Physica A 466 (2017) 211–223

Contents lists available at ScienceDirect

Physica A

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

Resistance maximization principle for defending networks against virus attack☆

Angsheng Li*, Xiaohui Zhang, Yicheng Pan

State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, 100190, PR China

HIGHLIGHTS

- We propose the metric of resistance of networks.
- The resistance quantitatively measures the force of a network.
- We propose the resistance maximization principle.
- There exist networks that are protected by a small number of controllers.

ARTICLE INFO

Article history:
Received 22 April 2016
Received in revised form 9 September 2016
Available online 22 September 2016

Keywords:
Network
Virus attack
Security
Resistance of networks
Robustness of networks

ABSTRACT

We investigate the defending of networks against virus attack. We define the *resistance of a network* to be the maximum number of bits required to determine the code of the module that is accessible from random walk, from which random walk cannot escape. We show that for any network G, $\mathcal{R}(G) = \mathcal{H}^1(G) - \mathcal{H}^2(G)$, where $\mathcal{R}(G)$ is the resistance of G, $\mathcal{H}^1(G)$ and $\mathcal{H}^2(G)$ are the one- and two-dimensional structural information of G, respectively, and that resistance maximization is the principle for defending networks against virus attack. By using the theory, we investigate the defending of real world networks and of the networks generated by the preferential attachment and the security models. We show that there exist networks that are *defensible* by a small number of *controllers* from cascading failure of any virus attack. Our theory demonstrates that *resistance maximization* is the principle for defending networks against virus attacks.

1. Introduction

It was shown that network topology is universal in nature, society, and industry [1]. Erdös–Rényi proposed the first model [2,3] (The ER model in short) to capture complex systems based on the assumption that real systems are evolved randomly. The ER model explores the well-known small-diameter property of networks, that the diameter of a network of n nodes is $O(\log n)$; this property is the essence of the small-world phenomenon, and is the first general property of networks. The small-world phenomenon of networks is simply guaranteed by some randomness in the sense that, for any graph, if we add a small number of edges randomly and uniformly in the graph, the diameter of the new graph is small with high probability.

Barabási and Albert [4] proposed a graph generator by introducing preferential attachment as an explicit mechanism, the model is thus called the preferential attachment (PA) model. Consequently, networks generated by the PA model naturally

☆ The authors are partially supported by the Grand Project "Network Algorithms and Digital Information" of the Institute of Software, Chinese Academy of Sciences by an NSFC Grant No. 61161130530, and by a China Basic Research Program (973) Grant No. 2014CB340302.

* Corresponding author.
E-mail address: angsheng@ios.ac.cn (A. Li).

http://dx.doi.org/10.1016/j.physa.2016.09.009
0378-4371/

follow a power law. It has been shown that most real networks follow a power law so that power law is the second universal property of networks [1].

In fact, the current highly connected world is assumed to be supported by numerous networking systems. Real networks are not only too important to fail, but also too complicated to understand. An immediate question is hence: Can we guarantee the security of activities in a highly connected network? It could be possible that a small number of attacks or even random errors of individuals may cause a global failure of the networks. And even worse, it seems that power law and small world property (the main advantages of networks) are obstacles of achieving security of networks.

Typical attacks have two types [5–10]. The first type is the physical attack of removal of some nodes or edges. It has been shown that in scale-free networks of the preferential attachment (PA) model [4], the overall network connectivity measured by the sizes of the giant connected components and the diameters does not change significantly under random removal of a small fraction of nodes, but vulnerable to removal of a small fraction of the high-degree nodes [10–12].

The second type is the cascading failure of attacks, which naturally appeared in rumor spreading, disease spreading, voting, and advertising [13,5,6]. One of the main features of networks in the current highly connected world is that failure of a few nodes of a network may cause a cascading failure of the whole network. For instance, the failure of a few US commercial banks caused the 2008 global financial crisis all over the world. It has been shown that in scale-free networks of the preferential attachment model even a weakly virulent virus can spread [14]. This explains a fundamental characteristic of security of networks [9].

For the physical attacks or random errors of removal of nodes, it was shown that the optimal networks resisting both physical attacks and random errors have at most three values of degrees for all the nodes of the networks [15], that networks having the optimal robustness resisting both high-degree nodes attacks and random errors, have a bimodal degree distribution [16]. To enhance the robustness of networks, it was proposed [17] the acquaintance immunization strategy, which calls for the immunization of random acquaintances of randomly chosen nodes, and more recently, a security enhancing algorithm was proposed in Ref. [18] by randomly swapping two edges for a number of pairs of edges. The results are all for security or robustness of physical attacks or random errors. More seriously, the graphs that are thus characterized as secure or robust are far from real graphs, because the graphs have only two or three choices of degrees for all the nodes, which never occurs in real networks. Li, Li, Pan and Zhang [19] proposed a dynamical model, the security model of networks, that generates secure networks. Li and Pan [20] showed that with appropriately chosen affinity exponent a, the networks generated by the security model are provably secure against small-scaled virus attacks.

Gao, Barzel and Barabási [21] proposed the notion of resilience to measure the ability of a network to resist various perturbations. The measure of resilience depends on appropriate choice of the functions that represent the dynamical laws of the network. Li and Pan [22] proposed the notion of structural information and established the fundamental theory of dynamical complexity of networks. Given a network G and a natural number K, the K-dimensional structural information of G is defined to be the number of bits required to determine the K-dimensional code of the node that is accessible from random walk in G. The structural information is a measure that represents the dynamical complexity of the network, and that determines and decodes the natural hierarchical structure of the network in which noises and random variations are maximally excluded. Based on the theory of structural information, Li et al. [23] proposed the notions of resistance and security index of networks, and showed that both the resistance and security index of a network measure the power of the network to resist cascading failure of intentional virus attacks.

An interesting open question is: how to control virus spreading in networks? What are the defending principles of networks? In the present paper, we will answer these questions.

We investigate the defending of networks based on our new notion of resistance of networks. The resistance of a network is the greatest number of bits required to determine the code of the module of the network that is accessible from random walk from which random walk cannot escape. The resistance characterizes the force of the network to resist the spreading of a virus that randomly infects its neighbors. We discover the resistance law of networks that the resistance of a graph G is the difference of the one-dimensional structural information and the two-dimensional structural information of the graph. This explores that resistance maximization is equivalent to the minimization of two-dimensional structural information and the maximization of the one-dimensional structural information of networks. According to this principle, if a network G is given, then the one-dimensional structural information is directly determined by the topology of the network, and the two-dimensional structural information is determined by a partition of the vertices of the network which must be found by an algorithm; if the network is constructed to realize the maximum security, we will need two requirements: the first is to maximize the one-dimensional structural information and the second is to minimize the two-dimensional structural information. The two requirements can be combined together, in which case, the new condition is to minimize the ratio of the two-dimensional structural information and the one-dimensional structural information, that is, to maximize the security index of the network, referred to Ref. [23].

Based on the theory, we propose a resistance maximization algorithm $\mathcal{E}$ and a controller defining algorithm $\mathcal{C}$ to find the controllers that prevent global failure from virus attacks. We show that for appropriately large affinity exponent, the networks generated by the security model are defensible by a small number of controllers, in the sense that, after protecting the small set of controllers, any virus attack on the network can only cause the infection of a small number of vertices of the network, and that the networks of the preferential attachment model cannot be protected by any small set of controllers. Our results demonstrate that there is a defending principle of networks, which consists of both the resistance maximization and the heterogeneity of the external degrees defined by the resistance maximization principle. We also show

that our defending principle ensures that there are certain real world networks that are defensible by a small number of controllers.

