Comparison of Failures and Attacks on Random and Scale-Free Networks

Jean-Loup Guillaume, Matthieu Latapy, Clémence Magnien

Comparison of Failures and Attacks on Random and Scale-free Networks

Jean-Loup Guillaume[1], Matthieu Latapy[1], and Clémence Magnien[2]

[1] LIAFA – CNRS – Université Paris 7 – 2 place Jussieu, 75251 Paris Cedex 05, France. Fax : 33 (0)1 44 27 68 49. (guillaume,latapy)@liafa.jussieu.fr
[2] CREA – CNRS – École Polytechnique – 1, rue Descartes, 75005 Paris, France. Fax : 33 (0)1 55 55 90 40. magnien@shs.polytechnique.fr

Abstract. It appeared recently that some statistical properties of complex networks like the Internet, the World Wide Web or Peer-to-Peer systems have an important influence on their resilience to failures and attacks. In particular, scale-free networks (*i.e.* networks with power-law degree distribution) seem much more robust than random networks in case of failures, while they are more sensitive to attacks.
In this paper we deepen the study of the differences in the behavior of these two kinds of networks when facing failures or attacks. We moderate the general affirmation that scale-free networks are much more sensitive than random networks to attacks by showing that the number of links to remove in both cases is similar, and by showing that a slightly modified scenario for failures gives results similar to the ones for attacks. We also propose and analyze an efficient attack strategy against links.

Keywords. Internet, Complex Networks, Random Graphs, Scale-Free Graphs, Resilience, Fault tolerence, Reliability, Network Topology

Introduction

In a random network [1,2] with n nodes, each of the $\frac{n\cdot(n-1)}{2}$ possible links exists with a given probability p. In other words, a random network is constructed from n nodes by choosing $m = p \cdot \frac{n\cdot(n-1)}{2}$ links at random. In such a network, the degree distribution p_k follows a Poisson law: $p_k = e^{-z}\frac{z^k}{k!}$ where z is the average degree. Intuitively, such a distribution means that most nodes have a degree close to the average, and that the number of nodes with a given degree decays exponentially fast away from the mean degree.

However, it has been shown recently that most real-world complex networks [3,4,5,6,7,8,9], in particular the Internet [10], the World Wide Web [7,11] and Peer-to-Peer systems [12], have a power-law degree distribution: $p_k \sim k^{-\alpha}$. In the cases we have cited, α is close to 2.5. Intuitively, such a distribution means that, despite most nodes have a low degree, the number of nodes with (very) high degree is not negligible.

Since this difference between random networks and real-world complex networks has been discovered, a strong effort has been put on the understanding

of its consequences. One of the most famous is that it significantly influences the robustness of networks [7,13,14,15,16,17], which can be observed as follows. Given a network, one can model a series of failures by a random removal of nodes (or links), whereas an attack is modeled by the targeted removal of a series of chosen nodes (or links). The way the nodes (or links) are chosen during an attack is called an *attack strategy*. The quality of the service provided by the network under consideration can be roughly evaluated by the size of its largest connected component (*i.e.* the number of machines which can communicate in the Internet, for instance). The resilience of the network to failures or attacks can then be analyzed by studying how the size of the largest connected component varies as a function of the number of removed nodes (or links). In particular, the network is said to have a *giant connected component* if it has a component of size linear with respect to the size of the network. In other words, a constant proportion (with respect to the network size) of the whole network is connected. Other criteria for measuring network efficiency have been proposed, see for instance [16,17,18,19].

The most widely studied attack strategy has been introduced independently in [7] and [13]. It consists in removing nodes by decreasing order of their degree. We will refer to this attack as the *classical* attack strategy. The effects of this attack strategy are plotted in Figure 1, together with the effect of failures.

