Impact of Random Failures and Attacks on Poisson and Power-Law Random Networks

CLÉMENCE MAGNIEN, MATTHIEU LATAPY, and JEAN-LOUP GUILLAUME, LIP6-CNRS, UPMC

It has appeared recently that the underlying degree distribution of networks may play a crucial role concerning their robustness. Previous work insisted on the fact that power-law degree distributions induce high resilience to random failures but high sensitivity to attack strategies, while Poisson degree distributions are quite sensitive in both cases. Then much work has been done to extend these results.

We aim here at studying in depth these results, their origin, and limitations. We review in detail previous contributions in a unified framework, and identify the approximations on which these results rely. We then present new results aimed at clarifying some important aspects. We also provide extensive rigorous experiments which help evaluate the relevance of the analytic results.

We reach the conclusion that, even if the basic results are clearly important, they are in practice much less striking than generally thought. The differences between random failures and attacks are not so huge and can be explained with simple facts. Likewise, the differences in the behaviors induced by power-law and Poisson distributions are not as striking as often claimed.

Categories and Subject Descriptors: A.1 [**Introductory and Survey**]; C.2.1 [**Computer-Communication Networks**]: Network Architecture and Design—*Network topology*; G.2.2 [**Discrete Mathematics**]: Graph Theory—*Network problems*

General Terms: Experimentation, Reliability, Security

Additional Key Words and Phrases: Network resilience, network failures and attacks, node and link removals

ACM Reference Format:
Magnien, C., Latapy, M., and Guillaume, J.-L. 2011. Impact of random failures and attacks on Poisson and power-law random networks. ACM Comput. Surv. 43, 3, Article 13 (April 2011), 31 pages.
DOI = 10.1145/1922649.1922650 http://doi.acm.org/10.1145/1922649.1922650

1. INTRODUCTION

It has been shown recently—see, for instance, Albert and Barabási [2002]; Dorogovtsev and Mendes [2002]; Newman [2003b]; Strogatz [2001]; Watts and Strogatz [1998]—that most real-world complex networks have nontrivial properties in common. In particular, the degree distribution (probability p_k that a randomly chosen node has k links, for each k) of most real-world complex networks is heterogeneous and well fitted by a power law, that is, $p_k \sim k^{-\alpha}$, with an exponent α between 2 and 3 in general. This property has been observed in many cases, including Internet and Web graphs [Faloutsos et al. 1999; Govindan and Tangmunarunkit 2000; Tangmunarunkit et al. 2001, 2002; Broido

This work was supported in part by the European MAPAP SIP-2006-PP-221003 project, the French ANR MAPE project, the MetroSec (Metrology of the Internet for Security and Quality of Services (http://www.laas/fr/~METROSEC)) project, and the GAP (Graphs, Algorithms and Probabilities) project.
Authors' addresses: C. Magnien, M. Latapy, and J.-L. Guillaume, LIP6, CNRS et UPMC, case 169, 4 place Jussieu, 75252 Paris Cedex 05, France; email: {clemence.magnien, matthieu.latapy, jean-loup.guillaume}@lip6.fr.

DOI 10.1145/1922649.1922650 http://doi.acm.org/10.1145/1922649.1922650

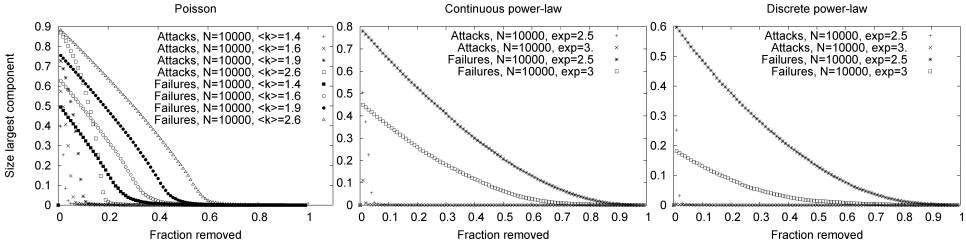

Fig. 1. Size of the largest connected component as a function of the fraction of randomly removed nodes (failures) and nodes removed in decreasing order of node degrees (classical attacks). We will define properly these different kinds of networks in Section 2.2. For technical details on our plots, see Section 2.4.

and Claffy 2001; Pastor-Satorras et al. 2001; Magoni and Pansiot 2001; Chang et al. 2001; Willinger et al. 2002; Chen et al. 2002; Yook et al. 2002; Bu and Towsley 2002; Albert et al. 1999; Adamic and Huberman 2000; Kumar et al. 1999; Broder et al. 2000; Laura et al. 2003], social networks [Liljeros et al. 2001; Newman 2001a; 2001b; Ebel et al. 2002; Jin et al. 2001], and biological networks [Kohn 1999; Uetz et al. 2000; Jeong et al. 2000; Farkas et al. 2003].

In most cases, the existence of a path in the network from most nodes to most others, called *connectivity*, is a crucial feature. For instance, in the case of the Internet, it means that computers can communicate; in the case of the Web it means that one may reach most pages from most others by following hyperlinks; and in the case of social networks it conditions the ability of information and diseases to spread. Note that connectivity may be a desirable feature (e.g., in the case of the internet), or an unwanted one (in the case of virus propagation, for instance), depending on the application under focus.

Networks are subject to damages (either accidental or not) which may affect connectivity. For instance, failures may occur on computers, causing removal of nodes in Internet graphs. Likewise, in social networks, people can die from a disease, or people deemed likely to propagate the disease can be vaccinated, which corresponds to node removals. For the study of these phenomena, accidental failures may be modeled by removals of random nodes or links in the considered network, while attacks may be modeled by removals following a given strategy.