2. Resistance of networks

We investigate the problem of defending networks against super virus spreading caused by any attack on the networks.

Given a network $G = (V, E)$, a *super virus* at a node v infects all the neighbors of v immediately. Therefore, if G is connected, then a single node that is infected by a super virus will eventually infect the whole network. To control the super virus spreading in a network, we find a small set of *controllers* C, a subset of V, such that the controllers cannot be infected and such that the controllers ensure that any super virus can infect only a small number of nodes of the network.

We investigate the problem of finding the small set of controllers.

In real world, we observe that in many cases, a few or even a single virus may infect a constant fraction of the nodes of networks. It has been a challenge to understand the laws of virus spreading and to control the virus spreading in networks. The challenge depends on a number of issues such as: (i) the strategy of attacks, (ii) the mechanisms of the viruses, and (iii) the number of viruses etc.

Considering the case that a virus randomly infects its neighbors, Li et al. [23] proposed the notions of resistance and security index of networks, and showed that both the metrics do measure the ability of a network to resist the cascading failure caused by a small number of intensional virus attacks.

However, our problem here is to control the spreading of super virus attacks, where a super virus infects all its neighbors immediately. The problem is harder than that in Ref. [23]. For solving the problem, we need some notions proposed by Li et al. [23]. In particular, we will need the notion of resistance of graphs.

Given a network $G = (V, E)$, suppose there is partition $\mathcal{P}$ of G such that random walk with stationary distribution in G easily goes to a small module X of $\mathcal{P}$ after which it is not easy to escape from the module X. In this case, we know that the spreading of the virus is restrained by the partition $\mathcal{P}$ of G. According to the hypothesis above, we define the *resistance of G given by a partition $\mathcal{P}$*.

Given a connected graph G, suppose that $\mathcal{P}$ is a partition of G. The *resistance of G by $\mathcal{P}$* [23] is defined as follows:

$$\mathcal{R}^{\mathcal{P}}(G) = -\sum_{j=1}^{L} \frac{V_j - g_j}{2m} \log_2 \frac{V_j}{2m}, \tag{1}$$

where V_j is the volume of the jth module X_j of $\mathcal{P}$, and g_j is the number of edges from X_j to nodes outside X_j.

In Eq. (1), in the jth term $-\frac{V_j-g_j}{2m} \log_2 \frac{V_j}{2m}$, $\frac{V_j-g_j}{2m} = \frac{V_j-g_j}{V_j} \cdot \frac{V_j}{2m}$ is the probability that a random walk goes to the jth module X_j after which the random walk fails to escape from the module X_j, and $-\log_2 \frac{V_j}{2m}$ is the number of bits required to determine the code of the jth module X_j in $\mathcal{P}$.

Therefore, $\mathcal{R}^{\mathcal{P}}(G)$ is the overall number of bits required to determine the code of the module X satisfying:

- X is accessible from random walk with stationary distribution in G, and
- Once the random walk arrives at X, it cannot escape from X.

The *resistance of a graph G* is defined as follows:

$$\mathcal{R}(G) = \max_{\mathcal{P}}\{\mathcal{R}^{\mathcal{P}}(G)\}, \tag{2}$$

where $\mathcal{P}$ runs over all the partitions of G.

According to the definition in Eq. (2), $\mathcal{R}(G)$ is the greatest overall number of bits required to determine the code of the module of G, which is accessible from random walk and from which random walk cannot escape.

Intuitively, the resistance of a network G measures quantitatively the force of G to block virus spreading in G by a partition of G.

The definition in Eq. (2) explores a defending principle of networks.

Resistance maximization principle of networks:

1. Given a network $G = (V, E)$, if there is a partition $\mathcal{P}$ such that the resistance of G defined by $\mathcal{P}$ is large, then G has strong ability to resist the cascading failure of any virus attack.

   In this result, we notice that we do not need to find the partition $\mathcal{P}$ such that $\mathcal{R}^{\mathcal{P}}(G)$ is large. The existence of such a partition $\mathcal{P}$ has already guaranteed the strong ability of the network to resist cascading failures of any virus attack.
2. However, to measure the security of the network, we need to find a partition $\mathcal{P}$ of G such that $\mathcal{R}^{\mathcal{P}}(G)$ is large.
3. Given a network G, the resistance maximization of G is the principle for defending virus spreading in G. That is, to control the virus spreading caused by attack on G, we have to find the partition $\mathcal{P}$ of G such that $\mathcal{R}^{\mathcal{P}}(G)$ is maximized among all the partitions of G. The defending strategy of G may be built based on the found partition $\mathcal{P}$.

Clearly, the metric of resistance is interesting in general. To see this, let us consider the following example. Suppose that $G = (V, E)$ is a connected graph such that the resistance $\mathcal{R}(G)$ of G is extremely small.

According to Eq. (1), for any vertex partition $\mathcal{P}$ of G,

$$\mathcal{R}^{\mathcal{P}}(G) = -\sum_{j=1}^{L} \frac{V_j - g_j}{2m} \log_2 \frac{V_j}{2m}$$

is very small.

This implies that for most j, $-\frac{V_j - g_j}{2m} \log_2 \frac{V_j}{2m}$ is small. For such a j, we have that either $\frac{V_j - g_j}{2m}$ is small or $-\log_2 \frac{V_j}{2m}$ is small. The former means that random walk easily leaves X_j, if it starts at some node in X_j, and the latter means that the V_j is large. Consequently, if each module X of $\mathcal{P}$ is small, then random walk easily goes from one module to another. This means that for any partition $\mathcal{P}$ of small sets of the vertices of G, random walk in G easily travels among the modules of $\mathcal{P}$. This argument indicates that G has some *small set expansion property*, in the sense that, for any small set X of V, with high probability, random walk from X quickly leaves X.

Therefore it is interesting to character the resistance of graphs and to explore the roles of the metric of resistance for both static graphs and evolving networks. The question is widely open, for which the foundation is the structural information theory founded by Li and Pan [22].

The resistance of networks introduced above is closely related to the one- and two-dimensional structural information of networks. To understand this, we introduce the metrics of one- and two-dimensional structural information proposed by Li and Pan in Ref. [22].

3. Structural information of networks

Given a connected graph $G = (V, E)$ and a natural number K, the K-dimensional structural information of G is the least number of bits required to determine the K-dimensional code of the node that is accessible from the random walk with stationary distribution in G. The definition explores that K-dimensional structural information minimization is the principle of the natural K-level structure of the graph. In particular, two-dimensional structural information minimization is the principle for discovering the natural community structure of a network.

3.1. Positioning entropy—one-dimensional structural information

Let $G = (V, E)$ be a connected graph with n nodes and m edges. For each node $i \in \{1, 2, \ldots, n\}$, let d_i be the degree of i in G, and let $p_i = d_i/2m$. Then the vector $\mathbf{p} = (p_1, p_2, \ldots, p_n)$ is the stationary distribution of a random walk in G.