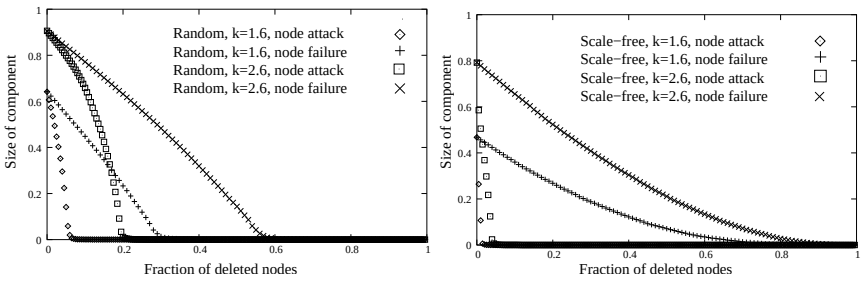

Fig. 1. Effects of random failures and attacks on random networks (left) and scale-free networks (right). The plots represent the size of the largest connected component as a function of the fraction of removed nodes. Different values of the mean degree k are considered.

From these experiments the following observations can be derived [7,13]. First, there is a qualitative difference in the behavior of random and scale-free networks in case of failures: for random networks, the size of the largest connected component drops to zero when a finite fraction of the nodes are removed (this fraction represents a *threshod value*), whereas for scale-free networks, it decreases very slowly, and reaches 0 only when most nodes have been removed. Thus scale-free networks appear to be much more resilient to failures than random networks. However, the opposite seems true for attacks: scale-free networks collapse much more quickly than random networks. The power-law distribution of degrees in the

Internet, which might therefore make it very resilient to failures but extremely sensitive to attacks, has even been called the *Achilles's heel of the Internet* [20].

Although attacks remove a very large fraction of the links, we show in Section 1 that this is not sufficient to explain the qualitative difference between failures and attacks for scale-free networks. We then investigate further this difference (Section 2) by proposing two new attack strategies, one against nodes and the other against links, and comparing their effects with those of the classical attacks and failures.

Before entering in the core of the paper, let us say a word on our plots. The plots for experimental results obtained by simulation are the average of simulations over a large number of samples. This is in general representative of the mean behavior, but it must be noted that the actual simulation result obtained on one instance may be significantly different in some cases (in particular in what concerns threshold values for scale-free networks).

Concerning the thresholds values, we considered that the threshold was reached whenever the size of the largest connected component of the network becomes smaller than 5% of the whole. The plots representing the thresholds are in function of the mean degree for random networks, and the degree exponent for scale-free networks, which are the main parameters in these contexts. A scale-free network is connected if $\alpha \leq 3.48$, therefore we will not be interested in the case where α is greater than this value.

For plots comparing the effect of different failures and attacks for random and scale-free networks, we have chosen to compare networks with the same average degree. The values we have chosen are 1.6 and 2.6, which corresponds to scale-free networks with exponents 3 and 2.5 respectively, representative of the values met in practice.

In several cases, we plot numerical evaluations for approximation formulæ. These formulæ have often been obtained under the continuous degree assumption. Because in our experimentations the degree is by essence discrete, empiric values may be quite different from the approximation values, which should therefore be taken as indicative.

All scale-free networks have been generated using the algorithm for obtaining networks with a prescribed degree distribution described in [21]. We have generated scale-free networks with N nodes and exponent α by drawing N degrees between 1 and N, following a power-law with exponent α. Then pairs of stubs are randomly connected. Some proofs in the following use the fact that links are pairs of randomly chosen stubs.

We also need to introduce a few notations: $\zeta(\alpha)$ is the Riemann ζ function, defined by $\zeta(\alpha) = \sum_{k=1}^{\infty} k^{-\alpha}$. the K-th harmonic number, denoted by $H_K^{(\alpha)}$, is equal to $H_K^{(\alpha)} = \sum_{k=0}^{K} k^{-\alpha}$. Finally, given a degree distribution p_k, we denote by $\langle k \rangle$ and $\langle k^2 \rangle$ the mean of the degree and the square degree respectively: $\langle k \rangle = \sum_{k=0}^{\infty} k p_k$ and $\langle k^2 \rangle = \sum_{k=0}^{\infty} k^2 p_k$.