Networks of different natures may behave differently when one removes nodes or links. The choice of the removed nodes or links may also influence significantly the obtained behavior. This has been confirmed in the famous article by Albert et al. [2000], in which the authors considered networks with Poisson and power-law degree distributions, and then removed nodes either randomly (failures) or in decreasing order of their degree (attacks).[1] They estimated the impact by the size of the largest connected component (i.e., the largest set of nodes such that there is a path in the network from any node to any other node of the set) when a given fraction of nodes are removed.

They obtained the results presented in Figure 1, which shows that networks with power-law degree distributions behave significantly differently from ones with Poisson distributions: in particular, they are very resilient to random failures, but very sensitive to attacks. This particularity is now referred as the *Achille's heel of the Internet* [Nature 2000; Barabási 2003]. In the case of a social network on which one wants to design vaccination strategies, it means that one may expect better efficiency by vaccinating people with the highest number of acquaintances than with random vaccination [Dezsö and Barabási 2002; Pastor-Satorras and Vespignani 2002; Cohen et al. 2003a; Holme 2004].

[1]Broder et al. [2000] pursued the same kind of idea earlier in the case of the Web.

Important efforts have then been made to give analytic results based on mean-field approximations completing the experimental ones; see, in particular, [Cohen et al. 2000, 2001a, 2003b; Callaway et al. 2000; Newman 2003a; Albert and Barabási 2002]. Our aim here is to present these results in detail, to deepen them with new results, and to discuss their implications. In particular, we give results concerning failures and attacks on links, as well as results on finite cases, which have received little attention. We also compare the two different approaches for the study of power-law networks proposed in Cohen et al. [2000, 2001a, 2003b] and Callaway et al. [2000] and Newman [2003a].

This article may therefore be considered as an in-depth survey of the main results of the field, with the aim of unifying the different approaches and questions that have been studied, which leads to the introduction of some new results.

This article is organized as follows. Section 2 is devoted to preliminaries, which consist of definitions and models, of methodological discussions, and of some preliminary results. Sections 3 and 4 deal with failures and attacks, respectively. We present classical results of the field, as well as several new results which aim at improving our understanding of the phenomena under consideration. Finally, we give an in-depth discussion and synthesis of our understanding of the field in Section 5.

A more complete and detailed version of this article is available online.[2] This extended version is in general more detailed than this current version. In particular, it contains the proofs of all results, which are omitted here for brevity. It also contains an additional section studying the behaviors of several real-world complex networks, and comparing them to expected behaviors from theory.

2. PRELIMINARIES

Before entering the core of the article, we need to cover some important preliminaries. They consist of definitions and results, mainly concerning probability distributions and models, but also methodological aspects. This section should be read carefully before the rest of the article since important notions are introduced and discussed.

Let us first insist on the fact that most results in this article are obtained using *approximations*, aimed at simplifying the computation. These approximations are valid in the limit of networks with large sizes. They typically consist of neglecting the difference between N and $N \pm n$ when n is small compared to N, or of supposing that random values are equal to the average. More subtle approximations are also done, belonging to the *mean-field* approximation framework, classical in statistical mechanics and widely used in the context of complex networks.

It is important to understand that the results are valid only in this framework, as we have no formal guarantee that all approximations are valid. This is why we will always explicitly point out the approximations we make, and compare analytic results to experiments.

2.1. Poisson and Power-Law Distributions

A probability distribution is given by the probability p_k, for all k, that the considered value is equal to k. The sum of all p_k must be equal to 1.

A Poisson distribution is characterized by $p_k = e^{-z}\frac{z^k}{k!}$, where z is the average value of the distribution. The probability of occurrence of a value x in such a distribution therefore decays exponentially with its difference $|z - x|$ to the average, which means that, in practice, all the values are close to this average.

A power-law distribution with exponent α is such that p_k is proportional to $k^{-\alpha}$. In such distributions, the probability of occurrence of a value x decays only polynomially

[2] http://arxiv.org/abs/0908.3154.

with x. This implies that, though most values are small, one may obtain very large values. In the whole article we will generally consider exponents between 2 and 4, which are the relevant cases in our context (see Section 1.2).

We will consider here two types of power-law distributions, which are the most widely used in the literature: *discrete* and *continuous* power-law distributions. They are both defined by their exponent α and their minimal value m.

A continuous power law is a Pareto distribution, that is, $\int_m^\infty Cx^{-\alpha}\mathrm{d}x = 1$. C is a normalization constant such that: $\int_m^\infty C\ x^{-\alpha}\mathrm{d}x = 1$, which gives $C = m^{\alpha-1}(\alpha-1)$. To obtain discrete values, we then take p_k equal to $\int_k^{k+1} Cx^{-\alpha}\mathrm{d}x$, which is proportional to $k^{-\alpha}$ in the limit where k is large. And finally, $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1})$. We will mainly use this form in the sequel but at some points we will switch back to the continuous form.

A discrete power-law distribution is $p_k = \frac{1}{C}\ k^{-\alpha},\ k \geq m$, where $C = \sum_{k=m}^\infty k^{-\alpha}$ is a normalization constant. In such a distribution, therefore, p_k is exactly proportional to $k^{-\alpha}$ for all $k \geq m$. In order to simplify the computation, we will always take $m = 1$ for discrete power-law distributions in this article. This implies that $C = \zeta(\alpha)$, where ζ is the Riemann zeta function defined for $\alpha > 1$ by $\zeta(\alpha) = \sum_{k=1}^\infty k^{-\alpha}$. Then, $p_k = \frac{1}{\zeta(\alpha)}\ k^{-\alpha}$.

