Robustness and decentralized search in complex networks

Xiao, Shi

2010

Xiao, S. (2010). Robustness and decentralized search in complex networks. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/42368>

<https://doi.org/10.32657/10356/42368>

**ROBUSTNESS AND DECENTRALIZED SEARCH IN
COMPLEX NETWORKS**

XIAO SHI

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in fulfillment of the requirement for the degree of
Doctor of Philosophy

2010

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other university or institution.

Date

Xiao Shi

ACKNOWLEDGEMENT

First of all, I would like to express my sincere gratitude and appreciation to my supervisor, Dr Xiao Gaoxi, for his valuable support and advice throughout the course of this study. Without his professional and dedicated guidance, this thesis would not have been possible.

I also give my sincere thanks to Ms Du Juan, Dr Shao Xuguang and Dr Jiang Lei for their friendship and support during my difficult time.

Thanks are also dedicated to all of my friends, faculty and technicians at Communication Lab IV and Network Technology Research Center. All of my experience and interactions with those people have shaped me and my work, and helped make my graduate experience worthwhile and unforgettable.

Finally, my deepest gratitude, in love and affection, belongs to my parents Xiao Shiwei and Li Zheng, to whom, I owe all that I have ever accomplished, and also to my dearest sister, Xiao Jun, for her encouragement and support for the completion of my study.

CONTENTS

Acknowledgement	i
Contents	ii
Summary	iv
List of Abbreviations	viii
List of Figures	ix
List of Tables	xi
Chapter 1: Introduction	1
1. 1 Motivations	1
1. 2 Contributions of the Thesis	5
1. 3 Thesis Outline	7
Chapter 2: Literature Review	9
2. 1 Complex networks	9
2. 1. 1 Erdős-Rényi model	9
2. 1. 2 Small-world model	11
2. 1. 3 Scale-free network	13
2. 2 Scale-free networks	15
2. 2. 1 Properties of scale-free networks	15
2. 2. 3 Generating random scale-free networks	20
2. 3 Robustness of scale-free networks	22
2. 3. 1 Damage to the network	22
2. 3. 2 Measuring the resilience of the network	23
2. 3. 3 Analysis on the resilience of networks	24
2. 3. 4 Strategies for enhancing network robustness	27
2. 4 Network assortativity	30
2. 4. 1 Assortativity coefficient	30
2. 4. 2 The relationship between assortativity and network robustness	31
2. 5 Network search	32
2. 6 Summary	34
Chapter 3: Robustness of Complex Communication Networks under Node Protection against Intentional Attacks	36
3. 1 Introduction	36
3. 2 Node protecting schemes	37
3. 3 An analytical model	38
3. 4 Numerical results and discussions	44

3. 5	Simulations on real-life network models	50
3. 6	Conclusions.....	57
Chapter 4: On Imperfect Node Protection in Complex Communication Networks		58
4. 1	Introduction.....	58
4. 2	Node protecting schemes	59
4. 3	An analytical model	60
4. 4	Numerical results and discussions	68
4. 5	Simulations on real-life network models	79
4. 6	Conclusions.....	85
Chapter 5: Tolerance of Local Information-Based Intentional Attacks in Complex Networks		86
5. 1	Introduction.....	86
5. 2	Distributed attack schemes	87
5. 3	Theoretical analysis	89
5. 4	Simulation results and discussion	97
5. 5	Simulations on real-life network models	103
5. 6	Conclusions.....	111
Chapter 6: Robustness of Scale-free Networks under Re-wiring Operations.....		112
6. 1	Introduction.....	112
6. 2	Random networks and their robustness	115
6. 3	Rewiring method.....	117
6. 4	Analysis.....	119
6. 5	Simulation results and discussions.....	122
6. 6	Conclusions.....	131
Chapter 7: Degree-Based Decentralized Search in Complex Communication Networks: An Empirical Study.....		132
7. 1	Introduction.....	132
7. 2	Degree-based decentralized search methods	135
7. 3	Evaluations of the effects of different information availabilities	138
7. 4	Conclusions.....	151
Chapter 8: Conclusions		153
8. 1	Summary of Main Contributions	153
8. 2	Future Research Directions.....	155
Author's Publications.....		159
BIBLIOGRAPHY		161

SUMMARY

Studies on complex networks have been motivated by the desire to better understand various real-life systems. Such systems, including Internet, world-wide web (WWW), airline transportation systems, food web, protein-protein reactions, co-authorship, and terrorist activities etc., when formulated into network models, display some nontrivial common features. For example, they all have heavy-tailed nodal-degree distributions, or in other words, all can be formulated into scale-free networks.

Extensive research efforts have been made to study the robustness of complex networks especially scale-free networks. One of the most important results is that while scale-free networks are strongly tolerant against random failures, they are fragile under intentional attack which crashes network nodes in a decreasing order of their nodal degrees.

Our research mainly focuses on network robustness under intentional attack. We first evaluate three different effects which have been largely ignored in existing studies.

- (i) A lot of real-life networks, especially communication networks, have their protection mechanisms against attacks. Due to the existence of such protection, some nodes, especially some heavily protected high-degree and medium-degree nodes, may survive even the carefully planned hostile attack. We show that in different scenarios, failing to crash a small

fraction of protected nodes has different impacts on the efficiency of the intentional attack;

- (ii) Still due to the existence of network protection, it may be difficult to totally crash a node though partially paralyzing it may still be feasible. We show that such “imperfect” protection nevertheless helps significantly enhance network robustness; and
- (iii) Many hostile attacks in complex systems, especially communication systems, are propagating from nodes to their adjacent nodes, utilizing only local network-topology information, if any. We demonstrate that such local information-based attacks can be very efficient in certain cases.

Our studies also reveal some interesting relationship between network robustness and network correlation profile. Specifically, existing studies claimed that scale-free networks with larger assortativity coefficients, where large-degree nodes tend to connect to large-degree nodes, generally are more robust. By proposing a simple rewiring method which does not change any nodal degree, we show that network robustness can be steadily enhanced at a slightly decreased assortativity coefficient. The tolerance against random failures meanwhile remains largely unaffected. Such studies reveal the more complicated relationship between network robustness and network assortativity.

Lastly, we study a different topic which also heavily relies on the existence of hub nodes: the degree-based decentralized search in complex networks. A few

different methods are proposed and evaluated for searching the target node(s) with only local information being available to each node.

Overall, our studies contribute to protection and target searching in large-scale complex systems, especially complex communication systems.

LIST OF ABBREVIATIONS

Abbreviations

AS	Autonomous system
----	-------------------

LIST OF FIGURES

Fig. 2. 1: Examples showing the differences between a random network and a scale-free network.	14
Fig. 3. 1: Crash threshold of random scale-free networks under intentional attacks where different node protection schemes are adopted.	47
Fig. 3. 2: Cluster diameters of random scale-free networks under intentional attacks where different node protection schemes are adopted.	49
Fig. 3. 3: The largest cluster size of the AS- and router-level Internet models under intentional attack where different node protection schemes are adopted.	54
Fig. 3. 4: Cluster diameters of the AS- and router-level Internet models under intentional attack where different node protection schemes are adopted.	56
Fig. 4. 1: The thresholds of link-removal probability to ensure that the networks never be totally crashed by intentional attack in random scale-free networks. For Scenarios II and III, the top 0.5% highest-degree nodes are protected.	69
Fig. 4. 2: The effects of link-removal probability on network crash threshold in the random scale-free network. We let $\alpha=2.5$. For Scenarios II and III, the top 0.5% highest-degree nodes are protected.	71
Fig. 4. 3: Comparisons between theoretical and simulation results of crash thresholds in random scale-free networks. The link-removal probability is 80% for all the three scenarios. For Scenarios II and III, the top 0.5% hub nodes are protected.	73
Fig. 4. 4: The crash thresholds of random scale-free networks in different scenarios. The link removal probability is 80% for all the three scenarios. For Scenarios II and III, the top 0.5% hub nodes are protected.	74
Fig. 4. 5: The relationship between the crash threshold and the portion of hub nodes being protected in random scale-free network. Let $\alpha=2.5$ and the link-removal probability be 80%.	75
Fig. 4. 6: The largest cluster sizes in random scale-free networks during the procedure of attack. The link removal probability is 80% for all the three scenarios. For Scenarios II and III, the top 0.5% hub nodes are protected.	77
Fig. 4. 7: The cluster diameters of random scale-free networks under attack. The link removal probability is 80% for all the three scenarios. For Scenarios II and III, the top 0.5% hub nodes are protected.	78
Fig. 4. 8: The largest cluster sizes of the AS-level model under attack with different link removal probabilities. For Scenarios II and III, the top 0.5% hub nodes are protected.	81
Fig. 4. 9: The largest cluster sizes of the router-level model under attack with different link removal probabilities. For Scenarios II and III, the top 0.5% hub nodes are protected.	83
Fig. 4. 10: The cluster diameters of the AS-level and router-level Internet models under attack. The link removal probability is 80%. The top 0.5% nodes are protected in Scenarios II and III.	84

Fig. 5. 1: The crash thresholds of the greedy sequential attack and coordinated attack in random scale-free networks.	99
Fig. 5. 2: The efficiencies of the lower-bounded parallel attack and flooding attack for random scale-free network ($\alpha = 2.5$) with nodal-degree lower bounds set at 4 and 10 and without lower bound (or equivalently with lower bound at 2), respectively.	102
Fig. 5. 3: The efficiencies of the greedy sequential attack and the coordinated attack in the two Internet models, respectively.	105
Fig. 5. 4: The efficiencies of the lower-bounded parallel attack with nodal-degree lower bounds being set at 1, 4 and 10, respectively.	106
Fig. 5. 5: The efficiencies of the lower-bounded flooding attack with nodal-degree lower bounds being set at 1, 4 and 10, respectively.	108
Fig. 5. 6: The cluster diameters of the AS-level Internet model under different types of attacks.	109
Fig. 5. 7: The largest cluster size of the AS-level Internet model with node protection Scenario I under coordinated and intentional attacks. The link-removal probability is 80%.	109
Fig. 6. 1: An illustration of the rewiring operation.	119
Fig. 6. 2: Simulation results of intentional attack on the random scale-free network where exponent equals to 2.5.	123
Fig. 6. 3: Simulation results of intentional attack on the random scale-free network where exponent equals to 3.	124
Fig. 6. 4: Simulation results of random failure on the random scale-free network where exponent equals to 2.5.	126
Fig. 6. 5: Simulation results of random failure on the random scale-free network where exponent equals to 3.	127
Fig. 6. 6: Simulation results of random failure on the router-level Internet model. ..	129
Fig. 6. 7: Simulation results of recalculated intentional attack on the random scale-free network where exponent equals to 2.5.	130
Fig. 7. 1: Performance of GLAN method [18] with different amounts of local information.	141
Fig. 7. 2: Performance evaluations of the Partial-3 method.	142
Fig. 7. 3: Performance evaluations of the DCB method.	145
Fig. 7. 4: Performance evaluations of the MPPS method with two parallel searching operations.	147
Fig. 7. 5: Evaluations of the GLAN method in networks with randomly-distributed duplicates of the target.	149
Fig. 7. 6: Evaluations of the GLAN method in networks with duplicates of the target on randomly-selected adjacent nodes.	150

LIST OF TABLES

Table 3. 1: Analytical and simulation results of the cutoff nodal degrees in random scale-free networks with multiple high-degree nodes being protected. Denote the simulation results as SR and the theoretical results as TR.....	48
Table 5. 1: Analytical and simulation results of the numbers of steps needed to remove all the nodes with high-enough degrees in different networks by the lower-bounded parallel attack.....	100
Table 5. 2: Comparison of the analytical and simulation results of the numbers of steps needed to remove all the candidate nodes in different networks by the lower-bounded flooding attack	101
Table 7. 1: Performance of the DCB Method	144

1 Introduction

1.1 Motivations

Studies on complex networks have been motivated by the desire to better understand various real-life systems. Such systems, including the Internet, world-wide web (WWW), airline transportation systems, food web, protein-protein reactions, co-authorship, and terrorist activities etc., when formulated into network models, display some nontrivial common features [1-2]. For example, all the systems listed above can be formulated into *scale-free* networks [3]. Specifically, term the number of links connected to a node as its degree. A scale-free network has its nodal degree distribution matching the following equation [3]

$$P(k) \propto k^{-\alpha}, \quad (1.1)$$

where $P(k)$ denotes the fraction of nodes with a degree k , and α is the exponent which typically lies between 2 and 3 in real-life systems [1]. A visible feature of scale-free networks is the existence of a relatively large number of high-degree hub nodes.

Real-life complex systems can suffer from various kinds of damages. For example, hackers may attack the Internet, and airports may be closed due to bad

weather, etc. Such damages can be loosely classified into two catalogs: failures and malicious attacks. The most popular models for describing such damages include *random failure* and *intentional attack* [4]. Under random failure, nodes in the network are randomly removed, while under intentional attack, nodes are removed in a decreasing order of their nodal degrees.

Extensive research efforts have been made to study the impacts of damages on complex networks especially scale-free networks. One of the most important results is that while scale-free networks are strongly tolerant against random failures, they are fragile under intentional attack [4].

Various strategies have been proposed to enhance network robustness against intentional attacks [5-11]. These solutions can be divided into two categories:

- *Preventive strategies*, which enhance network robustness *before* an attack takes place. The proposed methods include hiding or partially hiding network hub nodes [10], link insertions [5-6, 11], and link rewiring [8], etc.
- *Reactive strategies*, which take remedial actions during or after an attack to mitigate the damage. Existing methods include node re-generation and link repairing [7, 9], etc.

A detailed survey of these strategies can be found in Chapter 2. Note that though these methods appear to be effective, they may be difficult to be implemented in self-organized, extra-large networks such as the Internet.

Our research mainly focuses on network robustness. We try to figure out (arguably) more realistic strategies for enhancing scale-free networks against intentional attack. Specifically, we evaluate the effects of a few inherent properties of the real-life scale-free networks which have been largely ignored in existing studies.

- Some real-life networks, especially communication networks, have their protection mechanisms against attacks. For example, on gateway and backbone routers in the Internet, there are firewall and redundancy designs to prevent these routers from being crashed by hostile attacks. The existence of such protection may cause two effects: (i) due to the existence of such protection, some nodes, especially some important high-degree or medium-degree nodes may survive even the carefully planned hostile attacks; and (ii) it may be very difficult to totally crash a protected node though partially paralyzing it may still be feasible.
- To carry out accurate intentional attack, the global network topology information must be available to the attacker for figuring out the proper target in each step of attack, which is very difficult in large-scale systems. In many systems especially communications systems, the attacks may have to propagate from compromised nodes to their adjacent nodes, utilizing only local network-topology information.

We are motivated to evaluate the influences of these effects as well as the possibility of making use of these effects to enhance network robustness.

Another factor which may affect network robustness is the network connection pattern, also known as network correlation [12]. Different scale-free networks, though

with similar nodal-degree distributions, may have different patterns of connections between network nodes with different degrees. In some networks, high-degree nodes may tend to be connected to high-degree nodes, which form into an *assortative* network; and a *disassortative* network otherwise [12]. It is found that most social networks are assortative networks where most communication networks and biological systems are disassortative [12].

Existing studies claimed that scale-free networks with stronger assortativity generally are more robust [13-14]. Such conclusion however mostly comes from the empirical studies [13-14]. Our study was to reveal some more complicated relationship between network robustness and network assortativity by constructing a counterexample.

A topic which is closely related to network robustness is network decentralized search [15], i.e., the search for a target in a complex network without the help of a central controller. The close relationship between network robustness and decentralized search may be revealed in at least two aspects: (i) decentralized search typically also heavily relies on the existence of high-degree hubs; and (ii) how easy it is to find the needed information to some extent reflects how easy it is for hostile attack to find its target. Needless to say, network search has wide applications other than network robustness, e.g., in web search and peer-to-peer (P2P) services [16-17], etc.

Decentralized search mainly include degree-based search [18], which heavily utilizes network nodal-degree information, and similarity-based search [19], which uses content

analysis. Our research mainly focuses on degree-based search. We carry out an empirical study for evaluating the effects of information availability on searching efficiency.

1.2 Contributions of the Thesis

The major contributions of the thesis can be summarized as below:

1. We evaluate network robustness against intentional attack by considering an effect which has been largely ignored in past studies: due to the existence of protection mechanism, a limited set of high or medium-degree nodes may be perfectly protected against a certain type of intentional attack. Our theoretical analysis and simulation results show that though different choices of protected nodes have different impacts on network robustness, overall speaking protecting a limited set of high or medium-degree nodes helps significantly enhance network robustness. We observe that protecting a few highest-degree hubs turns out to be effective in networks with big hubs or large exponent values; in contrast, protecting a relatively large number of medium-degree nodes may be more favorable in networks with moderate-sized hubs.
2. The network robustness against intentional attack is evaluated by considering another effect caused by the existence of protection mechanism: it may be difficult to totally crash a protected node, but still be feasible to partially paralyze it. We term such case as with imperfect node protection. By studying a few different scenarios of imperfect node protection, we show that such protection nevertheless helps significantly strengthen network robustness.

3. In many real-life systems such as communication networks, hostile attacks typically have to “propagate” throughout the network, taking down those nodes adjacent to the already compromised nodes. Since such attacks typically use at most the local network-topology information, we term them as *distributed attacks*. We consider a few different cases where highest degree node(s) adjacent to compromised nodes are taken away in each step. Theoretical analysis and numerical simulations show that though different distributed attacks have different efficiencies, overall speaking they can be highly damaging. In some cases, they are almost as efficient as the global information-based intentional attack.
4. To study the relationship between network robustness and network assortativity, we propose a simple link rewiring method which does not change any nodal degree. The rewiring method steadily enhances the network robustness under intentional attack while slightly decreasing the assortativity coefficient, which is different from the general conclusion. The tolerance against random failures meanwhile remains largely unaffected.
5. We examine degree-based decentralized search methods which selectively forward a searching request from one node to one or some of its adjacent nodes based on their respective nodal degrees. By proposing and evaluating several different methods, we show that the efficiency of a degree-based search method is sensitive to the amount of local information each node has. And such high

sensitivity cannot be easily offset by having multiple parallel searching operations or by local information exchanges between a limited number of adjacent nodes. The challenge imposed by this high sensitivity, however, almost ceases to exist when there are multiple targets and only one of them needs to be found. An example case is where there are multiple copies of the same file on different servers in the Internet.

1.3 Thesis Outline

The rest part of this thesis is organized as follows.

Chapter 2 presents a brief survey of complex networks and scale-free networks, with a main focus on the network robustness against intentional attack. The existing studies on network assortativity are also briefly discussed, followed by a short summary of existing researches on network decentralized search.

In Chapter 3, a few node protection schemes against intentional attack are proposed. Theoretical analysis and simulation results show that though different choices of protected nodes have different impacts on network robustness, overall speaking protecting a small fraction of nodes does help enhance network robustness.

Chapter 4 examines the effects of various imperfect node protection scenarios. It is shown by theoretical and simulation results that such protection could be highly effective in some cases.

The efficiency of several distributed attack schemes is investigated in Chapter 5. We show that though different distributed attacks have different efficiencies, overall speaking they can be highly damaging. In some cases, they are almost as efficient as the global information-based intentional attack.

A simple rewiring method is proposed in Chapter 6, which does not change any nodal degree. Theoretical analysis and numerical studies show that by using this rewiring method, network robustness can be steadily enhanced while assortativity coefficient is slightly decreased. Such observation reveals the more complicated relationship between network robustness and network assortativity.

An empirical study on a few degree-based decentralized searching methods is carried out in Chapter 7. The results show that the amount of local information each node has strongly affects the efficiency of search. When there are multiple searching targets in the network and only one of them needs to be found, however, even the simplest method may perform satisfactorily.

Finally, Chapter 8 concludes the thesis and proposes some possible future research directions.

2 Literature Review

2.1 Complex networks

So far there is no specific definition of complex networks. Typically they are large-scale, distributed systems with relatively simple rules followed by each individual, yet leading to surprisingly complex behavior of the whole network. The increasing research interests on complex networks have been mainly motivated by the desire to better understand various real-life systems sharing such common features.

2.1.1 Erdős-Rényi model

A *random network* is defined as a network with random connections between different nodes, only subject to a given nodal-degree distribution of $P(k)$ [20]. In 1959, Paul Erdős and Alfréd Rényi proposed the first random network model, known as the Erdős-Rényi model [21].

In this model, there are a total of N nodes in the network. For any randomly chosen pair of nodes, the probability that they have a link in between is p . Therefore the Erdős-Rényi model has a Poisson degree distribution that [21]

$$p(k) = \binom{N}{k} p^k (1-p)^{N-k} \cong \frac{z^k e^{-z}}{k!}, \quad (2.1)$$

where z is the average nodal degree; $z = p(N-1)$.

Erdős-Rényi-- model dominated the research on complex networks for a long time. The most interesting finding of this model is that there exists a threshold value p_c that when $p < p_c$, the whole network is composed of no more than a large number of small *components* (A component is a set of connected nodes.) whereas when $p > p_c$, there exists a single *giant component* in the network. This phenomenon of such critical change is termed as *phase transition* [22-23].

The existence of the giant component makes the network workable for most nodes to communicate with each other. The nodes outside the giant component are known as *isolated* nodes.

Diameter and *average path length* represent the largest and average lengths of the shortest paths between all the node pairs in the giant component, respectively.

They measure the efficiency of communications in the network. The diameter d of the Erdős-Rényi model is scaled as $d = \frac{\ln(N)}{\ln(\langle k \rangle)}$; and the average path length l has

$l \sim \frac{\ln(N)}{\ln(\langle k \rangle)}$ [24]. The fact that both d and l scale as the logarithm functions of the

network size means that the path lengths are short and increase very slowly when the network size gets bigger. Networks with such a feature are called as with the *small-world* effect [25].

Clustering is another important parameter in measuring complex networks including random networks [25]. It reflects the phenomenon that in many real-life networks a node's two neighbors may be neighbors to each other. Such three nodes form into a triangle loop. It is not uncommon in social networks that the friends of a certain people are friends of each other. The clustering coefficient measures the probability that a node's two neighbors are directly connected. In the Erdős-Rényi model, this parameter equals to p [26].

With the developments of computer technology, it becomes easier to investigate the structures of large-scale real-life networks and capture some most important properties of them. Scientists found that Erdős-Rényi model, with its over-simplified connection pattern, does not correctly represent most real-life networks. Many complex network models have been proposed ever since.

2. 1. 2 Small-world model

As mentioned earlier, Erdős-Rényi model has the small-world property that the average path length is rather short compared to network size. A lot of real-life networks are found to be with this character, many of which are with high clustering coefficients as well. Such kind of networks is called as small-world networks. The first network model successfully reflecting these two features is Watts-Strogatz model [25].