1 The links point of view

The classical attack strategy removes high-degree nodes first. Since in a scale-free network there is a high heterogeneity between nodes, highest degree nodes have a very large number of links attached to them. Therefore, one may wonder if the efficiency of attacks on these networks is a consequence of the fact that the number of *links* removed is much larger than in the case of failures. Likewise, one may wonder if the fact that the attack results in the removal of much more links in a scale-free network than in a random one is the cause of the difference between the two. These explanations actually have been proposed by some authors to give an intuitive explanation of the results presented above.

The aim of this section is to evaluate these ideas by the study of classical attacks under the links point of view. Indeed, the classical attack strategy can be viewed as a strategy targeting links, where links adjacent to high degree nodes are removed first. Then, the size of the giant component can be plotted as a function of the number of removed links, see Figure 2. In this figure, the behavior of these networks under random link removal, *i.e. link* failure, is also plotted as a comparison.

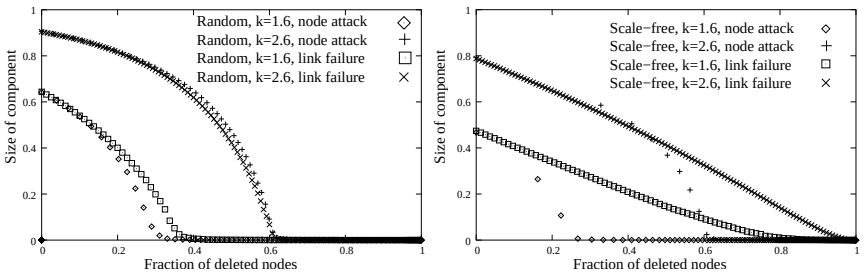

Fig. 2. The effects of the classical node attack when considering links, and of link failure, for random networks (left) and scale-free networks (right).

At first glance, this link attack strategy seems much more efficient than random removal. This can be confirmed formally with the same kind of arguments that have been developed in [14,15]. From this, one obtains that the threshold m_c of links that have to be randomly removed to break the network is:

$$m_c = 1 - \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle}$$

This result can be obtained by the following reasoning: when links are removed, this changes the degree distribution of the network. The new degree distribution can be explicitely computed. Since links are removed at random, the network is a random network with the new degree distribution. There exists a criterion [21]

for deciding if such a network has a giant component or not, and the above formula is obtained from the application of this criterion to the new degree distribution of the network.

It turns out that this quantity is the same as the threshold p_c for *nodes* failure [14,22]. This means in particular that link failures do not make scale-free networks collapse. Therefore, the fact that a scale-free network collapses using the classical attack means that the efficiency of this attack strategy is *not* due to the fact that it removes many links. If the same number of links are removed randomly, then the network does not collapse.

Let us now try to evaluate precisely the efficiency of this link attack. The fraction m_c of links that must be removed to break the network can be computed in the same manner as what has been done in [14,22] for the number of nodes. For any network, the fraction $m(p_c)$ of links removed in an attack is equal to $s(p_c)^2+2s(p_c)(1-s(p_c))$, where $s(p_c)$ represents the number of stubs (links' end-points) attached to the removed nodes. $s(p_c)$ can be evaluated by the following equations [15,14].

For scale-free networks:

$$s(p_c) - 2 = \frac{2-\alpha}{3-\alpha}\left(s(p_c)^{(3-\alpha)/(2-\alpha)} - 1\right), \tag{1}$$

or

$$s(p_c) = 1 - \frac{H_{K_c-1}^{(\alpha-1)}}{\zeta(\alpha-1)}, \text{ with } K_c \text{ satisfying } H_{K_c}^{(\alpha-2)} - H_{K_c}^{(\alpha-1)} = \zeta(\alpha-1). \tag{2}$$