We define the *positioning entropy of* G or *one-dimensional structural information of* G as follows:

$$\mathcal{H}^1(G) = H(\mathbf{p}) = H\left(\frac{d_1}{2m}, \ldots, \frac{d_n}{2m}\right) = -\sum_{i=1}^{n} \frac{d_i}{2m} \cdot \log_2 \frac{d_i}{2m}. \tag{3}$$

By definition, $\mathcal{H}^1(G)$ is the amount of information needed to determine the code of the node that is accessible from a step of random walk with stationary distribution in G.

$\mathcal{H}^1(G)$ is a dynamical notion about random walk in G. It is different from the Shannon entropy to determine the code of the node by a random selection among the nodes of the graph.

3.2. Structural information: two-dimensional structural information

The one-dimensional structural information of graph G can be naturally extended to the two-dimensional case.

Given a connected graph $G = (V, E)$, suppose that $\mathcal{P} = \{X_1, X_2, \ldots, X_L\}$ is a partition of V. By using the partition $\mathcal{P}$, we encode the nodes of G by two-dimensional vectors as follows: for a node v, we encode v by a pair (i, j) such that i is the code of v in its module X and j is the code of the module X that contains v.

Considering the information needed to determine the two-dimensional codes of the node that is accessible from the random walk with stationary distribution in G,

We define the *structural information of* G *by* $\mathcal{P}$, or the *two-dimensional structural information of* G *given by* $\mathcal{P}$ as follows:

$$\begin{aligned}\mathcal{H}^{\mathcal{P}}(G) &:= \sum_{j=1}^{L} \frac{V_j}{2m} \cdot H\left(\frac{d_1^{(j)}}{V_j}, \ldots, \frac{d_{n_j}^{(j)}}{V_j}\right) - \sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m} \\ &= -\sum_{j=1}^{L} \frac{V_j}{2m} \sum_{i=1}^{n_j} \frac{d_i^{(j)}}{V_j} \log_2 \frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L} \frac{g_j}{2m} \log_2 \frac{V_j}{2m},\end{aligned} \tag{4}$$

where L is the number of modules in partition $\mathcal{P}$, n_j is the number of nodes in module X_j, $d_i^{(j)}$ is the degree of the ith node in X_j, V_j is the volume of module X_j, and g_j is the number of edges with exactly one endpoint in module j.

According to the definition, $\mathcal{H}^{\mathcal{P}}(G)$ is the average number of bits needed to determine the code (i, j) of the node accessible from a step of random walk in G, where i is the code of the node in its own community, and j is the code of the community of the accessible node.

Now we are ready to define the *structural information*, or *two-dimensional structural information* of graphs.

Given a connected graph G, define the *structural information of* G or *two-dimensional structural information of* G as follows:

$$\mathcal{H}^2(G) = \min_{\mathcal{P}}\{\mathcal{H}^{\mathcal{P}}(G)\}, \tag{5}$$

where $\mathcal{P}$ runs over all the partitions of G.

According to the definition in Eq. (5), for a connected graph $G = (V, E)$, the two-dimensional structural information $\mathcal{H}^2(G)$ of G is the least overall number of bits needed to determine the two-dimensional code of the node that is accessible from the random walk with stationary distribution in G.

Li, Li and Pan [24] and Li et al. [25] have designed a community finding algorithm $\mathcal{E}$ on the basis of the minimization of the two-dimensional structural information of graphs, and shown that the algorithm $\mathcal{E}$ exactly identifies or precisely approximates the natural communities in many networks both generated by models and evolved in nature. Li, Yin and Pan [26] have developed a three-dimensional gene map for defining tumor types and subtypes based on the two- and three-dimensional structural information of cell sample networks.

4. Resistance law of networks

By the definition of the one-dimensional structural information and the additivity of the Shannon entropy function, we have the following *local resistance law of networks*:

For any graph G and any partition $\mathcal{P}$ of G,

$$\mathcal{R}^{\mathcal{P}}(G) = \mathcal{H}^1(G) - \mathcal{H}^{\mathcal{P}}(G). \tag{6}$$

By the definition in Eqs. (3) and (4), for the partition $\mathcal{P}$ of V,

$$\mathcal{H}^{\mathcal{P}}(G) = -\sum_{j=1}^{L}\frac{V_j}{2m}\sum_{i=1}^{n_j}\frac{d_i^{(j)}}{V_j}\log_2\frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L}\frac{g_j}{2m}\log_2\frac{V_j}{2m}, \tag{7}$$

and

$$\mathcal{H}^1(G) = H\left(\frac{d_1}{2m}, \ldots, \frac{d_n}{2m}\right) = -\sum_{i=1}^{n}\frac{d_i}{2m}\cdot\log_2\frac{d_i}{2m}. \tag{8}$$

By the additivity of the entropy function, for the partition $\mathcal{P}$,

$$\mathcal{H}^1(G) = -\sum_{j=1}^{L}\frac{V_j}{2m}\sum_{i=1}^{n_j}\frac{d_i^{(j)}}{V_j}\log_2\frac{d_i^{(j)}}{V_j} - \sum_{j=1}^{L}\frac{V_j}{2m}\log_2\frac{V_j}{2m}.$$

The resistance of G by $\mathcal{P}$ is

$$\mathcal{R}^{\mathcal{P}}(G) = -\sum_{j=1}^{L}\frac{V_j - g_j}{2m}\log_2\frac{V_j}{2m} = \mathcal{H}^1(G) - \mathcal{H}^{\mathcal{P}}(G).$$

The local resistance law follows.

According to Eqs. (2), (5) and (6), the resistance, one- and two-dimensional structural information satisfy the following:

Global resistance law of networks: For any graph G,

$$\mathcal{R}(G) = \mathcal{H}^1(G) - \mathcal{H}^2(G). \tag{9}$$

Therefore, the local resistance law in Eq. (6) and the global resistance law in Eq. (9) demonstrate that the following principle holds.

Defending principle of networks:

- For arbitrarily given graph G, the resistance maximization is equivalent to the structural information minimization.

  This is because for a given network G, the one-dimensional structural information has already been completely determined by the topology of G. However, for an evolving network, resistance maximization is both the maximization of one-dimensional structural information and the minimization of two-dimensional structural information.
- Resistance maximization is the principle for controlling virus spreading in the network G.

We show that the defending principle of networks above holds. For this, we develop a resistance maximization algorithm $\mathcal{E}$ and a controller algorithm $\mathcal{C}$. We show that for some networks, our algorithms find a small number of controllers that guarantee that the majority of the nodes or almost all the nodes of the networks cannot be infected by any virus attack, including the super virus attack, on any node of the networks. In this case, we say that the networks are *defensible*, and *non-defensible*, otherwise.

The resistance law provides a foundation for a theoretical direction about the metric of the resistance of networks, details are referred to Ref. [22].

5. Algorithms and methods for network defense

5.1. Algorithm $\mathcal{C}$ for defining controllers of networks

Given a network $G = (V, E)$, and a partition $\mathcal{P}$ of V, a controller of $\mathcal{P}$ in G is to prevent the infection of super viruses among different modules in $\mathcal{P}$. For this, we introduce an algorithm to define the global controllers of G on the basis of a partition $\mathcal{P}$.