The Watts-Strogatz model was proposed in 1998 by Watts and Strogatz. It starts from a ring lattice with N nodes, each node connecting to its k neighbors. Then each edge is randomly rewired with a probability p .

The Watts-Strogatz model cannot be easily analyzed since the rewiring process may cause isolated clusters [27]. Also it is rather difficult to get closed-form expressions of its cluster coefficient and average path length. Widely accepted

results include that the cluster coefficient is approximately $C = \frac{3(k-1)}{2(2k-1)}(1-p)^3$ [28];

and the average path length is $l = \frac{N}{k} f(Nkp)$ where $f(x)$ is approximated as

$$f(x) = \frac{1}{2\sqrt{x^2 + 2x}} \tanh^{-1} \sqrt{\frac{x}{x+2}} \quad [29].$$

A much-studied variant of Watts-Strogatz model was proposed in [30-31]. In the variant model, no original edges are rewiring, instead, links are added into the randomly chosen node pairs with probability p . Different from the Watts-Strogatz model, there is no isolated clusters in the variant model, and the cluster coefficient is

$$C = \frac{3(k-1)}{2(2k-1) + 4kp(p+2)} \quad [32].$$

The small-world network models properly reflect the clustering structure in many real-life systems, esp. social networks. However, their nodal-degree distributions do not match well with those of many real-life networks. While such is not a surprise since achieving a certain nodal-degree distribution was not a serious concern when these models were proposed, the mismatch stimulated the emergence

or re-emergence of many other network models in the next few years. Among them, the most influential one may be the scale-free networks, though the existence of such networks has been known since long time ago.

2. 1. 3 Scale-free network

It is found that the nodal-degree distribution of many real-life systems are not Poisson but follow a power-law distribution instead, i.e.,

$$p(k) \sim k^{-\alpha}, \quad (2.2)$$

where α is the degree exponent. Such networks are called *scale-free* networks [3]. It is confirmed that the Internet [11-13], World Wide Web [33-35], citation networks [36-37], movie actors network [3], co-authors network [38-39], sexual contacts [40-41], metabolic network [42], email networks [43], telephone call networks [44-45], and protein network [46] etc., all can be formulated into scale-free networks. The general characteristics of these networks can refer to Tables I and II in [27]. The exponent α of most real-life networks distributes between 2 and 3. Scale-free network models become a very hot research topic largely due to the surprising observations that so many different systems in different areas can be formulated into the same class of networks.

Fig. 2.1 compares the topologies of Erdős-Rényi network and scale-free network. The most visible difference is that there are more high-degree nodes in the scale-free network. Such high-degree nodes are also called as *hubs*.

In this thesis, our discussions mainly focus on scale-free networks.

(a) A 200-node Erdős-Rényi random network.

(b) A 200-node scale-free network.

Fig. 2. 1: Examples showing the differences between a random network and a scale-free network.

2.2 Scale-free networks

2.2.1 Properties of scale-free networks

As mentioned earlier, scale-free networks follow a power-law nodal degree distribution. Below we summarize a few important characters of such type of networks.

1. Giant component

It has been proved that for random scale-free network, there exists a giant component in the network when [47]

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} \geq 2. \quad (2.3)$$

Here $\langle k \rangle$ is the average nodal degree and $\langle k^2 \rangle$ the average square of nodal degree in the network. Note that this criterion is not only applicable to scale-free networks, but also all the other random networks.

Molloy and Reed derived the same conclusion in [48] with a different expression that $\sum_k k(k-2)P(k) \geq 0$, which can be proved to be equivalent to Eq. (2.3).

2. Average path length

The average path length of a random scale-free network is expressed as

$$l = \frac{2 \log \log N}{|\log(\alpha - 2)|} \quad [49] \text{ where } 2 < \alpha < 3. \text{ It shows that scale-free networks have the}$$

small-world effect, making them efficient in supporting various kinds of node-to-node communications.

3. Network Resilience

Existing results have proved that while scale-free networks are strongly tolerant against random failures, they are fragile under intentional attack which takes down network nodes in a decreasing order of their degrees [4]. This is called the ‘robust yet fragile’ property of scale-free networks. Detailed discussion on resilience of scale-free networks can be found later in Section 2.3.

4. Assortative mixing

The concept of assortativity was introduced by Newman [12] for describing the phenomenon in real-life systems that high-degree nodes may tend to connect high-degree or low-degree nodes. The network where high-degree nodes tend to connect high-degree nodes is known to have assortative mixing; and disassortative mixing if high-degree nodes tend to connect to low-degree ones. Both of these two types of networks are *correlated* networks. If a network shows no degree correlations, meaning that the degree of an edge end is independent of the degree at the other end of the edge, this network is an *uncorrelated* network.

The assortative coefficient, which measures the assortativity level of a network, lies in the range of $[-1, 1]$. It is positive for assortative networks and negative for disassortative networks [12]. For an uncorrelated network, the assortative coefficient

is 0. It is known that most social networks are assortative mixed where most communication and biological networks are disassortative. The most important exception is probably the assortative-mixed model of the Internet on router-level [50-51]. The assortative coefficients of some real-life networks can refer to Table I in [12].

Some further discussions on network assortativity can be found in Section 2.4.

2.2.2 Typical scale-free network models

Any network with a power-law nodal-degree distribution is a scale-free network. However, not all the scale-free networks are of the equal importance in research. In addition to the obviously important case of uncorrelated random scale-free networks, a few popular scale-free models are introduced below.

1. Barabási-Albert model

The Barabási-Albert (BA) model is the first widely accepted scale-free network model. It generates scale-free networks by “growth” and “preferential attachment” [27]. Specifically, network nodes are sequentially added, where each of the newly-added nodes is connected to a fixed number m of existing nodes. The probability that a newly-added node is connected to an existing node is proportional to the degree of the existing node. The minimal nodal degree in the BA model is apparently m . The exponent α of the BA model equals to 3 [27].

Though the assortativity coefficient of the BA model is 0, BA model is in fact not a random uncorrelated network: an easy observation is that a degree with the minimum degree m is never connected to another node with degree m . This is not random behavior.

There exist various extensions to the BA model. One of them is the fitness model proposed by Bianconi *et al.* [52], where the probability that a newly added node is connected to an existing node is proportional to a combination of the existing node's degree and its fitness. Here fitness measures the node's ability to attract links other than its nodal degree. The concept of fitness is further implemented in random networks [53], where the probability that a link is inserted to connect two node is a function of their fitness levels. It is shown that without “growth” and “preferential attachment”, a scale-free network can still be generated.

2. Correlated random scale-free networks

The degree correlations that we discussed previously can also be introduced into a random graph network [54]. Specifically, in a directed network, let e_{jk} denote the probability that a directed link goes from a node with an out-degree j to a node with an in-degree k . Such definition can be extended to an undirected network by simply replacing each undirected link by two directed ones leading in opposite directions [55]. Note that in some studies, e_{jk} denotes the probability that a randomly chosen link is connected to two nodes with *remaining degrees* (i. e. the total degree discounting the link through which the node is reached, and hence one

less than the original degree of this node) j and k respectively [12, 26, 55]. In a random undirected network, e_{jk} should satisfy [12]:

$$\begin{cases} \sum_{jk} e_{jk} = 1 \\ e_{jk} = e_{kj} \\ \sum_j e_{jk} = q_k \end{cases} . \quad (2.4)$$

Here q_k denotes the probability that an end node of a randomly selected link has a degree k . We have $q_k = \frac{kp_k}{\langle k \rangle}$ for the case where k is the original degree, and

$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$ for the case where it is the remaining degree [12].

The correlated random networks are defined as networks with random connections between network nodes only subjected to given $\{e_{jk}\}$. In literature, “random networks” generally refer to uncorrelated random networks. Unless otherwise specified, such will also be the case in this thesis.

3. HOT model

Among Internet scientists, there have been extensive debates on whether the Internet actually has the ‘robust yet fragile’ property. Doyle and his colleagues argued that though the Internet models follow the power-law degree distribution, their specific structures make them significantly different from the common scale-free network models [56]. They proposed the highly optimized/organized tolerance/tradeoffs (HOT) model. In this model, as a result of careful design for

maximizing network throughput, the hub nodes are typically the edge nodes with a large number of low-capacity connections. The removal of these hub nodes, though disastrous to the large number of low-capacity users connected to them, will not easily bring down the Internet.

2. 2. 3 Generating random scale-free networks

Network generation methods are of importance since they allow convenient studies on the properties of various networks and how such properties are affected by different parameters. While there are extensive researches on generating various scale-free networks, e.g., with different clustering and/or assortativity coefficients [8, 37-39], below we only introduce the simplest method for generating the uncorrelated random scale-free networks.

1. Generalized random scale-free networks

Molly and Reed proposed an algorithm for generating a random network with any given nodal-degree distribution [57]. This algorithm is widely used in generating random scale-free networks. The steps of this algorithm include

- Firstly, assign each node with a nodal degree according to the given degree distribution $p(k)$. The sum of the degrees must be even.
- Randomly connecting a pair of nodes. The probability of choosing a node is proportional to the node degree. Repeat this step until all the edges have been assigned.

Theoretically, with this construction process, the resulted network is a random network with no degree correlations. Simulation results however show that when $\alpha \leq 3$, the network is usually a disassortative network caused by the constraints of not having self-loop or multiple connections between any pair of nodes [58]. It is pointed out that, to avoid the possible degree correlations, for an N -sized network, the upper bound of the highest nodal degree, termed as the *cutoff*, should not exceed $\sqrt{N}$ [58].

2. Degree distribution: An explanatory note

Note that there are two kinds of nodal-degree distribution models used in research on scale-free networks: continuous and discrete power-law distributions [59].

In the continuous distribution, $p(k) = Ck^{-\alpha}$ where the degree k is a continues variable and the constant C satisfies $\int_m^K Ck^{-\alpha} dk' = 1$. Here m and K denote the minimum and cutoff degrees in the network, respectively. To generate a network for simulations, the degree distribution can be approximated as $p(k) = \int_{k-1/2}^{k+1/2} Ck'^{-\alpha} dk'$ or $p(k) = \int_k^{k+1} Ck'^{-\alpha} dk'$ (leading to no major differences in the output [59]). In the discrete distribution, $p(k) = Ck^{-\alpha}$ where the degree k is an integer variable and the constant C satisfies $\sum_m^K Ck^{-\alpha} = 1$.

The detailed comparisons between these two distributions can be found in [60-61]. Generally speaking, the continuous degree distribution is more widely used.

In our simulations, unless otherwise specified, we adopt the continuous power-law distribution model.

2.3 Robustness of scale-free networks

2.3.1 Damage to the network

Real-life networks are suffering from various types of damages, as we have mentioned in Chapter 1. Such damages have important influences on our life.

Accidental damages are usually modeled as random failures which randomly select nodes to be removed from networks. Damages caused by hostile attacks are typically evaluated by studying the effects of the intentional attack which takes down network nodes in a decreasing order of their degrees [4]. When a node is taken down, all the links connected to it are also removed.

Besides the most popular models of random failure and intentional attack, other important models for damage evaluations include i) cascading failure, where one failure can cause chain reactions of failures such as that in electricity power lines [62-63]; ii) hostile attacks which remove or congest links rather than crashing network nodes [64-66]; and iii) hostile attack on network nodes in a decreasing order of their betweenness (measured as the number of shortest paths going through the node) rather than nodal degrees [67], etc.

2.3.2 Measuring the resilience of the network

Both random failures and intentional attacks destroy the connections of the network. In both cases, there exists a threshold p_c that when the fraction of removed nodes is less than p_c , the network still has a giant component; once the fraction removal goes beyond p_c , the network is broken into small clusters. This p_c is called *crash threshold* [47].

Equation (2.3) is often used to test the existence of a giant component in a random network [47, 68]. Specifically, a typical method is to recalculate the nodal degree distribution and the corresponding value of $\langle k^2 \rangle / \langle k \rangle$ after some node/link removals until the criterion is no longer satisfied. The crash threshold p_c can therefore be calculated.

Another useful measure is the *largest cluster size*, defined as the number of nodes in the biggest connected component versus the overall number of nodes in the original network [4]. If the size is less than a certain threshold, the network may be regarded as crashed. Such measurement may be more meaningful in some applications where once a network is attacked to be too small, it is regarded as useless though the surviving part may be still largely connected.

The average path length of the largest cluster, which is termed as *cluster diameter*, is also quite commonly used, especially for evaluating the network's efficiency in supporting communications between nodes after attack [4]. A smaller cluster diameter means higher communication efficiency.

2.3.3 Analysis on the resilience of networks

Theoretical analysis works side-by-side with numerical simulations to help achieve better understanding of network resilience to various damages. Below we briefly summarize analytical results on a few most typical and important cases.

1. Random failures

Cohen *et al.* developed a closed-form analysis on crash threshold of random networks under random failure [47]. Below is the derivation.

Denote the fraction of the removed nodes as p . Then each link connected to a remaining node has a probability p leading to a removed node. After random failures, the degree distribution is changed from $p(k)$ to

$$p'(k) = \sum_{k_0=k}^K p(k_0) \binom{k_0}{k} (1-p)^k p^{k_0-k}, \quad (2.5)$$

where K denotes the highest nodal degree in the original network. The average nodal degree in remained network is $\langle k \rangle = \langle k_0 \rangle \times (1-p)$ and the average square of nodal degree is $\langle k^2 \rangle = \langle k_0^2 \rangle \times (1-p)^2 + \langle k_0 \rangle p(1-p)$ [47]. According to Eq. (2.3), the crash threshold p_c is

$$p_c = 1 - \frac{1}{\kappa_0 - 1}, \quad (2.6)$$

where $\kappa_0 = \langle k_0^2 \rangle / \langle k_0 \rangle$ is calculated from the original network. In scale-free networks, κ_0 takes a large value and hence p_c is close to 1, meaning that such networks are highly robust against random failures.

After random failures, the cluster diameter of the network is changed to $l' = \frac{2 \log \log P_\infty N}{\log(\alpha - 2)}$ for $2 < \alpha < 3$, where P_∞ is the largest cluster size. [69]

Further discussions in [70] clarifies the applicable range of Eq. (2.6) as *simple networks* with no self-loop or multi-links between any node pair. More importantly, it is pointed out that the cutoff of the N -sized network should not exceed $\sqrt{\langle k \rangle N}$.

2. Intentional attack

To the best of our knowledge, so far there is no closed-form solution for calculating the crash threshold under intentional attack. An approximate solution however has been proposed in [68] which we will briefly present below.

Denote the original cutoff nodal-degree in the network as K , which can be estimated as

$$\sum_{k=K}^{\infty} P(k) = \frac{1}{N}. \quad (2.7)$$

Here N denotes the network size. Denoting the new cutoff in the network after a fraction p of nodes is removed as $\tilde{K}$, we have $p = \sum_{\tilde{K}}^K P(k)$. If N is large enough, K

can be neglected. $\tilde{K}$ then can be approximately calculated as

$$\tilde{K} = mp^{1/(1-\alpha)}, \quad (2.8)$$

where m is the minimum degree in the original network.

Denote the corresponding link loss probability as $\tilde{p}$, $\tilde{p} = \sum_{k=\tilde{K}}^K \frac{kP(k)}{\langle k_0 \rangle}$. When K can be neglected, $\tilde{p}$ can be approximately expressed as

$$\tilde{p} = \left(\frac{\tilde{K}}{m} \right)^{2-\alpha} = p^{(2-\alpha)/(1-\alpha)}. \quad (2.9)$$

The above equation shows that when α is close to 2, a very small fraction of node removal will cause a heavy loss of links. Under such case, κ becomes

$$\kappa = \left(\frac{2-\alpha}{3-\alpha} \right) \frac{\tilde{K}^{3-\alpha} - m^{3-\alpha}}{\tilde{K}^{2-\alpha} - m^{2-\alpha}}. \quad (2.10)$$

Numerically solve the equation

$$1 - \tilde{p} = \frac{1}{\kappa - 1}, \quad (2.11)$$

the new cutoff $\tilde{K}$ and the corresponding p which is the crash threshold can both be obtained. Analytical and simulation results both show that the crash threshold p is of a very small value, which means that the scale-free networks are fragile under intentional attack.

The average path length after the intentional attack is $l'' = \frac{\log P_\infty N}{\log(\kappa - 1)}$ for the case where $2 < \alpha < 3$. Here $\kappa = \langle k^2 \rangle / \langle k \rangle$ [69].

The above analyses on network resilience to random failure and intentional attack prove the ‘robust yet fragile’ property of the scale-free networks.

There are also other studies on the network robustness under intentional attack. In [69] it is shown that, for either random failure or intentional attack, there exists a phase transition in the fraction of node loss, below which the network basically can still function well. Gallos *et al.* evaluated the case where the probability that a node gets crashed is a function of its degree [71]. They found that the networks remain to be fragile even with a rather low probability of node crash for high-degree nodes.

2.3.4 Strategies for enhancing network robustness

Extensive efforts have been made to study how to enhance network robustness [5-11, 72-75]. Existing results can be roughly classified into three categories:

1. Optimization of network topologies

The objective is to find network topologies with the best robustness. It was proved in [72] that the most robust network under random failures subject to a given $\langle k \rangle$ has only two different nodal degrees. A framework for finding network robustness under both random failures and intentional attacks was proposed in [74]. Parameters of nodal-degree distribution for reaching a good tradeoff between robustness against random failure and intentional attack are discussed in [73]. In [75], it is proved that the optimal design against a certain combination of failures and attacks has only three different nodal degrees.

Note that none of these proposed solutions are scale-free networks.

2. Preventive strategies:

Preventive strategies enhance network robustness *before* an attack takes place, which include hiding or partially hiding network hub nodes [10], link insertion [5-6, 11] and link rewiring [5, 8], etc.

It is pointed out in [10] that if due to incomplete information of network topology, some high or medium nodes are missed in the intentional attack, network robustness can be significantly enhanced.

Two link insertion schemes are compared with each other in [5]. The scheme of adding links between lowest-degree nodes outperforms the one randomly adding links. In [11], link insertions between low-degree nodes are compared with link insertions between high-degree nodes. The results still show that enhancing the links between low-degree nodes is more effective. The case where links are only inserted between far-away node pairs (e. g. the shortest path between the node pair is no less than a certain value) is discussed in [6], the network robustness can be significantly enhanced at a high cost of inserting a large-enough number of links.

Different from link insertion, link rewiring does not change the number of links in the network. Four rewiring schemes are proposed in [5]. Among them, repetitively rewiring the link originally connected to the highest-degree node to a randomly selected node performs the best.

3. Reactive strategies:

Reactive strategies take remedial actions during or after an attack to mitigate the damage. Typical methods include node re-generation [7] and link repairing [9].

A node re-generation scheme is discussed in [7] where nodes losing their neighbors can find replacements and the deleted nodes can rejoin into the network. It is found that the new network may lose its heavy-tailed feature in the degree distribution. A healing algorithm is hence proposed to repair the sharp cutoff in degree distribution.

Several link repairing schemes are proposed in [9]. When a node is crashed, each of its adjacent nodes will try to restore the lost link by connecting to a neighborhood node (based on local information) or a randomly selected node in the whole network, subject to some redundant resource constraints. Simulation results show that local information based repairing significantly enhances network robustness. In fact, a very interesting observation is that the case with minimum local information exchanges and limited redundant resources achieves the best performance.

2.4 Network assortativity

2.4.1 Assortativity coefficient

As we have discussed earlier, the assortativity coefficient measures in complex networks whether high-degree nodes tend to connect high-degree or low-degree nodes [12].

We have introduced the distribution e_{jk} in Section 2.3 which reflects the probability that a randomly chosen edge is connected to two nodes with degrees j and k respectively. The distribution q_k has also been introduced in the same section, defined as the probability that the node reached by following a randomly selected edge has a degree k . In an uncorrelated network, $e_{jk} = q_j q_k$.

The assortativity coefficient is defined as $\sum_{jk} jk(e_{jk} - q_j q_k)$. For the convenience of comparison between different networks, this coefficient is normalized as

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k), \quad (2.12)$$

where $\sigma_q^2 = \sum_k k^2 q_k - \left[\sum_k k q_k \right]^2$, defined as the assortativity coefficient of a perfectly assortative network (i.e., network with links only existing between nodes with same degrees [12]). Apparently, after normalization, $-1 \leq r \leq 1$. For uncorrelated networks, $r = 0$; for assortative networks, r is positive and for disassortative networks, it is negative [12].

An equation has also been proposed to calculate the assortativity coefficient in a practical network [12]:

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}. \quad (2.13)$$

where M is the total number of edges in the network, and j_i and k_i are the degrees of nodes at the two ends of i_{th} edge respectively.

2. 4. 2 The relationship between assortativity and network robustness

The robustness of correlated network under random failures is investigated in [76]. The robustness is influenced by the value of $\langle k^2 \rangle$ and the assortativity coefficient. When $\langle k^2 \rangle \rightarrow \infty$, the network is highly robust under random failure. Assortative network is highly robust even when $\langle k^2 \rangle$ is finite. Compared to that, disassortative networks are more fragile.

Generally speaking, assortative networks are also more robust than disassortative ones against intentional attack [13-14]. In [14], the authors generate a serial of correlated networks by rewiring the links between node pairs of an originally uncorrelated network. The rewiring operations do not change any nodal degree. Simulation results show that for networks with the same degree distribution, they are generally more robust when the assortativity coefficient is higher. Exceptions only happens when the assortativity coefficient comes close to its maximum value: occasionally network robustness is slightly weaker at a higher

value of assortative coefficient. In [13], the same rewiring method is adopted on some real-life biological networks. The same conclusion holds that a higher assortativity level generally leads to stronger network robustness where exceptions only occasionally happen.

2.5 Network search

An important operation in complex networks is network search. As we have mentioned in Chapter 1, network search is related to network attack and protection while it has its own wide applications. The existing network search methods mainly include *exhaustive search* and *guided search* [26]. In exhaustive search, all the contents found in the network are indexed and stored locally, e.g., with the implementations of network mining crawlers. With carefully indexed and locally stored information, all the search queries can be executed locally. Such methods are mainly used by powerful web search engine [26, 77]. In guided search, on the other hand, only a small fraction of the network is crawled by a limited number of crawlers to find a certain target in the network for each query.