For random networks:

$$s(p_c) = \sum_{k=K_c+1}^{\infty} \frac{k \cdot p_k}{z}, \text{ with } K_c \text{ satisfying } \sum_{k=0}^{K_c} k^2 \cdot p_k - \sum_{k=0}^{K_c} k \cdot p_k = z \tag{3}$$

These values are plotted in Figure 3, as well as some experimental values for the thresholds. We enter here in the details of the computation of the theoretical value of the threshold for random networks, obtained by solving Equation 3. Solving this equation gives the value of $K(p)$, the maximal degree in the network after the attack, in function of the mean degree z of the network. By definition, $K(p)$ can only take integer values. But since, in random networks, the degrees of the nodes are all gathered in a small set of values around n, it is not always possible to obtain values of $K(p)$ that statisfy exactly the equation. We have chosen the points obtained at the values of z that yield the least error, the other values of z forbidding any accurate computation of the theoretical threshold. It is nonetheless interesting to observe that the experimental values for the threshold follow the curve that is suggested by these few theoretical dots.

We can now conclude precisely on the efficiency of the classical attack strategy. First, althoug the number of links removed during such an attack on scale-free networks is huge, it is not sufficient to explain the collapse of the network:

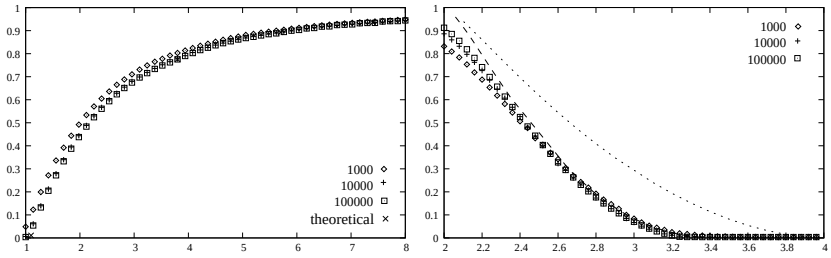

Fig. 3. Experimental values of the critical fraction $m(p_c)$ of links that must be removed in a classical node attack to disconnect random networks as a function of the mean degree (left), and scale-free networks as a function of the degree exponent (right). We have represented theoretical and experimental values. For scale-free networks, the values obtained from Equation 1 (dotted line) and from Equation 2 (dashed line) are plotted.

if the same number of links is randomly removed, then the network does not collapse. However, the number of removed links during a classical attack of a random network and of a scale-free network are very similar, for the values of the mean degree we are interested in. This moderates the conclusion that scale-free networks are particularly sensitive to classical attacks: in terms of links, they are as robust as random networks.

2 New attack strategies

In [21] a criterion for a network to almost surely have a giant component is given:

$$\langle k^2 \rangle - 2\langle k \rangle > 0 \iff p_1 < \sum_{k=3}^{\infty} k(k-2)p_k$$

The key point is therefore the proportion of nodes of degree 1 in the network. Therefore, it seems that any strategy aiming at increasing this proportion should quickly break the network. Using this remark, we propose two new attack strategies (one against nodes and the other against links) which give more insight on the actual efficiency of classical attacks.

2.1 Almost-failures attack

Our first attack strategy simply consists in randomly removing nodes of degree at least 2. This decreases the number of nodes of degree higher than 1 and increases the number of nodes of degree 0 or 1. The effect of this attack is shown in Figure 4.

Notice that this attack is barely different from node failure, and yet it is much more efficient. It actually is qualitatively different from failures, since it displays a threshold.

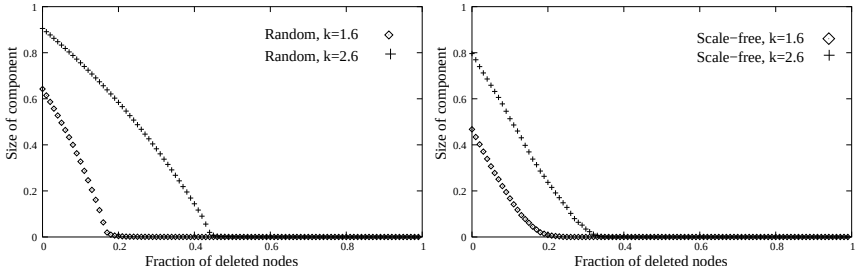

Fig. 4. The effect of the new node attack strategy on random networks (left) and scale-free networks (right).