Discrete and continuous power-law distributions each have advantages and drawbacks. For instance, continuous power laws are easier to use in experiments and discrete ones are more rigorous for small values. For a more complete discussion on the advantages and drawbacks of discrete and continuous distributions, see, for instance, Dorogovtsev and Mendes [2001], and Cohen et al. [2001b]. We will use both of them in the sequel, and discuss their differences.

2.1.1. Bounded Distributions. Given a distribution p_k as defined above, one may sample a finite number N of values from it. In such a sample, there is a maximal value K. Therefore, the actual distribution of the values in this sample, $p_k(N)$, is slightly different from p_k. In particular, for all $k > K$, $p_k(N) = 0$ (while in general $p_k > 0$). We will therefore call these distributions *bounded distributions*. We detail below important properties of bounded distributions.

The maximal value K among a sample of N values from a given distribution p_k is a random variable, and it is possible to give the exact formula for its expected value: let $X_1, \ldots, X_N$ be the values sampled from the distribution, and let $Y = \max_{i=1..N} X_i$. Y then has the following distribution:

$$P(Y = K) = \left(\sum_{k=0}^{K} p_k\right)^N - \left(\sum_{k=0}^{K-1} p_k\right)^N.$$

It is the probability that all values are lesser than or equal to K, minus the probability that all values are lesser than K, that is, the probability that all values are lesser than or equal to K and at least one is equal to K. However, deriving numerical values for the expected value of this random variable is too intricate. We will therefore use an approximation.

LEMMA 2.1 (COHEN ET AL. 2000). *For a given distribution p_k such that $p_k > 0$ for all k, the expected maximal value K among a sample of N values can be approximated by*

$$\sum_{0}^{K-1} p_k = 1 - \frac{1}{N}.$$

PROOF. The claim is equivalent to $\sum_K^\infty p_k = \frac{1}{N}$, which means that K is such that there is only one value larger than K in the sample. Moreover, this value must be

exactly equal to K; otherwise there would be only one value larger than $K+1$ and we would have $\sum_{K+1}^{\infty} p_k = \frac{1}{N}$, which is impossible since $p_k > 0$ for all k. □

This result may be used in practice to approximate iteratively the expected maximal value K among a sample of N values by setting $K = 0$ and increasing it until $\sum_{k=0}^{K} p_k \geq 1 - \frac{1}{N}$. For continuous power-law distributions, we can obtain the following approximation: $K = mN^{\frac{1}{\alpha-1}}$.

We performed experiments to compute the maximal value for samples of size $N = 100\,000$, for the three types of distributions we consider. Our approximations are quite accurate, and we will therefore use them in the rest of the article.

An important point to notice is that, for all the distributions we consider here, the expected maximal among N sampled values grows sublinearly with N. The above approximation explicitly states it for continuous power laws, and one may also check the discrete power-law and Poisson cases. As we will see, this is important for some approximations we will make in the following.

Sampling N values from a distribution also induces an expected distribution of the N values $p_k(N)$ which is different from the original distribution p_k.

LEMMA 2.2. *The expected distribution $p_k(N)$ of N values sampled from a given distribution p_k can be approximated, for all $k \leq K$, by*

$$p_k(N) = \frac{N}{N-1}\, p_k,$$

where K is the expected maximal value, related to p_k and N by Lemma 2.1.

The application of this result to the three cases of interest is straightforward.

2.1.2. First Moments of a Distribution. The average $\langle k \rangle = \sum_{k=0}^{\infty} k p_k$ of a distribution p_k is also called its *first moment*, the ith moment being defined as $\langle k^i \rangle = \sum_{k=0}^{\infty} k^i\ p_k$. In the continuous case, the ith moment is similarly defined as $\langle k^i \rangle = \int_{k=m}^{\infty} k^i\ p_k$. In the whole article, the first and second moments will play a central role. We present here the results for the three cases of interest.

LEMMA 2.3. *For a Poisson distribution with average value z, the first two moments of the expected distribution of a sample of N values can be approximated by*

$$\langle k \rangle = \left(\frac{N}{N-1}\right) \sum_{k=0}^{K} \frac{e^{-z} z^k}{(k-1)!} \quad \textit{and} \quad \langle k^2 \rangle = \left(\frac{N}{N-1}\right) \sum_{k=0}^{K} \frac{k e^{-z} z^k}{(k-1)!},$$

where K is the expected maximal value, related to p_k and N by Lemma 2.1.

LEMMA 2.4. *For a Poisson distribution with average value z, the first two moments are*

$$\langle k \rangle = z \textit{ and } \langle k^2 \rangle = z^2 + z.$$

LEMMA 2.5 (COHEN ET AL. 2000). *For a continuous power-law distribution with exponent α and minimal value m, the first two moments of the expected distribution of a sample of N values can be approximated by*

$$\begin{array}{lllll}
\langle k \rangle = m^{\alpha-1} K^{-\alpha+2} \frac{\alpha-1}{-\alpha+2} & \textit{and} & \langle k^2 \rangle = m^{\alpha-1} K^{-\alpha+3} \frac{\alpha-1}{-\alpha+3}, & \textit{if} & 1 < \alpha < 2, \\
\langle k \rangle = m \frac{\alpha-1}{\alpha-2} & \textit{and} & \langle k^2 \rangle = m^{\alpha-1} K^{-\alpha+3} \frac{\alpha-1}{-\alpha+3}, & \textit{if} & 2 < \alpha < 3, \\
\langle k \rangle = m \frac{\alpha-1}{\alpha-2} & \textit{and} & \langle k^2 \rangle = m^2 \frac{\alpha-1}{\alpha-3}, & \textit{if} & \alpha > 3,
\end{array}$$

where K is related to p_k and N by Lemma 2.1.

