Physica A 396 (2014) 173–184

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

The spread of computer viruses over a reduced scale-free network

Lu-Xing Yang [a], Xiaofan Yang [a,b,*]

[a] College of Computer Science, Chongqing University, Chongqing, 400044, China
[b] School of Electronic and Information Engineering, Southwest University, Chongqing, 400715, China

HIGHLIGHTS

- We propose the notion of reduced scale-free networks.
- We present an epidemic model of computer viruses on a reduced scale-free network.
- We show that this model has a unique globally asymptotically stable viral equilibrium.
- Under the proposed model, we examine the impact of different model parameters on virus spreading.
- We recommend some practicable measures to contain virus spreading.

ARTICLE INFO

Article history:
Received 29 September 2013
Available online 1 December 2013

Keywords:
Epidemic model of computer viruses
Reduced scale-free network
Viral equilibrium
Global stability

ABSTRACT

Due to the high dimensionality of an epidemic model of computer viruses over a general scale-free network, it is difficult to make a close study of its dynamics. In particular, it is extremely difficult, if not impossible, to prove the global stability of its viral equilibrium, if any. To overcome this difficulty, we suggest to simplify a general scale-free network by partitioning all of its nodes into two classes: higher-degree nodes and lower-degree nodes, and then equating the degrees of all higher-degree nodes and all lower-degree nodes, respectively, yielding a reduced scale-free network. We then propose an epidemic model of computer viruses over a reduced scale-free network. A theoretical analysis reveals that the proposed model is bound to have a globally stable viral equilibrium, implying that any attempt to eradicate network viruses would prove unavailing. As a result, the next best thing we can do is to restrain virus prevalence. Based on an analysis of the impact of different model parameters on virus prevalence, some practicable measures are recommended to contain virus spreading. The work in this paper adequately justifies the idea of reduced scale-free networks.

1. Introduction

Network viruses are loosely defined as harmful programs that can spread over the Internet [1]; email viruses and network worms are typical network viruses. Due to the subsequent enormous financial losses, network viruses have proved to be a grave threat to human society. What is more serious, the widespread popularization of the Internet has greatly enhanced the spreading ability of network viruses. There are two major means of resisting computer viruses: antivirus software and firewall. Unfortunately, neither of them is adept to containing the spread of viruses over the Internet; the former focuses

* Corresponding author at: School of Electronic and Information Engineering, Southwest University, Chongqing, 400715, China. Tel.: +86 23 65342815; fax: +86 23 65342815.
E-mail addresses: ylx910920@gmail.com (L.-X. Yang), xfyang1964@gmail.com (X. Yang).

0378-4371/$ – see front matter
http://dx.doi.org/10.1016/j.physa.2013.11.026

on detecting and cleaning up viruses staying in individual computers, whereas the latter attempts to deter the intrusion of viruses by detecting abnormal network traffic, at the cost of impeding the receipt of legal programs.

The key to effectively inhibiting virus spreading over the Internet is to determine factors that dominate virus spreading. Noting the appealing analogy between network viruses and infectious diseases, Cohen [2] and Murray [3] suggested to exploit the compartment modeling method developed in epidemic dynamics of infectious diseases to study the laws governing the propagation of network viruses. Following this idea, Kephart and White [4,5] initiated the study of epidemic models of network viruses. From then on, multifarious network virus spreading models, ranging from conventional models such as SIS models [6], SIR models [7], SIRS models [8–11], SEIR models [12], SEIRS models [13], SEIQRS models [14], SLBS models [15–19], SICS models [20,21], and some other models [22–24], to unconventional models such as delayed models [25–31] and stochastic models [15,32], have been proposed based on the homogeneously mixed assumption, i.e., the assumption that every computer on the Internet is equally likely to be accessed by any other computer on the Internet. As was pointed out by Balthrop et al. [33], this assumption is well suited to the spread of some viruses, such as the Nimda and SQLSlammer worms, over the IP network.

