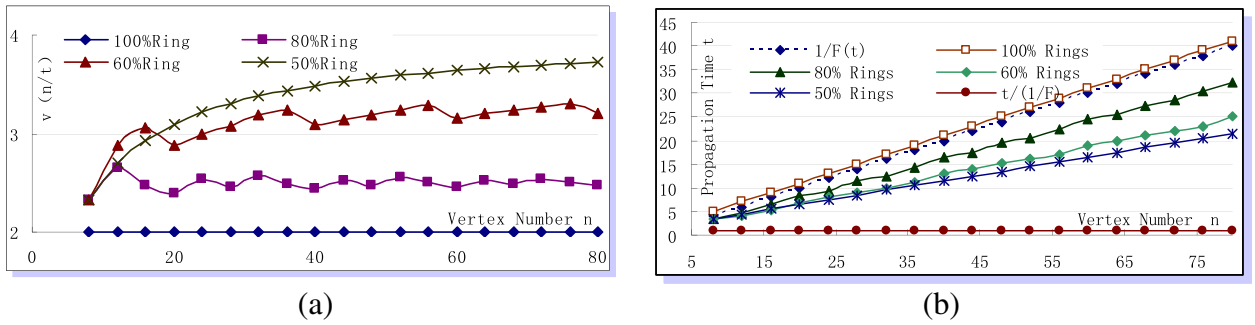

Figure 9 Simulation of virus propagation in RPS (a, b).

reciprocal value F^{-1} of propagation power are almost linearly related in the ring structure, which is consistent with the theoretical analysis in Section 4.

6.2.3. Relationship between propagation power and viral infection in SPS

We respectively select 100%, 80%, 60%, and 50% of the nodes from the network to constitute a star structure. The remaining nodes are randomly connected to the networks. The relationships between propagation speed v and the number of nodes n in the star structure are observed under different proportions by experiment as shown in Figure 10a.

As shown in Figure 10a, when a network is closer to a star structure, the propagation speed is higher. Thus, we conclude that the star structure may facilitate the spread of the virus. Figure 10b shows that the value of propagation power F is not only related to propagation time; it is also influenced by the number of nodes in the ring structure. When the number of nodes increases, reciprocal value F^{-1} of the propagation power will also increase.

We can also find that propagation time n/v (the time t_0 taken by a node to infect another node is assumed to be 1) and reciprocal value F^{-1} of propagation power are almost linearly related in the star structure.

When n approaches infinity, the ratio between n/v and reciprocal value F^{-1} of propagation power is gradually fixed, suggesting that the relationship between the reciprocal value F^{-1} of the propagation power and the propagation time n/v approaches linear in the star structure, which is consistent with the theoretical analysis in Section 4.

6.2.4. Range of changes of community propagation power

In addition to the changes in basic structures, for a change in the community, the value of propagation power changes within a certain range in a dynamic community. When a community is a complete subgraph, we find that the risk of virus propagation is highest based on our definition of propagation power. The value of propagation power F in a complete graph is $F = \left(\left(\sum_1^n 1\right)/n\right)^{-1} = 1$, according to Section 4.1. When a community is transformed into a linear structure, the propagation power F is lowest and the reciprocal value of propagation power is highest, according to Section 3.2, as shown in Figure 11a.

The reciprocal values of propagation power F^{-1} in all communities change within a certain range. In all communities, the propagation speed is highest in the complete subgraph and is lowest in linear structure. The experimental results are shown in Figure 11b.

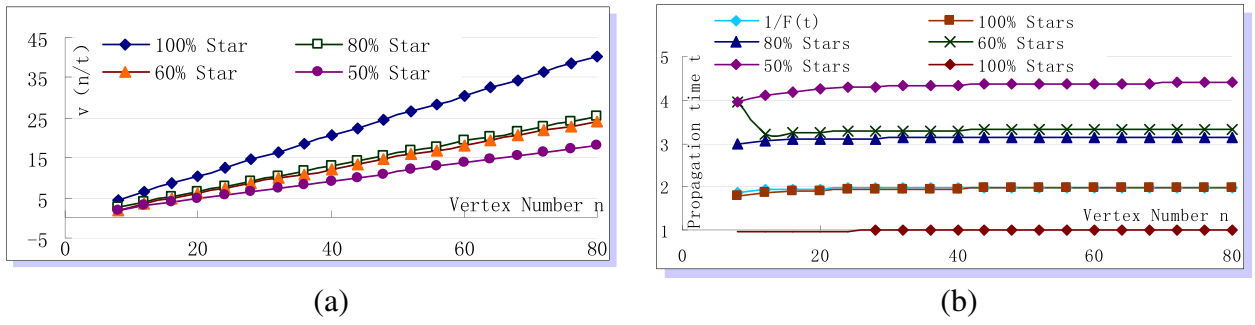

Figure 10 Simulation of virus propagation in SPS (a, b).

