A SIMPLE MODEL FOR CASCADING FAILURES IN SCALE-FREE NETWORKS

V. H. Pereira*, P. S. Martins*, A. F. de Angelis*, V. S. Timóteo*

*School of Technology, University of Campinas - UNICAMP
13484-332 - Limeira, São Paulo, Brazil

Emails: `<vanessa, pmartins, andre, varese>@ft.unicamp.br`

Abstract— Cascading failures are very disruptive events that can damage critical infrastructure such as electrical and telecommunication networks. They start with a single-point failure and spread fast through the networks, sometimes causing the collapse of the whole system. We study and classify three containment strategies for cascading failures. Additionally, we propose a simple dynamic model for cascading failures in scale-free networks considering random failures and targeted attack scenarios. Our results allowed us to mathematically describe the behavior of the process and to test three strategies to protect the network against the cascade.

Keywords— complex networks; scale-free networks, cascading failures; network attacks

1 Introduction

Cascading failures (CFs) are a severe problem in systems like power grids, telecommunications systems and other critical infrastructure networks. General solutions to this problem are not available and thus more research is still required. Such sort of investigation is mainly based in mathematical models, since it is not possible to execute trials in the real-world without incurring the high risk of injuries, damages, and several other prejudicial effects.

Scale-free Networks (SFNs) (i.e. networks that follow a power-law degree distribution) are a well suited representation of a range of systems in several distinct fields such as biology, chemistry, physics and technology. For this reason, SFNs have been the subject of intense research efforts and there are many models to generate their ensembles.

In this work, the SFN Barabasi-Albert model (Barabási and Albert, 1999) was chosen to generate the network topology. A simple cascade spreading model (SCSM) was used to describe the failure process across the network. The SCSM is a probabilistic model that acquires information about node degrees in order to spread the failure.

To investigate the cascading process, we generated a set of SFNs realizations and these ensembles were used to simulate the propagation of the failure. Once created, the network is no longer modified: nodes and links are not added or removed. The SCSM defines the behavior of the failure in the following way: if a node downs, then the failure propagates to the neighbors of this node according to a global parameter of vulnerability applied on a node-degree basis. A failed node is not repaired and remains dead for the rest of the simulation. As the topology of the network is never destroyed or changed, the metric used to assess the spreading of the cascade and the efficacy of the containment approaches is the *number of survivor nodes*.

Two simulations scenarios were addressed in this work: 1) accidental failures, where the nodes fail at random, and 2) hub (or targeted) attacks, where an external malicious agent chooses the most connected node to down. It is well known that SFNs are more resilient to random problems than to hub attacks and the results we have obtained corroborate this fact. The main result of this work is the study and classification of three containment strategies for cascading failures, and the analytical modeling of the cascade itself. From a mathematical perspective, we noted an interesting similarity between the behavior of the model for both random failures and targeted attack scenarios.

The remainder of this paper is organized as follows: Section 2 describes the main concepts and the computational model. In Section 3 we review recent related work; Section 4 shows the main aspects of the simulation framework and the results. Finally, in Section 5 we present our main conclusions.

2 Computation Model

The cascading model employed in this work is the one based on the work of Lewis (Lewis, 2006) and embodied by the *Critical Infrastructure Protection - Attacker Defender (AD)* software, which was also adopted as a simulation tool in this work (Lewis, 2013).

We consider a scale-free network subject to malicious attacks and modeled as a directed graph $G(V, E)$ defined by the sets V and E, where V is a finite set of vertices and E a finite set of edges. An edge e_{ij} connects node i with node j. A cascade can be contained by specific strategies or containment mechanisms. Attacks can be random or directed to some topological features of the network.

V. Timóteo and P. Martins are with Grupo de Ótica e Modelagem Numérica (GOMNI), UNICAMP

DOI: 10.5540/03.2013.001.01.0182

Each node has a vulnerability v, which measures the probability of that node spreading the failure to other elements. Following an attack, a number of nodes are deemed to be survival nodes. The remaining nodes are dead nodes, i.e. they have lost their processing capacity. We assume that the attackers know the topology of the network.

The network is subject to a disruptive process characterized by an initial fail in a single point somewhere in the network that scatters to neighbors and starts a number of new cycles of fail and spreading. Such process it also known as *avalanche* and it is found in several domains. From a social perspective, rumor dispersion and collective opinion changes can be studied as CFs. The spreading of diseases in a population, information in a system, or malicious pieces of software in a computer network can be traced by CFs templates (Lewis, 2006) (Lewis, 2009).