We can easily prove this by providing an upper bound for this threshold: when all nodes that had initially a degree higher that one have been removed, then the network surely does not have a giant component anymore, since all nodes have degree at most 1. Therefore the giant component is destroyed when a fraction $1 - p_1 - p_0$ of the nodes has been removed.

For scale-free networks with exponent α, this quantity is equal to $1 - 1/\zeta(\alpha)$. For random networks with mean degree z, it is equal to $1 - e^{-z}(z+1)$. The plots for these quantities are shown in Figure 5, together with experimental values.

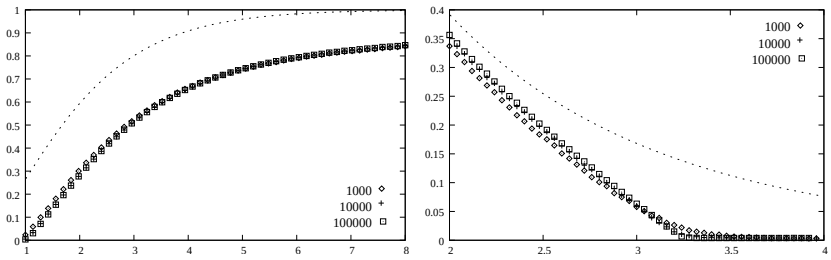

Fig. 5. The plots for the upper bound for the new node attack strategy (lines), and for experimental values of the threshold for networks of size 10^3, 10^4 and 10^5, for random networks (left) and scale-free networks (right). The lines represent the theoretical upper bound.

Notice that the values of the threshold are quite large (one has to remove a large fraction of the nodes do destroy the network). Our aim here, though, is not to obtain an efficient attack strategy, but to show that the qualitative difference between the classical attack strategy and node failures on scale-free networks relies on the fact that, in an attack, no nodes of degree 1 are removed: if nodes of degree higher than 1 are randomly removed, then the same qualitative behavior is recovered.

2.2 Efficient link attack

We have seen in Section 1 that, although the classical attack displays a threshold when considered from the links point of view, it is not efficient in this regard. Still based on the fact that increasing the proportion of nodes of degree 1 collapses the network, we now propose the following attack strategy on links: we remove at random links between nodes of degree at least 2. The effect of this attack is shown in Figure 6.

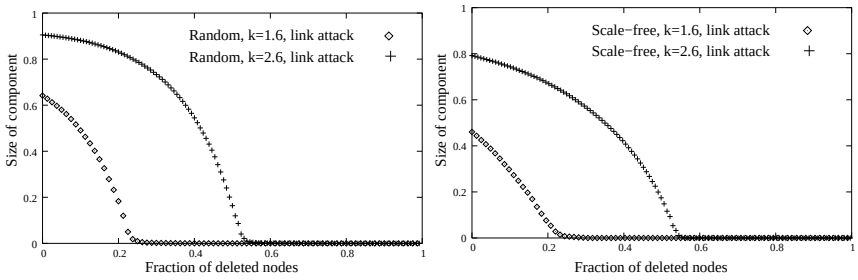

Fig. 6. The effect of the new link attack strategy on random networks (left) and scale-free networks (right).

As expected, this attack strategy displays a threshold m_c. Again, we can show this by providing an upper bound as follows.

When all the links between nodes of degree at least 2 have been removed, the network is decomposed in a set of disjoint stars (each central node is connected to nodes of degree 1). Since the maximal degree of a node in a finite scale-free network with N nodes can be evaluated as $N^{\frac{1}{\alpha-1}}$ [22], the size of the largest connected component (*i.e.* the largest star) is sublinear with respect to N whenever $\alpha > 2$.