Guided search is typically *decentralized search* [15]. In other words, the operations of finding and/or reaching a certain target are carried out with the coordination of a central control. While centralized search is becoming increasingly powerful, driven by giant companies including Google and Microsoft, decentralized search is also gaining increasing research interests, thanks to its potential and existing wide applications in peer-to-peer networks [16-17], web crawlers for

specific groups of users [78-79], sensor networks [80-81], and dynamic supply chains [82], etc. The existing guided search methods can be mainly classified into following three different categories:

- Methods with minimum assumptions of network topologies and content distributions. Typical methods include the breadth-first searching method with limited flooding [83-84], random walks [85], and various *informed* searching methods in which the information available to each node may vary from simple forwarding hints to exact object locations [86-87], etc.
- Methods that explore the special features of complex-network topologies, most notably the well-known “six-degree separation” in small-world networks [88] and the relatively large number of hubs in scale-free networks. For scale-free networks, the most important method is the *degree-based* searching method [18, 89] which forwards the searching requests to selected adjacent nodes based on their respective nodal degrees until the target is found or the searching request is dropped. By using such methods, higher-degree nodes, which are connected to more adjacent nodes and therefore have a better chance to locate the target, are generally more likely to be selected [15, 18, 89]. In fact, the classic degree-based method simply forwards the request to the largest-degree adjacent node [18]. Other existing results include those methods that explore the hierarchical structure

of social networks [88, 90], or the betweenness of each node in weighted scale-free networks [91], etc.

- Methods that explore content distributions on different network nodes. The most important one is probably the *similarity-based* method [19, 92-93], which models and resembles the human behaviors of tending to forward the searching requests to the adjacent node(s) with most similar content to the content that is being searched. For example, it may be plausible to forward a request for soccer-team information to basket-ball fans rather than to opera addicts. The major challenge in developing similarity-based methods for telecommunication applications, however, is to devise proper measurements of similarity [19, 92]. Closely related topic includes identifying community structures in complex networks [94-95]. A community is a set of nodes with higher densities of connections among themselves than to the outside nodes and typically with certain content similarity as well.

Recently, there is an effort to combine the degree-based and similarity-based methods [89, 96]. The initial results turn out to be quite encouraging.

2.6 Summary

In this chapter, to provide a general background for the later discussions on enhancing the network robustness of complex networks under intentional attack and improving the efficiency of network search, we gave a brief survey of the existing

studies on related topics in complex networks, especially scale-free networks. In the following chapters, we shall investigate the effects of a few factors which have been largely missed in existing studies on network robustness (Chapters 3-5); propose some conclusions contradict to the existing ones on the relationship between robustness and network assortativity (Chapter 6), and carry out a careful empirical study on the efficiency of degree-based decentralized search in scale-free networks (Chapter 7).

3 Robustness of Complex Communication Networks under Node Protection against Intentional Attacks

3.1 Introduction

As described in Chapter 2, scale-free networks are strongly tolerant against random failures yet fragile under intentional attack [4]. To enhance network robustness, different methods have been proposed including link insertion [5-6, 11], link rewiring [8] and link recovery [7], etc. Though these methods appear to be effective, they may be difficult to be implemented in self-organized, extra-large networks such as the Internet.

In this chapter we study enhancing the robustness of complex networks by providing protections over a limited set of high or medium-degree nodes. Such a protection mechanism may actually already exist in some real-life systems. For example, on gateway and backbone routers in the Internet, there are firewall and redundancy designs to prevent these routers from being crashed by hostile attacks. Compared to operations such as link insertion or link rewiring, protecting a small set of nodes arguably may be easier and more realistic. Our theoretical analysis and simulation results show that though different choices of protected nodes have

different impacts on network robustness, overall speaking protecting a limited set of high or medium-degree nodes helps enhance network robustness.

The rest of this chapter is organized as follows. In Section 3.2, A few node protection schemes are proposed. Theoretical analysis and simulations results on random scale-free networks are proposed in Sections 3.3 and 3.4, respectively. Numerical evaluations on a few real-life network models are proposed in Section 3.5. Finally, Section 3.6 concludes the chapter.

3.2 Node protecting schemes

We consider three node protection schemes as follows:

- Only the single highest-degree node is protected. Such a scheme can serve as a benchmark for the other node protection schemes.
- A few highest-degree nodes are protected. Studies on such a scheme help reveal the enhancements to network robustness when more high-degree nodes are protected.
- The high-degree nodes are all taken down, yet some medium-degree nodes are protected. Such a scheme may be easier to be implemented in certain circumstances since medium-degree nodes are generally less obvious targets than highest-degree ones. We assume that a certain fraction of the medium-sized nodes are protected, where the so-called medium-degree

nodes refer to those with their degrees varying in a certain range from T_1 to T_2 .

The effects of these schemes in the random scale-free networks will be analyzed in the next section.

3.3 An analytical model

As that has been defined earlier in Chapter 2, a *random network* has random connections between different nodes, only subject to a given nodal-degree distribution of $P(k)$ [21]. It is well known that a random network loses its global connectivity when [47]

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} < 2, \quad (3.1)$$

where $\langle k \rangle$ denotes the average nodal degree, and $\langle k^2 \rangle$ the average square of nodal degree. Under random node failures, the *crash threshold*, defined as the fraction of nodes to be removed for the network to lose its global connectivity, is given as [47]:

$$p_c = 1 - \frac{1}{\kappa_0 - 1}. \quad (3.2)$$

Here $\kappa_0 = \langle k_0^2 \rangle / \langle k_0 \rangle$ where $\langle k_0 \rangle$ and $\langle k_0^2 \rangle$ denote the average nodal degree and average square of nodal degree in the original network respectively. For the case under intentional attack, an approximate analytical solution of the crash threshold is

presented in [68], which has been presented in Section 2.3.3. Our analysis on intentional attack remains nearly the same, with only slight modifications. For the convenience to understand our analysis, we nevertheless present our analysis in detail.

If the network is crashed after a fraction of p_c of all the nodes is removed, p_c is its crash threshold. The network cutoff, which denotes the highest original degree of the node which has *not* been removed, is reduced to $\tilde{K}$. As pointed out in [68], the intentional attack on a network can be viewed as composed of two different steps: (i) removal of all the nodes with degrees higher than $\tilde{K}$; and then (ii) random removal of links connected to the other nodes at a probability of $\tilde{p}_c$. We have

$$p_c = \sum_{\tilde{K}}^K p(k) = \frac{K^{1-\alpha} - \tilde{K}^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}}. \quad (3.3)$$

Note that Eq. (3.3) is slightly different from Eq. (7) in [68] since when the network size is not large enough, the original cutoff K may not be safely ignored as what was done in Eqs. (8) and (10) of [68]. Also note that the sum of p_c starts from $\tilde{K}$ rather than $\tilde{K} + 1$ since the continuous model which we have introduced in Chapter 2 is adopted. The corresponding probability $\tilde{p}_c$ of *link* loss therefore approximately equals to

$$\tilde{p}_c = \sum_{\tilde{K}}^K \frac{kp(k)}{\langle k \rangle} = \frac{K^{2-\alpha} - \tilde{K}^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}}. \quad (3.4)$$

After the intentional attack, the nodal-degree distribution is changed to

$$\tilde{p}(k) = \sum_{k_0=k}^{\tilde{K}} p(k_0) \binom{k_0}{k} (1 - \tilde{p}_c)^k \tilde{p}_c^{(k_0-k)} \quad (3.5)$$

Under such case, κ becomes

$$\kappa = \left(\frac{2-\alpha}{3-\alpha} \right) \frac{\tilde{K}^{3-\alpha} - m^{3-\alpha}}{\tilde{K}^{2-\alpha} - m^{2-\alpha}}. \quad (3.6)$$

Since in random node failure, link loss probability equals to node loss probability, replacing p_c by $\tilde{p}_c$ in Eq. (3.2) yields

$$1 - \tilde{p}_c = \frac{1}{\kappa - 1} \quad (3.7)$$

From Eqs. (3.3), (3.4), (3.6) and (3.7), p_c can be numerically solved.

Now we study the cases with node protection. Firstly we consider the benchmark case where only the single highest-degree node is protected. The degree of this node in the original network is K . In an N -node network, we shall have

$$p(K) = \frac{1}{N}. \quad (3.8)$$

Denote the crash threshold under such case as p'_c and the corresponding cutoff value (excluding the protected hub) as $\tilde{K}'$. We have

$$p'_c + \frac{1}{N} = \sum_{\tilde{K}'}^K p(k) = \frac{K^{1-\alpha} - \tilde{K}'^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}}. \quad (3.9)$$

The corresponding probability of link loss, denoted as $\tilde{p}'_c$, can be expressed as

$$\tilde{p}'_c = \frac{\sum_{\tilde{K}'}^{K-1} kp(k)}{\langle k \rangle} = \frac{(K-1)^{2-\alpha} - \tilde{K}'^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}}. \quad (3.10)$$

And the network nodal-degree distribution is changed to:

$$\tilde{p}(k)' = \frac{1}{N} \binom{K}{k} (1 - \tilde{p}_c')^k \tilde{p}_c'^{K-k} + \sum_{k_0=k}^{\tilde{K}'} p(k_0) \binom{k_0}{k} (1 - \tilde{p}_c')^k \tilde{p}_c'^{(k_0-k)}. \quad (3.11)$$

Therefore κ of the after-attack network can be approximated as

$$\kappa = \frac{\frac{C}{3-\alpha} (\tilde{K}'^{3-\alpha} - m^{3-\alpha}) + \frac{1}{N} K^2}{\frac{C}{2-\alpha} (\tilde{K}'^{2-\alpha} - m^{2-\alpha}) + \frac{1}{N} K}. \quad (3.12)$$

Using Eqs. (3.7) and (3.9), (3.10) and (3.12), p_c' can be numerically solved.

For the case where a total of N_0 highest-degree nodes are protected, a similar analysis can be carried out in what follows.

Denote the lowest original degree of the N_0 protected nodes as T in an N -node network. Denote the crash threshold under such case as p_c'' and the corresponding cutoff value (excluding the protected hubs) as $\tilde{K}''$. We have

$$p_c'' + \frac{N_0}{N} = \sum_{k=T}^K p(k) = \frac{K^{1-\alpha} - \tilde{K}''^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}}. \quad (3.13)$$

The corresponding probability of link loss, denoted as $\tilde{p}_c''$, has,

$$\tilde{p}_c'' = \sum_{k=T}^{\tilde{K}''} \frac{kp(k)}{\langle k \rangle} = \frac{T^{2-\alpha} - \tilde{K}''^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}}. \quad (3.14)$$

And the network degree distribution is changed to

$$\tilde{p}(k)'' = \sum_{k_0=T}^K p(k_0) \binom{k_0}{k} (1 - \tilde{p}_c'')^k \tilde{p}_c''^{k_0-k} + \sum_{k_0=k}^{\tilde{K}''} p(k_0) \binom{k_0}{k} (1 - \tilde{p}_c'')^k \tilde{p}_c''^{k_0-k} \quad (3.15)$$

Therefore κ of the after-attack network can be approximated as

$$\kappa = \frac{\frac{C}{3-\alpha} \left[(\tilde{K}^{n^{3-\alpha}} - m^{3-\alpha}) + (K^{3-\alpha} - T^{3-\alpha}) \right]}{\frac{C}{2-\alpha} \left[(\tilde{K}^{n^{2-\alpha}} - m^{2-\alpha}) + (K^{2-\alpha} - T^{2-\alpha}) \right]}. \quad (3.16)$$

By using Eqs. (3.7), (3.13), (3.14) and (3.16), p_c'' can be numerically solved.

In independent realizations of numerical simulations, it is observed that there exist strong fluctuations in the value of p_c'' , which can be easily understood: the degrees of the N_0 protected nodes vary from one realization to another though statistically they satisfy the power-law distribution. When multiple high-degree nodes are protected, even when the value of N_0 is still relatively small, a scale-free network can be enhanced to have strong robustness; or in other words, the cutoff $\tilde{K}''$ is of a rather low value. The fluctuations in the specific degrees of the protected nodes therefore lead to big differences in $\langle k^2 \rangle / \langle k \rangle$ in the after-attack networks, which in turn significantly changes the value of p_c'' . A more meaningful measurement of network robustness under such case may be $\tilde{K}''$, which reveals the necessary “depth” of the attack in order to crash a network. Note that the value of $\tilde{K}''$ can also be numerically solved from Eqs. (3.7), (3.13), (3.14) and (3.16).

Finally we consider the case where largest-degree nodes are all removed yet a fraction p_f of medium-degree nodes are protected. The degrees of the medium-degree nodes lies in the region of $[T_1, T_2]$.

Denote under such case the crash threshold to be p_c''' and the corresponding cutoff value (excluding the protected nodes) as $\tilde{K}'''$. Denote the fraction of nodes that has been protected as p_t . We have

$$\begin{cases} p_c''' + p_t = \sum_{\tilde{K}'''}^K p(k) = \frac{K^{1-\alpha} - \tilde{K}'''^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}} \\ p_t = p_f \cdot \sum_{T_1}^{T_2} p(k) \end{cases} \quad (3.17)$$

The corresponding probability of link loss, denoted as $\tilde{p}_c'''$, can be expressed as,

$$\tilde{p}_c''' = \sum_{\tilde{K}'''}^K \frac{kp(k)}{\langle k \rangle} - p_f \cdot \sum_{T_1}^{T_2} \frac{kp(k)}{\langle k \rangle} = \frac{(K^{2-\alpha} - \tilde{K}'''^{2-\alpha}) - p_f(T_2^{2-\alpha} - T_1^{2-\alpha})}{K^{2-\alpha} - m^{2-\alpha}}. \quad (3.18)$$

The network degree distribution is changed to

$$\tilde{p}(k)''' = p_f \sum_{k_0=T_1}^{T_2} p(k_0) \binom{k_0}{k} (1 - \tilde{p}_c''')^k \tilde{p}_c'''^{k_0-k} + \sum_{k_0=k}^{\tilde{K}'''} p(k_0) \binom{k_0}{k} (1 - \tilde{p}_c''')^k \tilde{p}_c'''^{k_0-k} \quad (3.19)$$

Therefore κ of the after-attack network approximately equals to

$$\kappa = \frac{\frac{C}{3-\alpha} \left[(\tilde{K}'''^{3-\alpha} - m^{3-\alpha}) + p_f (T_2^{3-\alpha} - T_1^{3-\alpha}) \right]}{\frac{C}{2-\alpha} \left[(\tilde{K}'''^{2-\alpha} - m^{2-\alpha}) + p_f (T_2^{2-\alpha} - T_1^{2-\alpha}) \right]}. \quad (3.20)$$

By using Eqs. (3.7), (3.17), (3.18) and (3.20), p_c''' can be numerically solved.

In addition to the crash threshold, another parameter useful for measuring the robustness of complex networks, especially complex communication networks, is their *cluster diameter*, defined as the average hop length of the shortest paths

between all the source-destination node pairs in the largest connected component [4]. A network with a large cluster diameter, though may be still connected, cannot efficiently support communication applications. Accurate analysis on cluster diameter under intentional attack, however, largely remains an open issue, as mentioned earlier in Section 2.3.3. In this chapter, we evaluate cluster diameter by numerical simulations.

3.4 Numerical results and discussions

In our simulations, random scale-free networks are generated by adopting the algorithm proposed in [58], with exponent value distributing between 2 and 3. Unless otherwise specified, the network size N is 10,000; the minimal nodal degree m is 1 and the cutoff degree K is 100, which equals to $\sqrt{N}$. For the case with multiple protected hubs, we let N_0 vary from 1 to 5. For the medium-side node protection scheme, we consider the cases that the top 1% of the biggest hubs is removed and after that, 10% and 50% of the next 1% of largest-degree nodes are protected respectively. We adopt the definition in [47] that a network is crashed once $\langle k^2 \rangle / \langle k \rangle \leq 2$ is achieved.

Figure 3.1 shows the theoretical and simulation results of crash threshold in random scale-free networks under intentional attack with different node protection schemes.. Each curve comes from an average of 50 independent realizations. For comparison, the results for the case without any protection are also presented. Note that for such case, error bars are not plotted to avoid making the figures too crowded.

Generally the simulation results for the case without protection have much smaller fluctuation ranges than those of the other cases.

The single-hub protection scheme is evaluated in Fig. 3.1(a). We see that by protecting a single biggest hub with a moderate cutoff value of 100 in a 10,000-node network, we can steadily increase the crash threshold by 3 ~ 5% , meaning that 300~500 additional nodes have to be removed in order to crash a network. When the degree of the biggest hub is higher, the enhancement effects shall become even more significant.

Fig. 3.1(b) illustrates the effects of protecting 10% of the medium-sized nodes. Under such case we put 0.1% of all the network nodes under protection. We see that the enhancement effects are more significant where α is of a smaller value. For example, when $\alpha = 2.1$, the threshold increases from 10.5% to 14% ; yet when $\alpha = 2.9$, it increases only marginally from 3.25% to 3.5%. This can be explained as follows. Though the networks are generated with the same minimal nodal degree and the same cutoff, different exponent values lead to different nodal-degree distributions of the networks. Specifically, degrees of the protected medium-sized nodes become lower in networks with larger exponents. For example, when $\alpha = 2.1$, the degrees of the protected nodes vary from 27 to 45; whereas when $\alpha = 2.9$, they vary from 7 to 12. The differences become even more significant when 50% of medium-size nodes are protected. As we can see in Fig. 3.1(c), for the cases where $\alpha \leq 2.2$, protecting 50% of medium-size nodes enhances the networks to virtually

never crashed; yet when $\alpha = 2.9$, the enhancement is still marginal. The basic observation is that when only a small fraction of all the nodes are protected, their degrees have to be high enough; otherwise the enhancement to network robustness would not be significant.

Another observation is that there exist larger fluctuations in simulation results from different realizations when α is of smaller values, which can also be explained by the larger ranges the degrees of the protected nodes have.

For the case with multiple hubs being protected, as discussed earlier, there exist strong fluctuations in the simulation results of crash threshold. For example, for the case where $\alpha = 2.5$, in our simulations the calculated crash thresholds vary from 38% to almost 100% in different realizations. However, simulation results on the cutoff nodal degree $\tilde{K}''$ turn out to be very stable and match well with analytical results, as presented in Table I. As we can observe, protecting four to five biggest hubs can already ensure the networks to be virtually never crashed by intentional attack: to crash the networks, the attack has to go all the way to remove even some degree-1 nodes.

(a) Single biggest hub protected

(b) 10% medium-sized nodes protected

(c) 50% medium-sized hubs protected

Fig. 3. 1: Crash threshold of random scale-free networks under intentional attacks

where different node protection schemes are adopted.

Table 3. 1: Analytical and simulation results of the cutoff nodal degrees in random scale-free networks with multiple high-degree nodes being protected. Denote the simulation results as SR and the theoretical results as TR.

$\alpha \backslash N_0$	1		2		3		4		5	
	SR	TR	SR	TR	SR	TR	SR	TR	SR	TR
2.1	6	6	4	5	3	3	2	2	1	1
2.2	5	5	4	4	2	2	1	1	1	1
2.3	5	5	3	3	2	2	1	1	1	1
2.4	5	5	3	3	2	2	1	1	1	1
2.5	4	5	3	3	2	2	1	1	1	1
2.6	4	4	3	3	2	1	1	1	1	1
2.7	4	4	3	3	2	2	1	1	1	1
2.8	4	4	3	3	2	2	1	1	1	1
2.9	4	4	3	3	2	2	2	2	1	1

Simulation results of cluster diameters of random scale-free networks under intentional attacks with different types of node protection are presented in Fig. 3.2. Due to space limitation, we plot only the results for the cases where $\alpha = 2.1$, 2.5 and 2.9, respectively. The basic observation is that all the three protection methods help reduce the cluster diameter under attack. Among them, the most effective one when $\alpha = 2.1$ is to protect 50% of medium-degree nodes. When $\alpha = 2.5$ or 2.9, the most effective methods is to protect a few highest-degree nodes.

(a) $\alpha = 2.1$

(b) $\alpha = 2.5$

(c) $\alpha = 2.9$

Fig. 3. 2: Cluster diameters of random scale-free networks under intentional attacks where different node protection schemes are adopted.

To summarize, for random scale-free networks with small exponents, either protecting hub nodes or medium-degree nodes significantly enhances network robustness, whereas for networks with large exponents, due to the relatively low degrees of medium-sized nodes, it is probably more effective to protect a small number of hub nodes.

3.5 Simulations on real-life network models

To evaluate the effects of the proposed node protection schemes in real-life communication networks, we carry out simulations on two different models as follows:

- A real-life Internet model on the Autonomous System (AS) level as measured by the Applied Network Research (NLNR) Project on January 2, 2000 [97], which contains 6470 inter-connected nodes and 12,566 links. We have verified that it is a scale-free network.
- A real-life Internet model on the router level as measured by the Cooperative Association for Internet Data Analysis (CAIDA), which contains 192,244 nodes and 609,066 links [98]. Its power-law nodal-degree distribution has also been verified.

Note that the real-life Internet model on the AS-level is an Internet graph resembling the relationship between Autonomous Systems (ASs), while the model on the router level resembles the relationship between routers. These two models are

of importance since (i) they are from real-life data and about important Internet topologies; and (ii) they are both of scale-free nodal-degree distributions yet with different statistical features. Specifically, AS-level model is a disassortative model where high-degree nodes tend to connect to low-degree nodes while the router-level model is an assortative model with the opposite feature.

Both network models have exponent values close to 2 in power-law curve approximations.

As that in most existing studies (e.g., [4, 47, 68]), we evaluate during the procedure of intentional attack (i) *the largest cluster size*, defined as the number of nodes in the biggest connected component versus the number of nodes in the original network; and (ii) the cluster diameter.

The earlier definition of crash threshold p_c where $\langle k^2 \rangle / \langle k \rangle \leq 2$ does not apply to these two networks since neither of them is a random network. To differentiate, we evaluate the *threshold of crash* (TOC), defined as the percentage of network nodes that has to be removed to reduce the largest cluster size to be no more than 5%. It is shown in Fig. 3.3 that without any protection, the two real-life models have their TOC values at 3.2% and about 14% respectively.

We still consider the three different protection schemes as follows:

- Only the biggest hub is protected;

- Several highest-degree nodes are protected. Specifically, we let the top 3 hubs be protected in the AS-level model and top 50 hubs be protected in router-level model respectively.
- The big hubs are all taken down, yet some medium-sized nodes are protected. Specifically, in the router-level model, we let the top 1% of biggest hubs be removed and after that, 10% and 50% of the next 1% highest-degree nodes are protected respectively. For the AS-level model, since it is crashed much faster than the router-level model under the intentional attack, we let the top 0.5% of biggest hubs be removed and after that, 10% and 50% of the next top 1% highest-degree nodes are protected respectively. The degrees of the protected medium-size nodes vary from 26 to 61 in the AS-level model and 38 to 55 in the router-level model, respectively.