All the nodes have a fixed processing capacity C (equation (1)). The *capacity of a node* is a measure of its ability to handle the traffic load, proportional to its initial computational load. The Capacity is allocated based on the following equation:

$$C_{A(i)} = \left(1 + \alpha \frac{B_i}{\lambda N D + \lambda}\right) L_i \tag{1}$$

where α is the tolerance factor, L_i is the load on the node, λ is the average traffic generation rate, N is the network size, and D is the average size of the minimum path.

Each node j in the network has a capacity threshold, which is the maximum flow that the edge can transmit. Since the node capacity on real-life networks is generally limited by cost, it is natural to assume (for simplicity) that the capacity C_j of the node j is proportional to its initial load: C_j = T * L_j, j =1,2, 3,...N, where constant T (≥ 1) is the tolerance parameter that describes the network tolerance. As each node has a limited capacity to handle the load, if L + ΔL_j > C for node j, then node j crashes and it further induces the redistribution of the additional load - what may lead other nodes to a breakdown.

In addition, the following parameters are addressed in this work:

- *Betweenness centrality of a node*: The betweenness is a centrality measure of a vertex within a graph. The betweeness of a node is defined as:

$$B_i = \sum_{j,l \epsilon N, j \neq l} \frac{n_{jl}(i)}{n_{jl}} \tag{2}$$

  where n_{jl} is the total number of shortest paths from node j to l and $n_{jl}(i)$ is the number of those paths that pass through i.

- *Local Cluster Coefficient*: This is a measure of degree to which nodes tend to cluster together. It is given by:

$$C_i = \frac{2|e_{jk}|}{k_i(k_i - 1)} : v_j, v_k \ \epsilon \ N_i, e_{jk} \ \epsilon \ E \tag{3}$$

  and the average cluster coefficient is given by:

$$C = \frac{1}{n} \sum_{i=1}^{n} C_i \tag{4}$$

- *The spectral radius $r(G)$*: The spectral radius is measured from the adjacency matrix of a graph. It is the largest non-trivial eigenvalue of det[A(G)-λI]=0, where A is the adjacency matrix and I is the identity matrix. The eigenvalues are the diagonals (i.e. λ_1, λ_2, λ_3, ... λ_n) of λI.

We now present an overview of recent work on cascading failures in scale-free networks.

3 Related Work

Sun *et al.* have proposed the strategic allocation of capacity (amount of traffic allowed by a node) in the nodes of a network, to gain robustness and reduce the size of a cascading failure in the network (Sun et al., 2008). Wu *et al.* have also considered node capacity in their work. They simulated the CFs in scale-free networks with community structure, in which local and global level obey the power-laws. They have focused on the study of the cascade at different removal strategies to understand the influence of the avalanches in such networks. These results suggest that modularity and large coefficients of reserve capacity are necessary to avoid CFs in community structures (Wu et al., 2006).

Zhao and Xu have showed how to increase the robustness of a network adding new links between the nodes of low degree. They noted that because of the loop formed between these nodes, the network maintains its operation even after failure of a node of high degree (Zhao and Xu, 2009).

In this work we simulate, combine, compare and rank these previous approaches, i.e. the addition of links and capacity, under the model and context described in Sections 2 and 4 respectively.

4 Simulation

The goal of this simulation is to assess the effectiveness of a set of defense (i.e. containment) strategies against network attacks. The simulation model is illustrated in Fig. 1. We systematically expose the network to different types of attacks. The performance metric we elect to assess the the containment approach is the percentage of survival nodes S resulting from an attack. Clearly,

DOI: 10.5540/03.2013.001.01.0182

the larger the number of survival nodes, the more successful is the strategy under consideration.

Regarding the node-selection policy, our simulations consider two types of attacks: 1) *Random attacks*, where a random network node (with a low degree) is selected as the target; 2) *Hub attacks*, where only nodes with the highest degrees (i.e. hubs) are targeted instead. In any case, once a node is chosen it is subject to the attack and the failure spreads across the network according to the vulnerability v of the affected nodes.

