Enhancing network robustness against malicious attacks

An Zeng* and Weiping Liu

Department of Physics, University of Fribourg, Chemin du Musée 3, CH-1700 Fribourg, Switzerland

In a recent work [Schneider et al., Proc. Natl. Acad. Sci. USA 108, 3838 (2011)], the authors proposed a simple measure for network robustness under malicious attacks on nodes. Using a greedy algorithm, they found that the optimal structure with respect to this quantity is an onion structure in which high-degree nodes form a core surrounded by rings of nodes with decreasing degree. However, in real networks the failure can also occur in links such as dysfunctional power cables and blocked airlines. Accordingly, complementary to the node-robustness measurement ($R_n$), we propose a link-robustness index ($R_l$). We show that solely enhancing $R_n$ cannot guarantee the improvement of $R_l$. Moreover, the structure of an $R_l$-optimized network is found to be entirely different from that of an onion network. In order to design robust networks that are resistant to a more realistic attack condition, we propose a hybrid greedy algorithm that takes both the $R_n$ and $R_l$ into account. We validate the robustness of our generated networks against malicious attacks mixed with both nodes and links failure. Finally, some economical constraints for swapping the links in real networks are considered, and significant improvement in both aspects of robustness is still achieved.

I. INTRODUCTION

The security of the infrastructure in modern society is of great importance. Systems such as the Internet, power grids, transportation, and fuel distribution networks need to be robust and capable of enduring random failures or intentional attacks [1]. Many processes taking place in networks could be catastrophic cascade of failures in real systems.

In a recent work, a new measure was proposed for network robustness under malicious attack on the nodes [25]. This measurement, which we call node robustness in this paper, considers the size of the largest component during all possible malicious attacks, namely $R_n = \frac{1}{N} \sum_{q=1}^{N} S(q)$, where $N$ is the number of nodes in the network and $S(q)$ is the relative size of a giant component (i.e., the fraction of nodes in the largest connected cluster) after removing $qN$ largest degree nodes. The normalization factor $1/N$ makes the robustness of networks with different sizes comparable. A robust network generally corresponds to a large $R_n$ value. With this measurement, a greedy algorithm is designed to enhance the node robustness in real systems, and a large improvement is observed even though a small number of links are modified. Moreover, the optimal structure for node robustness is found to be an onion structure in which high-degree nodes are highly connected with surrounding rings of nodes of decreasing degree. Recently, a simple method was also proposed to generate such robust onion networks [26].

However, the analysis in Ref. [25] is only based on targeted attacks on nodes. In reality, failures can happen in connections between nodes as well [18]. For example, power cables may be dysfunctional, and some airlines may be forced to cease operations due to terrible weather or terrorist attacks. In this paper, we propose a link-robustness index ($R_l$) to measure the ability of a network to resist link failures. We find that solely enhancing $R_n$ does not always improve $R_l$, and the network structure for optimal $R_l$ is far different from that of the onion network. In addition, we find that the graph spectrum index [21,22] only measures robustness against an attack on nodes, but it cannot reflect the link robustness of networks. In order to design robust networks that are resistant to different kinds of malicious attacks, we propose a greedy algorithm that is aimed at both $R_n$ and $R_l$ improvement. To validate the robustness of the resulting networks, we examined them against a more realistic attack strategy which combines both nodes and links failure. Because the manipulation of a real network always involves certain economical constraints, we took this into account in our method and some significant improvement in both $R_l$ and $R_n$ are still obtained. Finally, our study suggests that robustness improvement depends strongly on the considered attack strategy. Therefore, each real system should have its own optimal structure for robustness based on the attack it receives.

II. LINK ROBUSTNESS OF NETWORKS

Since a robust network should be able to resist the most destructive attack, we begin our analysis by comparing the harm caused by different malicious attack strategies on links. The most destructive attack is supposed to destroy the most “important” links in the networks. As in Ref. [25], we monitor the size of the giant component to estimate how the network gets destroyed after these “important” links are removed step by step. There are many methods to measure the “importance” of links. Here we mainly consider three indexes to identify the most important link to delete. The indexes include edge-betweenness, link clustering coefficient, and degree product.
The edge-betweenness of a link is the fraction of shortest paths that pass through it [27]. In this strategy, the link with the highest edge-betweenness is removed in each step. The link cluster coefficient is the number of triangles to which a given link belongs divided by the number of triangles that might potentially include it, given the degrees of the adjacent nodes [28]. In this strategy, the link with the lowest link cluster coefficient is removed in each step. The degree product of a link is simply calculated by multiplying the degree of the nodes on the two ends of the link. In this strategy, the link with the largest degree product is removed in each step. Moreover, we also use the random link removal as a benchmark for comparison. In order to simulate a more harmful strategy, we apply a dynamical approach in which the “importance” of the given link belongs divided by the number of triangles that might potentially include it, given the degrees of the adjacent nodes [28]. In this strategy, the link with the lowest link cluster coefficient is removed in each step. The degree product of a link is simply calculated by multiplying the degree of the nodes on the two ends of the link. In this strategy, the link with the largest degree product is removed in each step. Moreover, we also use the random link removal as a benchmark for comparison.