For the cases with medium-degree node protections, the presented simulation results for the AS-level model come from average values of 100 independent realizations. For the router-level model, due to its extra-large size, it is prohibitively time-consuming to carry out extensive realizations. Therefore only 5 realizations have been conducted.

Fig. 3.3 illustrates the simulation results of the largest cluster size of these two real-life models under intentional attack with different node protection schemes. We observe that protecting the single biggest hub does not enhance network robustness

significantly in the router-level model. In the AS-level model, however, the TOC value jumps from a rather low 3.2% to a very high 60.3%. The TOC value can be further increased to 73.3% when 3 hub nodes are protected. The fragileness of the AS-level model under intentional attack and the significant enhancement by largest-hub protections can be explained by the same reason: the existence of big hubs in the relatively small network. The biggest hub in the AS-level modal has a degree of 1458. The top 3 hubs have combined degrees of 2899. Putting these three nodes under protection therefore strongly enhances network tolerance against intentional attack. In the router-level model, on the contrary, the biggest hub has a degree of only 1071, a small value compared to the network size. The top 50 hubs combined together have total degrees of only 23,969, still moderate considering the extra-large network size.

(a) AS-level Model

(b) Router-level Model

Fig. 3. 3: The largest cluster size of the AS- and router-level Internet models under intentional attack where different node protection schemes are adopted.

The effects of protecting medium-degree nodes turn out to be quite different. In the AS-level model, protecting medium-degree nodes are not as effective as protecting a few biggest hubs. In the router level, medium-degree node protection performs better. The TOC is increased from 14% to 25% when 50% of medium-degree nodes are protected. Such differences can again be explained by combined degrees of the protected node: the combined degrees of *all* the medium-sized nodes are 2,696 and 91,901 in the AS-level and router-level model respectively.

To summarize, the basic observation is that generally better performances can be achieved by putting a set of nodes with larger combined nodal degrees under protection. Considering the fact that it is difficult to identify all the medium-sized nodes in extra-large real-world networks, we argue that in certain circumstances it may be a more effective strategy to protect or hide some medium-size nodes rather than the obvious biggest hubs.

Finally, simulation results of the cluster diameter of these two models under intentional attack with different node protection schemes are presented in Fig. 3.4. We see that in most cases, the cluster diameters increase significantly before the networks are crashed, making the networks, though still largely connected, highly inefficient in supporting data communications. Such impacts can also be degraded when some high or medium-degree nodes are protected. Once again, the effects of largest-hub protections are more significant in AS-level model. In the router-level

model, it helps maintain a relatively short cluster diameter by protecting 50% of its medium-degree nodes.

(a) AS-level Model

(b) Router-level Model

Fig. 3. 4: Cluster diameters of the AS- and router-level Internet models under intentional attack where different node protection schemes are adopted.

3.6 Conclusions

In this chapter, we examined the effects of a few node protection schemes in enhancing network robustness against intentional attack. It is observed that though different schemes have different performances in different networks, overall speaking protecting a limited set of high or medium-degree nodes help significantly enhance network robustness. While protecting a few highest-degree hubs turns out to be effective in networks with big hubs or large exponent values, protecting a relatively larger number of medium-degree nodes may be more favorable in networks with moderate-size hubs. Such insights, as we believe, will be helpful for developing effective attacking/protecting strategies for complex communication systems.

4 On Imperfect Node Protection in Complex Communication Networks

4.1 Introduction

As we have pointed out in Chapter 3, in real-life complex systems especially communication systems, usually there exist protections over critical components against hostile attack. Due to the existence of such protections, a limited set of high or medium-degree nodes may be perfectly protected. Such effects have been evaluated in Chapter 3. In this chapter, we consider different cases where some protected nodes are partially paralyzed rather than totally crashed. We term such protection as imperfect node protection. Theoretical analysis and extensive simulation results show that such protection nevertheless helps significantly enhance network robustness.

The rest of this chapter is organized as follows. A few different scenarios of imperfect node protection are defined in Section 4.2. Theoretical analysis and simulation results on the effects of such protections in random scale-free networks are presented in Section 4.3 and Section 4.4, respectively. Numerical evaluations of the

proposed scenarios in a few real-life network models are presented in Section 4.5. Finally, Section 4.6 concludes the chapter.

4.2 Node protecting schemes

We consider a few different scenarios of imperfect node protection where the protected nodes are partially paralyzed rather than totally crashed under intentional attack.

- Scenario I: all the nodes in the network are imperfectly protected. When a node is attacked, each of the links connected to it is blocked or removed at a probability of P_f .
- Scenario II: only high-degree nodes in the network have protection. Specifically, we assume that there exists a threshold value T . Each node with its degree no less than T is imperfectly protected. When such a node is attacked, each of the links connected to it is blocked or removed at a probability of P_f . For any node with its degree lower than T , it is totally crashed once attacked.
- Scenario III: similar to the Scenario II, only that links between high-degree nodes are perfectly protected. Specifically, when nodes with degrees no less than T are attacked, they do not lose any links between them, meanwhile however the links between these nodes and those nodes

with degrees lower than T will be lost at a probability of P_f . For nodes with degrees lower than T , they are totally crashed once attacked.

None of these three scenarios may easily happen in real-life systems since (i) there is hardly any system with equal protection over all its nodes; and (ii) perfect protection probably does not exist at all. However, these scenarios nevertheless provide some useful benchmarks for estimating the effects of more realistic cases. Specifically, the three scenarios loosely resemble the real-life cases where (i) most nodes have some protection; (2) only important nodes (usually hubs) are heavily guarded; and (3) important nodes are protected and the links between them are also carefully guarded, respectively.

4.3 An analytical model

As we have presented previously, a random network loses its global connectivity when [47]

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} \leq 2, \quad (4.1)$$

where $\langle k \rangle$ denotes the average nodal degree and $\langle k^2 \rangle$ the average of the square of nodal degree.

For a random scale-free network, denote its lowest nodal degree as m and its cutoff nodal degree as K . To facilitate the discussions, we term a node that has ever been attacked when a network is crashed as an *attacked* node; and *un-attacked* node

otherwise. Below we analyze the three different scenarios in random scale-free networks separately.

Scenario I:

Assume that the network is crashed after a fraction p_c of all its nodes has been attacked. Hereafter we term the value of p_c as the *crash threshold*. Also assume that when a node is attacked, each link connected to it has a probability of P_f to get removed. In this chapter, we name P_f as the *link-removal probability*. Denote the lowest *original* degree of the attacked nodes as $\tilde{K}$. We have [68]

$$p_c = \sum_{\tilde{K}}^K p(k) = \frac{K^{1-\alpha} - \tilde{K}^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}}. \quad (4.2)$$

The fraction of links connected to all the attacked nodes, denoted as $\tilde{p}$, can be expressed as [68]

$$\tilde{p} = \sum_{\tilde{K}}^K \frac{kp(k)}{\langle k \rangle} = \frac{K^{2-\alpha} - \tilde{K}^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}}. \quad (4.3)$$

For a link connecting an attacked node and an un-attacked node in the original network, it has a probability of P_f to be removed; for a link connecting two attacked nodes in the original network, since both of its two end nodes are to be attacked, it has a probability of $[1 - (1 - P_f)^2]$ to be removed. Therefore, denoting the link loss probability of an attacked node and an un-attacked node as $\tilde{p}^+$ and $\tilde{p}^-$ respectively, we have

$$\begin{cases} \tilde{p}^+ = P_f(1 - \tilde{p}) + \tilde{p}[1 - (1 - P_f)^2] \\ \tilde{p}^- = P_f \cdot \tilde{p} \end{cases} \quad (4.4)$$

After attack, the network nodal-degree distribution is changed from $p(k)$ to $\tilde{p}(k)$ where

$$\tilde{p}(k) = \begin{cases} \sum_{k_0=\tilde{K}}^K p(k_0) \binom{k_0}{k} (1 - \tilde{p}^+)^k (\tilde{p}^+)^{k_0-k}, & k \geq \tilde{K} \\ \sum_{k_0=\tilde{K}}^K p(k_0) \binom{k_0}{k} (1 - \tilde{p}^+)^k (\tilde{p}^+)^{k_0-k} + \sum_{k_0=m}^{\tilde{K}} p(k_0) \binom{k_0}{k} (1 - \tilde{p}^-)^k \tilde{p}^{-k_0-k}, & k < \tilde{K} \end{cases} \quad (4.5)$$

For the attacked nodes, their average nodal degree in the original network, denoted as $\langle k \rangle^+$, is

$$\langle k \rangle^+ = \frac{\sum_{k_0=\tilde{K}}^K k p(k)}{p_c} = \frac{\langle k \rangle \cdot \tilde{p}}{p_c} \quad (4.6)$$

After attack, their average nodal degree is reduced to

$$\langle k' \rangle^+ = \frac{(1 - \tilde{p}^+) \cdot \langle k \rangle \cdot \tilde{p}}{p_c} \quad (4.7)$$

As to the average square of nodal degree of the attacked nodes, before attack, it is

$$\langle k^2 \rangle^+ = \frac{\sum_{k_0=\tilde{K}}^K k^2 p(k)}{p_c} = \frac{C}{p_c(3-\alpha)} (K^{3-\alpha} - \tilde{K}^{3-\alpha}) \quad (4.8)$$

And after attack, it becomes

$$\begin{aligned}\langle k'^2 \rangle^+ &= (1 - \tilde{p}^+)^2 \langle k^2 \rangle^+ + \langle k \rangle^+ \cdot \tilde{p}^+ \cdot (1 - \tilde{p}^+) \\ &= \frac{1 - \tilde{p}^+}{p_c} \left[\frac{(1 - \tilde{p}^+)C}{3 - \alpha} (K^{3-\alpha} - \tilde{K}^{3-\alpha}) + \tilde{p} \cdot \langle k \rangle \cdot \tilde{p}^+ \right].\end{aligned}\quad (4.9)$$

For the un-attacked nodes, before attack, their average nodal degree is

$$\langle k \rangle^- = \frac{\langle k \rangle (1 - \tilde{p})}{1 - p_c} \quad (4.10)$$

After attack, it becomes

$$\langle k' \rangle^- = \frac{(1 - \tilde{p}^-) \cdot \langle k \rangle \cdot (1 - \tilde{p})}{(1 - p_c)}. \quad (4.11)$$

Before attack, the average square of nodal degree of the un-attacked nodes is

$$\langle k^2 \rangle^- = \frac{\sum_{k=1}^{\tilde{K}} k^2 p(k)}{1 - p_c} = \frac{C}{(1 - p_c)(3 - \alpha)} (\tilde{K}^{3-\alpha} - m^{3-\alpha}) \quad (4.12)$$

After attack, it becomes

$$\begin{aligned}\langle k'^2 \rangle^- &= (1 - \tilde{p}^-)^2 \langle k^2 \rangle^- + \langle k \rangle^- \cdot \tilde{p}^- \cdot (1 - \tilde{p}^-) \\ &= \frac{1 - \tilde{p}^-}{1 - p_c} \left[\frac{(1 - \tilde{p}^-)C}{3 - \alpha} (\tilde{K}^{3-\alpha} - m^{3-\alpha}) + \langle k \rangle (1 - \tilde{p}) \cdot \tilde{p}^- \right].\end{aligned}\quad (4.13)$$

From Eqs. (4.6)-(4.13), we have that after attack the value of $\tilde{\kappa}$ is

$$\tilde{\kappa} = \frac{\langle k'^2 \rangle}{\langle k' \rangle} = \frac{p_c \langle k'^2 \rangle^+ + (1 - p_c) \langle k'^2 \rangle^-}{p_c \langle k' \rangle^+ + (1 - p_c) \langle k' \rangle^-} \quad (4.14)$$

From Eqs. (4.1)-(4.4), (4.7), (4.9), (4.11), (4.13) and (4.14), the crash threshold p_c can be obtained.

Consider the extreme case where all the nodes in the network are attacked. We have

$$\begin{aligned}\tilde{\kappa} &= \frac{(1-P_f)^4 \langle k^2 \rangle + [1 - (1-P_f)^2](1-P_f)^2 \langle k \rangle}{(1-P_f)^2 \langle k \rangle} \\ &= (1-P_f)^2 \kappa + 1 - (1-P_f)^2.\end{aligned}\quad (4.15)$$

It shows that when $P_f < 1 - \sqrt{\frac{1}{\kappa-1}}$, or in other words if each node can protect at least a fraction of $\sqrt{\frac{1}{\kappa-1}}$ of all the links connected to it, in the Scenario I a random scale-free network will never be totally crashed by the intentional attack.

Scenario II:

In such scenario, only nodes with degrees no less than T have protection. Denote the crash threshold under such case as p'_c and the lowest original degree of the attacked nodes as $\tilde{K}'$. When $T \leq \tilde{K}'$, the analysis remains the same as that for Scenario I. Hereafter we only consider the case where $T > \tilde{K}'$. Denote the fraction of links connected to the attacked nodes as $\tilde{p}'$. We have

$$\begin{cases} p'_c = \sum_{\tilde{K}'}^K kp(k) = \frac{K^{1-\alpha} - \tilde{K}'^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}} \\ \tilde{p}' = \sum_{\tilde{K}'}^K \frac{kp(k)}{\langle k \rangle} = \frac{K^{2-\alpha} - \tilde{K}'^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}} \end{cases} \quad (4.16)$$

Denote the fraction of nodes with degrees no less T as p_t , and the fraction of links connected to these nodes as $\tilde{p}_t$. They can be expressed as

$$\begin{cases} p_t = \sum_T^K kp(k) = \frac{K^{1-\alpha} - T^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}} \\ \tilde{p}_t = \sum_T^K \frac{kp(k)}{\langle k \rangle} = \frac{K^{2-\alpha} - T^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}} \end{cases} \quad (4.17)$$

Now we calculate link loss probabilities of different nodes with different degrees.

For those nodes with degrees between $\tilde{K}'$ and T , since they have no protection, they will be totally crashed. Their link loss probability is therefore 100%. Denote the link loss probability of the nodes with degrees no less than T as $\tilde{p}'^+$, and the link loss probability of un-attached nodes as $\tilde{p}'^-$. It can be derived that

$$\begin{cases} \tilde{p}'^+ = P_f(1 - \tilde{p}') + [1 - (1 - P_f)^2] \cdot \tilde{p}_t + (\tilde{p}' - \tilde{p}_t) \\ \tilde{p}'^- = \tilde{p}' - (1 - P_f)\tilde{p}_t \end{cases} \quad (4.18)$$

Note that the first equation of (4.18) contains three different parts, calculating the link-loss probabilities between high-degree (no less than T) and low-degree (lower than $\tilde{K}'$) nodes, between high-degree and high-degree nodes, and connected to the in-between nodes, respectively.

After attack, the network degree distribution is changed to

$$\tilde{p}(k) = \begin{cases} \sum_{k_0=T}^K p(k_0) \binom{k_0}{k} (1 - \tilde{p}'^+)^k \tilde{p}'^{+k_0-k}, & k \geq \tilde{K}' \\ \sum_{k_0=T}^K p(k_0) \binom{k_0}{k} (1 - \tilde{p}'^+)^k \tilde{p}'^{+k_0-k} + \sum_{k_0=m}^{\tilde{K}'} p(k_0) \binom{k_0}{k} (1 - \tilde{p}'^-)^k \tilde{p}'^{-k_0-k}, & k < \tilde{K}' \end{cases} \quad (4.19)$$

For the nodes with original degrees no less than T , after attack, their average nodal degree, denoted as $\langle k'' \rangle^+$, becomes

$$\langle k'' \rangle^+ = \frac{(1 - \tilde{p}'^+) \cdot \langle k \rangle \cdot \tilde{p}_t}{p_t}. \quad (4.20)$$

And their average square of nodal degree is

$$\langle k''^2 \rangle^+ = \frac{1 - \tilde{p}'^+}{p_t} \left[\frac{(1 - \tilde{p}'^+)C}{3 - \alpha} (K^{3-\alpha} - T^{3-\alpha}) + \tilde{p}_t \cdot \langle k \rangle \cdot \tilde{p}'^+ \right]. \quad (4.21)$$

For the un-attacked nodes, after attack, their average nodal degree, denoted as $\langle k'' \rangle^-$, is

$$\langle k'' \rangle^- = \frac{(1 - \tilde{p}'^-) \cdot \langle k \rangle \cdot (1 - \tilde{p}')}{(1 - p'_c)}; \quad (4.22)$$

and their average square of nodal degree is

$$\langle k''^2 \rangle^- = \frac{1 - \tilde{p}'^-}{1 - p'_c} \left[\frac{(1 - \tilde{p}'^-)C}{3 - \alpha} (\tilde{K}^{3-\alpha} - m^{3-\alpha}) + \langle k \rangle (1 - \tilde{p}') \cdot \tilde{p}'^- \right]. \quad (4.23)$$

Hence,

$$\tilde{\kappa} = \frac{\langle k''^2 \rangle}{\langle k'' \rangle} = \frac{p_t \langle k''^2 \rangle^+ + (1 - p'_c) \langle k''^2 \rangle^-}{p_t \langle k'' \rangle^+ + (1 - p'_c) \langle k'' \rangle^-}. \quad (4.24)$$

From Eqs. (4.16)-(4.18), (4.20)-(4.24), the crash threshold p'_c can be numerically solved.

For each given value of T , there exists a threshold value of the link-removal probability that when P_f is lower than this threshold, the network will never be crashed. Unlike the case of Scenario I, however, there is no easy closed-form solution

for this threshold. Instead, it can be numerically solved from Eqs. (4.16)-(4.18), (4.20)-(4.24).

Scenario III:

Similar to that in Scenario II, only those nodes with degrees no less than T have protection. The difference is that the links between these high-degree nodes will never be removed. Denote the crash threshold for such case as p_c'' , and the lowest degree of the attacked nodes as $\tilde{K}''$. Similar to that for Scenario II, we only need to consider the case where $T > \tilde{K}''$.

Denote the fraction of the links connected to the attacked nodes as $\tilde{p}''$. The derivations of all the equations keep very similar to those for Scenario II. The only difference lies in the calculations of the link loss probability of those nodes with degrees no less than T , Denoting such probability as $\tilde{p}''^+$, we have

$$\tilde{p}''^+ = P_f(1 - \tilde{p}'') + \tilde{p}'' - \tilde{p}_t. \quad (4.25)$$

Replace $\tilde{p}''^+$ with $\tilde{p}''^+$ in Eqs. (4.20) and (4.21), p_c'' can be numerically solved.

Again, there exists a threshold value of the link-removal probability below which the network is never crashed. The threshold value can be numerically solved, like that for Scenario II.

Same as that in Chapter 3, the cluster diameter, as another important parameter for measuring network robustness, will be evaluated by numerical simulations.

4.4 Numerical results and discussions

In our simulations, random scale-free network are generated by adopting the algorithm proposed in [58], with exponent ranging from 2 to 3. The network size N is 10,000. The minimal degree is 1 and the cutoff degree is 100, which equals to $\sqrt{N}$. Unless otherwise specified, the simulation results come from 50 independent realizations.

All the three scenarios have been simulated. We adopt the definition in [47] that a network is crashed once $\langle k^2 \rangle / \langle k \rangle \leq 2$ is achieved.

(a) Scenario I

(b) Scenario II

(c) Scenario III

Fig. 4. 1: The thresholds of link-removal probability to ensure that the networks never be totally crashed by intentional attack in random scale-free networks. For Scenarios II and III, the top 0.5% highest-degree nodes are protected.

As discussed in Section 4.3, for all the three Scenarios there exists a threshold value of P_f , below which the intentional attack can never completely crash a random scale-free network. Fig. 4.1 shows this threshold for networks with different exponents. The simulation results are calculated by the trial-and-error method with a step length of 1%. As we can see, subject to the same minimum and cutoff nodal degrees, networks with smaller exponents have higher values of κ and consequently higher thresholds of link-removal probability. This conclusion applies to all the three scenarios.

The influences of link-removal probability on network robustness are further evaluated for all the three different scenarios. The results for the case where $\alpha = 2.5$ are plotted in Fig. 4.2. For Scenarios II and III, we let the top 0.5% highest-degree nodes be protected. An interesting observation is that there exists a phase transition in the network crash threshold: once the link-removal probability is low enough, the threshold quickly jumps up from a low value to 1. This can be explained: for both Scenarios II and III, similar to that of Scenario I, there exists a threshold of the link-removal probability below which the network will never be crashed. The network remains to be fragile under intentional attack until the link-removal probability gets close to the threshold.

(a) Scenario I

(b) Scenario II

(c) Scenario III

Fig. 4. 2: The effects of link-removal probability on network crash threshold in the random scale-free network. We let $\alpha=2.5$. For Scenarios II and III, the top 0.5% highest-degree nodes are protected.

Fig. 4.3 makes the comparisons between analytical and simulation results of crash thresholds in random scale-free networks. The link-removal probability is set as 80% for all the three scenarios. For Scenarios II and III, we still let the top 0.5% hub nodes be protected. It can be observed that the analytical results match well with the simulation results. Also we can see that even by protecting only 20% of the links connected to an attacked node, the network crash threshold can be significantly increased. Note that since we set the same minimum and cutoff nodal degrees for all the different networks, networks with smaller exponents have higher average nodal degrees and consequently higher crash thresholds. A higher average nodal degree also tends to make imperfect protection more effective in enhancing network robustness.

The comparisons between the effects of the three scenarios on crash thresholds in random scale-free networks are presented in Fig. 4.4. We adopt the same parameters of link-removal probability and node protection percentage as those in Fig. 4.3. Since simulation results match well with theoretical analysis results, to avoid making the figure too crowded, only the theoretical analysis are plotted. We see that Scenario I slightly outperforms the other two scenarios when the exponent is of a high value. When the exponent value is low, however, Scenario III leads to a much larger crash threshold. In fact, as we can easily observe, the networks will never be crashed when $\alpha \leq 2.4$, which can be explained as follows: in Scenario III, the sub-network composed of a small number of hub nodes and the connections between them are strictly protected. The survival of this sub-network helps keep $\langle k^2 \rangle / \langle k \rangle > 2$ in networks under attack. In networks with smaller threshold values, the numbers of

moderate-degree nodes are relatively larger. The protected sub-network together with these moderate-degree nodes make a network extremely difficult to be totally crashed.

(a) Scenario I

(b) Scenario II

(c) Scenario III

Fig. 4. 3: Comparisons between theoretical and simulation results of crash thresholds in random scale-free networks. The link-removal probability is 80% for all the three scenarios. For Scenarios II and III, the top 0.5% hub nodes are protected.

Fig. 4. 4: The crash thresholds of random scale-free networks in different scenarios.

The link removal probability is 80% for all the three scenarios. For Scenarios II and III, the top 0.5% hub nodes are protected.