Two sets of simulations were considered:

- *Simulation 1*: The following network containment approaches were evaluated in terms of the percentage number (%) of survival nodes (S): 1) No containment mechanisms, 2) Increasing the number of links between nodes with a lower degree, 3) Increasing the node capacity and 4) Combined approach, where we increase the number of links and the node capacity simultaneously. Networks with various sizes (i.e. 10,20..100 nodes) were subject to attacks and examined. We considered a fixed vulnerability for all nodes ($v = 0.5$).

- *Simulation 2*: No containment approaches were in place. Instead, for each type of attack (i.e. hub or random), we performed simulations measuring the number of survival nodes while varying the size of the network from 10 to 1000 nodes (i.e. 10,20,70,100,200,500,1000). Additionally, for each network size, we varied the node vulnerability from 0.1 to 1. For example, if a node affected by the failure has four connections and its vulnerability is 0.5, the fault is transmitted to two connections, meaning that two adjacent nodes are affected.

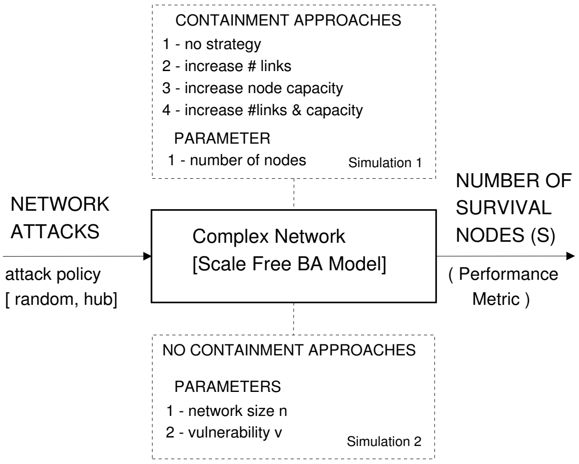

Figure 1: Simulation model.

We now turn to the analysis of the results.

4.1 Results: Classification of Containment Strategies

4.1.1 Simulation 1

Figs. 2 and 3 illustrate the percentage of survival nodes S as a function of the number of network nodes for each strategy employed (simulation 1). Based on these results, we can rank the four containment approaches according to the selected performance criteria (i.e percentage of survival nodes S) as follows:

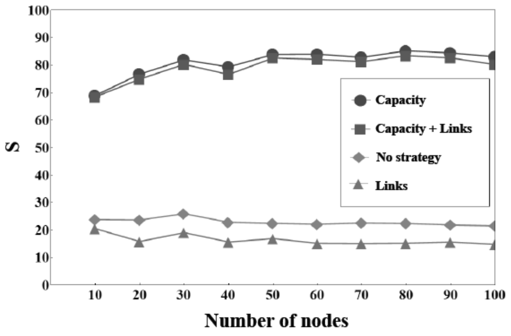

Figure 2: Percentage of Survival Nodes for Hub Attacks (simulation 1)

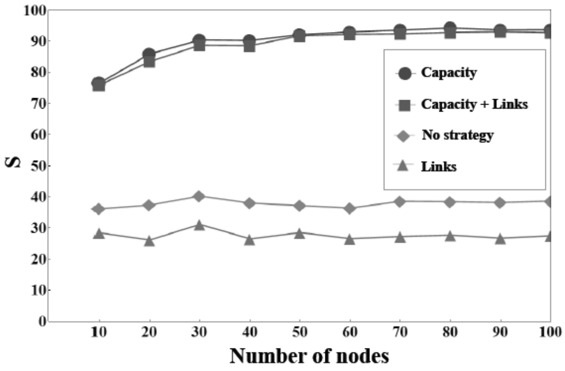

Figure 3: Percentage of Survival Nodes for Random Attacks (simulation 1)

1. *Increase in network capacity.* By far this is the best approach of all alternatives considered, since it provides the largest number of survival nodes for all types of attacks.

2. *Combined approach.* The simultaneous addition of links and capacity in the network comes as a second alternative approach, i.e. whereas it increases the number of survival nodes in comparison with the no-containment strategy, it does not outperform the exclusive increase in network capacity.

3. *No containment strategy.* Simply leaving the network as it is, i.e. not changing the number of links or capacity, does not improve on the results achieved with the alternatives above.

However, it does present better results than the increase in the number of links.

4. *Increase in the number of new links.* This strategy is outperformed by the previous one, since it reduces the number of survivors in all scenarios examined. This may be a consequence of increased spectral radius of the networks after the addition of new links. In general, the increase of the spectral radius has implied in a reduction of the robustness of the network, and thus an increase in their vulnerability to attacks. More connections enhance the persistence of the spreading failure, which cross the network epidemically with one rate of infection.