Suppose that $\mathcal{P} = \{X_1, X_2, \ldots, X_N\}$ is a partition of $G = (V, E)$. We will give an algorithm to find k global controllers of G, denoted by $\mathcal{C}$.

$\mathcal{C}$ proceeds as follows.

(1) For every node y, define the external degree of y to be the number of edges from y to nodes outside y's own community, denoted by $d_E(y)$.
(2) Let x be the node with greatest external degree. Then:
- enumerate x into X,
- for every y, if $y \notin X$, there is an edge between x and y and x and y are in distinct communities, then set $d_E(y) \leftarrow d_E(y) - 1$.

(3) If $|X| = k$, then output X and terminate.
(4) Otherwise. Then go back to step (2) above.

5.2. Resistance maximization algorithm $\mathcal{E}$

According to the resistance law, that is, $\mathcal{R}(G) = \mathcal{H}^1(G) - \mathcal{H}^2(G)$, the maximization of the resistance is equivalent to the minimization of the structural information of G. We design our community detection algorithm on the basis of structural information minimization. We will use a simple greedy algorithm to find a partition which minimizes the structural information of the network G introduced in Refs. [24–26].

Suppose that $\mathcal{P} = \{X_1, X_2, \ldots, X_L\}$ is a partition of V. For i, j with $1 \le i, j \le L$, by definition in (4), if we obtain a partition $\mathcal{P}'$ from $\mathcal{P}$ by merging X_i and X_j, the difference of structure entropies of the two partitions is given by

$$\begin{aligned}\Delta^{\mathcal{P}}_{i,j}(G) &= -\frac{V_i}{2m}\sum_{k=1}^{n_i}\frac{d_k^{(i)}}{V_i}\log\frac{d_k^{(i)}}{V_i} - \frac{V_j}{2m}\sum_{k=1}^{n_j}\frac{d_k^{(j)}}{V_j}\log\frac{d_k^{(j)}}{V_j} + \frac{V_X}{2m}\sum_{k=1}^{n_i+n_j}\frac{d_k^{(i,j)}}{V_X}\log\frac{d^{(i,j)}}{V_X} \\ &\quad - \frac{g_i}{2m}\log\frac{V_i}{2m} - \frac{g_j}{2m}\log\frac{V_j}{2m} + \frac{g_X}{2m}\log\frac{V_X}{2m} \qquad (10)\\ &= \frac{1}{2m}[(V_i - g_i)\log V_i + (V_j - g_j)\log V_j - (V_X - g_X)\log V_X + (g_i + g_j - g_X)\log 2m], \qquad (11)\end{aligned}$$

where $X = X_i \cup X_j$, V_X is the volume of X, g_X is the number of edges from X to nodes outside of X, $d_k^{(i,j)}$ is the degree of the kth node in X.

If there is no edge between X_i and X_j, then $g_X = g_i + g_j$. In this case,

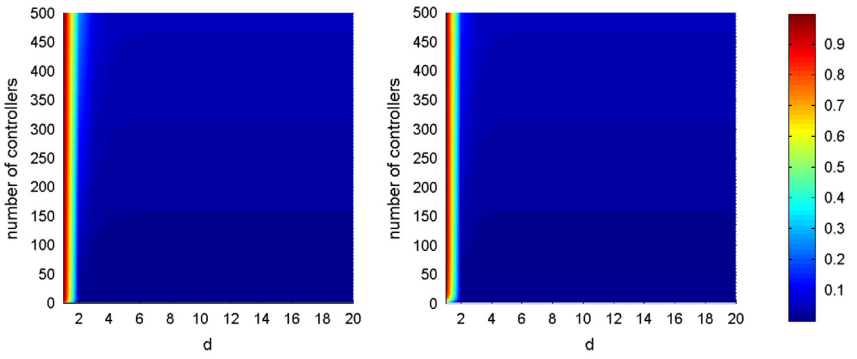
$$\begin{aligned}\Delta^{\mathcal{P}}_{i,j}(G) &= \frac{1}{2m}[(V_i - g_i)\log V_i + (V_j - g_j)\log V_j - (V_X - g_X)\log V_X] \\ &= \frac{1}{2m}\left[(V_i - g_i)\log\frac{V_i}{V_i + V_j} + (V_j - g_j)\log\frac{V_j}{V_i + V_j}\right] < 0. \qquad (12)\end{aligned}$$

Therefore, $\Delta^{\mathcal{P}}_{i,j}(G)$ is locally computable, and if there is no edge between X_i and X_j, then $\Delta^{\mathcal{P}}_{i,j}(G) \le 0$.

The algorithm, written by $\mathcal{E}$, proceeds as follows.

Given a network G:

(1) Set the initial partition such that each module contains a single node,
(2) recursively merge the modules X_i and X_j such that the corresponding $\Delta^{\mathcal{P}}_{i,j}(G)$ is maximized, until there is no such merging operation, in which case, output the corresponding partition $\mathcal{P}$.



(a) PA, Mean. (b) PA, Min.

Fig. 1. The fraction of nodes saved on the networks of the preferential attachment model by our defending algorithm $\mathcal{E}$ and $\mathcal{C}$. The parameter d is the number of average edges in the PA model. In this experiment, the number of nodes of the networks is 10,000, d is from 1 to 20 with unit 1, and the number of controllers k is from 1 to 500 with unit 1. For each type of the model, we generate 100 networks. The average and minimum are taken over the 100 networks for each type. (a) and (b) are the average fraction and the minimal fractions of nodes saved, respectively.

It has been shown that the algorithm $\mathcal{E}$ exactly identifies or precisely approximates true communities in many networks either generated by models or real world networks [24–26].

Clearly, the algorithm $\mathcal{E}$ is a greedy algorithm to find a partition $\mathcal{P}$ towards the maximization of the resistance of G.

Our method for defending networks consists of the resistance maximization algorithm $\mathcal{E}$ and the controller defining algorithm $\mathcal{C}$. The experimental method for models of networks is given in the next subsection.

5.3. Experimental methods

Given a network $G = (V, E)$ and a natural number k, we find a partition $\mathcal{P}$ of G by the resistance maximization algorithm $\mathcal{E}$, we define k controllers of G by the algorithm $\mathcal{C}$ on the basis of partition $\mathcal{P}$. Let C be the set of defined controllers. Suppose that G has been defined C as the sets of controllers of G. Our experiments of virus spreading on G proceeds as follows:

(i) Every node in C is a controller of G.
(ii) For every $x \in V \setminus C$, attack x by a super virus. Let I_x be the set of nodes that are infected by the super virus attack on x, and S_x be the complement of I_x in G.
(iii) Let S^G_{avg} be the average size of S_x for all $x \in V \setminus C$, and let S^G_{min} be the least size of S_x for all $x \in V \setminus C$.

Our experiments for the security model, the preferential attachment model and the small world model proceed as follows:

(1) For each type, we generate 100 networks $G_1, G_2, \ldots, G_{100}$.
(2) For each network G_i, we define the partition of G_i by the resistance maximization algorithm $\mathcal{E}$ and the controller algorithm $\mathcal{C}$ to define the controllers of G_i.
(3) We define S_{avg} to be the average $S^{G_i}_{\text{avg}}$ for all i from 1 to 100.
(4) We define S_{min} to be the minimal $S^{G_i}_{\text{min}}$ for all i from 1 to 100.