Fig. 4.5 shows in Scenarios II and III the improvements to network robustness when more hub nodes are protected. The simulation results are for the case where $\alpha = 2.5$ and the link-removal probability is 80%. We see that protecting a larger number of hub nodes only gradually increases network crash threshold in Scenario II, whereas in Scenario III since the links between the protected hubs are also protected, the threshold value boosts up with the number of protected hubs. In other words, protecting more hub nodes may not be effective unless the links between these hubs are also protected.

(a) Scenario II

(b) Scenario III

Fig. 4. 5: The relationship between the crash threshold and the portion of hub nodes being protected in random scale-free network. Let $\alpha=2.5$ and the link-removal probability be 80%.

Figs. 4.4 and 4.5 may leave the impression that Scenario III is clearly the winner over the other two in enhancing network robustness. While this may be true in measuring the crash threshold, especially in random scale-free networks with small exponents, Fig. 4.6 shows that the situation is actually not that simple. In this figure, we plot *the* largest cluster size, defined as the number of nodes in the biggest connected component versus the number of nodes in the original network [4], during the procedure of attack. Due to space limitation, we show only the results where $\alpha = 2.1, 2.5$ and 2.9 , respectively. The link-removal probability is 80% and top 0.5% hub nodes are protected in Scenarios II and III. It can be observed that though Scenario III pushes the crash threshold to be very high where $\alpha = 2.1$ and 2.5 , most of time during the procedure of intentional attack, it does not make the largest connected component be much larger than that in Scenario I. The small set of protected hubs makes the network to stay with $\langle k^2 \rangle / \langle k \rangle > 2$ even when the connected component size is already quite small.

Finally we show in Fig. 4.7 the cluster diameters in all the three different scenarios. All the parameters remain the same as those in Fig. 4.6. The basic observation is that all the three scenarios help reduce the cluster diameter under attack. Among them, the most effective one is Scenario III. The existence of the links between hubs not only keeps network connected but also effectively lowers the cluster diameter in the largest connected component. Scenario III may indeed become the favorable option when keeping a small cluster diameter is of critical importance.

(a) $\alpha=2.1$

(b) $\alpha=2.5$

(c) $\alpha=2.9$

Fig. 4. 6: The largest cluster sizes in random scale-free networks during the procedure of attack. The link removal probability is 80% for all the three scenarios. For Scenarios II and III, the top 0.5% hub nodes are protected.

(a) $\alpha=2.1$

(b) $\alpha=2.5$

(c) $\alpha=2.9$

Fig. 4. 7: The cluster diameters of random scale-free networks under attack. The link removal probability is 80% for all the three scenarios. For Scenarios II and III, the top 0.5% hub nodes are protected.

4.5 Simulations on real-life network models

To evaluate the effects of the proposed node protection schemes in communication networks, we carry out simulations on the same two network models which we have used in Chapter 3:

- A real-life Internet model on the AS-level as measured by the Applied Network Research (NLNR) Project on January 2, 2000 [97], which contains 6470 inter-connected nodes and 12,566 links.
- A real-life Internet model on the router level as measured by the Cooperative Association for Internet Data Analysis (CAIDA), which contains 192,244 nodes and 609,066 links [98].

As that in most existing studies (e.g., [4, 47, 68]), we evaluate during the procedure of intentional attack the largest cluster size and the cluster diameter. Same as that in Chapter 3, since $\langle k^2 \rangle / \langle k \rangle \leq 2$ does not apply to non-random networks, we calculate the threshold of crash (TOC) defined as the percentage of network nodes that has to be removed to reduce the largest cluster size to be no more than 5%. It is shown in Fig. 4.8 and Fig. 4.9 that without any protection, the two real-life models have their TOC values at 3.2% and about 14% respectively.

We still consider all the three different scenarios. For each scenario, we consider different cases where the link removal probability is 50%, 70% and 80%, respectively. For Scenarios II and III, we still let the top 0.5% of hub nodes be protected. For all

the cases, the presented simulation results for the AS-level model come from average values of 100 independent realizations. For the router-level model, due to its extra-large size, it is prohibitively time-consuming to carry out extensive realizations. Therefore only 5 realizations have been conducted.

The largest cluster sizes of the AS-level model under attack with different link removal probabilities are shown in Fig. 4.8. The three subfigures are for the three different cases where link removal probabilities are 50%, 70% and 80%, respectively. We observe that Scenario I steadily leads to the best performance among the three. In fact, since the threshold value of link removal probability to ensure that the network never be crashed in Scenario I is 78%, the network never gets crashed when the link removal probabilities are 50% and 70% respectively. Even when the link removal probability is 80%, TOC remains to be high at 60.3%.

Scenarios III performs only slightly better than Scenario II. It increases TOC from 75.1% to 77.9%, from 58.4% to 64%, and from 46.4% to 55.6% in the three different cases with different link removal probabilities, respectively. As we have observed in the last section, the extra protection over the links between a small number of hubs does not significantly increase the largest cluster size though it does decrease the cluster diameter. The TOC value therefore may not be significantly increased by such extra protection unless the number of the protected hubs is large enough to increase the largest cluster size to be higher than the threshold value (5% in our case) with the existence of the links between them.

(a) Link-removal probability = 50%

(b) Link-removal probability = 70%

(c) Link-removal probability = 80%

Fig. 4. 8: The largest cluster sizes of the AS-level model under attack with different link removal probabilities. For Scenarios II and III, the top 0.5% hub nodes are protected.

Fig. 4.9 shows the largest cluster sizes of the router-level model under attack with different link removal probabilities where all the parameter values remain the same as those in Fig. 4.8. Once again we can easily observe that Scenario I leads to the best performance. The threshold value of the link removal probability to crash the

network in Scenario I is 75%. When the link-removal probability is 80%, the TOC value of Scenario I is 48.9%.

Similar to that in the AS-level model, Scenario III only slightly outperforms Scenario II. The TOC values under different link removal probabilities are 28.1% vs. 28.1%, 22.9% vs. 25%, and 19.8% vs. 22.9%, respectively. The most visible difference between Fig. 8 and Fig. 9 is probably that Scenarios II and III become much less effective in enhancing network robustness in the router-level model. This difference mainly comes from the degrees of their hub nodes: The biggest hub in the AS-level model has a degree of 1458. The top 0.5% hubs combined together have a total degree of 6937. Putting these nodes under protection therefore strongly enhances network robustness even when such protection is imperfect. In the router-level model, on the contrary, the biggest hub has a degree of only 1071, a small value compared to the network size. The top 0.5% hubs have total degrees of only 129,681, still moderate considering the extra-large network size. The protection effects therefore become less significant.

Finally, simulation results of the cluster diameters of the two real-life models under attack are presented in Fig. 4.10. We present only the results for the case with a link removal probability of 80% while the conclusion holds for all the other cases we have simulated: all the three different scenarios, especially Scenario III, help to significantly reduce the cluster diameter. Protecting the links between a small number of hubs effectively help shorten the paths between different node pairs.

(a) Link-removal probability = 50%

(b) Link-removal probability = 70%

(c) Link-removal probability = 80%

Fig. 4. 9: The largest cluster sizes of the router-level model under attack with different link removal probabilities. For Scenarios II and III, the top 0.5% hub nodes are protected.

(a) AS-level

(b) Router-level

Fig. 4. 10: The cluster diameters of the AS-level and router-level Internet models under attack. The link removal probability is 80%. The top 0.5% nodes are protected in Scenarios II and III.

Overall, we have the conclusion that protecting hubs is an effective strategy in networks with high-degree hubs, e.g., the earlier random network models and the AS-level Internet model. For networks with moderate hub node sizes, putting more nodes under protection may be necessary. For both cases, protecting the links between hubs helps to reduce the cluster diameter.

4.6 Conclusions

In this chapter, we evaluated the network robustness against intentional attack where some or all of the network nodes are with some imperfect protection. It is found that even a rather weak protection help significantly enhance the chance that the network survives the most hostile attack. While protecting a small number of big hubs or a large number of moderate-size nodes perform differently in different networks, protecting the links between hubs may be worth the efforts if maintaining a short cluster diameter is of priority. Such insights shall be helpful for the future developments of more efficient network attack and protection schemes.

5 Tolerance of Local Information-Based Intentional Attacks in Complex Networks

5.1 Introduction

In the previous two chapters, we evaluated the effects of having protection mechanisms in complex networks. In this chapter, we consider another effect which has been largely ignored in existing studies: in many real-life systems including communication networks, hostile attacks typically have to “propagate” throughout the network, taking down those nodes adjacent to the already compromised nodes. Since such attacks typically use at most the local network-topology information, we term them *distributed attacks*. Theoretical analysis and numerical simulations show that though different distributed attacks perform differently in different networks, overall speaking they can be highly efficient. In some cases, they are almost as damaging as the global information-based intentional attack. Effective methods therefore have to be developed to protect systems against such attacks.

The rest of this chapter is organized as follows. In Section 5.2, A few distributed attack schemes are proposed. Theoretical analysis and simulation results

for evaluating these schemes on random scale-free networks are proposed in Sections 5.3 and 5.4, respectively. Numerical evaluations on a few real-life network models are presented in Section 5.5. Finally Section 5.6 concludes the chapter.

5.2 Distributed attack schemes

We define a node that has already been taken down as a *crashed* node; and otherwise a *live* node. As mentioned earlier, distributed attacks basically target on some or all of the live nodes adjacent to the crashed nodes in each step; and the selections of the targets depend on only the local network-topology information.

We study a few different distributed attack schemes as follows:

- **Greedy sequential attack:** the largest-degree node adjacent to the node crashed in the *last* step is selected as the next-step target. If no adjacent node exists, a live node is randomly selected as the target.
- **Coordinated attack:** the largest-degree live node adjacent to any crashed node is selected as the next-step target. If no adjacent node exists at all, a live node is randomly selected as the target.
- **Lower-bounded parallel attack:** instead of taking down only a single node in each step, the attack starts from *every* crashed node to randomly take down one of its adjacent live nodes as long as the live node's

degree is no less than a certain threshold. If no such adjacent node exists, the attack stops.

- **Lower-bounded flooding attack:** the attack starts from every crashed node to take down *all* the adjacent live nodes as long as the live node's degree is no less than a certain threshold. If no such adjacent node exists, the attack stops.

The studied four attack schemes may not easily happen in real-life systems. However, they provide some useful benchmarks for estimating the performance of other cases. Specifically, the first two schemes are for evaluating distributed attacks which take down only one node in each step. While the first one is greedy yet memoryless (meaning that the attack always start from the last node it has crashed), the second one with more sophisticated coordination is probably among the worst of this class of attacks. The third and fourth schemes, on the other hand, take down multiple nodes in each step: the third one takes down at most one node starting from each crashed node, while the fourth one adopts brutal force to crash more. Note that the cases without a lower bound of nodal degree can be viewed as the special case where the lower bound value equals to the lowest nodal-degree of the network.

The effects of these distributed attack schemes in the random scale-free networks will be analyzed in the next section.

5.3 Theoretical analysis

As has been mentioned a few times in previous chapters, a random network loses its global connectivity when [47]

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} < 2, \quad (5.1)$$

where $\langle k \rangle$ denotes the average nodal degree, and $\langle k^2 \rangle$ the average of nodal-degree squares; and the crash threshold is [47]:

$$1 - p_c = \frac{1}{\kappa_0 - 1}. \quad (5.2)$$

Here $\kappa_0 = \langle k_0^2 \rangle / \langle k_0 \rangle$ where $\langle k_0 \rangle$ and $\langle k_0^2 \rangle$ denote the average nodal degree and average nodal-degree square in the original network respectively.

For analyzing the robustness of random networks under intentional attack, an approximate solution was proposed in [68], as has been reviewed in Section 2.3.3. For easy reference and convenient understanding of our analysis, we briefly introduce it again.

Specifically, denote the original cutoff in the network as K and the new cutoff after the intention attack as $\tilde{K}$, the fraction of removed nodes as p , and the corresponding link loss probability as $\tilde{p}$. We have

$$\begin{cases} p = \sum_{\tilde{K}}^K p(k) = \frac{K^{1-\alpha} - \tilde{K}^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}} \\ \tilde{p} = \sum_{\tilde{K}}^K \frac{kp(k)}{\langle k \rangle} = \frac{K^{2-\alpha} - \tilde{K}^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}} \end{cases} \quad (5.3)$$

For such a case, κ becomes

$$\kappa = \frac{\langle k^2 \rangle}{\langle k \rangle} = \left(\frac{2-\alpha}{3-\alpha} \right) \frac{\tilde{K}^{3-\alpha} - m^{3-\alpha}}{\tilde{K}^{2-\alpha} - m^{2-\alpha}}, \quad (5.4)$$

where m denotes the minimum nodal degree in the network. The new cutoff $\tilde{K}$ can be calculated and the corresponding p , which is the crash threshold, can then be obtained by numerically solving the equation

$$1 - \tilde{p} = \frac{1}{\kappa - 1}. \quad (5.5)$$

We now proceed to analyze the four distributed attacks separately.

Greedy Sequential Attack

In a random scale-free network, high-degree nodes are generally connected into a dense sub-network. Starting from any node in the network, a greedy sequential attack can reach a highest-degree node in at most a few steps, and starting from there reaches most or all the other hub nodes. When virtually all the hubs have been crashed, the efficiency of each step of the attack drops drastically to be approximately equivalent to that of a random failure. Therefore the greedy sequential attack can be approximately viewed as composed of two different stages:

i) an intentional attack that takes down all the hubs; and ii) a random failure of nodes in the remaining network.

Assume that in the first stage, the hub nodes with degrees no less than a certain threshold value K_T are all removed. Denote the corresponding fraction of crashed nodes and link loss probability in this stage as p_i and $\tilde{p}_i$, respectively. We have

$$\begin{cases} p_i = \sum_{K_T}^K p(k) = \frac{K^{1-\alpha} - K_T^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}} \\ \tilde{p}_i = \sum_{K_T}^K \frac{kp(k)}{\langle k \rangle} = \frac{K^{2-\alpha} - K_T^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}} \end{cases} \quad (5.6)$$

The ratio of average nodal degree versus the average square of nodal degree after the first stage therefore becomes

$$\kappa' = \left(\frac{2-\alpha}{3-\alpha} \right) \frac{K_T^{3-\alpha} - m^{3-\alpha}}{K_T^{2-\alpha} - m^{2-\alpha}}, \quad (5.7)$$

Assume that in the second stage a fraction p_r of nodes is crashed before the network loses its global connectivity. The nodal degree distribution after the second-stage attack is

$$\tilde{p}(k) = \sum_{k_0=k}^{K_T} p(k_0) \binom{k_0}{k} (1 - (p_r + \tilde{p}_i))^k (p_r + \tilde{p}_i)^{k_0-k}. \quad (5.8)$$

In random networks, from Eq. (5.2) we have

$$\begin{cases} 1 - (p_r + \tilde{p}_i) = \frac{1}{\kappa' - 1} \\ p_r = \frac{\kappa' - 2}{\kappa' - 1} - \tilde{p}_i \end{cases} \quad (5.9)$$

The crash threshold of the original network is $p_r + p_i$.

Achieving a good estimation of the threshold value K_T may not be easy in some networks. It shall be the value where the remaining network resembles a random network. In our experience it is a good practice to set K_T as $2 \times \langle k_0 \rangle$ in random scale-free networks with exponent values between 2 to 3.

Coordinated Attack

Coordinated attack can be almost as efficient as the global information-based intentional attack. Specifically, assume that the network is crashed by the intentional attack when the cutoff degree is reduced to K_c . If the network nodes with degrees no less than K_c (i.e., the set of nodes to be removed by the intentional attack in order to crash a network) form into a single connected component, once the coordinated attack reaches any node of this component, it thereafter becomes equivalent to the intentional attack; otherwise, the efficiency of the coordinated attack is slightly lower. By figuring out the number of disconnected components in the sub-network composed by those nodes with degrees no less than K_c , we can get a good estimation of the *difference* between the efficiencies of coordinated attack and intentional attack.

Direct calculation of the number of disconnected components in a network however remains as a challenging problem. In this chapter, we calculate a *lower bound* of this number instead. Specifically, as we can see, the subnet composed by those nodes with degrees no less than K_c in the original network still form into a

random network, though the nodal-degree distribution is no longer scale-free. In this subnet, nodes with degree 0 are apparently isolated. Nodes with degree 1, if it is connected to another node also with degree 1, these two nodes form into an isolated component. The sum of the numbers of these two types of isolated components safely serves as a lower bound of the number of disconnected components in the subnet. Below we derive the equations for calculating this lower bound.

Denote the fraction of nodes with degrees no less than K_c as p_c and the corresponding portion of links connected to these nodes as $\tilde{p}$. Both of these two parameters can be calculated using Eq. (5.3), with $\tilde{K}$ being replaced by K_c .

For a node with degree k ($k \geq K_c$) in the original network, the probability that it has a degree 0 in the subnet equals to the probability that all of its links connected to the nodes with degrees lower than K_c in the original network, which is $(1 - \tilde{p})^k$; the probability that it has a degree 1 in the sub-network is the probability that it has exactly one link connected to another node with degree no less than K_c , which equals to $k \cdot \tilde{p} \cdot (1 - \tilde{p})^{k-1}$. Therefore the fraction of nodes in the sub-network with degree 0 and degree 1 can be expressed as:

$$\left\{ \begin{array}{l} p_0 = \frac{\sum_{k \geq K_c} p(k)(1 - \tilde{p})^k}{\sum_{k \geq K_c} p(k)} \\ p_1 = \frac{\sum_{k \geq K_c} p(k) \cdot k \cdot \tilde{p} \cdot (1 - \tilde{p})^{k-1}}{\sum_{k \geq K_c} p(k)} \end{array} \right. \quad (5.10)$$

The probability that a degree-1 node in the subnet is connected to another degree-1 node, denoted as q_1 , is $p_1/\langle k' \rangle$, where $\langle k' \rangle$ is the average degree in the subnet, i.e., $\langle k' \rangle = \frac{\langle k_0 \rangle \cdot \tilde{p} \cdot \tilde{p}}{p_c}$. Hence

$$q_1 = \frac{p_c \cdot p_1}{\langle k_0 \rangle \cdot \tilde{p} \cdot \tilde{p}}. \quad (5.11)$$

The lower bound of the difference between the crash thresholds of intentional attack and coordinated attack is $p_0 + \frac{p_1 \cdot q_1}{2}$. Though the lower bound may not appear to be very tight, as we can see later in Sec. 5.4, it is actually good enough for most cases. The lower bound of the crash threshold of the coordinated attack is

$$p_{ca} = p_c \left(1 + p_0 + \frac{p_1 \cdot q_1}{2} \right). \quad (5.12)$$

Lower-Bounded Parallel Attack

For the cases with multiple nodes being crashed in each step, since generally a scale-free network loses its global connectivity in just a few steps and then virtually all its nodes, instead of analyzing the crash threshold, we study the steps needed to crash all the nodes in the networks. Arguably such analysis may be more meaningful for estimating the network robustness and the time for reaction while under attack.

In the lower-bounded parallel attack, starting from each crashed node the largest-degree adjacent live node is crashed if that node's degree is no less than a preset threshold. Denote the threshold value as T . In an N -node network, denote the total number of nodes with original degrees no less than T as N' ,

$$N' = N \cdot \sum_{k=T}^K p(k) = N \cdot \frac{K^{1-\alpha} - T^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}}. \quad (5.13)$$

The probability that a crashed node has an original degree k is

$$p'(k) = \frac{p(k)}{\sum_{k \geq T} p(k)} \quad (k \geq T). \quad (5.14)$$

Denote the probability that a link leading to a node with degree no less than T as p' . We have

$$p' = \sum_{k=T}^K \frac{kp(k)}{\langle k \rangle} = \frac{K^{2-\alpha} - T^{2-\alpha}}{K^{2-\alpha} - m^{2-\alpha}}. \quad (5.15)$$

Assume a just-crashed node with degree k ($k \geq T$) has an average of n'_k live adjacent nodes with degree no less than T , We have

$$n'_k = k \times p' - 1 \quad (k \geq T). \quad (5.16)$$

Denote the number of crashed nodes after l steps as a_l . We derive the relationship between a_l and a_{l-1} as follows. Let $a_1 = 1$. If in a certain step l , starting from every crashed node a live adjacent node is further crashed, then $a_l = 2a_{l-1}$. Now we take into account the probability that all nodes adjacent to a crashed node and with degrees no less than T have been crashed. While accurate analysis of this probability is rather complicated, an approximate model can be easily derived: for a crashed node with degree k , averagely all its adjacent live nodes with degrees no less than T will be crashed in the next n'_T steps. In other words, if a degree- k node was

crashed in step $l - n'_k$, averagely it has no live adjacent node to be crashed in step l . Since the probability that a degree- k node is crashed in step $l - n'_k$ equals to $p'(k) \cdot a_{l-1-n'_k}$ and the lowest degree of the nodes to be crashed is T , we have

$$a_l = \begin{cases} 1 & l = 1 \\ 2a_{l-1} & 2 \leq l \leq n'_T \\ 2a_{l-1} - \sum_{k \geq T} p'(k) \times a_{l-1-n'_k} & n'_T < l \end{cases} \quad (5.17)$$

From Eqs. (5.13) - (5.17), we can calculate the number of steps needed to remove all the nodes with degrees no less than T . Note that Eq. (5.17) is an approximate calculation since it utilizes average value rather than detailed nodal-degree distribution. As we will see later in Section 5.4, however, the accuracy turns out to be satisfactory.

Lower-Bounded Flooding Attack

In the lower-bounded flooding attack, starting from a crashed node all its live adjacent nodes will be crashed as long as their degrees are no less than a certain threshold value T . Denote the average degree of those nodes with degrees no less than T as K''

$$K'' = \frac{\sum_{k=T}^K kp(k)}{\sum_{k=T}^K p(k)} = \langle k \rangle \times \frac{(K^{2-\alpha} - T^{2-\alpha})(K^{1-\alpha} - m^{1-\alpha})}{(K^{2-\alpha} - m^{2-\alpha})(K^{1-\alpha} - T^{1-\alpha})}. \quad (5.18)$$

Starting from each crashed node an average of n'' nodes can be crashed in the next step,

$$n'' = K'' \cdot p' - 1 = \langle k \rangle \cdot \frac{(K^{2-\alpha} - T^{2-\alpha})^2 (K^{1-\alpha} - m^{1-\alpha})}{(K^{2-\alpha} - m^{2-\alpha})^2 (K^{1-\alpha} - T^{1-\alpha})} - 1, \quad (5.19)$$

where p' denotes the probability that a link leads to a node with a degree no less than T , which can be calculated from Eq. (5.15).