Regarding the node selection policy, it is clear that random attacks yield a larger percentage of unaffected nodes (in all sizes) compared with hub attacks. This is due to the random nature of the attack, which affects a smaller portion of the network. Random attacks also present a relatively smoother decrease in the number of survivor nodes compared to the attack on hubs, which increase the likelihood of spreading the failure.

In addition to the measurement of survival nodes, we have also collected information on the structural impact caused by the addition of new links (Table 1):

Table 1: Simulation 1: Structural Impact by the addition of links (B=before, A=after)

N	Links		Betweeness		Cluster Coefficient		Spectral Radius	
–	B	A	B	A	B	A	B	A
10	17	18	21	18.5	0,46	0,40	3,91	4,04
20	37	42	37	62	0,50	0,31	4,92	5,16
30	55	61	57	104	0,28	0,19	4,94	5,10
40	76	87	82	148	0,33	0,26	5,92	6,11
50	96	105	44	109,5	0,06	0,06	5,53	5,62
60	117	135	116	345	0,11	0,08	6,01	6,14
70	135	152	198	310	0,10	0,05	6,27	6,41
80	157	176	201	435	0,10	0,07	6,28	6,41
90	176	196	190	368	0,07	0,06	6,61	6,73
100	197	223	200	376	0,18	0,13	7,1	7,25

- *Betweenness:* The maximum centrality increases in all cases, due to the increase of the number of possible paths between nodes. The exception is for the network of 10 nodes, due to the addition of one single link.

- *Cluster Coefficient:* The value of the cluster coefficient reduced in all networks subject to the addition of links. As mentioned in Section 2, this measure indicates how grouped are the neighbors for each node.

- *Spectral Radius:* The spectral radius increases with the addition of new links in all cases. This means a reduction in the network's robustness, and therefore an increase in network vulnerability. More connections imply in an increase of the likelihood of spreading the fault epidemically.

4.1.2 Simulation 2

Fig. 4 shows the percentage of survivor nodes as functions of the vulnerability, for random and hub attacks. Notice that the random attacks are less damaging to the network than the hub attacks, as they leave the network with a larger number of survival nodes. It becomes clear from the consequence graphs the impact of the vulnerability on the network, where it is shown that any value larger than 0.6 imply a quite small percentage of survival nodes. This is due to the increased spreading of the failure, which strengthens the cascading effect. It is also worth pointing out that the hub attacks affect almost all the network (resulting in no survival nodes), whereas the random attacks leave some nodes intact. These nodes survived probably because they were completely isolated from the rest of the network, i.e. their neighboring nodes were downed before the failures could propagate any further and eventually reach them.

4.2 Results: Modeling the Behavior of the Cascading Failure

Finding the functions that define the cascade is crucial for understanding, forecasting and analyzing their behavior. For the analysis of the aforementioned cascade, we observed the graphs in logarithmic scale (Fig. 4), which allowed us to find an equation that models them. We considered the possibility of obtaining this equation using an exponential function. However, each type of failure requires a model of exponential function that fits the curve.

An inspection of the graphs of the survivor nodes (log scale) with the vulnerability suggests a simple mathematical model to describe the behavior of survival nodes. For the attacks on hub nodes, we propose the following model for the number of survival nodes S:

$$S(v)_{\text{hub}} = a\left(1 - e^{bv^2}\right) \tag{5}$$

and for the failures caused by random attacks we use:

$$S(v)_{\text{random}} = a\left(1 - e^{bv^c}\right) \tag{6}$$

where a, b and c are parameters to be determined as a function of both the size of the network and the type of attack.

In order to find the values of the parameters a, b and c, we have fitted the simulation results with our models using a least-squares algorithm. The results are displayed in Tables 2 and 3 for each type of attack.