LEMMA 2.6 (COHEN ET AL. 2000). *For a continuous power-law distribution with exponent α and minimal value m, the first two moments are*

$$\langle k \rangle = m \frac{\alpha - 1}{\alpha - 2} \text{ if } \alpha > 2 \quad \text{and} \quad \langle k^2 \rangle = m^2 \frac{\alpha - 1}{\alpha - 3} \text{ if } \alpha > 3,$$

and they diverge in all the other cases.

LEMMA. *For a discrete power-law distribution with exponent α, the first two moments of the expected distribution of a sample of N values can be approximated by*

$$\langle k \rangle = \frac{H_K^{(\alpha-1)}}{H_{K-1}^{(\alpha)}} \quad \text{and} \quad \langle k^2 \rangle = \frac{H_K^{(\alpha-2)}}{H_{K-1}^{(\alpha)}},$$

where $H_K^{(\alpha)} = \sum_{k=1}^{K} k^{-\alpha}$ is the Kth harmonic number for α, where K is the expected maximal value, related to p_k and N by Lemma 2.1.

LEMMA 2.8 (NEWMAN 2003a). *For a discrete power-law distribution with exponent α, the first two moments are*

$$\langle k \rangle = \frac{\zeta(\alpha - 1)}{\zeta(\alpha)} \text{ and } \langle k^2 \rangle = \frac{\zeta(\alpha - 2)}{\zeta(\alpha)}.$$

Although the difference between the distributions themselves is small (the ratio between a bounded and an unbounded distribution is approximately $N/(N-1)$), this is not the case for the moments of these distributions. In practice, we can notice that, for Poisson distributions, the values of the first and second moments are almost identical for bounded and unbounded distributions, while there is a noticeable difference for power-law distributions. This can be understood as follows: these differences are strongly related to the quantities $\sum_{k=K+1}^{\infty} k p_k$ and $\sum_{k=K+1}^{\infty} k^2 p_k$. In both cases, values of p_k for $k > K$ are quite small: $\sum_{k=K}^{\infty} p_k = 1/N$. For Poisson networks, this $1/N$ is distributed among p_k which decrease exponentially, and K is small. Therefore the values of kp_k and $k^2 p_k$ for $k > K$ are small. For power-law networks, on the other hand, K is large, and the probabilities p_k decrease only polynomially; therefore values of kp_k and $k^2 p_k$ for $k > K$ are much larger.

These observations will explain in the following why in some cases theoretical predictions for the finite case and for the infinite limit are almost identical for Poisson networks and quite different for power-law networks.

2.2. Modeling Issues

In this section we detail the models of networks we will consider, then discuss the modeling of failures and attacks we will use. We finally present results concerning the connectivity of random networks, which will play a key role in the sequel.

2.2.1. Random Networks. Given an integer N and a distribution p_k, one can easily generate a network taken uniformly at random among the ones having N nodes and degree distribution p_k. Indeed, it is sufficient to sample the degree of each node with respect to p_k, then to attach to it as many *stubs* as its degree, and finally to construct links by choosing random pairs of stubs.[3] This model is known as the *configuration model* [Bender and Canfield 1978] and is widely used in the literature; see, for instance, Bollobás

[3]If the sum of degrees is odd, then one just has to sample again the degree of a random node until the sum becomes even.

[1985]; Molloy and Reed [1995, 1998], and Aiello et al. [2000]. We will call *random networks* all networks obtained using it.[4]

If one chooses a Poisson distribution of average z then one obtains an equivalent of the Erdös-Rényi model [Erdös and Rényi 1959] in which the network is constructed from N initially disconnected nodes by adding $M = \frac{z\,N}{2}$ links between randomly chosen pairs of nodes. One then obtains a network taken uniformly at random among the ones having N nodes and M links.

As already discussed in the Introduction, the degree distribution of a network may be seen as responsible for some of its most important features (like robustness). Studying random networks with prescribed degree distributions is therefore a key issue. These networks are particularly well suited for formal analysis, and most formal results obtained on complex networks in the literature, including the ones on robustness, rely on this modeling; this is why we use it here.

We will focus on three classes of networks, namely, the ones with Poisson, continuous power-law, and discrete power-law degree distributions, which we will call *Poisson networks*, *continuous power-law networks*, and *discrete power-law networks*, respectively.

In our experiments, we will consider Poisson networks with average degree z between 1 and 8, because for $z < 1$ the networks do not have a giant component (see Section 2.2.3 and Lemma 2.10), and we have observed that the behaviors for $z \geq 8$ are very similar to and easily predictable from the ones observed for $z = 8$. Concerning power-law networks, we will always take the minimal degree m equal to 1, which fits most real-world cases. We will consider exponents between 2 and 4 because below 2 the average degree is infinite (see Lemmas 2.6 and 2.8) and above 4 the network has only small connected components, as we will see below (see Lemmas 2.11 and 2.12). Moreover, most real-world cases fit in these ranges.

2.2.2. Failures and Attacks. There are many ways to model various kinds of failures and attacks. We will focus here on removals of nodes or links. We will suppose that failures are random, in contrast to attacks, which follow strategies.

Random node failures are series of removal of nodes chosen at random. Equivalently, one may choose a fraction of the nodes at random and then remove them all. Likewise, *random link failures* consist of the removal of links chosen at random.