6. Results

6.1. PA model

In Fig. 1, we depict the color codes of the average and minimal fractions of nodes saved by our defending algorithms $\mathcal{E}$ and $\mathcal{C}$ for the networks of the preferential attachment model.

According to Fig. 1, we observe the following results:

(1) For the mean, according to Fig. 1(a), we have:
  (a) For $d = 1$, there is a column of color codes ≈ 0.4 for controllers of sizes from very small to 500.
  (b) For $d > 1$, the color codes are almost the same for all numbers of controllers from 1 to 500, and the uniform color codes are as small as 0.1.
(2) For the minimum, according to Fig. 1(b), we have:
  (a) For $d = 1$, there is a column of color codes ≈ 0.4 for controllers of sizes from very small to 500.
  (b) For $d > 1$, the color codes are almost the same for all numbers of controllers from 1 to 500, and the uniform color codes are as small as 0.1.

The results demonstrate that the networks of the preferential attachment model with $d > 1$, are non-defensible, at least by our defending method consisting of resistance maximization algorithm $\mathcal{E}$ and controller defining algorithm $\mathcal{C}$, and that the networks of the PA model with $d = 1$ actually are trees, in which a small constant fraction of nodes can be saved by a small number of controllers (less than 1% of nodes of the networks) the defined by our method.

The clear color codes for $d = 1$ and $d > 1$ in Fig. 1 suggest some interesting theoretical problems. For example, we conjecture that for a network of the PA model with $d > 1$, any small set ($o(n)$, say) of controllers can save at most $o(n)$ many nodes of the network. Clearly, theoretically proving such kind of results would be very interesting in both network theory, computer science, mathematics and even physics.

6.2. Networks generated by the security model

We have seen that the networks of the PA model are non-defensible for super virus spreading. Are there networks that are defensible?

Li et al. [19,20] proposed the following security model of networks. We introduce it below.

The security model proceeds as follows:

Given an *affinity exponent* $a \geq 0$ and a natural number d,

(1) Let G_d be an initial d-regular graph such that each node has a distinct *color* and called *seed*.
For each step $i > d$, let G_{i-1} be the graph constructed at the end of step $i - 1$, and $p_i = 1/(\log i)^a$.
(2) At step i, we create a new node, v.
(3) With probability p_i, v chooses a new color, in which case,
  (i) we call v a seed,
  (ii) (preferential attachment) create an edge (v, u) where u is chosen with probability proportional to the degrees of nodes in G_{i-1}, and
  (iii) (randomness) create $d - 1$ edges (v, u_j), where each u_j is chosen randomly and uniformly among all seed nodes in G_{i-1}.
(4) Otherwise, then v chooses an old color, in which case,
  (i) (randomness) v chooses uniformly and randomly an old color as its own color and
  (ii) (homophyly and preferential attachment) create d edges (v, u_j), where u_j is chosen with probability proportional to the degrees of all nodes of the same color as that of v in G_{i-1}.

We verify that the defending method here controls super virus spreading in the networks of the security model.

6.3. Varying the affinity exponent a of the security model

Fig. 2 depicts the color codes of the average fraction of nodes that are saved by our defending method from the infection of super virus spreading in the networks of the security model, as the affinity exponent a increases.

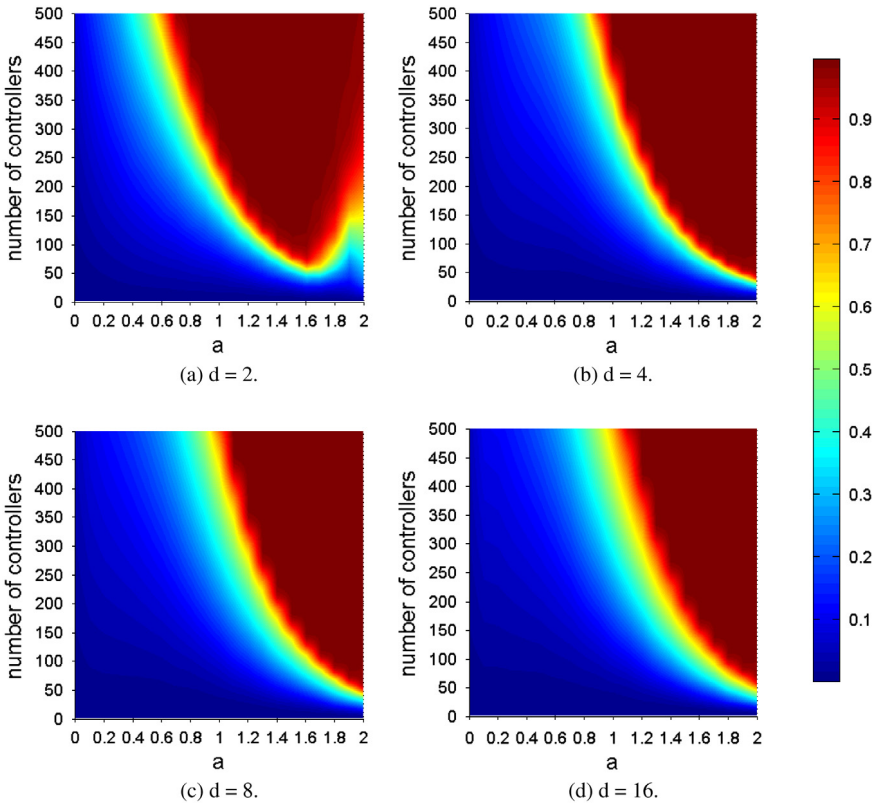
According to Fig. 2, we observe the following results:

(1) For each d, there is a golden boundary B_d such that it is narrow and it divides the square into two areas, the dangerous area, colored blue, and the secure area, colored red.
This shows that there is a phase transition from dangerous area to secure area for the networks of security model, which is given by the golden curve in each of Fig. 2(a)–(d).
(2) The golden boundary B_2 for $d = 2$ in Fig. 2(a) decreases as the affinity exponent a increases up to $a \approx 1.6$, and then increases as a increases after ≈ 1.6. In particular, for $d = 2$, if $a = 1.6$, then 50 controllers globally protect the networks, on the average.
(3) For $d \geq 4$, the golden belt B_d decreases as the affinity exponent a increases. Furthermore, for each d, for $a = 2$, 50 controllers protect the networks from the global infection of any super virus, on the average.
(4) By comparing Fig. 2(a)–(d), the affinity exponent required by the secure area increases as d increases.

Fig. 3 depicts the color codes of the minimal fraction of nodes that are saved by our defending method from the infection of super virus spreading in the networks of the security model as the affinity exponent a increases.

According to Fig. 3, we observe the following results:

(1) The same as that in Fig. 2, for every d, there is a golden curve C_d which distinguishes the dangerous area and the secure area of the whole area, and the curve C_d is similar to the golden boundary B_d in Fig. 2.
(2) For every d, and every a, the number of controllers is larger than $C_d(a)$, then the small set of controllers ensures that for any network of the security model, and any super virus attack, almost of the nodes of the network are protected from the infection of the super virus. On the other hand, if the number of controllers is slightly less than $C_d(a)$, then the least number of nodes saved by the controllers is only a small fraction of the nodes of the network.
(3) If $d > 2$, and a is appropriately large, equal to 2 say, then 50–70 controllers are sufficient to protect the global failure of the networks from the infection of any super virus attack.