In Ref. [25], it is found that the most robust structure against node attack is the onion-like network, which corresponds to the topology with maximum $R_n$. However, it is still unclear whether this structure is tolerant to the link attack as well. We therefore report the $R_l$ in BA networks and the corresponding onion networks in Fig. 2. Interestingly, despite the fact that the onion networks are resistant to malicious node attack, they are weaker than the original BA networks with respect to the intentional link attack. More specifically, the $R_l$ in onion networks is 19.9% lower than that in the BA model (for detailed values, see Table I). One typical onion network is shown in Fig. 3(a). As we can see, nodes with almost the same degree are connected to form a layer, and different layers rely on several links to communicate. Since the edge-betweenness of these intralayer links is relatively high, they will be removed early when the network is attacked on the links. Consequently, some isolated layers can be formed quickly, which makes the onion structure sensitive to the link attack.

Therefore, it is necessary to design a structural manipulating method to enhance the link robustness for the networks. Since changing the degree of a node is commonly assumed to be more expensive than changing the connections, we keep the degree of each node invariant in our algorithm. Starting from an original network, we swap the connections of two randomly chosen edges, i.e., we randomly select two edges $ab$ and $cd$ (which connect node $a$ with node $b$, and node $c$ with node $d$, respectively), then change them to $ad$ and $bc$ only if $R_{l_{\text{new}}} > R_{l_{\text{old}}}$. We then repeat this procedure with another randomly chosen pair of edges until no further substantial improvement
TABLE I. Properties in the different networks: node-robustness index ($R_n$), link-robustness index ($R_l$), the spectrum of the adjacency matrix ($\lambda_1/\lambda_2$), degree assortativity ($r$), average shortest path length ($\langle d \rangle$), and clustering coefficient ($\langle C \rangle$).

<table>
<thead>
<tr>
<th>Network</th>
<th>Algorithm</th>
<th>$R_n$</th>
<th>$R_l$</th>
<th>$\lambda_1/\lambda_2$</th>
<th>$r$</th>
<th>$\langle d \rangle$</th>
<th>$\langle C \rangle$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>Original</td>
<td>0.201</td>
<td>0.429</td>
<td>1.856</td>
<td>−0.181</td>
<td>2.576</td>
<td>0.142</td>
</tr>
<tr>
<td></td>
<td>$R_n$-optimized</td>
<td>0.352</td>
<td>0.343</td>
<td>2.579</td>
<td>0.158</td>
<td>2.828</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>$R_l$-optimized</td>
<td>0.200</td>
<td>0.497</td>
<td>1.891</td>
<td>−0.162</td>
<td>2.584</td>
<td>0.137</td>
</tr>
<tr>
<td></td>
<td>Hybrid-optimized</td>
<td>0.219</td>
<td>0.491</td>
<td>1.898</td>
<td>−0.153</td>
<td>2.583</td>
<td>0.133</td>
</tr>
<tr>
<td>USAir</td>
<td>Original</td>
<td>0.110</td>
<td>0.244</td>
<td>2.382</td>
<td>−0.208</td>
<td>2.738</td>
<td>0.625</td>
</tr>
<tr>
<td></td>
<td>$R_n$-optimized</td>
<td>0.293</td>
<td>0.245</td>
<td>5.054</td>
<td>−0.148</td>
<td>2.875</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td>$R_l$-optimized</td>
<td>0.111</td>
<td>0.319</td>
<td>2.631</td>
<td>−0.315</td>
<td>2.492</td>
<td>0.480</td>
</tr>
<tr>
<td></td>
<td>Hybrid-optimized</td>
<td>0.196</td>
<td>0.298</td>
<td>3.018</td>
<td>−0.237</td>
<td>2.593</td>
<td>0.429</td>
</tr>
<tr>
<td>Grid</td>
<td>Original</td>
<td>0.111</td>
<td>0.093</td>
<td>1.122</td>
<td>0.001</td>
<td>6.588</td>
<td>0.123</td>
</tr>
<tr>
<td></td>
<td>$R_n$-optimized</td>
<td>0.240</td>
<td>0.173</td>
<td>1.404</td>
<td>0.356</td>
<td>6.128</td>
<td>0.015</td>
</tr>
<tr>
<td></td>
<td>$R_l$-optimized</td>
<td>0.125</td>
<td>0.248</td>
<td>1.192</td>
<td>0.019</td>
<td>4.974</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>Hybrid-optimized</td>
<td>0.161</td>
<td>0.237</td>
<td>1.272</td>
<td>0.110</td>
<td>5.017</td>
<td>0.031</td>
</tr>
</tbody>
</table>

is achieved for a given large number of consecutive swapping trials (here, we set it as $10^4$).