Assumed that in the first step, only a single node is crashed, and after l'' steps, all the N' nodes as defined in Eq. (5.13) are taken down. Approximately we have

$$(n'')^{l''-1} \approx N'. \quad (5.20)$$

Hence,

$$l'' = \log_{n''}^{N'} + 1, \quad (5.21)$$

which further yields

$$l'' = \frac{\log \left(N \times \frac{K^{1-\alpha} - T^{1-\alpha}}{K^{1-\alpha} - m^{1-\alpha}} \right)}{\log \left[\langle k \rangle \times \frac{(K^{2-\alpha} - T^{2-\alpha})^2 (K^{1-\alpha} - m^{1-\alpha})}{(K^{2-\alpha} - m^{2-\alpha})^2 (K^{1-\alpha} - T^{1-\alpha})} - 1 \right]} + 1. \quad (5.22)$$

5.4 Simulation results and discussion

In our simulations, random scale-free networks are generated by adopting the algorithm proposed in [58], with α varying from 2 to 3. Unless otherwise specified, the network size N is 10,000; the minimal nodal degree m is 2 and the cutoff degree K is 100, which equals to $\sqrt{N}$. For the lower-bounded parallel attack and lower-bounded flooding attack schemes, we consider the cases where the lower

bound T equals to 4 and 10 respectively. The case with no lower bound is also simulated for comparison purpose. We adopt the definition in [47] that a network is crashed once $\langle k^2 \rangle / \langle k \rangle \leq 2$ is achieved. Unless otherwise specified, the presented simulation results come from an average of 50 independent realizations.

In Fig. 5.1, we compare the analytical and simulation results of crash thresholds for the greedy sequential attack and coordinated attack, respectively. For comparison, the results for the case of classic intentional attack are also presented. In the analysis, the threshold degree K_T for the greedy sequential attack is set as $2 \cdot \langle k_0 \rangle$. Analytical and simulation results basically match well, and the lower bound of the crash threshold for coordinated attack turns out to be reasonably tight. For the efficiency of the two types of attacks, we observe that the greedy sequential attack is not very effective in crashing networks. The networks are only crashed after around half of all the nodes are removed. The coordinated attack, on the other hand, performs nearly as efficient as the intentional attack, especially when α is of small values. The main drawback of the coordinated attack, however, is the rather complicated collaborations between all the attacking frontiers in order to find the best next-step target.

Fig. 5. 1: The crash thresholds of the greedy sequential attack and coordinated attack in random scale-free networks.

Table 5.1 lists the number of steps needed for the lower-bounded parallel attack to crash all the nodes with degrees no less than the lower bound. The analytical and simulation results achieve a reasonable match. We see that the lower-bounded parallel attack is very efficient in not only crashing but eliminating a large network as long as the lower bound T is small enough. Note that since we set the same minimum and cutoff nodal degrees for all the networks, the link density is higher when α is of a smaller value, which allows the parallel attack to become even more efficient.

Table 5. 1: Analytical and simulation results of the numbers of steps needed to remove all the nodes with high-enough degrees in different networks by the lower-bounded parallel attack

	$T = 2$		$T = 4$		$T = 10$	
	Analytical result	Simulation result	Analytical result	Simulation result	Analytical result	Simulation result
$\alpha = 2.1$	17	18	15	16	12	13
$\alpha = 2.3$	18	18	15	16	12	13
$\alpha = 2.5$	18	19	17	17	12	13
$\alpha = 2.7$	19	19	17	17	13	14
$\alpha = 2.9$	20	20	18	18	14	15

The results of lower-bounded flooding attack are presented in Table 5.2. The analytical and simulation results match quite well. Comparing Table 5.1 and Table 5.2, we see that compared to that of the lower-bounded parallel attack, the lower-bound flooding attack, by removing more nodes in each step, gives the defending side a much shorter time to react.

Table 5. 2: Comparison of the analytical and simulation results of the numbers of steps needed to remove all the candidate nodes in different networks by the lower-bounded flooding attack

	$T = 2$		$T = 4$		$T = 10$	
	Analytical result	Simulation result	Analytical result	Simulation result	Analytical result	Simulation result
$\alpha = 2.1$	7	7	5	5	4	5
$\alpha = 2.3$	7	7	6	6	5	6
$\alpha = 2.5$	8	8	6	6	5	6
$\alpha = 2.7$	9	9	7	7	6	7
$\alpha = 2.9$	10	10	8	8	7	7

In Fig. 5.2, the procedures of the two attacks are plotted with further details for the case where $\alpha = 2.5$. We show in each step of attack the largest cluster size, defined previously as the number of nodes in the biggest connected component versus the number of nodes in the original network [4]. An interesting observation we can make in Fig. 5.2(a) is that for the lower-bounded parallel attack, the largest cluster size is smaller when the lower bound is higher until the attack stops. This can be easily explained: since in each step a live adjacent node with degree no less than the lower bound is randomly selected as the next-step target, a higher lower bound makes the high-degree nodes to be crashed faster, which reduces the largest cluster size. In the lower-bounded flooding attack, however, the situation is totally reversed:

a lower value of T makes more nodes crashed in each step and consequently a smaller largest cluster size, as we observe in Fig. 5.2(b).

(a) lower-bounded parallel attack

(b) lower-bounded flooding attack

Fig. 5. 2: The efficiencies of the lower-bounded parallel attack and flooding attack for random scale-free network ($\alpha = 2.5$) with nodal-degree lower bounds set at 4 and 10 and without lower bound (or equivalently with lower bound at 2), respectively.

Another interesting observation is when the threshold is set at a relatively high value of 10, the attack is terminated when $\langle k^2 \rangle / \langle k \rangle$ is still much higher than 2 and the network is still largely connected, as can be seen in Fig. 5.2(a).

5.5 SIMULATIONS ON REAL-LIFE NETWORK MODELS

To evaluate the effects of the proposed node protection schemes in real-life communication networks, we carry out simulations on the same two network models as those in Chapters 3 and 4:

- A real-life Internet model on the AS-level as measured by the Applied Network Research (NLNR) Project on January 2, 2000 [97], which contains 6470 inter-connected nodes and 12,566 links.
- A real-life Internet model on the router level as measured by the Cooperative Association for Internet Data Analysis (CAIDA), which contains 192,244 nodes and 609,066 links [98].

Network robustness is still evaluated by measuring the (i) largest cluster size; and (ii) cluster diameter. The definition of the threshold of crash (TOC) also remains unchanged.

The presented simulation results for the AS-level model come from an average of 100 independent realizations. For the router-level model, due to its extra-large

size, it is prohibitively time-consuming to carry out extensive realizations. Therefore only 10 realizations have been conducted.

Fig. 5.3 shows the largest cluster sizes under greedy sequential and coordinated attack, respectively. For comparison purpose, results under the intentional attack are also presented. We observe that though the two Internet models are not random networks, they confirm the conclusion that the greedy sequential attack is not quite efficient in crashing scale-free networks, whereas the coordinated attack, though based on local information only, performs nearly as efficient as intentional attack.

For the two lower-bounded attack schemes, we set up the lower-bound degree of the crashed nodes at 1, 4 and 10 respectively. The simulation results of lower-bounded parallel attack are reported in Fig. 5.4. For the AS-level model, the attack is highly efficient when the lower bound is 10. In fact, the TOC value is only slightly increased from 3.2% of the intentional attack to 4%. If we set the lower bound at 4, more nodes have to be removed before the network is crashed: the TOC becomes 9.6%. With no lower bound (or equivalently by setting the lower bound at 1), 25.3% of all the nodes have been removed when the network is finally crashed. The conclusion remains the same as that in the random scale-free networks; i.e., higher lower-bound values make the largest cluster size decrease faster. In the router-level model, setting the lower bound at 10 makes the distributed attack almost as efficient as the intentional attack -- the TOC only increased slightly from 13.6% to 13.84%. The efficiency drops when the lower bound gets lower.

(a) Internet AS-level model

(b) Internet Router-level model

Fig. 5. 3: The efficiencies of the greedy sequential attack and the coordinated attack in the two Internet models, respectively.

(a) AS-level Internet model

(b) Router-level Internet model

Fig. 5. 4: The efficiencies of the lower-bounded parallel attack with nodal-degree lower bounds being set at 1, 4 and 10, respectively.

The above observations show that in real-life complex networks where high-degree nodes largely form a *connected* sub-network, taking down these nodes does not need global information of network topology; local information-based attacks can be almost equally effective. To prevent a complex system from being crashed by such highly effective distributed attacks, it is crucial to identify and stop the attack at an early stage: as we can see, the networks' decomposition is slow at the first several steps and then quickly speeds up.

Fig. 5.5 shows the largest cluster sizes under the lower-bounded flooding attack. In both of the two Internet models, the attack quickly eliminates a large portion of the network in just a few steps. Different lower bound values lead to only slight differences and the reaction time is very short.

Lastly we present the results of cluster diameters under attack. Since it is extremely time-consuming to carry out even a moderate number of realizations of diameter calculations in the router-level model in order to achieve an acceptable error range, we present only the results for the AS-level model. As can be observe in Fig. 5.6(a), the coordinated attack quickly increases the cluster diameter; and shortly after that, the network is totally crashed. The greedy sequential attack, though not quite efficient in crashing the whole network, also quickly increases the cluster diameter, showing that it may nevertheless make the networks to be inefficient in supporting some application, e.g., data communications.

(a) Internet AS-level model

(b) Internet Router-level model

Fig. 5. 5: The efficiencies of the lower-bounded flooding attack with nodal-degree lower bounds being set at 1, 4 and 10, respectively.

(a) Greedy sequential attack and coordinated attack

(b) Lower-bounded parallel attack

(c) Lower-bounded flooding attack

Fig. 5. 6: The cluster diameters of the AS-level Internet model under different types of attacks.

For the lower-bounded parallel attacks, as can be seen from Fig. 5.6(b), the cluster diameter increases slowly at the beginning and then speeds up quickly. Note that when we set the lower bound at 10, the cluster diameter remains at a constant value, averaged about 6.65, after 12 steps of attacks. This large diameter belongs to a sparsely-connected cluster with an average size of 62 nodes, in which there is no node with a degree higher than 10. The error bar for such a case is smaller than the symbol size.

Fig. 5. 7: The largest cluster size of the AS-level Internet model with node protection Scenario I under coordinated and intentional attacks. The link-removal probability is 80%.

Finally, it is worth mentioning that in random scale-free networks with imperfect protection, generally speaking the efficiency of the distributed attack schemes remains as comparable to that of the global information-based intentional

attack. To avoid having a large number of figures and lengthy discussions, we show in Fig. 5.7 only the comparisons between coordinated and intentional attacks in a network where a node averagely loses 80% of its links when it is attacked (i.e., the imperfect node protection Scenario I as defined in Sec. 4.2 with a link removal probability of 80%). We see that the two different types of attacks perform nearly the same. . Such conclusions can be extended to all the other cases we have tested.

5.6 Conclusions

In this chapter, we examined the robustness of complex networks under a few different types of distributed attacks. Analytical and extensive simulation results showed that such local information-based distributed attacks can be highly effective, sometimes almost as efficient as the global information-based intentional attack, and they typically leave rather limited time for reaction. Such insights, as we believe, will be helpful for developing effective attacking/protecting strategies for future complex systems.

6 Robustness of Scale-free Networks under Re-wiring Operations

6.1 Introduction

In the previous chapters, we have discussed the effects of protection mechanisms and the efficiency of distributed attack in scale-free networks. In this chapter, we study the relationship between network robustness and its connection patterns. More specifically, our research mainly focuses on revealing the complicated relationship between network robustness and assortativity.

As that has been discussed in Chapter 2, scale-free networks are strongly robust under random failures yet fragile against intentional attack. The fragileness of the scale-free networks under the intentional attack comes from their heavy-tailed property, causing loss of a large number of links when a hub node is crashed. The heavy loss of network links quickly makes the network to be sparsely connected and then fragmented. A random failure, on the other hand, usually removes a low-degree node. The network connectivity therefore would not be severely impaired.

It has been found that the robustness of the scale-free networks under intentional attack can be affected by several different factors:

- Exponent α . It is proven that the fragileness only exists in scale-free networks with exponents smaller than 3 [68]. However, most of the real-life scale-free networks have their exponents within this range;
- Density of connections (measured by the average nodal degree) [5]. Networks having the same exponent may have different densities of connections. Those with higher densities have stronger robustness (and usually higher costs as well);
- Assortative mixing [12], which measures the probability that a hub node is connected to another hub node. As reviewed in Chapter 2, it is known that most social networks are assortative mixed where hub nodes tend to connect to hub nodes; whereas most communication and biological networks are disassortative where hub nodes tend to connect to low-degree nodes [12]. The most important exception is probably the assortative-mixed model of the Internet on router level [50]. For a given nodal-degree distribution, generally speaking, assortative-mixed networks are more robust than disassortative ones [12-14], though exceptions do occasionally exist. A more detailed survey can be found in Section 2.4.2.

The assortative coefficient of a network can be calculated as [12],

$$\frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2} \quad (6.1)$$

where M is the total number of links, and j_i and k_i are the degrees of the nodes at each end of the i_{th} link. A positive coefficient value reflects an assortative mixed network model, while a negative value reflects a disassortative one

In this chapter, we reveal some more complicated relationship between network robustness and assortativity. Specifically, we show that by proper link rewiring which does not change any nodal degree, in random uncorrelated networks or random correlated networks (both have been defined in Sec 2.2) satisfying some simple conditions stronger network robustness can be steadily achieved at a slightly lower assortativity level. The observations also suggest some new possibilities of network enhancement and protection.

The rest of this chapter is organized as follows. In Section 6.2, detailed analysis on the robustness of random correlated network is presented. A simple link rewiring scheme is proposed in Sections 6.3. In Section 6.4, theoretical analysis is developed to show that the rewiring operations enhance network robustness while decreasing its assortativity level. Numerical results are presented in Section 6.5. Finally, Section 6.6 concludes the chapter.

6.2 Random networks and their robustness

As discussed in Chapter 2, the simplest random network model is the *uncorrelated* network model, which has random connections between any node pair, only subject to given nodal-degree distribution.

Degree correlations can also be introduced into the random network. Specifically, the uncorrelated random network model is extended to have random connections only subject to the given probability distribution e_{jk} , defined as the probability that a randomly chosen link lies between two nodes with degrees j and k respectively. To differentiate, we term such network model as *random correlated* networks.

As mentioned in Chapter 2, in some studies, e_{jk} denotes the probability that the *remaining* degrees of the two end nodes are j and k , respectively (e.g., [12, 26, 55]). In this chapter, we adopt the former definition.

Our discussions still start from the well-known equation that [47]

$$\kappa \equiv \frac{\langle k^2 \rangle}{\langle k \rangle} < 2 \quad (6.2)$$

where $\langle k \rangle$ denotes the average of nodal degrees and $\langle k^2 \rangle$ the average of the squares of nodal degrees. This equation actually also applies to random correlated networks where loops can be neglected. In fact, the derivation of Eq. (6.2) in [47] does not depend on any non-correlation assumption. As an alternative approach, below we derive Eq. (6.2) by explicitly including e_{jk} into equations.

In random correlated networks where loops can be ignored, a giant component starts to emerge when a node i connected to a node j in a spanning cluster is also connected to at least one other node. Hence [47]

$$\langle k_i | i \leftrightarrow j \rangle = \sum_{k_i} k_i P(k_i | i \leftrightarrow j) = 2 \quad (6.3)$$

where k_i denotes the degree of node i and $P(k_i | i \leftrightarrow j)$ the conditional probability that node i has a degree k_i given that it is connected to node j . Same as that in [47], we have

$$P(k_i | i \leftrightarrow j) = P(i \leftrightarrow j | k_i) P(k_i) / P(i \leftrightarrow j) \quad (6.4)$$

Where $P(i \leftrightarrow j)$ denotes the probability that two randomly chosen node i and j are connected to each other.

Define $P(i \leftrightarrow j | k_i, k_j)$ as the conditional probability that two nodes i and j are connected given that their nodal degrees are k_i and k_j , respectively. We have

$$P(i \leftrightarrow j | k_i, k_j) = \begin{cases} \frac{k_i \cdot e_{k_i k_j}}{q_{k_i} \cdot N \cdot P(k_j)}, & k_i \neq k_j \\ \frac{k_i \cdot e_{k_i k_j}}{q_{k_i} \cdot [NP(k_j) - 1]}, & k_i = k_j \end{cases} \quad (6.5)$$

Note that $\frac{e_{k_i k_j}}{q_{k_i}}$ denotes the conditional probability that a link is connected to a

degree- k_j node given that it is connected to a degree- k_i node. $\frac{k_i e_{k_i k_j}}{q_{k_i}}$ therefore

calculates the expected number of degree- k_j nodes connected to a degree- k_i node.

Here N denotes the number of nodes in the network, and

$$q_{k_i} = \sum_{k_j} e_{k_i k_j}$$

Therefore,

$$\begin{aligned} P(i \leftrightarrow j | k_i) &= \frac{1}{N-1} \left[(NP(k_i) - 1)P(i \leftrightarrow j | k_i, k_j) + \sum_{k_j, k_j \neq k_i} NP(k_j)P(i \leftrightarrow j | k_i, k_j) \right] \\ &= \frac{k_i}{N-1} \left[\frac{e_{k_i k_i}}{q_{k_i}} + \sum_{k_j, k_j \neq k_i} \frac{e_{k_i k_j}}{q_{k_i}} \right] \\ &= \frac{k_i}{N-1} \end{aligned} \quad (6.6)$$

Finally,

$$\begin{aligned} P(i \leftrightarrow j) &= \sum_{k_i} P(i \leftrightarrow j | k_i)P(k_i) \\ &= \frac{\sum_{k_i} k_i \cdot P(k_i)}{N-1} \\ &= \frac{\langle k \rangle}{N-1} \end{aligned} \quad (6.7)$$

From Eqs. (6.3), (6.4), (6.6) and (6.7), we shall then have Eq. (6.2) holds in random correlated networks.

6.3 Rewiring method

Under intentional attack, network nodes with highest degrees in the original network are removed one after another. When there are multiple nodes with the same degree in the original network, one of them is randomly removed. The procedure is repeated until the network is crashed, or in other words, until

$$\kappa \equiv \frac{\langle k^2 \rangle_T}{\langle k \rangle_T} < 2 \quad (6.8)$$

where T denotes the degree of the last removed node in the original degree, $\langle k \rangle_T$ is the average of nodal degrees once the network has been crashed, and $\langle k^2 \rangle_T$ the average of the squares of nodal degrees at that time [75]. Hereafter we term such a network just crashed by the intentional attack as a *post-attack network*.

The proposed rewiring method repetitively applies the rewiring operation in the original network as follows:

Divide the nodes with degrees strictly lower than T into two non-overlapping parts, with degrees $\{1, 2, \dots, T_1\}$ and degrees $\{T_1 + 1, T_1 + 2, \dots, T - 1\}$ respectively.

Denote the degree of a certain node i as k_i . Randomly choose two links A-B and C-D which satisfy the condition that

$$k_A > T > k_B > T_1 \geq \max(k_C, k_D) \quad (6.9)$$

If there are no links A-C and B-D, the rewiring operation replaces links A-B and C-D with links A-C and B-D as long as such does not generate a loop. An example of the rewiring operation is illustrated in Fig. 6.1.

The rewiring operation obviously does not change any nodal degree. It changes a random network into another one with the same $P(k)$ but different probability distributions $\{e_{jk}\}$. Consider its effects on the average nearest neighbor degree, defined as the average degree of all the neighbors of the nodes with a certain degree k [58]. For nodes with degrees higher than T_1 in the original network, the average nearest neighbor degree is decreased, while for the other nodes, it is increased. In the

example case in Fig. 6.1, it can be observed that the average nearest neighbor degree is decreased for nodes A and B, and increased for nodes C and D.

Fig. 6. 1: An illustration of the rewiring operation.

In this chapter, we evaluate the simple case in which nodes A to D are randomly selected as long as their degrees satisfy Eq. (6.9) and the rewiring does not generate a loop. The optimal rewiring method which achieves the best network robustness without changing any nodal degree remains as an open problem which is out of scope of this chapter.

6. 4 Analysis

For each node i with a certain degree k_i in the original network, the k_i links attached to it can be categorized into two groups, those connecting to nodes which have been removed when network is crashed, and those connecting to nodes

remaining in the post-attack network. Denoting the number of links belonging to these two groups as k_i^+ and k_i^- respectively, we have

$$k_i = k_i^+ + k_i^- \quad (6.10)$$

Now we analyze the influence of the rewiring operations. As we can see in Fig. 6.1, since node A will be removed during attack, the rewiring operations actually replace link C-D with link B-D in the post-attack network. We calculate how the value of $\langle k^2 \rangle / \langle k \rangle$ changes in the post-attack network. The value of $\langle k \rangle$ apparently does not change in the post-attack network. As to the value of $\langle k^2 \rangle$, it is changed by

$$(k_B^- + 1)^2 + (k_C^- - 1)^2 - (k_B^-)^2 - (k_C^-)^2 = 2(k_B^- - k_C^-) + 2 \quad (6.11)$$

In random uncorrelated networks, since $k_B > k_C$, we have that

$$E(k_B^-) > E(k_C^-) \quad (6.12)$$

where $E(.)$ denotes the expected value of all the nodes involved. In random correlated networks with positive correlation between k_i and k_i^- where $k_i < T$, or in other words,

$$E(k_i^-) > E(k_j^-) \quad \text{where } T > k_i > k_j \quad (6.13)$$

We shall still have

$$E(k_B^-) > E(k_C^-)$$

In both cases, we have that the expected value of $\langle k^2 \rangle / \langle k \rangle$ is increased. With a large enough number of rewiring operations, $\langle k^2 \rangle / \langle k \rangle$ may be pushed up to be higher than 2, allowing the emergence of a giant component. Therefore statistically the rewiring operations enhance the network robustness.

We then show that the proposed rewiring method slightly decreases network assortativity coefficient. According to Eq. (6.1), each rewiring operation changes the value of the network assortativity coefficient by

$$\begin{aligned} & \frac{1}{K} (k_A \times k_C + k_B \times k_D - (k_A \times k_B + k_C \times k_D)) \\ &= \frac{1}{K} (k_C - k_B)(k_A - k_D) \end{aligned} \tag{6.14}$$

where K is a positive constant number. From Eq. (6.9), $(k_A - k_D) > 0$, $(k_C - k_B) < 0$. Therefore, the value of Eq. (6.14) is negative. The overall network assortativity coefficient is decreased. Such effects can also be observed from the fact that the rewiring operations decrease the average nearest neighbor degrees of high-degree nodes, and increase them for low-degree nodes, as we have pointed out earlier in the section of "Rewiring method".