(a) d = 2. (b) d = 4. (c) d = 8. (d) d = 16.

Fig. 2. The color codes of the average fraction of nodes saved by our defending method on the networks of the security model with $n = 10{,}000$, affinity exponent a from 1 to 2 with unit 0.1. The number of controllers is from 1 to 500 with unit 1. For each type of the network, we generate 100 networks. The average is computed over the 100 networks generated for each type. (a), (b), (c) and (d) are for $d = 2$, 4, 8 and 16, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The results demonstrate that for every nontrivial d, the number of controllers required to protect the network from global failure of super virus attack is decreasing as the affinity exponent a increases, so that for appropriately large affinity exponent a, a small number of controllers defined by our defending method protect the security of the networks from super virus attack. Equally interesting, there is a phase transition phenomenon from the dangerous area to the secure area of the networks generated by the security model. The phase transition is determined by a function $K_d(a)$ such that if the number k of controllers k is larger than $C_d(a)$, then the network is protected from super virus attack.

6.4. Varying d of the security model

Fig. 4 depicts the color codes of the average fraction of nodes saved by our defending method on the networks of the security model as d varies.

According to Fig. 4, we observe the following results:

(1) For $a = 0$. In this case, the security model becomes a dynamical random model, which generates connected and random graphs. According to Fig. 4(a), we have:
   (a) If $d = 1$, then for almost all small k of controllers, the color codes are approximately equal, and are around 0.4.
   (b) If $d > 1$, then for all k from 1 to 500, the color codes are almost all less than 0.1.
   (c) The color codes are similar to that in Fig. 1(a) for the preferential attachment model.

(2) If $a = 0.5$, then the color codes in Fig. 4(b) are similar to that in Fig. 4(a).

(3) If $a = 1$, according to Fig. 4(c), we have:
   (a) there is a secure area, which is the read left-upper corner cut by the line through points (0, 200) and (5, 500), and
   (b) there is a dangerous area, colored blue, which is approximately the lower part of line $k = 200$.

(4) For $a = 1.5$, according to Fig. 4(d), we have:
   (a) there is a golden curve g such that if the number of controllers is larger than $g(d)$, then the networks are protected.
   (b) if the number of controllers is less than $g(d)$, then the majority of nodes of the networks are probably infected by a super virus.

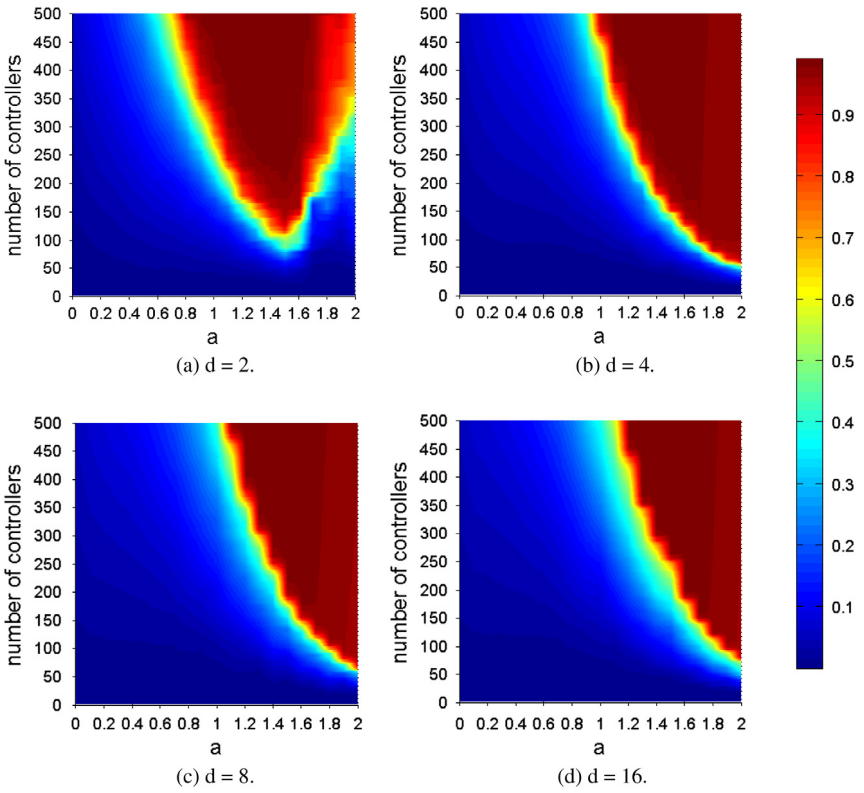


Fig. 3. The color codes of the minimal fraction of nodes saved by our defending method on the networks of the security model with $n = 10{,}000$, affinity exponent a from 1 to 2 with unit 0.1. The number of controllers is from 1 to 500 with unit 1. For each type of the network, we generate 100 networks. The minimum is computed over the 100 networks generated for each type. (a), (b), (c) and (d) are for $d = 2$, 4, 8 and 16, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 5 depicts the color codes of the minimal fraction of nodes saved by our defending method on the networks of the security model as d varies.

According to Fig. 5, we observe that the color codes in Fig. 5 are similar to that in Fig. 4.

Figs. 4 and 5 show that for appropriately large affinity exponent a, the networks of the security model are defensible by a small set of controllers, for which the small number of controllers of the networks is robust to the varying of d, and that the size of controllers of the networks is largely determined by the affinity exponent a of the model.

6.5. Real world networks

In this subsection, we look at the application of our algorithm to find controllers for various real world networks.

Fig. 6 depicts the curves of the fractions of nodes saved for four real world networks by our algorithms $\mathcal{E}$ and $\mathcal{C}$.

The four networks are the following:

(1) US airports

It is the directed graph of flights between US airports in 2010. An edge represents a connection from one airport to another and the weight of an edge shows the number of flights on that connection in the given direction. The graph contains 1572 nodes, and 28,235 edges. The graph can be found in US airport network data, 2015, [http://konect.uni-koblenz.de/networks/opsahl-usairport].

(2) The US power grid.

It contains the information of the power grid of the Western States of the United States of American. A node is either a generator, or a transformator, or a substation. An edge represents a power supply line. The Network contains 4941 nodes, and 6594 edges. The network can be found in [http://konect.uni-koblenz.de/networks/opsahl-powergrid].

(3) Gnutella peer-to-peer network, August 9, 2002

This is a snapshot of the Gnutella peer-to-peer file sharing network at August 9, 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts.

The graph contains 8104 nodes, and 26,008 edges. We use P2P No 9 to denote the graph.