Actually, the link swapping greedy algorithm has been commonly applied to achieve the optimal or near-optimal network functions such as node robustness [25], immunization [31], synchronization [32], and so on. In our case, although we cannot guarantee that this algorithm will obtain the global optimum, we have checked that the results from this algorithm are relatively stable in different swap trials. Moreover, it yields similar results to those obtained by the simulated annealing algorithm in improving link robustness.

In Fig. 2, we can clearly see that the $R_l$ can be significantly improved by the algorithm. Compared to the original BA network, $R_l$ can be increased by 15.8% (see Table I for detailed values). In Fig. 3(b), we also show the structure of the $R_l$-optimized network. Different from the “onion” network obtained in Ref. [25], the $R_l$-optimized network exhibits roughly the prickle-covered “urchin” structure in which no obvious community exists and nodes with small degree are not inclined to connect to each other but mainly attach to the nodes with higher degree. In this way, each pair of nodes has many paths to communicate with one another. Even many highest edge-betweenness links are removed, the network can stay connected.

III. IMPROVING ROBUSTNESS IN REAL NETWORKS

In real systems, failures can occur not only in nodes but also in links. For example, heavy snow can cause some power cables to snap, and mechanical aircraft problems can result in flight delays or cancellations. Therefore, when designing robust networks, we should take both $R_n$ and $R_l$ into account.

In order to achieve this objective, we propose a hybrid greedy algorithm to manipulate the network structure for better robustness. Different from the process in the previous section, we swap the connections of two randomly chosen edges only if there is no improvement in a certain number of consecutive swapping trials, which is set to $10^4$ here.

In addition to the BA network model, we consider two real systems: (i) USAir: the US air transportation system [33] that contains 332 airports and 2126 airlines. (ii) Grid: an electrical power grid in part of western Europe (mainly Portugal and Spain) [34], with nodes representing generators, and links corresponding to the high-voltage transmission lines between them. This network contains 217 nodes and 320 links. Both real networks are well connected and without any isolated component.

For each network mentioned above, we obtained the corresponding $R_n$-optimized, $R_l$-optimized, and hybrid-optimized networks using the greedy algorithms, and the related results are given in Table I. As we can see from the BA model and the USAir network, optimizing $R_n$ cannot guarantee the improvement of $R_l$, and optimizing $R_l$ cannot always increase $R_n$ either. However, the hybrid method can improve both $R_n$ and $R_l$ from the original networks. More specifically, $R_n$ and $R_l$ are increased by 78.2% and 22.1%, respectively, in the USAir network. In the Grid network, the improvement of $R_n$ is 46.4% and the increment of $R_l$ can reach even 154.8%.

FIG. 3. (Color online) Simple examples of (a) the $R_n$-optimized network (the onion network), (b) the $R_l$-optimized network (the urchin network), and (c) the hybrid-optimized network. The size of the nodes is proportional to their degree. Both networks are obtained by using the corresponding greedy algorithm in a BA model with $N = 100$ and $\bar{k} = 6$. 
Compared with $R_n$-optimized and $R_l$-optimized networks, the hybrid-optimized networks do not have the advantage of a single aspect of robustness, but they are maintained with a reasonable balance between $R_n$ and $R_l$.

The spectrum of the adjacency matrix, namely the ratio of the largest and second largest eigenvalues $\lambda_1/\lambda_2$, was formerly used to characterize network robustness [21,22]. However, we observe that the spectrum index has a certain positive correlation with $R_n$ but no obvious relation to $R_l$. Therefore, it actually only represents the node robustness but cannot reflect the network robustness against link attacks. The topology properties of the resulting networks are also analyzed. The result in Table I shows that the hybrid-optimized networks usually have larger assortativity, a smaller average shortest path length, and a lower cluster coefficient than the original networks. It has been revealed that the optimal structure for $R_n$ is the onion structure in which nodes with almost the same degree are connected, so the most significant feature for the $R_n$-optimized network is the large assortativity. For the aspect of $R_l$, the most destructive attack strategy is based on the highest load (edge-betweenness), so the less significant the community structure is, the higher $R_l$ will be. Consequently, the network that is robust against link attack should have a small average shortest path length and a small cluster coefficient. Unlike the onion networks, the $R_l$-optimized networks usually do not have a large assortativity, which explains why the onion networks do not have a high $R_l$. For the resulting networks from the hybrid algorithm, they will finally carry these topology properties from both $R_n$-optimized and $R_l$-optimized networks. One example of a hybrid-optimized network is shown in Fig. 3(c). The structure is between the “onion” network and the “urchin” network. Although the hybrid-optimized network looks like an “urchin” network, it still has some links connecting the nodes with small degree.