To summarize, we showed that the proposed rewiring method enhances network robustness in random uncorrelated networks as well as random correlated networks satisfying Eq. (6.13) while decreasing their assortativity.

6.5 Simulation results and discussions

As that in previous chapters, we evaluate the robustness of a network by measuring its (i) largest cluster size, and (ii) cluster diameter.

We present the simulation results on two random scale-free networks generated by adopting the algorithm proposed in [58], with exponent values of 2.5 and 3, respectively. In both networks, there are 10,000 nodes. The minimum nodal degree is 2 and the cutoff nodal degree is 400. Calculations show that the two networks have assortativity coefficient values of -0.0146 and -0.0062, respectively. This means that both networks are slightly disassortative, which matches what has been discussed in [58]. To carry out rewiring operations, in the former network, $T = 6$ and we set $T_1 = 3$, while in the latter one, $T = 5$ and $T_1 = 3$. We test network robustness when the maximum numbers of 2920 and 1084 rewiring operations have been carried out in the two networks respectively.

(a) Largest cluster size

(b) Cluster diameter

Fig. 6. 2: Simulation results of intentional attack on the random scale-free network

where exponent equals to 2.5.

(a) Largest cluster size

(b) Cluster diameter

Fig. 6. 3: Simulation results of intentional attack on the random scale-free network

where exponent equals to 3.

The simulation results for the case under intentional attack are reported in Fig. 6.2 and Fig. 6.3, in which we show the biggest cluster size and the cluster diameter in the two networks before and after rewiring operations, respectively. For each network, 100 independent realizations have been carried out. The figures show the average values of all the realizations and the error bars. If we regard a network as being crashed once $\langle k^2 \rangle / \langle k \rangle < 2$, the percentage of the network nodes that have to be removed to crash the network (hereafter termed as *threshold of crash (TOC)*) is increased from 15% to 26% in the first network and from 13% to 16% in the second one. Also, we see that most of time during the attack until the network is crashed, the cluster diameter is smaller after rewiring. Note that when networks are attacked to be highly sparsely connected, there is a large range of standard error in network diameter. This is not a surprise: different connection patterns of a small number of relatively high-degree (e.g., degree-4) nodes and a large number of low-degree (mostly degree-1 and degree-2) nodes can make big differences to network diameter.

Fig. 6.2 and Fig. 6.3 evidently show that the rewiring method enhances network robustness in scale-free networks with different exponent values. And the enhancements are more significant when more link pairs are rewired.

We then evaluate whether the rewiring operations affect the network tolerance against random failures. Figs. 6.4-6.5 show that most of time no major changes are made, except that the cluster diameter may be slightly different.

(a) Largest cluster size

(b) Cluster diameter

Fig. 6. 4: Simulation results of random failure on the random scale-free network where exponent equals to 2.5.

(a) Largest cluster size

(b) Cluster diameter

Fig. 6. 5: Simulation results of random failure on the random scale-free network where exponent equals to 3.

For the interest of potential applications, we also test a real-life router-level Internet model which contains 192,244 nodes and 609,066 links [98]. Our calculations show that it is indeed an assortative-mixing network with an assortativity coefficient value of 0.02498. In this network, $T = 6$ and we set $T_1 = 3$. A total of 26,017 link pairs can be rewired. Since it is prohibitively time-consuming to carry out 100 realizations on such a large network, only 10 realizations have been conducted and the average results are shown in Fig. 6.6. It can be seen that though the router-level Internet is not strictly a random network, the rewiring operations nevertheless significantly enhance the network robustness: the TOC value is increased from 27% to 34%.

It is known that recalculated intentional attack which removes the largest-degree node in the *remaining* network is more efficient in crashing down scale-free networks than the one based on nodal degrees in the original network [67]. We carry out numerical simulations for testing the effects of rewiring under such attack. The results for the random scale-free network with exponent value of 2.5 are presented in Fig. 6.7, while the conclusions hold in all the other networks we have simulated. It can be observed that the rewiring method can still significantly enhance network robustness though the TOC values with or without rewiring both become lower than those under original degree-based intentional attack. Specifically, the TOC value is increased from 13% to 21%. Theoretical analysis on the case of recalculated intentional attack will be carried out in our future research.

(a) Largest cluster size

(b) Cluster diameter

Fig. 6. 6: Simulation results of random failure on the router-level Internet model.

(a) Largest cluster size

(b) Cluster diameter

Fig. 6. 7: Simulation results of recalculated intentional attack on the random scale-free network where exponent equals to 2.5.

Finally, note that assortativity level is indeed decreased by the rewiring operations. According to our calculations, after the rewiring, the assortativity coefficient is slightly decreased from -0.0146 to -0.0268 in the first model and from -0.0062 to -0.0114 in the second one. In the router-level network model, the coefficient value is decreased from 0.02498 to 0.02406.

6.6 Conclusions

In this chapter, we proposed a simple rewiring method which enhances network robustness without changing any nodal degree. Interesting observation is that it slightly decreases network assortativity coefficient. The rewiring method meanwhile does not significantly affect network tolerance against random failure. Such observations reveal the more complicated relationship between network robustness and assortativity level, and some new possibilities of enhancing network robustness by properly interconnecting the low-degree nodes.

7 Degree-Based Decentralized Search in Complex Communication Networks: An Empirical Study

7.1 Introduction

Similar to the network robustness, network decentralized search [15], defined as the operations of finding and/or reaching a certain target without any central control, also heavily relies on the existence of hub nodes. As mentioned in Section 2.5, this problem is becoming increasingly important, partly due to the advances in peer-to-peer networks [16-17], web crawlers [78-79], sensor networks [80-81], and dynamic supply chains [82], etc. The existing methods can be roughly classified into following three different categories:

- Methods with minimum assumptions of network topologies and content distributions [83-86].
- Methods that explore the special features of complex-network topologies, like the *degree-based* searching method [18] which forwards the search request to the largest-degree adjacent node. To

facilitate later discussions, hereafter we will call this method the *Greedy Largest-Adjacent-Node* (GLAN) method.

- Methods that explore content distributions on different network nodes, like the *similarity-based* method [19, 92-93].

A brief survey of existing studies on decentralized search can be found in Section 2.5. In this chapter, we concentrate on degree-based methods with the main objectives of achieving an in-depth understanding of such methods and studying the possibility of improving their performances. Since achieving accurate theoretical analysis of degree-based methods remains a challenge even for the simplest GLAN method (A simple analytical model was proposed in [18], for qualitatively studying the properties of degree-based methods rather than achieving quantitative accuracy matching with numerical simulation results.), an empirical study is conducted. Our main observations and contributions can be summarized as follows:

- We evaluate the performance of the simplest GLAN method [18] with different amounts of local information assumed to be available to each node. Three different cases are tested and in each of these cases, each node can identify the target within one-, two- or three-hop distance from itself (We say the node has one-, two-, and three-hop information in this chapter.), respectively. We find that the performance of the method is significantly better when more information is available.

- Considering the fact that acquiring complete two-hop information may be difficult in some complex systems, let alone three-hop information, we evaluate the possibility of achieving good performance without three-hop information. Specifically, we propose and evaluate several different methods applying (i) *local information exchanges* between a limited number of adjacent nodes (which can be implemented in many communications networks), and (ii) parallel searching operations, respectively. It is shown that, though these methods help to improve the performance of search operations, they cannot easily achieve comparable performance to that of the case with complete three-hop information.
- Finally, we investigate the case in which there exists duplicates of the target in the complex networks, such as having copies of the same file on different servers in the Internet. We show that even the simplest GLAN method turns out to be highly effective when there are only two or three copies of each target in a large network with several thousand nodes.

The rest of this chapter is organized as follows. In Section 7.2, we define in details the degree-based search methods to be investigated. Extensive simulation results and discussions are presented in Section 7.3. Finally, Section 7.4 concludes the chapter.

7.2 Degree-based decentralized search methods

A. The GLAN method [18]

The simplest (and probably most important) degree-based search method is proposed in [18], of which the operation procedure can be defined as follows:

Upon the arrival of a searching request, each node checks the local information it has. If the target is identified among its neighbors, the search stops; otherwise, there are two different possible courses of action:

- 1) If some adjacent nodes have never been visited by the searching request before, forward the searching request to the largest-degree node among them. When there is a tie, break it randomly.
- 2) If all the adjacent nodes have been visited by the searching request, deflect the searching request to the node where it came from.

The searching operation will be terminated if the request has been forwarded/deflected many times yet still cannot find the target.

A slightly different method is reported in [18]: in case 2), if all the adjacent nodes have been visited by the searching request before, find among them those nodes whose adjacent nodes have not been *exhaustively* visited by the searching request, then forward the request to the highest-degree node among them. Our experiences show that the two methods perform nearly the same. The latter one only

occasionally helps to shorten the searching procedure. Therefore in this chapter we discuss only the former method.

B. The revised methods with local information exchanges or parallel operations

As mentioned above, we will evaluate the performance of the GLAN method with one-hop, two-hop and three-hop information, respectively (In [18], it has always been assumed that two-hop information is available.). As we will see in Section 7.3, the performance becomes significantly better when more information is available. In many real-world systems, however, it may not be easy to acquire two-hop information, let alone three-hop information. For example, in the Internet, a big hub may have hundreds of thousands of neighboring nodes within three-hop distance from itself. To test the possibility of relaxing the request of having complete three-hop information without significantly sacrificing searching performance, we consider the following methods which are based on information exchanges between a limited number of adjacent nodes, and parallel searching operations, respectively.

- *Partial Three-Hop Information-Based* (Partial-3) method: This method is similar to the GLAN method. Only that, when the target cannot be found by a node within its two-hop neighborhood, the node will “consult” a certain number of adjacent nodes on whether the target is within their two-hop neighborhood. When a positive reply is received, the search stops; otherwise, the search request is forwarded (or

deflected) in the same manner as that in the GLAN method. In our simulations, we let each node consult its largest-degree adjacent nodes that have never been visited or consulted before.

- *Degree-Constrained Broadcasting (DCB) Method:* It is very similar to the Partial-3 method. The only difference is that, when the target cannot be found by a node within its two-hop neighborhood, this node will “consult” *all* its adjacent nodes with nodal degrees higher than a certain threshold value.
- *Multi-Path Parallel Search (MPPS) Method:* In this method, a certain number of searching requests are generated to look for the same target. Specifically, we let the source node generate the required number of searching requests and send them to its adjacent nodes with the largest, second-largest, and third-largest nodal degrees, etc. If the required number of parallel searching operations is larger than the number of adjacent nodes the source node has, searching requests will be sent to all the adjacent nodes, and be further “split” on the largest-degree adjacent nodes. This procedure will be repeated until the required number of parallel operations is achieved. Each of the parallel searching operations would then be conducted in the same way as that in the GLAN method.

To summarize, we test the Partial-3 method to see whether having partial three-hop information can achieve a performance comparable to that of the case of having complete three-hop information. The DCB method is tested to see whether consulting nodes with high degrees rather than consulting a fixed number of adjacent nodes in each step will lead to a higher efficiency. The MPPS method is simulated to reveal the improvements that can be achieved by having multiple parallel searching operations.

7.3 Evaluations of the effects of different information availabilities

The decentralized search methods proposed in the last section will be tested on two different network models as follows;

- Random scale-free networks are generated by adopting the algorithm proposed in [58], with exponent value equals to 2.5. Unless otherwise specified, the network size N is 10,000; the minimal nodal degree m is 2 and the cutoff degree K is 100, which equals to $\sqrt{N}$.
- A real-life Internet model on the AS-level as measured by the Applied Network Research (NLNR) Project on January 2, 2000 [97], which contains 6470 inter-connected nodes and 12,566 links.

For random scale-free networks, the result comes from an average of 50 independent realizations. In each realization, 500 source-target pairs are randomly

selected. In Internet AS-level network, we also test 500 source-target pairs. We record the *hop length* (i.e., the number of links each searching request goes through before it reaches the target) of each of these searching requests and display their distributions. For the MPPS method, we record the hop length of the *first* successful searching request. In all our simulations, unless otherwise specified, a searching request is dropped once the number of forward/deflection operations is equal to the number of nodes in the network, which seldom happens from our experience.

A. Performance of the GLAN method

We first simulate the GLAN method with one-, two-, and three-hop information, respectively. The simulation results are presented in Fig. 7.1. Evidently, having more local information helps to significantly improve the performance of the searching method. For example, for the random scale-free model, with single-hop information, the average hop length of each searching request is as large as 870.057. This number can be reduced to 78.337 when two-hop information is available; and further reduced to 8.1134 when three-hop information is available (To provide a benchmark, the average hop length of the shortest paths between all the node pairs is 4.71.). The searching requests with hop length less than 10 account for 6.4%, 41.9%, and 89.4% of all the requests in the three different cases, respectively. Similar conclusions hold for the Internet AS-level model.

B. Performance of the methods with local information exchanges or parallel operations

The results in Fig. 7.1 show that, having complete two-hop information is not sufficient to achieve comparable performance to that of the centralized search in which the shortest path between each node pair can be calculated. Considering the fact that it may not be feasible to acquire three-hop information in many real-world systems, we test the performance of the Partial-3, DCB and MPPS methods, respectively.

Partial-3 method has been tested where each node can consult at most two, three or five of its adjacent nodes respectively. The simulation results are presented in Fig. 7.2. We see that compared to the case with complete two-hop information, the Partial-3 method performs much better. However, it does not improve the performance to an extent comparable to that of the case with complete three-hop information. For the random scale-free model, even when each node can consult up to five adjacent nodes, the average hop length of each search operation is still 21.334 (With complete three-hop information, the average path length would be a much shorter 8.1134.). For the Internet AS-level model, the average hop length of Partial-3 method with five adjacent nodes being consulted is 14.266, not close to that of the case with complete three-hop information either.

(a) SF model

(b) Internet AS-level model

Fig. 7. 1: Performance of GLAN method [18] with different amounts of local information.

(a) SF model

(b) Internet AS-level model

Fig. 7. 2: Performance evaluations of the Partial-3 method

The results for the DCB method are shown in Fig. 7.3, where each node can consult all its adjacent nodes with degrees higher than 5, 10, and 20 respectively. We see that in both the random scale-free network model and the Internet AS-level model, when the threshold value of nodal degree is set too high, few adjacent nodes would be consulted; as a result, the performance of this method is not significantly better than that of the case with two-hop information only. For example, for the random scale-free model with the threshold value set as 20, the average number of adjacent nodes consulted in each step is only 0.915 while the average hop length of each searching request is reduced from 78.337 to 55.4724. The performance, however, can be significantly improved when we lower the threshold value. For the Internet AS-level model, when the threshold value is lowered to 10, by consulting 2.25 adjacent nodes on average in each step, the average hop length of each searching request is reduced to 12.482, outperforming the Partial-3 method which consults up to 5 adjacent nodes in each step. When the threshold value is further lowered to 5, the progress is less impressive: the average hop length is reduced to 9.994 at a higher cost of consulting 5.17 adjacent nodes on average in each step. However, the earlier conclusion still holds, i.e., that compared to the Partial-3 method, selectively consulting large-degree adjacent nodes leads to better performance at a lower cost of neighborhood information exchanges. Table 7.1 provides a detailed summary of all the results for the DCB method.

Table 7. 1: Performance of the DCB Method

	Scale-free model			Internet AS-level model		
	Threshold values			Threshold values		
	5	10	20	5	10	20
Average hop length	19.361	38.616	55.4724	9.994	12.482	26.604
Average no. nodes consulted in each step	7.043	2.1971	0.915	5.168	2.253	0.696

An interesting observation from Fig. 7.3 and Table 7.1 is that, applying the DCB method in the Internet AS-level model with threshold value 5 achieves comparable performance to the GLAN method with complete three-hop information (average hop length 9.994 vs. 7.262), which is not the case for the random scale-free model (average hop length 19.3612 vs. 8.1134). Meanwhile, the improvement achieved by the DCB method over the Partial-3 method is more significant for the Internet AS-level model than the random scale-free model. This can be explained as follows: hub nodes are more “concentrated” in the Internet AS-level model than in the random scale-free model. Specifically, nodes with degrees higher than 5 account for 10% in the Internet AS-level model, each of which with an average of 1,120 two-hop neighborhood nodes. A small fraction of these high-degree nodes, together with their two-hop neighborhoods, literally cover the whole networks. For the DCB method, by flooding searching requests to high-degree nodes whenever they are adjacent, therefore has a good chance to identify the target quickly. On the other hand, for the random scale-free model, nodes with degrees higher than 5 account for

21.7% among all the network nodes, with a much-lower average of 154 nodes within their two-hop neighborhood; therefore, it is not a surprise that broadcasting to neighborhood hubs becomes less effective.

(a) SF model

(b) Internet AS-level model

Fig. 7. 3: Performance evaluations of the DCB method.

In the simulation of the MPPS method, we assume there are two parallel searching operations and every network node has complete two-hop information. As we can see in Fig. 7.4, the improvements over the classic GLAN method are not significant. The main reason, as we observed, is that one of the two searching requests can easily get into a dead-end when all the adjacent nodes have either been visited by itself or by the other searching request arrived earlier. In other words, one of the two parallel searching operations is spending a lot of time re-visiting the node where it came from and, in some cases, even all the way back to the source node. Unless a more sophisticated scheme is designed to avoid such a “trap”, e.g., to encourage one of the two searching operations to go to low-degree rather than high-degree nodes (which turns out to be not very efficient also according to our preliminary experiments because the chance of identifying the target within the neighborhood of a low-degree node is low), major improvements may not be easily achieved.

(a) SF model

(b) Internet AS-level model

Fig. 7. 4: Performance evaluations of the MPPS method with two parallel searching operations.

C. Evaluations of the GLAN method in networks with duplicates of the target

Consider that in many cases, especially world-wide web (WWW) and peer-to-peer networks, the information or the file being searched may be available on not one, but multiple different nodes. To study the effect of such duplications on the efficiency of the searching methods, we simulate the case in which the simplest GLAN method is applied with two-hop information being known to each node. The objective is to figure out how quickly at least one of the duplicates would be found.

The results reported in Fig. 7.5 turns out to be very good. When there are two duplicates of each target randomly distributed in the Internet AS-level model, the average hop length of each searching operation drops from 42.97 to 4.114, which can be further lowered to 3.346 when there are three duplicates of each target.

Close examination reveals the reasons behind this significant improvement. Take the Internet AS-level model as an example: when there is only a single target in the network, about 60% of search operations go through less than five hops. When there are two duplicates of each target, rough analysis shows that the percentage can be increased to $1 - (1 - 60\%) \times (1 - 60\%) = 84\%$, which actually matches well with the simulation results. When there are three duplicates, similar analysis shows that the percentage would be further increased to about 93%, which again roughly matches the simulation results. Similar conclusions hold for the random scale-free model.

(a) SF model

(b) Internet AS-level model

Fig. 7. 5: Evaluations of the GLAN method in networks with randomly-distributed duplicates of the target.

(a) SF model

(b) Internet AS-level model

Fig. 7. 6: Evaluations of the GLAN method in networks with duplicates of the target on randomly-selected adjacent nodes.

Finally, we take into account a phenomenon that exists in many real-world systems, i.e., network nodes with duplicates are usually connected. This does not come as a surprise considering the fact that duplicates in many systems actually come from a single source in the first place, spreading from one node to its neighbor(s). We test different cases in which there are duplicates of the target on two or three randomly selected adjacent nodes respectively. Fig. 7.6 shows that, though a few searching requests have to go through a long path before finally reaching their targets (with hop length larger than 100), which seldom happens where duplicates are randomly distributed, the average hop length of all the searching requests remains low. This observation is very encouraging as it suggests that the high sensitivity of the degree-based method to the amount of local information each node has may be offset by having duplicates of each target on different nodes. A degree-based decentralized search method, if properly designed and developed, may actually work very well in networks such as WWW or peer-to-peer file-sharing systems in which duplicates of the same file do widely exist.

7.4 Conclusions

We studied the degree-based decentralized searching methods in complex networks. Extensive simulations results show that the performance of such methods is quite sensitive to the amount of local information each node has. Such high sensitivity imposes a serious challenge to the design of efficient searching methods,

which cannot be easily resolved by allowing limited information exchanges between neighborhood nodes or by parallel search operations. Among the three proposed methods, namely Parital-3, DCB and MPPS, we found that consulting large-degree adjacent nodes (i.e., DCB) performs the best; and the least effective one is to generate multiple parallel searching requests at the source node (MPPS). Finally, we showed that the unfavorable high sensitivity can be largely offset by having duplicates of the target on different network nodes. Even the classic GLAN method performs quite well when there are only two duplicates of the target in large networks with thousands of nodes.

8 Conclusions

8.1 Summary of Main Contributions

Complex networks are gaining increasing research interests in the past few years, thanks to the wide existences of real-life complex systems which can be formulated into such network models. Scale-free network is safely among the most important types of complex networks, if not the single most important one. While enjoying high efficiency in communication and strong tolerance against failures, scale-free networks however are fragile under intentional attack.

In this thesis, we studied on robustness of scale-free networks against intentional attack, mainly focusing on various factors which may affect network robustness and possible strategies for enhancing networks. A closely related topic, degree-based decentralized search, which relies on the existence of high-degree hubs, has also been investigated.

Our main contributions can be summarized as follows:

- We investigated the performances of different network node protections schemes.

It is observed that though different schemes have different performances in different networks, generally speaking protecting a limited set of high or medium-degree nodes help significantly enhance network robustness. While protecting a few highest-degree hubs turns out to be effective in networks with big hubs or large exponent values, protecting a relatively larger number of medium-degree nodes may perform better in networks with moderate-size hubs.

- We then evaluated the network robustness against intentional attack where some or all of the network nodes are with certain imperfect protection. It is found that even a weak protection helps significantly enhance network robustness under some most hostile attacks. Protecting the links between hubs on the other hand helps maintain a short cluster diameter, which is important for telecommunication networks. .
- The robustness of complex networks under a few different types of distributed attacks was examined. Analytical and extensive simulation results showed that the local information-based distributed attacks can be highly effective in destroying large-scale complex networks, in some cases almost as efficient as the global information-based intentional attack.
- To investigate the relationship between network robustness and network assortativity, we proposed a simple rewiring method which does not change any nodal degree. It was proved that though the rewiring operations slightly decrease

network assortativity coefficient, they steadily enhance network robustness. Network tolerance against random failure meanwhile is barely affected. Such observations contradict most of the existing results, revealing a more complicated relationship between network robustness and assortativity. The study also demonstrates the possibilities of enhancing network robustness by properly interconnecting the low-degree nodes.