In reality, however, some kinds of network viruses spread over a network other than the IP network. For instance, ILoveYou and Klez email viruses spread over an email address book network [33]. Previous empirical analysis reveals that a realistic email address book network is far from homogeneously mixed; rather, its degree distribution is something like an exponential form [34] or a stretched exponential form [34] or a heavy-tailed form [35]. Hence, it is of great importance to understand the impact of network topology on virus spreading [36,37]. The overwhelming majority of previous work towards this direction was carried out based on a general scale-free network [36–47]. In every such model, two or three compartments are introduced for each possible node degree, leading to an explosion in the total number of compartments. As a result, it is very difficult to have a profound understanding of such models of medium complexity; in particular, it is extremely difficult to prove the global stability of the viral equilibrium, if any [48–50].

To study the impact of network topology on the spread of computer viruses in a more theoretical way, we suggest to simplify the topology of a general scale-free network by partitioning all nodes in the network into two classes: higher-degree nodes and lower-degree nodes, assuming that the degrees of all higher-degree nodes are equal to their average degree k_1 and that the degrees of all lower-degree nodes are equal to their average degree k_2. As with general scale-free networks, we assume that a node in the simplified network is connected to a k-degree node with a probability that is proportional to the value of k. Thus, a *reduced scale-free network* is formed. In reality, higher-degree nodes represent the most influential nodes on the Internet, such as websites for famous universities or large corporations, in the sense that they usually maintain touch with a larger number of other nodes, whereas lower-degree nodes stand for less influential nodes on the Internet, ranging from personal computers to smart phones and netbooks, in the sense that they usually keep in touch with only a small number of other nodes. In this context, higher-degree nodes are on the Internet all along, whereas lower-degree nodes leave and enter the Internet frequently.

The major objective of this paper is to understand the spread of computer viruses on a reduced scale-free network. For that purpose, we propose a novel epidemic model of computer viruses. Our theoretical analysis reveals that this model always admits a globally stable viral equilibrium. Hence, any attempt to eradicate network viruses would prove infructuous. As a result, the next best thing we can do is to restrain virus prevalence. The impact of different model parameters on virus prevalence is analyzed, thereby some practicable measures are recommended to contain virus spreading.

The subsequent materials are organized in this fashion: Section 2 elaborates the new model, Section 3 makes a close study of the model, and Section 4 examines the impacts of different model parameters on virus prevalence. Finally, Section 5 summarizes this work and looks ahead at some issues that are worthy of study.

2. Model description

As usual, in this paper we shall neglect both the nature of computer viruses and the details of their infection, and simply assume that, at any time, every computer in the world is in one of two possible states: *susceptible* and *infected*. For our purposes, the contact relationship at any time between computers on the Internet can be represented by a graph, where nodes represent computers on the Internet at that time, and there is an edge between two nodes if and only if there is a connection at that time between their corresponding computers. We shall approximately assume that the Internet owns N_1 nodes of degree k_1, N_2 nodes of degree k_2, and no nodes of any other degree, where $k_1 \gg k_2$, N_1, N_2, k_1 and k_2 are unvaried as time. As a result, the Internet always has exactly $N = N_1 + N_2$ nodes. For $j = 1, 2$, let P_j denote the probability that a node chosen randomly from the Internet is of degree k_j, then $P_j = \frac{N_j}{N}$.

For $j = 1, 2$, let $S_j(t)$ (resp. $I_j(t)$) denote the average number of k_j-degree susceptible (resp. infected) nodes at time t. Let $S(t)$ (resp. $I(t)$) denote the average number of susceptible (resp. infected) nodes at time t. Then,

$$S(t) = S_1(t) + S_2(t), \qquad I(t) = I_1(t) + I_2(t).$$

For $j = 1, 2$, let $s_j(t) = S_j(t)/N_j$ (resp. $i_j(t) = I_j(t)/N_j$) denote the average relative density of k_j-degree susceptible (resp. infected) nodes at time t. Let $s(t) = S(t)/N$ (resp. $i(t) = I(t)/N$) denote the average density of susceptible (resp. infected) nodes at time t. Then,

$$i(t) = P_1 i_1(t) + P_2 i_2(t).$$