Attacks, on the other hand, follow a *strategy* for removing nodes or links. We then say that we observe an *attack following this strategy*. For instance, we presented in the Introduction the most famous strategy, which consists of removing nodes in decreasing order of their degrees. We will call this the *classical attack*, and we will define other strategies in Section 4.

Notice, moreover, that, when one removes a node, one also removes all the links attached to it. This leads to the *link point of view* of node failures and attacks, which consists of observing the fraction of *links* removed during *node* removals.

For these various strategies, we want to observe the resilience of networks, which requires one to use a criterion to capture the impact of failures or attacks on a network. We will here consider the fraction of nodes in its largest connected component as a function of the number of nodes or links removed. This captures the ability of nodes to communicate: the smaller this fraction, the greater the impact of the removals.

2.2.3. Largest Connected Component. In many cases, the largest connected component of a random network contains the most nodes of the network. More precisely, depending on the underlying degree distribution, the size of the largest connected component may

[4]These networks may contain loops (links from one node to itself) and multiple links (more than one link between two given nodes) in small quantities, which we will neglect in our reasoning as explained in Section 2.3.

scale linearly with the size of the network. The network is then said to have a *giant component*.

There actually exists a precise and simple criterion on the degree distribution of a random network to predict if this network will have a giant connected component or not. Since most of the results we will discuss later in this contribution rely on an appropriate use of this criterion, we recall it here.

THEOREM 2.9 (MOLLOY AND REED 1995). *A random network with size N tending toward infinity and with degree distribution p_k such that it has maximal value $K < N^{1/4}$ almost surely has a giant component if and only if*

$$\langle k^2 \rangle - 2\langle k \rangle = \sum_{k=0}^{K} k(k-2)p_k > 0.$$

This theorem has been rigorously proved in Molloy and Reed [1995] and Aiello et al. [2000] and has been proved in the mean-field approximation framework in Cohen et al. [2000] and Newman [2003a].

This result can be applied to the three kinds of networks we consider here (since their maximal degree is sublinear, as explained in Section 2.1), which gives the following results which are very simple conditions under which the random networks we consider have a giant component.

LEMMA 2.10. *A Poisson network with size tending toward infinity and average degree z almost surely has a giant component if and only if $z > 1$.*

LEMMA 2.11. *A continuous power-law network with size tending toward infinity, exponent α, and minimal degree $m = 1$ almost surely has a giant component if and only if $\alpha < 4$.*

LEMMA 2.12. *A discrete power-law network with size tending toward infinity and exponent α almost surely has a giant component if and only if α is such that $\frac{\zeta(\alpha-2)}{\zeta(\alpha-1)} > 2$ (a numerical evaluation gives $\alpha < 3.48$).*

2.3. Mean-Field Framework and Generating Functions

As already emphasized at the beginning of Section 2, most results in this article are made using *approximations*, valid in the mean-field framework. We detail below some which deserve attention. We then present the generating function framework, which makes it possible to embed these approximations in a powerful formalism.

2.3.1. Mean-Field Approximations in Random Networks. The fact that stubs are linked fully at random in a random network is a feature which has important consequences in our context. In particular, when one removes a link chosen at random in such a network, this is equivalent to the removal of two stubs at random, and so the obtained network is still random (with a different degree distribution in general). Likewise, when one removes a node, the obtained network is also random.

Mean-field approximations are very helpful in the study of random networks since they allow one to neglect some correlations which would be very hard to handle.

Consider a random network that is large and sparse (the probability for two randomly chosen nodes to be linked together is almost zero) and with small maximal degree compared to its size (this will always be the case in our context). Then, given any node, called the *source*, the probability that two of its neighbors are directly linked together is negligible. Likewise, if we take all the nodes at distance 2 of the source then the probability of having a link between two of them is very small and may also be neglected. So does the probability of having a link from a node at distance 2 to more

than one node at distance 1, or to the source. Continuing this reasoning, the network may be considered locally as a tree: any subnetwork composed of the nodes at a distance lower than a given finite value is a tree if the size of the network tends toward infinity.

The approximation above neglects very small probabilities, or equivalently considers the limit where the size of the network tends toward infinity.

In the same manner, it is known that any random network with a maximal degree lower than $\sqrt{\langle k \rangle N}$ almost surely has no loops or multiple links [Chung and Lu 2002; Burda and Krzywicki 2003]. Likewise, Theorem 2.9 is formally true only for networks with maximal degree less than $N^{1/4}$. These conditions might not be true for all the networks under consideration; however, we will consider that the networks do not possess loops and multiple links and that Theorem 2.9 can be applied.

The mean-field framework allows another important approximation which comes from the fact that there is no distinction between choosing a stub at random and following a link at random from a random starting node. Indeed, since links are formed by pairs of randomly chosen stubs, it makes in principle no difference.

One consequence is that we suppose that there is no correlation between the degree of a node and the degrees of its neighbors, that is, that the random starting node we choose has no impact on the neighboring node we will reach. This is indeed true when the maximal degree is below $N^{\min(1/2,1/(\alpha-1))}$ [Burda and Krzywicki 2003]. We will neglect the possible correlations, even if the above condition is not fulfilled.

This approximation may be used to describe the degree distribution of neighbors of nodes, in other words, the degree of a node reached by starting from a randomly chosen node and following one of its links chosen at random. According to the mean-field approximation above, this is equivalent to choosing a random stub, and therefore the probability that a random stub belongs to a given node is proportional to this node's degree, that is, the probability of reaching a node of degree k is proportional to $k\ p_k$. The sum of these probabilities must be equal to 1; we therefore obtain the following probability: $\frac{k\ p_k}{\sum_{j=0}^{\infty} j p_j} = \frac{k\ p_k}{\langle k \rangle}$.