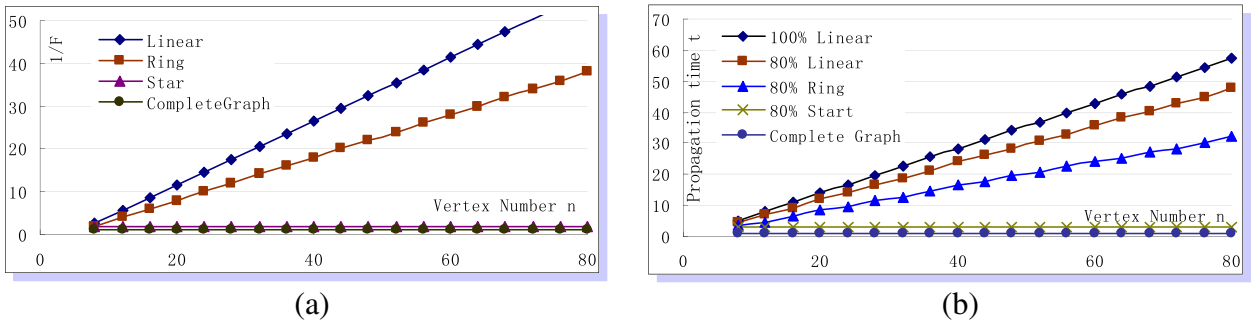

Figure 11 Range of propagation power (a, b).

As shown in Figure 11b, propagation time n/v in any community change within a certain range. The propagation time is shortest in the complete subgraph and is longest in the linear structure.

The above results are consistent with Theorem 2 which suggests that the ratio of the reciprocal of propagation energy of any network structure to time (n/v) has upper and lower limits, i.e.,

$$\sup\ n/v = F^{-1} \times f(n,a),$$

where $f(n,\ \alpha)$ is the linear function of n (when F^{-1} approaches the lower limit and n/v approaches the upper limit);

$$\inf(n/v) = F^{-1} \times f(n,a),$$

where $f(n,\ \alpha)$ increases when n decreases and vice versa (when F^{-1} approaches the upper limit and n/v approaches lower limit).

7. Case studies

With the concept of propagation power F, we can better understand the risk of virus propagation. In this section, the examples in Figures 1, 2, and 3 are analyzed to illustrate how to apply this propagation power model to actual cases and how to use the proposed algorithms.

Figure 1 in Section 2 shows the topology of two microblog communities. From the figure, it is difficult to tell the differences of risk of virus propagation between communities a and b. But through the calculation, we know the propagation power of the first community F_a = 0.1919 and the propagation power of the second community F_b = 0.2786. Therefore, the virus propagation risk of the second community is bigger than that of the first community. According to our calculation, it takes 5.0222 s to infect all nodes in the first community and 3.458 s in the second community through virus propagation (assuming the average time for one node to infect another node is 1). Therefore, a precise threshold value of risk of virus propagation can be set so that preventive measures can be taken to manage infection risk based on the calculation results if the calculated result is bigger than the threshold value.

The propagation risk is not only useful in the analysis of the propagation risk of virus in the static network. It may also be used to optimize network connection to restrain virus propagation. For example, in the four schemes of network connection in Figure 2, the networks are connected through two lines in each scheme. The following table presents the values of propagation power and the time taken to infect all nodes through simulation.

It is clear that the bigger the propagation power F, the shorter the propagation time of virus. According to Table 2, scheme 4 should be used. Therefore, propagation power can be used to optimize the network structure for restraining virus propagation.

It is not necessary to recalculate the value of propagation power in the real network by dynamic algorithm. As shown in Figure 3, when adding an edge to graph a, which consists of three communities, graph a will become graph b. Since the function value of community ratings remains unchanged, the structure of the original community also remains unchanged and it is not necessary to recalculate F values of any communities. When adding two edges to graph b, graph b will become graph c. We need to calculate the partial propagation power of community II and III. But it is not necessary to calculate the propagation power of community I.