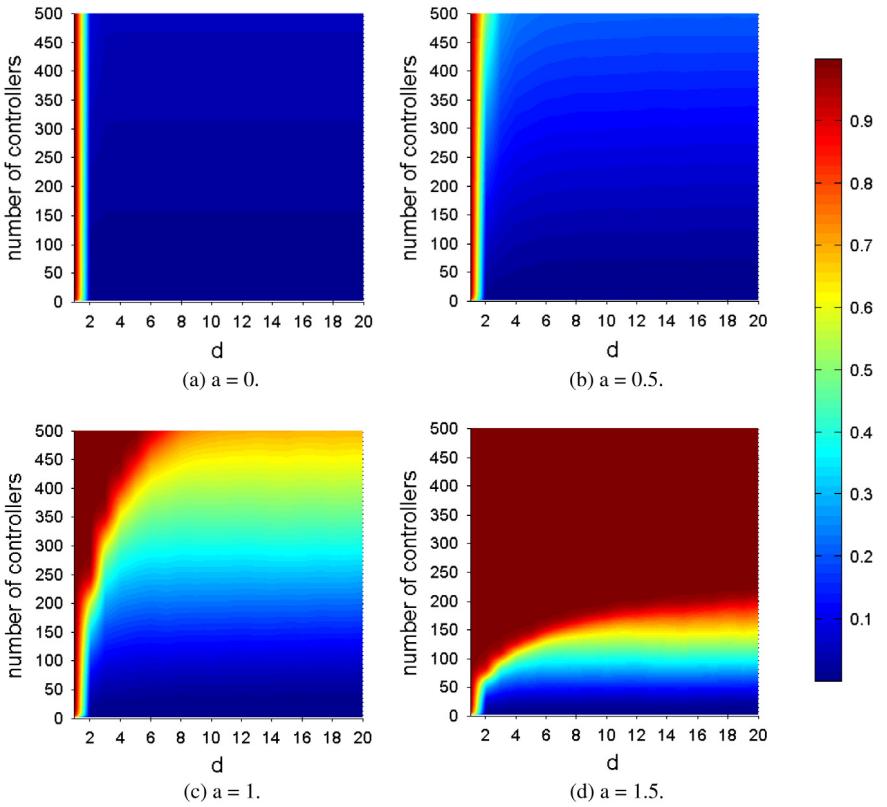


Fig. 4. The color codes of the average fraction of nodes saved by our defending method on the networks of the security model as d varies. For each type, we generate 100 networks. The average is computed over the 100 networks. (a), (b), (c) and (d) are for $a = 0$, 0.5, 1 and 1.5, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(4) Google+

This directed network contains Google+ user to user links. A node represents a user, and a directed edge denotes that one user has the other user in his circles. The graph contains 23,613 nodes and 39,230 edges.

The graph can be found in Google+ network dataset—KONECT, May 2015. [http://konect.uni-koblenz.de/networks/ego-gplus] and in Ref. [27].

According to Fig. 6, we observe the following results:

(1) Google+ is protected by a set of controllers of size less than 0.05% nodes of the network.
(2) The power grid is protected by a set of controllers of 5% nodes of the network.
(3) For both the P2P and the US airports, the networks cannot be protected by a set of controllers of 5% of the size of the networks.

7. Defending principles of networks

Fig. 1 shows that the networks of the preferential attachment model are non-defensible by a small number of controllers. Figs. 4(a) and 5(a) show that if the affinity exponent $a = 0$, then the networks of the security model cannot be protected by a small number of controllers, in which case, the networks of the security model are basically random graphs. The results demonstrate that the networks generated by the classical PA model or dynamical random model cannot be protected by a small number of controllers.

The results in Figs. 2, 3, 4(d) and 5(d) demonstrate that for appropriately large affinity exponent a, the networks of the security model are defensible through a small number of controllers. Therefore, there exist networks that can be protected by a small number of controllers.

Our defending method consists of two algorithms, the resistance maximization algorithm $\mathcal{E}$ and the controller defining algorithm $\mathcal{C}$. Algorithm $\mathcal{E}$ assumes the following:

Resistance hypothesis: Resistance maximization is the principle for security and defending of networks.

This hypothesis assumes that virus randomly spreads. According to this hypothesis, the resistance maximization algorithm $\mathcal{E}$ finds a partition $\mathcal{P}$ of G such that a virus may easily walk to a (small) module X of $\mathcal{P}$ after which it is hard to escape.

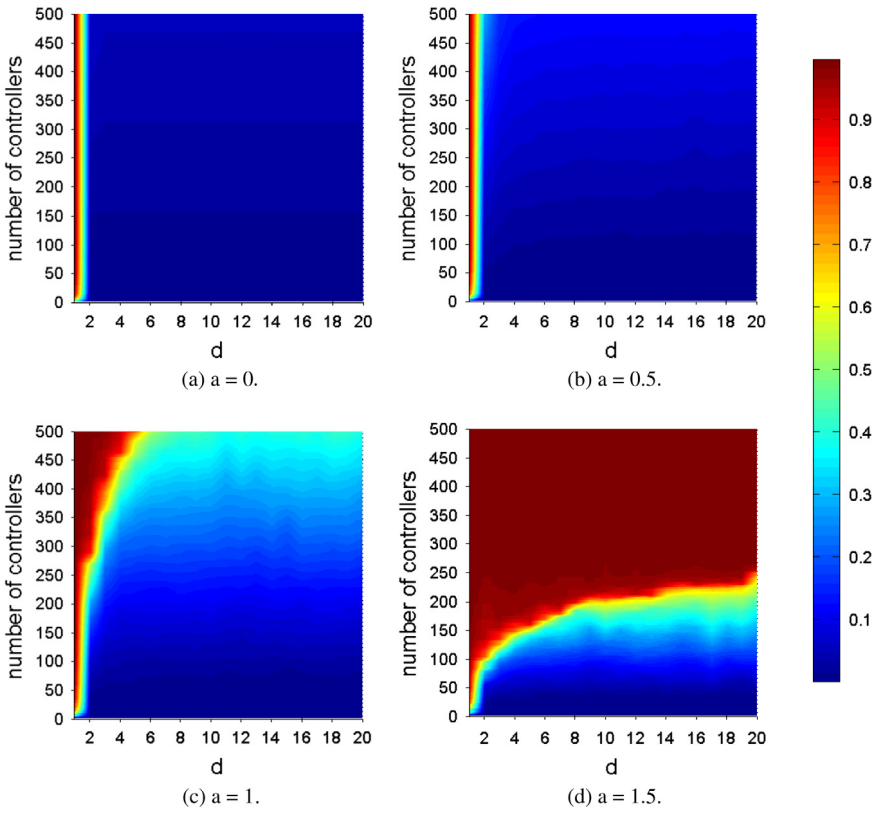


Fig. 5. The color codes of the minimum fraction of nodes saved by our defending method on the networks of the security model as d varies. For the each type, we generate 100 networks. The minimum is computed over the 100 networks. (a), (b), (c) and (d) are for $a = 0, 0.5, 1$ and 1.5, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

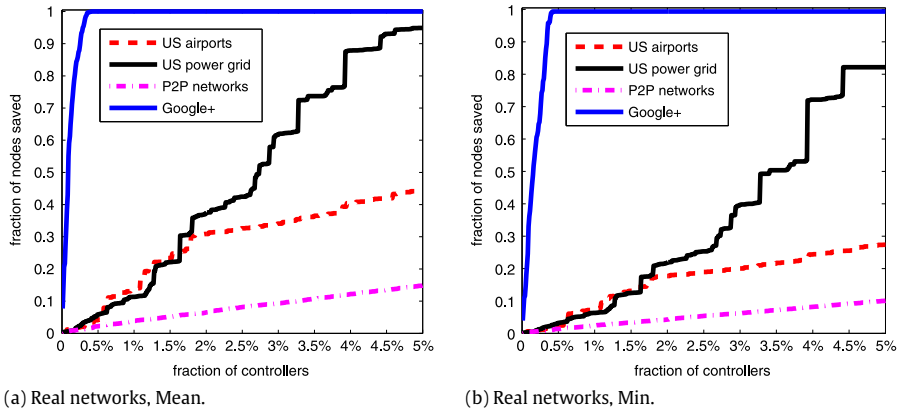


Fig. 6. The fraction of nodes saved on the networks of four real world networks by our defending algorithm $\mathcal{E}$ and $\mathcal{C}$. In this experiment, number of controllers is up to 5% of the nodes of the networks. The average and minimum are taken over all possible super virus attacks of the networks. (a) and (b) are the average fraction and the minimal fractions of nodes saved, respectively.