An upper bound for m_c is therefore given by the fraction of links between nodes of degree at least 2. This quantity is 1 minus the fraction of links incident to at least one node of degree 1. The number of such links is given by the number of nodes of degree 1, minus the number of links between two nodes of degree 1.

This last number can be computed as follows. There are Np_1 nodes of degree 1, each of them having a probability $Np_1/2|E|$ of being connected to another node of degree 1[3] ($|E| = N\langle k\rangle/2$ denotes the number of links in the network). Therefore the number of *nodes* of degree 1 adjacent to another node of degree 1 is $N^2p_1^2/2|E| = Np_1^2/\langle k\rangle$ on average. Finally, the number of links between two such nodes is therefore $Np_1^2/2\langle k\rangle$.

From this we have that the number of links adjacent to at least one node of degree 1 is: $Np_1 - Np_1^2/2\langle k\rangle$, and the number of links *not* adjacent to any node

[3] This is accurate in the limit of large N.

of degree 1 is: $|E| - Np_1 + Np_1^2/2\langle k\rangle$. The fraction of such links therefore is:

$$1 - \frac{2p_1}{\langle k\rangle} + \frac{p_1^2}{\langle k\rangle^2}.$$

For scale-free networks, this quantity is equal to:

$$1 - \frac{2}{\zeta(\alpha-1)} + \frac{1}{\zeta^2(\alpha-1)} = 1 - \frac{2\zeta(\alpha-1)-1}{\zeta^2(\alpha-1)}.$$

For random networks, it is equal to:

$$1 - 2e^{-z} + e^{-2z}.$$

This upper bound can be evaluated numerically. The result of this evaluation is shown in Figure 7, together with experimental values.

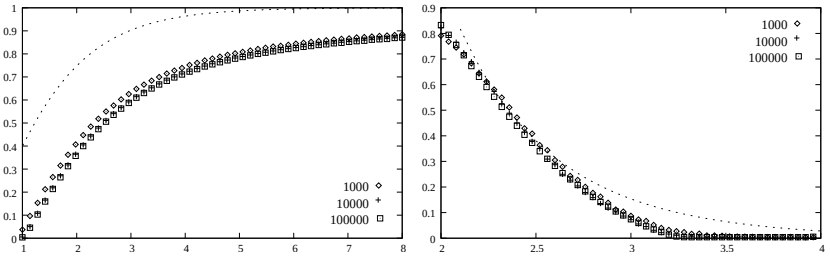

Fig. 7. Experimental values for the threshold for the new link attack strategy, for networks of size 10^3, 10^4 and 10^5, for random networks (left) and scale-free networks (right). The lines represent the upper bounds.

If we compare these results to the ones obtained in Section 1, then we can observe that our attack strategy is more efficient than the classical one, viewed from the links point of view. This is not surprising since in the classical attack strategy one may remove many links attached to nodes of degree 1, which does not help in destroying the network. Our strategy, on the opposite, focuses on those links which really disconnect the network.

Conclusion and discussion

In this contribution, we provided a detailed comparison of the impact of failures and classical attacks on random and scale-free networks. Our aim was to give a more precise insight on the actual efficiency of attacks on scale-free compared to random networks, and compared to failures.

To achieve this, we investigated the often claimed affirmation that the efficiency of attacks on scale-free networks is due to the large number of links they remove. We show that removing the same number of links at random has much less impact, contradicting this affirmation. However, when the number of removed links is considered, scale-free networks are not more fragile than random ones. Finally, we used a classical criterion for network connectivity to design two new attack strategies. The first one is very close to a series of failures but behaves qualitatively like classical attacks (there is a threshold for scale-free networks). This tends to show that the presence of a threshold for classical attacks is not due to a high efficiency, but rather to the fact that they do not remove nodes of degree 1. The second strategy we propose, based on links removal, shows that one can design attack strategies more efficient than the classical one, with respect to the fraction of removed links.