- In the last part of our research, we studied the degree-based decentralized searching methods in complex networks. Extensive simulations results showed that the performances of such methods are highly sensitive to the amount of local information available to each node. Such high sensitivity imposes a serious challenge to the design of efficient searching methods, which cannot be easily resolved by allowing limited information exchanges between neighborhood nodes or by having parallel search operations. The unfavorable high sensitivity, however, can be drastically offset by having duplicates of the target on different network nodes. In fact, even the classic GLAN method performs very well when there are two duplicates of the target in large networks with thousands of nodes.

8.2 Future Research Directions

To design more efficient and realistic strategies for enhancing network robustness and more effective decentralized search strategies, several research topics are of our future research interest. They include

- Theoretical studies on robustness of correlated scale-free networks. The robustness of uncorrelated scale-free networks has been theoretically analyzed but not for correlated networks. It is known that different nodal-degree correlations lead to different network robustness levels though the crash threshold of any scale-free network would probably remain to be low [13-14]. Almost all these results and conclusions however come from empirical studies. Our studies in Chapter 6 reveal some complicated relationship between network robustness and assortativity. Network robustness may also be influenced by other factors such as clustering effects and community structure, etc. Theoretical analyses on the largest cluster size and cluster diameter, which are of high importance to research on communication systems, are not available. All these topics request in-depth studies.
- Implementations of network protection and distributed attack schemes in correlated scale-free network. Considering almost all the real-world scale-free networks are correlated, it would be interesting to figure out whether different correlations make significant differences to network robustness with different protection schemes under different types of distributed attacks.
- Strategies for maintaining a partially functioning network under intentional attack. It is found that some protection strategies, though keeping networks connected under attack, fail to maintain a reasonably small cluster diameter. An arguably more meaningful protection strategy may be to keep short connections between as

many node pairs rather than keeping all the nodes connected through rather long paths. There is hardly any existing result for fulfilling this objective to the best of our knowledge.

- Extensions to weighted network. Link weight has not been introduced in this thesis: a link is assumed to be either connected or removed. Introducing link weight will allow studies on the cases where connections are weakened/enhanced rather than removed/added. For example, in communication systems, a “weakened” link may be a congested link with limited yet nonzero capacity. It is of our future research interests to study attacks, protections and rewiring in weighted networks.
- Network decentralized search making use of both of the nodal degree and content similarity information. We may have reasons to expect in some cases, if not all, WebPages with similar contents are more likely to be connected. Making use of content similarity therefore might make sense in such cases. There are some existing studies on this topic [89, 96]. A lot of issues are still open, including some very fundamental ones such as the definition of similarity of contents.
- Network search based on criteria other than nodal degree, e.g., node betweenness defined as the number of shortest paths passing through the node. Though finding accurate node betweenness typically requires global information for shortest-path calculations, such information, once generated, may be distributed to all the nodes, say, as a routing table to be stored on each node. It would be interesting to study

decentralized search utilizing node betweenness information. It may be even more interesting to study on the case when such information is inaccurate or incomplete.

We believe that studies on the above topics shall contribute to further enhancing network protection and network search efficiency.

Author's Publications

1. S. Xiao and G. Xiao, "On Intentional Attacks and Protections in Complex Communication Networks," *Proc. IEEE Globecom'2006*.
2. S. Xiao and G. Xiao, "On Degree-Based Decentralized Search in Complex Networks," *Proc. European Conference on Complex Systems 2006 (ECCS'06)*, Sept. 2006.
3. S. Xiao, G. Xiao, and T. H. Cheng, "Robustness of Complex Communication Networks under Rewiring Operations," *Proc. IEEE ICCS'2006*, Oct.-Nov. 2006.
4. S. Xiao, G. Xiao, and Y.-W. Leung, "A Network Flow Approach for Static and Dynamic Traffic Grooming in WDM Networks," *Computer Networks* (Elsevier), Vol. 50, No. 17, pp. 3400-3415, Dec. 2006.
5. S. Xiao and G. Xiao, "On Local Link Repairing in Complex Communication Networks under Intentional Attack," *Proc. ICICS'2007*, Dec. 2007.
6. S. Xiao, G. Xiao, and T. H. Cheng, "Tolerance of Intentional Attacks in Complex Communications Networks," *IEEE Communications Magazine*, vol. 46, no. 1, pp. 146-152, Jan. 2008.

7. S. Xiao, G. Xiao, T. H. Cheng, S. Ma, X. Fu, and H. Soh, "Robustness of Complex Communication Networks under Rewiring Operations," *EPL*, vol. 89, 38002, Feb. 2010.
8. S. Xiao, G. Xiao, and T. H. Cheng, "Tolerance of Local Information-Based Intentional Attacks in Complex Networks," *Journal of Physics A*, vol. 43, 335101, Aug. 2010.
9. S. Xiao, and G. Xiao, "On Imperfect Node Protection in Complex Communication Networks," submit to *Journal of Physics A*.

BIBLIOGRAPHY

- [1] E. Ben-Naim, H. Frauenfelder, and Z. E. Toroczkai, *Complex Networks*: SpringerVerlag Berlin Heidelberg, 2004.
- [2] S. Bornholdt, and H. G. E. Schuster, *Handbook of Graphs and Networks: From the Genome to the Internet*: Wiley-VCH, 2003.
- [3] A.-L. Barabási, and R. Albert, “Emergence of Scaling in Random Networks,” *Science*, vol. 286, no. 5439, pp. 509-512, 1999.
- [4] R. Albert, H. Jeong, and A.-L. Barabási, “Error and attack tolerance of complex networks,” *Nature*, vol. 406, pp. 378-382, 2000.
- [5] A. Beygelzimer, G. M. Grinstein, R. Linsker *et al.*, “Improving network robustness by edge modification,” *Physica A*, vol. 357, no. 3-4, pp. 593-612, 2005.
- [6] L. d. F. Costa, “Reinforcing the resilience of complex networks,” *Phy. Rev. E*, vol. 69, pp. 066127, 2004.
- [7] B. A. Rezaei, N. Sarshar, V. P. Roychowdhury *et al.*, “Disaster management in scale-free networks: Recovery from and protection against intentional attacks,” *Physica A*, vol. 81, no. 497-514, 2007.

- [8] S. Xiao, G. Xiao, and T. H. Cheng, "Robustness of Complex Communication Networks under Rewiring Operations," in Proc. IEEE ICCS 2006, 2006.
- [9] S. Xiao, G. Xiao, and T. H. Cheng, "Robustness of Complex Communication Networks under Repiring Operations," in Proc. IEEE ICCS'2007, 2007.
- [10] S. Xiao, G. Xiao, and T. H. Cheng, "Tolerance of Intentional Attacks in Complex Communications Networks," *IEEE Communications Magazine*, vol. 46, no. 1, pp. 146-152, 2008.
- [11] J. Zhao, and K. Xu, "Enhancing the robustness of scale-free networks," *J. Phys. A: Math. Theor.*, vol. 42, 2009.
- [12] M. E. J. Newman, "Assortative mixing in networks," *Phys. Rev. Lett.*, vol. 89, pp. 208701, 2002.
- [13] P. Holme, and J. Zhao, "Exploring the assortativity-clustering space of a network's degree sequence," *Phys. Rev. E*, vol. 75, pp. 046111, 2007.
- [14] J. Zhao, L. Tao, H. Yu *et al.*, "The effects of degree correlations on network topologies and robustness," *Chinese Physics*, vol. 16, pp. 3571-3580, 2007.
- [15] J. Kleinberg, "Complex Networks and Decentralized Search Algorithms," in Proceedings of the International Congress of Mathematicians (ICM), Madrid, Spain, 2006.

- [16] A. Oram, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies*: O'Reilly Media, 2001.
- [17] S. Waterhouse, D. M. Doolin, G. Kan *et al.*, "Distributed Search in Peer-to-Peer Networks," *IEEE Internet Computing*, vol. 6, no. 1, pp. 68-72, 2002.
- [18] L. A. Adamic, R. M. Lucose, A. R. Puniyani *et al.*, "Search in Power-Law Networks," *Phy. Rev. E*, vol. 64, pp. 046135, 2001.
- [19] J. Kleinberg, "Small-world Phenomena and the Dynamics of Information," *Advances in Neural Information Processing Systems 14*, T. G. Dietterich, ed., pp. 431-438, Cambridge, MA, USA: MIT Press, 2002.
- [20] P. Erdős, and A. Rényi, "On random graphs I," *Publ. Math. Debrecen*, vol. 6, pp. 290-297, 1959.
- [21] B. Bollobás, *Random Graphs*: Academic Press, 1985.
- [22] P. Erdős, and A. Rényi, "On the evolution of random graphs," *Publications of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, pp. 17-61, 1960.
- [23] P. Erdős, and A. Rényi, "On the strength of connectedness of a random graph," *Acta Mathematica Scientia Hungary*, vol. 12, pp. 261-267, 1961.

- [24] F. Chung, and L. Lu, "The average distances in random graphs with given expected degrees," *Proc. Natl. Acad. Sci. USA*, vol. 99, pp. 15879-15882, 2002.
- [25] D. J. Watts, and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature Physics*, vol. 393, pp. 440-442, 1998.
- [26] M. E. J. Newman, "The structure and function of complex networks," *SIAM Review*, vol. 45, no. 2, pp. 167-256, 2003.
- [27] R. Albert, and A.-L. Barabási, "Statistical mechanics of complex networks," *Review of modern physics*, vol. 74, pp. 47-98, 2002.
- [28] A. Barrat, and M. Weigt, "On the properties of small-world networks," *Eur. Phys. J. B*, vol. 13, no. 547-560, 2000.
- [29] M. E. J. Newman, C. Moore, and D. J. Watts, "Mean-field solution of the small-world network model," *Phys. Rev. Lett.*, vol. 84, pp. 3201-3204, 2000.
- [30] M. E. J. Newman, and D. J. Watts, "Renormalization group analysis of the small-world network model," *Phys. Lett. A*, vol. 263, no. 341-346, 1999.
- [31] R. Monasson, "Diffusion, localization and dispersion relations on 'small-world' lattices," *Eur. Phys. J. B*, vol. 12, pp. 555-567, 1999.
- [32] M. E. J. Newman, "The structure and function of networks," *Computer Physics Communications*, vol. 147, pp. 40-45, 2002.

- [33] R. Albert, H. Jeong, and A.-L. Barabási, “Diameter of the world-wide web,” *Nature*, vol. 401, pp. 130-131, 1999.
- [34] A.-L. Barabási, R. Albert, and H. Jeong, “Scale-free characteristics of random networks: The topology of the World Wide Web,” *Physica A*, vol. 281, no. 69-77, 2000.
- [35] A. Z. Broder, S. R. Kumar, F. Maghoul *et al.*, “Graph structure in the web,” *Computer Networks*, vol. 33, no. 1-6, pp. 309-320, 2000.
- [36] S. Redner, “How popular is your paper? An empirical study of the citation distribution,” *Eur. Phys. J. B*, vol. 4, pp. 131-134, 1998.
- [37] P. O. Seglen, “The skewness of science,” *J. Amer. Soc. Inform. Sci.*, vol. 43, pp. 628-638, 1992.
- [38] M. E. J. Newman, “Scientific collaboration networks: I. Network construction and fundamental results,” *Phy. Rev. E*, vol. 64, pp. 016131, 2001.
- [39] M. E. J. Newman, “Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality,” *Phy. Rev. E*, vol. 64, pp. 016132, 2001.
- [40] F. Liljeros, C. R. Edling, L. A. N. Amaral *et al.*, “The web of human sexual contacts,” *Nature*, vol. 411, pp. 907-908, 2001.

- [41] J. H. Jones, and M. S. Handcock, "An assessment of preferential attachment as a mechanism for human sexual network formation," *Proc Biol Sci.*, vol. 270, no. 1520, pp. 1123-1128, 2003.
- [42] H. Jeong, B. Tombor, R. Albert *et al.*, "The large-scale organization of metabolic networks," *Nature*, vol. 407, pp. 651, 2000.
- [43] H. Ebel, L.-I. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," *Phy. Rev. E*, vol. 66, pp. 035103, 2002.
- [44] W. Aiello, F. Chung, and L. Lu, "A random graph model for massive graphs," in Proceedings of the 32nd Annual ACM Symposium on Theory of Computing, New York, 2000, pp. 171-180.
- [45] W. Aiello, F. Chung, and L. Lu, "Random evolution of massive graphs," *Handbook of Massive Data Sets*, J. Abello, P. M. Pardalos and M. G. C. Resende, eds., pp. 97-122, Dordrecht: Kluwer, 2002.
- [46] H. Jeong, S. Mason, A.-L. Barabási *et al.*, "Lethality and centrality in protein networks," *Nature*, vol. 411, pp. 41-42, 2001.
- [47] R. Cohen, K. Erez, D. ben-Avraham *et al.*, "Resilience of the Internet to random breakdown," *Phys. Rev. Lett.*, vol. 85, no. 21, pp. 4626-4628, 2000.

- [48] M. Molloy, and B. Reed, “The size of the giant component of a random graph with a given degree sequence,” *Combin. Probab. Comput.*, pp. 295-305, 1998.
- [49] R. Cohen, and S. Havlin, “Scale-Free Networks Are Ultrasmall,” *Phys. Rev. Lett.*, vol. 90, no. 5, pp. 058701, 2003.
- [50] P. Echenique, J. Gómez-Gardeñes, Y. Moreno *et al.*, “Distance-d covering problems in scale-free networks with degree correlations,” *Phy. Rev. E*, vol. 71, pp. 035102, 2005.
- [51] L. K. Gallos, C. Song, and H. A. Makse, “Scaling of Degree Correlations and Its Influence on Diffusion in Scale-Free Networks,” *Phys. Rev. Lett.*, vol. 100, pp. 248701, 2008.
- [52] G. Bianconi, and A.-L. Barabási, “Competition and multiscaling in evolving networks,” *Europhys. Lett.*, vol. 54, pp. 436-442, 2001.
- [53] G. Caldarelli, A. Capocci, P. De Los Rios *et al.*, “Scale-Free Networks from Varying Vertex Intrinsic Fitness,” *Phy. Rev. Lett.*, vol. 89, pp. 258702, 2002.
- [54] D. S. Callaway, J. E. Hopcroft, J. Kleinberg *et al.*, “Are randomly grown graphs really random?,” *Phy. Rev. E*, vol. 64, pp. 041902, 2001.
- [55] M. E. J. Newman, “Mixing patterns in networks,” *Phy. Rev. E*, vol. 67, pp. 026126, 2003.

- [56] J. C. Doyle, D. L. Alderson, L. Li *et al.*, “The ‘robust yet fragile’ nature of the Internet,” *Proc. Nation. Acad. Science*, vol. 102, no. 41, pp. 14497-14502, 2005.
- [57] M. Molloy, and B. Reed, “A critical point for random graphs with a given degree sequence,” *Random Structures and Algorithms*, vol. 6, no. 161-179, 1995.
- [58] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, “Generation of uncorrelated random scale-free networks,” *Phy. Rev. E*, vol. 71, pp. 027103, 2005.
- [59] C. Magnien, M. Latapy, and J.-L. Guillaume, "Impact of random failures and attacks on passion and power-law random networks," <http://arxiv.org/abs/0908.3154>, 2009].
- [60] R. Cohen, R. Erez, D. b. Avraham *et al.*, “Reply to the comment on ‘breakdown of the internet under intentional attack’,” *Phys. Rev. Lett.* , vol. 87, 2001.
- [61] S. N. Dorogovtsev, and J. F. F. Mendes, “Comment on ‘Breakdown of the internet under intentional attack’,” *Phys. Rev. Lett.*, vol. 87, 2001.
- [62] P. Crucitti, V. Latora, and M. Marchiori, “A model for cascading failures in complex networks,” *Phy. Rev. E*, vol. 69, pp. 045104, 2004.

- [63] L. Zhao, K. Park, and Y.-C. Lai, “Attack vulnerability of scale-free networks due to cascading breakdown,” *Phy. Rev. E*, vol. 70, pp. 035101, 2004.
- [64] S. He, S. Li, and H. Ma, “Effect of edge removal on topological and functional robustness of complex networks,” *Physica A*, vol. 388, pp. 2243-2253, 2009.
- [65] G. Zhang, D. Wang, and L. G., “Enhancing the transmission efficiency by edge deletion in scale-free networks,” *Phys. Rev. E* vol. 76, pp. 017101, 2007.
- [66] P. Holme, “Edge overload breakdown in evolving networks,” *Phy. Rev. E*, vol. 66, pp. 036119, 2002.
- [67] P. Holme, B. J. Kim, C. N. Yoon *et al.*, “Attack vulnerability of complex networks,” *Phy. Rev. E*, vol. 65, pp. 056109, 2002.
- [68] R. Cohen, K. Erez, D. ben-Avraham *et al.*, “Breakdown of the Internet under intentional attack,” *Phys. Rev. Lett.*, vol. 86, no. 16, pp. 3682-3685, 2001.
- [69] E. López, R. Parshani, R. Cohen *et al.*, “Limited path percolation in complex networks,” *Phys. Rev. Lett.*, vol. 99, pp. 188701, 2007.
- [70] G. Paul, S. Sreenivasan, and H. E. Stanley, “Resilience of Complex Networks to Random Breakdown,” *Phy. Rev. E*, vol. 72, no. 5, pp. 056130.1-056130.6, 2005.

- [71] L. K. Gallos, R. Cohen, P. Argyrakis *et al.*, “Stability and topology of scale-free networks under attack and defense strategies,” *Phys. Rev. Lett.*, vol. 94, pp. 188701, 2005.
- [72] G. Paul, S. Sreenivasan, S. Havlin *et al.*, “Optimization of network robustness to random breakdowns ” *Physica A*, vol. 370, no. 2, pp. 854-862, 2006.
- [73] G. Paul, T. Tanizawa, S. Havlin *et al.*, “Optimization of robustness of complex networks,” *Euro. Phys. J. B*, vol. 38, pp. 187-191, 2004.
- [74] B. Shargel, H. Sayama, I. R. Epstein *et al.*, “Optimization of robustness and connectivity in complex network,” *Phy. Rew. Lett.*, vol. 90, no. 6, pp. 068701, 2003.
- [75] A. X. C. N. Valente, A. Sarkar, and H. A. Stone, “2-peak and 3-peak optimal complex networks,” *Phys. Rev. Lett.*, vol. 92, pp. 118702, 2004.
- [76] A. Vázquez, and Y. Moreno, “Resilience to damage of graphs with degree correlations,” *Phys. Rev. E*, vol. 67, pp. 015101, 2003.
- [77] O. Giraud, B. Georgeot, and D. L. Shepelyansky, “Delocalization transition for the Google matrix,” *Phys. Rev. E*, vol. 80, pp. 029107, 2009.

- [78] S. Chakrabarti, M. v. d. Berg, and B. Dom, "Focused Crawling: A New Approach to Topic-specific Web Resource Discovery," *Computer Networks*, vol. 31, no. 11-16, pp. 1623-1640, 1999.
- [79] M. Diligenti, F. Coetzee, S. Lawrence *et al.*, "Focused Crawling Using Context Graphs," in Proceedings of the 26th International Conference on Very Large Database (VLDB), Cairo, Egypt, 2000, pp. 527-534.
- [80] C. Y. Chong, and S. P. Kumar, "Sensor Networks: Evolution, Opportunities, and Challenges," *Proc. IEEE* vol. 91, no. 8, pp. 1247-1256, 2003.
- [81] U. N. Raghavan, H. P. Thadakamalla, and S. R. T. Kumara, "Phase Transitions and Connectivity in Distributed Wireless Sensors Networks," in Proceedings of the Thirteenth International Conference on Advanced Computing & Communications (ADCOM), Amrita Vishwa Vidyapeetham, India, 2005, pp. 10-15.
- [82] H. P. Thadakamalla, U. N. Raghavan, S. R. T. Kumara *et al.*, "Survivability of Multiagent-Based Supply Networks: A Topological Perspective," *IEEE Intell. Syst.*, vol. 19, no. 5, pp. 24-31, 2004.
- [83] V. Kalogeraki, D. Gunopulos, and D. Zeinalipour-Yazti, "A Local Search Mechanism for Peer-to-Peer Networks," in Proceedings of the 11th international Conference on Information and Knowledge Management (CIKM), McLean, Virginia, USA., 2002, pp. 303-307.

- [84] B. Yang, and H. Garcia-Molina, "Improving Search in Peer-to-Peer Networks," in Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS), Vienna, Austria, 2002, pp. 5-14.
- [85] C. Lv, P. Cao, E. Cohen *et al.*, "Search and Replication in Unstructured Peer-to-Peer Networks," in Proceedings of the 16th annual ACM International Conference on supercomputing (ICS), New York, NY, USA, 2002, pp. 84-95.
- [86] A. Crespo, and H. Garcia-Molina, "Routing Indices for Peer-to-Peer Systems," in Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS), Vienna, Austria, 2002, pp. 23-32.
- [87] D. Tsoumakos, and N. Roussopoulos, *Adaptive Probabilistic Search (APS) for Peer-to-Peer Networks*, University of Maryland, 2003.
- [88] J. Travers, and S. Milgram, "An Experimental Study of the Small World Problem," *Sociometry*, vol. 32, no. 4, pp. 425-443, 1969.
- [89] O. Simsek, and D. Jensen, "Decentralized Search in Networks Using Homophily and Degree Disparity," in Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI), Edinburgh, Scotland, UK, 2005, pp. 304-310.
- [90] D. J. Watts, P. S. Dodds, and M. E. J. Newman, "Identity and Search in Social Networks," *Science*, vol. 296, no. 5571, pp. 1302-1305, 2002.

- [91] H. P. Thadakamalla, R. Albert, and S. R. T. Kumara, "Search in Weighted Complex Networks," *Phy. Rev. E*, vol. 72, pp. 066128, 2005.
- [92] J. Kleinberg, "Navigation in a small world," *Nature*, vol. 406, no. 6798, pp. 845, 2000.
- [93] J. Kleinberg, "The Small-World Phenomena: An Algorithmic Perspective," in Proceedings of the ACM Symposium on Theory of Computing, Portland, Oregon, USA, 2000, pp. 163-170.
- [94] A. Clauset, M. E. J. Newman, and C. Moore, "Finding Community Structure in Very Large Networks," *Phys. Rev. E*, vol. 70, pp. 066111, 2004.
- [95] G. W. Flake, S. Lawrence, and C. L. Giles, "Efficient Identification of Web Communities," in Proceedings of the ACM Conference on Knowledge and Data Discovery (KDD), Boston, MA, USA, 2000, pp. 150-160.
- [96] M. Boguñá, D. Krioukov, and K. Claffy, "Navigability of Complex Networks," *Nature Physics*, vol. 5, pp. 74-80, 2009.
- [97] <http://moat.nlanr.net/Routing/rawdata>.
- [98] http://www.caida.org/tools/measurement/skitter/router_topology/.