Our algorithm $\mathcal{C}$ assumes the following:

Top external degree hypothesis: The nodes of the top external degrees are the controllers of networks from cascading failure of virus attacks.

Intuitively, the hypothesis is correct, because the controllers defined by this way prevent the virus to spread from a module to another. Therefore, viruses can spread in only the small communities of the network.

According to the two hypotheses above, we know that a network G can be protected by a small number of controllers, if the following properties hold:

(1) The resistance maximization principle ensures that there is a partition $\mathcal{P}$ of G such that the resistance of G by $\mathcal{P}$ is maximal or large.
(2) Using $\mathcal{P}$, we define the external degree $d^E(x)$ for every node x.
(3) If:
  (i) for most x, $d^E(x) = 0$,
  (ii) there is only a small number of nodes x such that $d^E(x)$ is nontrivially large.

(1)–(3) ensure that a small number of controllers guarantee that the cascading failure set of any virus, including the super virus infect only a small number of nodes in G.

Therefore, the defending principle of networks consists of two hypotheses, the first is the resistance maximization principle, and the second is the external degree principle in (1)–(3) above.

We remark that resistance maximization is a necessary, but insufficient condition for defending the networks. However, nevertheless, resistance maximization is the principle for finding the partition that may block virus spreading in networks.

8. Conclusions and discussions

We proposed the notion of resistance of networks, and established the resistance law of networks. The notion of resistance explores that resistance maximization is the principle for the security of networks against cascading failures of viruses that randomly spread. The resistance law implies that resistance maximization is equivalent to the minimization of two-dimensional structural information of the networks. We proposed the algorithm $\mathcal{E}$ to maximize resistance of networks, and the algorithm $\mathcal{C}$ to define controllers of networks. We showed that for the networks generated by the preferential attachment model, there is no a small set of controllers that prevent global cascading failure of virus attacks. We showed that for appropriately large affinity exponent a, for the networks generated by the security model, there is a small set of controllers which ensure that any super virus attack never cause a global failure of the networks. Therefore, there exist networks that are defensible by a small number of controllers. According to the results, we proposed a defending principle for networks, which consists of two hypotheses, the first is the resistance maximization, and the second is the heterogeneity of the external degrees of the nodes defined by the partition found by the resistance maximization principle. We also showed that the networks of the security model have a phase transition determined by a *golden curve* as a function $g_c(a)$ of the affinity exponent such that if the number of controllers is slightly larger than $g_c(a)$, then the networks are controlled, and if the number of controllers is slightly smaller than $g_c(a)$, then single virus may cause a global failure of the network. We show that for some real world networks, there is a small set of controllers that protect the networks from global failure of any super virus attack.

References

[1] A. Barabási, Scale-free networks: A decade and beyond, Science 325 (5939) (2009) 412–413.
[2] P. Erdős, A. Rényi, On random graphs, Publ. Mat. 6 (1959) 290–297.
[3] P. Erdős, A. Rényi, On the evolution of random graphs, Magy. Tud. Akad. Mat. Kutató Int. Közl. 5 (1960) 17–61.
[4] A. Barabási, R. Albert, Emergence of scaling in random networks, Science 286 (5439) (1999) 509–512.
[5] R.M. Anderson, R.M. May, Infectious Diseases of Humans: Dynamics and Control, Oxford Science Publications, 1992.
[6] S. Morris, Contagion, Rev. Econom. Stud. 67 (1) (2000) 57–78.
[7] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2003, pp. 137–146.
[8] D. Kempe, J. Kleinberg, É. Tardos, Influential nodes in a diffusion model for social networks, in: Automata, Languages and Programming: 32nd International Colloquium, ICALP 2005, 2005, pp. 1127–1138.
[9] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, A. Vespignani, D.R. White, Economic networks: The new challenges, Science 325 (5939) (2009) 422–425.
[10] R. Albert, H. Jeong, A. Barabási, Error and attack tolerance of complex networks, Nature 406 (6819) (2000) 378–382.
[11] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, Breakdown of the Internet under intentional attack, Phys. Rev. Lett. 86 (2001) 3682–3685.
[12] A.E. Motter, Cascade control and defense in complex networks, Phys. Rev. Lett. 93 (2004) 098701.
[13] D.J. Watts, A simple model of global cascades on random networks, Proc. Natl. Acad. Sci. 99 (9) (2002) 5766–5771.
[14] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200–3203.
[15] A.X. Valente, A. Sarkar, H.A. Stone, Two-peak and three-peak optimal complex networks, Phys. Rev. Lett. 92 (2004) 118702.
[16] T. Tanizawa, G. Paul, R. Cohen, S. Havlin, H.E. Stanley, Optimization of network robustness to waves of targeted and random attacks, Phys. Rev. E 71 (2005) 047101.
[17] R. Cohen, S. Havlin, D. Avraham, Efficient immunization strategies for computer networks and populations, Phys. Rev. Lett. 91 (2003) 247901.
[18] C.M. Schneider, A.A. Moreira, J.S. Andrade, S. Havlin, H.J. Herrmann, Mitigation of malicious attacks on networks, Proc. Natl. Acad. Sci. 108 (10) (2011) 3838–3841.
[19] A. Li, X. Li, Y. Pan, W. Zhang, Strategies for network security, Sci. China - Inf. Sci. 58 (1) (2015) 1–14.
[20] A. Li, Y. Pan, A theory of network security: Principles of natural selection and combinatorics, Internet Math. 12 (3) (2016) 145–204.
[21] J. Gao, B. Barzel, A.L. Barabási, Universal resilience patterns in complex networks, Nature 530 (7590) (2016) 307–312.
[22] A. Li, Y. Pan, Structural information and dynamical complexity of networks, IEEE Trans. Inform. Theory 62 (6) (2016) 3290–3339.
[23] A. Li, Q. Hu, J. Liu, Y. Pan, Resistance and security index of networks: Structural information perspective of network security, Sci. Rep. 6 (26810) (2016) 1–23.
[24] A. Li, J. Li, Y. Pan, Discovering natural communities in networks, Physica A 436 (2015) 878–896.
[25] A. Li, J. Li, Y. Pan, X. Yin, X. Yong, Homophyly/kinship model: Naturally evolving networks, Sci. Rep. 5 (15140) (2015) 1–26.
[26] A. Li, X. Yin, Y. Pan, Three-dimensional gene map of cancer cell types: Structural entropy minimisation principle for defining tumour subtypes, Sci. Rep. 6 (20412) (2016) 1–25.
[27] J. Leskovec, J.J. Mcauley, Learning to discover social circles in ego networks, Adv. Neural Inf. Process. Syst. 25 (2012) 539–547.