These results lead us to the conclusion that, despite failures and classical attacks clearly behave differently and although the random or scale-free nature of the network strongly influences this, one should be careful in driving conclusions from this. The sensitivity of scale-free networks to attacks relies on the fact that they have many low-degree nodes. Their robustness relies on the fact that when we choose a node at random, we choose such a node with high probability. Moreover, the fact that a classical attack on a scale-free network removes many links may be considered as partly but not fully responsible for its rapid breakdown.

This work may be pursued in many directions. First, the accuracy of the evaluation of the various thresholds should be improved. Likewise, the impact of the finite size of real-world network is in general not understood and should be studied. Moreover, other properties of real-world complex networks, like clustering or degree correlations, should be taken into account. From a more general point of view, the impact of failures and attacks on the actual networks of interest, like the Internet, the World Wide Web and Peer-to-Peer systems, but also biological or social networks, should be deepened. It is likely that these networks have some hidden properties which render them very resilient to failures, and maybe sensitive to certain attack strategies.

Acknowledgments. This work was partly funded by the *Metrosec : Metrology for Security and Quality of Service* project. (`http://www.laas.fr/~owe/METROSEC/`)
We warmly thank Alessandro Vespignani for useful comments and discussions.

References

1. B. Bollobás. *Random Graphs*. Academic Press, 1985.
2. P. Erdös and A. Rényi. On random graphs I. *Publ. Math. Debrecen*, 6:290–297, 1959.
3. M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.
4. A.-L. Barabási, Z. Deszo, E. Ravasz, S.H. Yook, and Z. Oltvai. Scale-free and hierarchical structures in complex networks. In *Sitges Proceedings on Complex Networks*, 2004.

5. S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. *Adv. Phys. 51, 1079-1187*, 2002.
6. S.N. Dorogovtsev and J.F.F. Mendes. *Evolution of Networks: From Biological Nets tou the Internet and WWW*. Oxford University Press, 2000.
7. A.Z. Broder, S.R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J.L. Wiener. Graph structure in the web. *WWW9 / Computer Networks*, 33(1-6):309–320, 2000.
8. A.Z. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J.L. Wiener. Graph structure in the web. In *Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications netowrking*, pages 309–320. North-Holland Publishing Co., 2000.
9. M.E.J. Newman. Random graphs as models of networks. In Stefan Bornholdt and Heinz Georg Schuster, editors, *Hankbook of Graphs and Networks: From the Genome to the Internet*. Wiley-vch, 2003.
10. M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–262, 1999.
11. L. Adamic and B. Huberman. Power-law distribution of the world wide web. *Science*, 287, 2000.
12. M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing Journal special issue on peer-to-peer networking*, 6(1), 2002.
13. R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance in complex networks. *Nature*, 406:378–382, 2000.
14. D.S. Callaway, M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett.*, 85:5468–5471, 2000.
15. R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.*, 86:3682–3685, 2001.
16. S.-T. Park, A. Khrabrov, D.M. Pennock, S. Lawrence, C. Lee Giles, and L.H. Ungar. Static and dynamic analysis of the internet's susceptibility to faults and attacks. In *IEEE Infocom 2003*, San Francisco, CA, April 1–3 2003.
17. A. Broido and K. Claffy. Topological resilience in ip and as graphs. 2002. `http://www.caida.org/analysis/topology/resilience/`
18. V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Phys. Rev. Lett.*, 87, 2001.
19. P. Crucitti, V. Latora, M. Marchiori, and A. Rapisarda. Efficiency or scale-free networks: error and attack tolerance. *Physica A*, 320:622–642, 2003.
20. R. Pastor-Satorras and A. Vespignani. *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press, 2003. To appear.
21. M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161, 1995.
22. R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Resilience of the internet to random breakdown. *Phys. Rev. Lett.*, 85:4626, 2000.