8. Conclusions

8.1. Summary of the study

This paper analyzes the virus propagation and attempts to develop virus defense mechanisms by taking into account the network topology and the dynamic change

Table 2 Propagation power table

Scheme	Propagation power F	Propagation time t
1	0.22	4.33
2	0.19	4.97
3	0.20	4.82
4	0.17	5.45

of the network structure. Unlike traditional approaches to restraining computer virus propagation, we propose the concept of propagation power F to quantify the risk of virus propagation which is influenced by network structures. Preventive measures can then be taken in advance.

The main contributions of this paper are as follows:

1. A quantitative framework is proposed to assess the influence of the structure to virus propagation. We present three basic propagation structures (linear, ring, and star structures) and the concept of propagation power F to determine how a structure affects the speed of virus propagation.
2. The quantitative relationships between propagation power and network structures are found and verified. It is theoretically proven that propagation power F and propagation speed $(1/v)$ are related in the three basic structures. The relationships are also verified through experiments. Therefore, propagation power F is a valid indicator of the risk of virus propagation.
3. We develop an algorithm to compute the propagation power F for the dynamic community structure. Without calculating the entire community, this algorithm can be used to compute propagation power in large-scale network efficiently.

The feasibility of the framework and the algorithms has been verified through experiments and case studies.

8.2. Future work

To further improve the proposed approach, efforts should be made to improve the algorithm for computing propagation power so that the efficiency of calculation can increase without compromising the accuracy. In addition, the characteristics of network virus should be taken into consideration to build a more integrated mechanism for restraining virus propagation.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

The paper is supported by the National Natural Science Foundation of China (60903175, 61272405, 61272033, 61272451) and University Innovation Foundation (2013TS102, 2013TS106). Mr. Ho Simon Wang at HUST Academic Writing Center has provided tutorial assistance to improve the manuscript.

Author details

[1]School of Computer, Huazhong University of Science and Technology, Wuhan 430074, China. [2]Network and Computer Center, Huazhong University of Science and Technology, Wuhan 430074, China. [3]Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