We can derive from this the probability q_k that a neighbor of a node has k *other* neighbors, which will be useful in the sequel. It is nothing but the probability that a node obtained by following a link has $k+1$ neighbors, and so

$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}. \tag{1}$$

2.3.2. Basics on Generating Functions. Generating functions, also called *Formal Power Series*, are powerful formal objects widely used in mathematics, computer science, and physics. They encode series of numbers $(s_k)_{k\geq 0}$ as functions $f(x) = \sum_{k=0}^{\infty} s_k x^k$. Operations on the series of numbers then correspond to operations on the associated functions. See Wilf [1994] for an introduction.

The application of generating functions to the random network context is presented in detail in Newman et al. [2001]. Using them to encode series of probabilities (like for instance degree distributions), the authors showed how mean-field approximations may be embedded with the help in this formalism. Once this is done, it is possible to manipulate the associated notions efficiently and easily. We give an overview of this approach below, following the notations in the above reference.

We begin by encoding the degree distribution p_k by the generating function G_0:

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k. \tag{2}$$

This function is an encoding of the whole distribution: $p_k = G_0^{(k)}(0)/k!$. Moreover, we have $G_0(1) = \sum_{k=0}^{\infty} p_k = 1$ (this is true for all generating functions encoding distributions of probabilities), and the average is given by $\langle k \rangle = \sum_{k=1}^{\infty} k\ p_k = G_0'(1)$.

Going further, let us consider the generating function G_1 for the number of other neighbors of a node chosen by following one random link of a randomly chosen node. This number is distributed according to q_k, defined in Equation (1). We then have

$$G_1(x) = \sum_{k=0}^{\infty} q_k x^k = \frac{\sum_{k=0}^{\infty}(k+1)p_{k+1}x^k}{\langle k \rangle} = \frac{\sum_{k=1}^{\infty} k p_k x^{k-1}}{\langle k \rangle} = \frac{G_0'(x)}{\langle k \rangle}. \tag{3}$$

We give now a few results on generating functions which will play an important role. These results are rewritings of results in Callaway et al. [2000] and Newman [2003a]. They aim at expressing the existence of a giant component in terms of generating functions.

Let us consider a random network with degree distribution p_k encoded in G_0. Let us suppose that some of its nodes (respectively, links, i.e., pairs of stubs) are marked. All marked nodes are to be removed; we are therefore interested in *clusters* of unmarked nodes, that is, sets of unmarked nodes such that there exists a path composed only of unmarked nodes (respectively, links) between any two of them. We are interested in the existence of a giant such cluster.

Let us consider a node reached by following a random link, that is, a node obtained by picking a random stub. Consider the number of unmarked nodes that can be reached from this node by following links between unmarked nodes (respectively, unmarked links) only. Two cases may occur: either the chosen node (respectively, stub) is marked, in which case the cluster is of size zero, or it is unmarked. Let us denote by r_k the probability that it is unmarked and has k other stubs, and by $F_1(x)$ the corresponding generating function: $F_1(x) = \sum_{k=0}^{\infty} r_k x^k$. Note that the case where the chosen node (respectively, stub) is marked plays no role in $F_1(x)$. Note also that $F_1(1)$ is the fraction of unmarked nodes (respectively, links) in the network.

THEOREM 2.13 (CALLAWAY ET AL. 2000; NEWMAN 2003A). *If τ is the fraction of marked nodes (respectively, links) such that removing all these marked nodes (respectively, links) gives a network with no giant component; then τ is such that $F_1'(1) = 1$.*

This result is very powerful and general. To compute the fraction of nodes (respectively, links) to remove from a network in order to ensure that the resulting network contains no giant component, it is sufficient to give an expression for $F_1(x)$ and then to determine the fraction which leads to $F_1'(1) = 1$. One must, however, keep in mind that it relies on mean-field approximations, and that the formalism sometimes makes it difficult to see exactly when approximations are performed. However, in the current state of our knowledge, these approximations are necessary to derive the results we seek. It is important to pursue the development of exact methods.

2.4. Plots and Thresholds

The first main kind of plots we will consider represents the fraction of nodes in the largest connected component of a network as a function of the fraction of removed nodes or links.[5] Figure 1 provides an example. In order to be able to compare the various kinds of networks, we selected two typical exponents for the power law, namely 2.5 and 3, and produced continuous and discrete power-law networks with these exponents, as well

[5]All plots are averaged over 1 000 realizations.

Table I. Exponents We Consider in Our Experiments on Power-Law Networks, and Average Degrees They Induce (Obtained in Practice with Minimal Value $m = 1$ and $N = 100{,}000$ Nodes)

	Average degree	
Exponent	Continuous power-law	Discrete power-law
2.5	2.6	1.9
3	1.6	1.4

as Poisson networks with the same average degrees. These values are summarized in Table I.

In our context, it is usual to witness a *threshold* phenomenon: there exists a critical value p_c such that, whenever the fraction of removed nodes (or links, depending on the context) is lower than p_c, the network almost surely still has a giant component, whereas whenever the fraction of removed nodes (or links) is greater than p_c the network almost surely does not have a giant component anymore. In other words, the threshold is reached when the fraction of nodes in the largest connected component goes to zero (there is no giant component anymore).

For a given finite-size network, the notion of threshold does not make sense: the fraction of nodes in the largest connected component will never be zero. Therefore we have considered that the threshold is reached when the largest connected component contains less than 5 % of all the nodes.