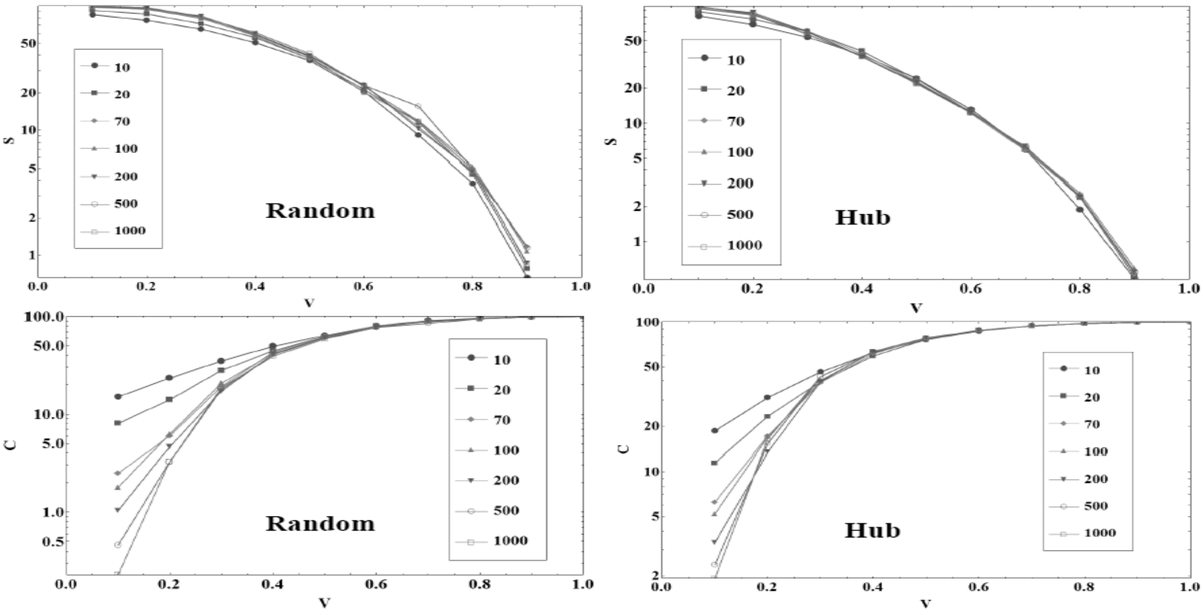

Figure 4: Percentage of survivors (top panels) and consequence (bottom panels) as functions of the vulnerability for random (left) and hub (right) attacks for different number of network nodes (simulation 2).

Table 2: Model parameters for Hub attacks. In this case $\mathbf{c} = 2$.

Nodes	parameter **a**	parameter **b**
10	4.45141	1.17071
20	4.55630	1.18756
70	4.62180	1.27201
100	4.63104	1.31698
200	4.65923	1.33271
500	4,65369	1.35289
1000	4.65175	1.35710

Table 3: Model parameters for Random attacks.

Nodes	parameter **a**	parameter **b**	parameter **c**
10	4.44183	1.17263	2.51885
20	4.53832	1.17263	2.52544
70	4.61326	1.26296	2.60895
100	4.61499	1.25503	2.59426
200	4.62226	1.31230	2.70485
500	4.64025	1.14506	2.51106
1000	4.63083	1.25573	2.65949

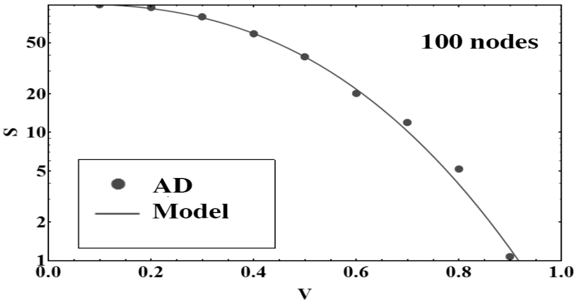

Figure 5: Percentage of survivors as a function of the vulnerability for random attacks.

The curves of the original failure and the curves of the exponential function were plotted on the same graph: in Figs. 5 and 6 we show the fitting of the survivability curve to the data points, considering 100 nodes under random and hub attacks respectively. The circles represent the simulations with the AttackerDefender (Lewis, 2006) and the line is the fit with our model. The results are very similar when considering other number of network nodes.

Figs. 7 and 8 illustrate the curve fitting for the parameter Consequence (C, where C = 100 - S). The failure has a behavior that is close to the exponential function parameterized with variable values reasonable for each failure scenario and each network size.

One interesting result is that the values of the parameters a and b are approximately independent of the number of nodes and they are about the same for both Hub and Random attacks provided $c \sim 2.5$ for the Random case. This result makes the simple model even more general for the prediction of the survivors upon cascading failures.

Analyzing the attack on the hubs, where we plot the nodes in the graph versus vulnerabilities, it is observed that as the vulnerability and likelihood of spreading increases, the percentage of survivors tends to be equal for all sizes of networks.