Received: 19 February 2013 Accepted: 1 August 2013
Published: 17 August 2013

References

1. J Leyden, *Mobile malware menace hits high - McAfee*, 2007. http://www.theregister.co.uk/2007/02/12/mobile_malware/. Accessed 2 Mar 2007
2. D-H Shi, B Lin, H-S Chiang, M-H Shih, Security aspects of mobile phone virus: a critical survey. Ind. Manage. Data Syst. **108**(4), 478–494 (2008)
3. H Kim, J Smith, G Shin Kang, *Detecting energy-greedy anomalies and mobile malware variants. Proceedings of the 6th International Conference on Mobile Systems Applications and Services, Breckenridge, 17–20 June 2008* (Association for Computing Machinery, New York, 2008), pp. 239–252
4. L Xie, H Song, T Jaeger, S Zhu, *A systematic approach for cell-phone worm containment. Proceedings of the 17th International World Wide Web Conference (WWW08), Beijing, 21–25 April 2008* (Association for Computing Machinery, New York, 2008), pp. 1083–1084
5. N Xu, F Zhang, Y Luo, W Jia, W Xuan, J Teng, *Stealthy video capturer: a new video-based spyware in 3G smartphones. Proceedings of the 2nd ACM Conference on Wireless Network Security(WiSec09), Zurich, 16–18 March 2009* (Association for Computing Machinery, New York, 2009), pp. 69–78
6. M Salathe, M Kazandjieva, JW Lee, P Levis, MW Feldman, A high-resolution human contact network for infectious disease transmission. PNAS **107**(51), 22020–22025 (2010)
7. EP Fenichel, C Castillo-Chavez, MG Ceddia, G Chowell, PA Parra, GJ Hickling, G Holloway, R Horan, B Morin, C Perrings, M Springborn, L Valazquez, C Villalobos, Adaptive human behavior in epidemiological models. PNAS **108**(15), 6306–6311 (2011)
8. P Wang, MC Gonzalez, CA Hidalgo, Understanding the spreading patterns of mobile phone viruses. Science **324**(5930), 1071–1076 (2009)
9. C Gao, J Liu, Modeling and restraining mobile virus propagation. T. Mobile Comput. **99**, 1–3 (2012)
10. K Channakeshava, K Bisset, VSA Kumar, M Marathe, S Yardi, *High performance scalable and expressive modeling environment to study mobile malware in large dynamic networks, 25th IEEE International Parallel and Distributed Processing Symposium, Anchorage, 16–20 May 2011* (IEEE, Piscataway, 2011), pp. 770–781
11. Z Yajin, J Xuxian, *Dissecting android malware: characterization and evolution. Proceedings of the 33rd IEEE Symposium on Security and Privacy, San Francisco, 20–23 May 2012* (IEEE, Piscataway, 2012), pp. 95–109
12. A Gopalan, S Banerjee, AK Das, S Shakkottai, *Random mobility and the spread of infection. IEEE INFOCOM 2011, Shanghai, 10–15 April 2011* (IEEE, Piscataway, 2011), pp. 999–1007
13. R Pastor-Satorras, A Vespignani, Immunization of complex networks. Physical **65**(3), 6104 (2002)
14. Z Dezso, AL Barabasi, Halting viruses in scale-free networks. Physical **65**(5), 5103 (2002)
15. P Holme, BJ Kim, Vertex overload breakdown in evolving networks. Physical **65**(6), 6109 (2002)
16. Y Chen, G Paul, S Havlin, F Liljeros, HE Stanley, Finding a better immunization strategy. Physical **101**(5), 8701 (2008)
17. R Cohen, S Havlin, D Ben-Averaham, Efficient immunization strategies for computer networks and populations. Physical **91**(24), 7901 (2003)
18. P Holme, Efficient local strategies for vaccination and network attack. Europhys. Lett. **68**(6), 908–914 (2004)
19. LK Gallos, F Liljeros, P Argyrakis, A Bunde, S Havlin, Improving immunization strategies. Physical **75**(4), 5104 (2007)
20. J Gomez-Gardenes, P Echenique, Y Moreno, Immunization of real complex communication networks. Eur. Phys. **49**(2), 259–264 (2002)
21. P Echenique, J Gomez-Gardenes, Y Moreno, A Vazquez, Distance-d covering problem in scale-free networks with degree correlation. Physical **71**(3), 5102 (2005)
22. C Gao, J Liu, N Zhong, Network immunization with distributed autonomy-oriented entities. IEEE Trans. Parallel Distrib. Syst. **22**(7), 1222–1229 (2011)
23. G Zyba, GM Voelker, M Liljenstam, A Mehes, P Johansson, *Defending mobile phones from proximity malware. IEEE INFOCOM 2009, Rio de Janeiro, 19–25 April 2009* (IEEE, Piscataway, 2009), pp. 1503–1511
24. P Hui, J Crowcroft, E Yoneki, BUBBLE rap: social-based forwarding in delay tolerant networks. IEEE Trans. Mob. Comput. **10**(11), 1576–1589 (2011)
25. F Li, Y Yang, J Wu, *CPMC: An efficient proximity malware coping scheme in smartphone-based mobile networks. IEEE INFOCOM 2010, San Diego, 14–19 March 2010* (IEEE, Piscataway, 2010), pp. 1–9
26. J Jackson, S Creese, Virus propagation in heterogeneous Bluetooth networks with human behaviors. IEEE T. Depend. Secure **9**(6), 930–943 (2012)

27. B Abhijit, H Xin, S Kang, P Taejoon, *Behavioral detection of malware on mobile handsets, in MobiSys '08. Proceedings of the 6th International Conference on Mobile Systems Applications and Services, Breckenridge, 17–20 June 2008* (Association for Computing Machinery, New York, 2008), pp. 225–238
28. H Kim, J Smith, KG Shin, On detecting energy-greedy anomalies. IEEE Trans. Mob. Comput. **10**(7), 968–981 (2011)
29. P-S Romualdo, V Alessandro, Epidemic dynamics in finite size scale-free networks. Phys. Rev. E **65**(3), 035108 (2002)
30. YJ Zhou, Z Wang, W Zhou, XX Jiang, Hey, you, get off of my market: detecting malicious apps in official and alternative android markets, in *Proceedings of the 19th Network and Distributed System Security Symposium NDSS 2012* (San Diego, 2012)
31. R Sedgewick, K Wayne, *Algorithms* (Addison-Wesley Press, New Jersey, 2011), pp. 412–445

doi:10.1186/1687-1499-2013-210
Cite this article as: Cai *et al.*: Virus propagation power of the dynamic network. *EURASIP Journal on Wireless Communications and Networking* 2013 **2013**:210.