The second main kind of plots represents, for a given node or link removal strategy, the threshold as a function of the average degree for Poisson networks, and the exponent of the power law for power-law networks. We plot experimental results for different sizes of networks (see footnote 5). To help in the comparison between different kinds of networks, we add on to these plots vertical lines at the values quoted in Table I. We also plot the theoretical predictions we obtain, together with experimental results, to make it possible to compare them.

We will see that the experimental results do not always fit analytic predictions very closely. This is influenced in part by the choice considering that a giant component must contain at least 5 % of the nodes, as explained above. But other factors impact this. In the case of random failures, for instance, there is a significant difference between the infinite limit and the finite case, even for large sizes. This is why we present results for both finite cases and the infinite limit when possible. This makes it possible to observe the error due to the asymptotic approximation. More generally, the difference between predictions and numerical values are due to the approximations made in the derivations of the analytic results.

2.5. Toward a More Realistic Modeling

Modeling large networks is a complex task and many parameters have to be taken into account. We present here a quick overview of the different types of graph modeling, together with a brief discussion on their advantages and drawbacks. We also present other attack or failures strategies, as well as other definitions of network resilience. Studying all this in detail is, however, out of the scope of this article.

The modeling approach we use here relies on *random sampling*: given a set of properties, the goal is to choose with uniform probability a graph among the set of all graphs having these properties. This approach has the advantage of allowing one to study precisely the impact of a given property: if a certain behavior is observed on graphs obtained with such a model, then we can conclude that this behavior is a consequence of the studied property. This type of model is also well suited for exact proofs. However, it also suffers from limitations: some properties cannot currently be reproduced by this type of model, for instance, the clustering coefficient.

In this article, we focus on the degree distribution of graphs, and use models producing random graphs with given degree distributions. However, the sole degree distribution cannot reproduce the complexity of networks. It has also been shown that, when the sole degree distribution is taken into account, high-degree nodes tend to be connected to each other, which might not be realistic [Doyle et al. 2005].

Efforts have therefore been made to study other properties taking into account the tendency of nodes to be linked to nodes of the same degree or not, and incorporate them into random models. This can be captured by the assortativity parameter [Newman 2002], degree correlations, which are the probabilities $P(d|d')$ that a node of degree d is linked to another node of degree d' [Boguná et al. 2003; Vázquez and Moreno 2003], or by the $s(g)$ parameter which is the normalized sum of $d_i d_j$ for all links (i, j) [Doyle et al. 2005; Li et al. 2006].

Some authors have studied network resilience to failures and attacks, in a similar way as what we present in this article, on random networks with degree correlations, see for instance [Vázquez and Moreno 2003].

Another, and orthogonal, modeling approach consists of taking into account properties that play a role in the construction of a network. For instance, in the case of the Internet, one can consider the way routers work, or other technological or economic constraints. One then iterates an evolution process which respects a set of properties and eventually produces a graph similar to the original network.

This type of approach has the advantage of being able to take into account many properties that cannot be considered in random modeling, for instance, clustering [Watts and Strogatz 1998; Kleinberg 2000; Dorogovtsev et al. 2000; Dorogovtsev and Mendes 2002; Klemm and Eguiluz 2002; Holme and Kim 2002]. This makes it relevant for producing graphs for simulation purposes. However, the evolution rules can create graph structures that are hard to characterize, and in the end the properties of the obtained graphs are not always fully understood. A simple example of this is the preferential attachment model [Albert and Barabási 1999], which produces trees (if new nodes create a single link) or graphs with no nodes of degree 1 (if new nodes create more than one link). These models therefore do not yet allow us to study formally the impact of some topological properties of the networks. Also, in the case of optimization models, it is not always easy to find interesting parameters to optimize. Therefore, this approach is complementary to the first one.

The HOT framework, for *Heuristically Optimal Topology*, belongs to this approach. This framework has been mainly used in the context of the Internet [Doyle et al. 2005; Li et al. 2004]. The authors of these papers proposed rewiring a network with a given degree distribution in a nonrandom fashion which preserves the degree sequence, aiming at optimizing some properties of the networks to mimic real ones, for instance, technological or economical ones. This kind of optimization can also be found in the context of biological systems [Csete and Doyle 2002].

Concerning network robustness, we only consider here the size of the largest component as an indicator of the state of the network after failures or attacks. However, other approaches have been introduced.

Latora and Marchiori [2001] and Holme et al. [2002] used the efficiency, also called *average inverse geodesic length*, which is very similar to the average distance but allows one to consider disconnected networks. In a similar fashion, Park et al. [2003] introduced the *diameter-inverse-K* (DIK) measure, which also allows one to take into account disconnected graphs and for connected graphs allows one to distinguish between short or large average distance. Crucitti et al. [2003], considered that using shortest paths can put a higher load on some nodes, which can increase when failures or attacks occur. In consequence, each node is associated with a given capacity that cannot be exceeded without a loss of efficiency of the node, which forces the

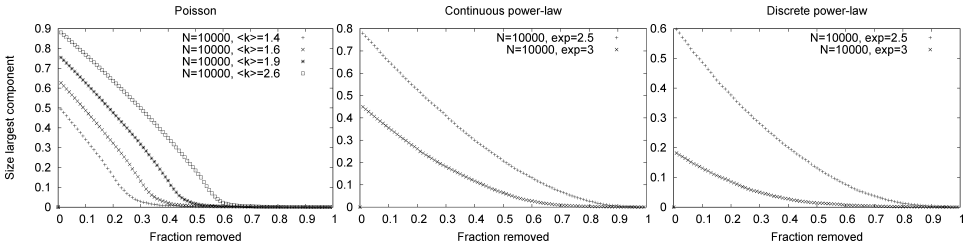

Fig. 2. Size of the largest connected component as a function of the fraction of randomly removed nodes. For technical details on our plots, see Section 2.4.

use of longer paths. This concept has mainly been used for cascading failures and attacks.