Since the attacks in nodes and links can happen simultaneously, one interesting aspect to consider is to see how the networks in Table I react to the attack combining node failures and link failures. Accordingly, we designed a mixed attack strategy in which the largest degree nodes will be removed with probability $f$ and the links with highest edge-betweenness will be cut with probability $1 - f$. The procedure goes on until the size of the giant component reaches 0. We first set $f = 0.5$ as an example and report in Figs. 4 and 5 the performance of the networks in Table I. The results show that the hybrid-optimized networks preserve the giant component most effectively.

We then consider the mixed attack process with $f$ varying from 0 to 1. When $f = 0$, the process is just the pure highest load (edge-betweenness) attack on links. When $f = 1$, it returns to the largest degree attack on nodes. Here, we are mainly interested in the situation when $0 < f < 1$. In order to estimate in which range of $f$ the hybrid-optimized network has an advantage, we generalize the definition of robustness to a quantity $Q$ in the mixed attack process,

$$Q = \frac{1}{M} \sum_{m=1}^{M} S(m),$$

(2)

where $M$ is the total number of steps to reduce the size of the giant component to 0. $Q$ measures how tolerant a network is against a malicious attack (which can be a node attack, a link attack, or a combination of the two). According to Eq. (2), $Q = R_l$ when $f = 0$ and $Q = R_n$ when $f = 1$.

The $Q$ value of the networks in Table I under different $f$ are reported in Figs. 6 and 7. Obviously, the original networks perform worst under any $f$. The $R_n$-optimized networks can indeed improve the $Q$ value when $f$ is large. However, they do not have too much of an advantage when $f$ is small. More specifically, in the USAir network (see Fig. 6), the $R_n$-optimized network has almost the same $Q$ when $f$ is smaller than 0.4. The $R_l$-optimized network can significantly improve the $Q$ value when $f$ is small, but $Q$ drops nearly back to the original network level when $f$ is large. A similar trend can be observed also in the Grid network (Fig. 7). These phenomena are reported in Figs. 4 and 5 the performance of the networks in Table I. The results show that the hybrid-optimized networks preserve the giant component most effectively.

![FIG. 4. (Color online) The change of the relative size of giant components $S$ with attack step $m$ when different networks are attacked by the mixed strategy. The original network is USAir and the fraction of node failure $f$ is set as 0.5. The results are averaged over 100 independent realizations.](image)

![FIG. 5. (Color online) The change of the relative size of giant components $S$ with attack step $m$ when different networks are attacked by the mixed strategy. The original network is USAir and the fraction of node failure $f$ is set as 0.5. The results are averaged over 100 independent realizations.](image)
Learning how to enhance the robustness of networks is an important topic that is related to protecting real systems from random failures and malicious attacks. In the literature, most of the works have been focused on proposing methods to improve network robustness against an attack on nodes. However, the connections between nodes can also be damaged due to unexpected accidents or situations. This requires us to take link failure into account when designing robust networks. In this paper, based on the highest load attack strategy, we propose the link-robustness index to estimate how a network can resist the most destructive targeted attack on its links. Moreover, we designed a hybrid greedy algorithm to enhance both node robustness and link robustness. When attacked with the strategy combining node and link failure, the resulting networks from the hybrid method outperform the networks in which only $R_n$ or $R_l$ was improved. Finally, some economical constraints are considered when enhancing the robustness of real networks, and some significant improvement is observed.

As shown in our results, different attack strategies require different optimal network structures to be tolerant to the damage. From a practical point of view, the hybrid method can create a reliable network which is generally robust to an attack combining node failures and link failures. In reality, the probability of node failure and link failure can hardly be known, especially when systems undergo malicious attacks. Since the hybrid-optimized networks perform very stably under different attack situations, they can be the most suitable structures when designing real systems. Finally, we remark that many possible extensions of this work can be made in the future. For example, there is a family of problems in which the goal is to minimize robustness in order to design an effective immunization strategy [31,36], and the hybrid immunization on both links and nodes can be considered in this case. Moreover, link failure should also be taken into consideration when studying interdependent networks, and the idea of a hybrid-optimized method can be extended to design a robust structure for interdependent systems.

ACKNOWLEDGMENTS

We would like to thank Yi-Cheng Zhang, Ciulio Cimini, Chi Ho Yeung, and Xiao-Pu Han for helpful suggestions. This work is supported by the Swiss National Science Foundation (No. 200020-132253).
[35] In the USAir network, \( \Delta R_e = 5 \times 10^{-4} \) and \( \Delta R_l = 5 \times 10^{-4} \). In the Grid network, \( \Delta R_e = 10^{-3} \) and \( \Delta R_l = 5 \times 10^{-4} \).