The results for random failures can be compared to the results for hub attacks: as expected, the number of survivors is always larger in random failures than the number of survivors in hub

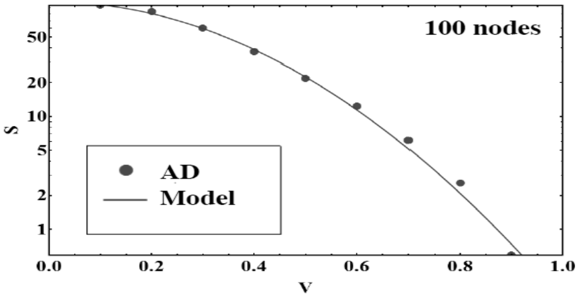

Figure 6: Percentage of survivors as a function of the vulnerability for hub attacks.

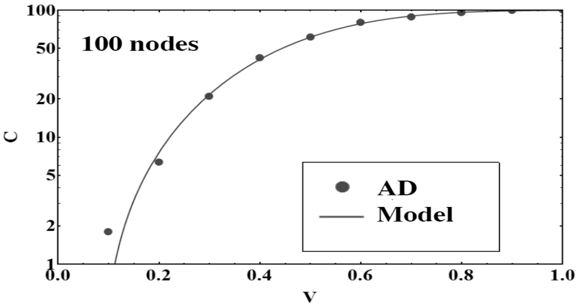

Figure 7: Percentage consequence as a function of the vulnerability for random attacks.

(i.e. target) failures, for all network sizes. This is explained by the fact that hub attacks have the ability to reach more nodes, since the hubs have more connections.

5 Summary and Conclusions

The understanding of the behavior of failures in a scale-free complex network that is subject to intentional attacks is critical to allow for preparedness in face of real-world attacks.

In this work we have analyzed the impact of both random and hub attacks on scale-free networks subject to cascading failures. In addition, we have also mathematically modeled the behavior of the cascade and observed the effectiveness of defensive strategies against these attacks.

This work also paves the way for future research on complex networks in the presence of failures. In particular, an interesting proposal considers the analysis of other containment strategies as well as other parameters in the model, such as the cost function of resources (nodes and links), and the propagation speed of faults across the network.

Although we have primarily considered intentional, malicious attacks on the network infrastructure, we believe that the results presented in this work are by far more general and can be used to analyze and describe other scenarios such as the behavior of regular failures in complex networks. These results may be helpful to protect real-life networks and avoid cascading-failure-induced disasters.

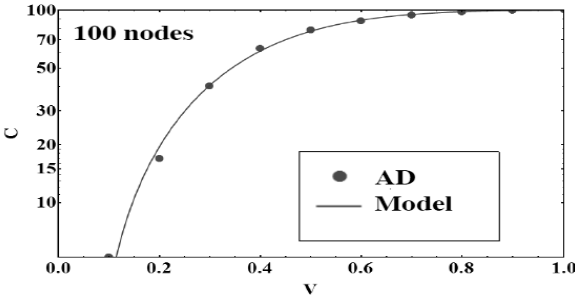

Figure 8: Percentage consequence as a function of the vulnerability for hub attacks and different number of network nodes.

References

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks, *Science* **286**: 509–512. DOI: 10.1126/science.286.5439.509

Lewis, T. (2006). *Critical Infrastructure Protection in Homeland Security: Defending a Net-worked Nation*, // John Wiley and Sons. DOI: 10.1002/0471789542

Lewis, T. (2013). Critical infrastructure protection software, `http://www.chds.us`.

Lewis, T. G. (2009). *Network Science: Theory and Applications*, John Wiley and Sons. DOI: 10.1002/9780470400791

Sun, H., Zhao, H. and Wu, J. (2008). A robust matching model of capacity to defense cascading failure on complex networks, *Physica A: Statistical Mechanics and its Applications* **387**(25): 6431 – 6435.

Wu, J.-j., Gao, Z.-y. and Sun, H.-j. (2006). Cascade and breakdown in scale-free networks with community structure, *Phys. Rev. E* **74**: 066111. DOI: 10.1103/PhysRevE.74.066111

Zhao, J. and Xu, K. (2009). Enhancing the robustness of scale-free networks, *Journal of Physics A: Mathematical and Theoretical* **42**: 195003. DOI: 10.1088/1751-8113/42/19/195003