More specific measures have also been introduced. In particular, in the case of the Internet, nodes and links carry some demand and the efficiency of the network can be estimated by its capacity to carry it. Li et al. [2004] and Doyle et al. [2005] defined the maximal throughput based on the bandwidth of nodes, a routing matrix, and the traffic demand for all nodes. In the case of the Internet, high-degree nodes are at the periphery of the network and the fragility of the network lies more on the low-degree core nodes than on these periphery nodes.

Some attack strategies based on these measures of importance have also been introduced. Holme et al. [2002], for instance, compared the removal of highest-degree nodes with the removal of highest-betweenness nodes. Furthermore, the order of removal can be either chosen before any removal occurs, or recalculated after each removal.

Finally, more complex types of failures and attacks, like cascading failures, have been considered in the literature, for instance, in Crucitti et al. [2004a, 2004b]; Lee et al. [2005]; Newth and Ash [2004]; Motter and Lai. [2002]; Motter [2004]; Pertet and Narasimhan [2005]; and Zhao et al. [2004].

3. RESILIENCE TO RANDOM FAILURES

The aim of this section is to study the resilience of random networks to random failures. Recall that random node (respectively, link) failures consist of the removal of randomly chosen nodes (respectively, links).

We will first consider random node failures on random networks (Section 3.1). Then, in order to deepen our understanding, we will consider these failures from the *link* point of view (Section 3.2): what fractions of the *links* are removed during random node failures? Finally, we will consider random *link* failures (Section 3.3).

3.1. Random Node Failures

In this section, we first present a general result on random node failures, independent of the type of underlying network, as long as it is a *random* network. We sketch the two main proofs of this result and apply it to the cases under consideration. Figure 2 displays the behaviors observed.

There is a fundamental difference between Poisson and power-law networks: in the Poisson case the giant component is destroyed when a fraction of the nodes significantly lower than 1 has been removed, whereas in the power-law cases one needs to remove almost all nodes. The aim of this section is to formally confirm this, and give both formal and intuitive explanations of this phenomenon.

3.1.1. General Results. Our aim here is to prove the following general result, which gives the value of the threshold for random node failures.

THEOREM 3.1 (CALLAQAY ET AL. 2000; NEWMAN 2003a; COHEN ET AL. 2000, 2003b). *The threshold p_c for random node failures in large random networks with degree distribution p_k is given by*

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

Notice that this theorem states that in some cases p_c might be less than zero. But we have.

$$p_c = 1 - \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \leq 0 \iff \langle k^2 \rangle - 2\langle k \rangle \leq 0.$$

According to Theorem 2.9, this implies that the network almost surely has no giant component. In this case, the notion of threshold therefore has no meaning.

Theorem 3.1 has been derived in different ways in the literature. The two main methods were proposed in Cohen et al. [2000, 2003b] and in Callaway et al. [2000] and Newman [2003a]. We detail both approaches below.

The proof in Cohen et al. [2000, 2003b] relies on the fact that random node failures on a random network lead to a network which may still be considered as random (with a different degree distribution). Therefore, by computing the degree distribution of this network and its first and second moments, one can use the criterion in Theorem 2.9 to decide if there is still a giant component or not.

PROPOSITION 3.2 (COHEN ET AL. 2000, 2003B). *In a large random network with degree distribution p_k, after the removal of a fraction p of the nodes during random node failures the degree distribution $p_k(p)$ is given by*

$$p_k(p) = \sum_{k_0=k}^{\infty} p_{k_0} \binom{k_0}{k} (1-p)^k p^{k_0-k},$$

and the first and second moments of the degree distribution $p_k(p)$ are

$$\langle k(p) \rangle = (1-p)\langle k \rangle \quad \text{and} \quad \langle k^2(p) \rangle = (1-p)^2\langle k^2 \rangle + p(1-p)\langle k \rangle.$$

The proof of Theorem 3.1 is direct using using Theorem 2.9 and Proposition 3.2.

The other proof [Callaway et al. 2000; Newman 2003a] relies on the use of generating functions (see Section 2.3), each node being marked as *absent* with probability p and as *present* with probability $1-p$.

Recall that $F_1(x)$ is the generating function for the probability of finding an unmarked (i.e., present) node with k (marked or unmarked) other neighbors at the end of a randomly chosen link. In our case, $F_1(x)$ therefore is

$$F_1(x) = \sum_{k=0}^{\infty} (1-p) q_k x^k = (1-p) G_1(x),$$

where $G_1(x) = \sum_{k=0}^{\infty} q_k x^k$ is the generating function for the probability of finding a node with k others neighbors at the end of a randomly chosen link, defined in Section 2.3. We can then prove Theorem 3.1 as a direct consequence of Theorem 2.13: from Theorem 2.13, the threshold p_c is reached when $F_1'(1) = 1$, which is equivalent here to $(1-p_c)G_1'(1) = 1$. From the expression of $G_1(x)$ (see Section 2.3) we can then derive the result.

3.1.2. The Cases of Poisson and Power-Law Networks. Theorem 3.1 is valid for any random network, whatever its degree distribution. To study the behavior of Poisson and power-law networks in case of random node failures, we therefore only have to apply it

to these cases, both for finite networks with N nodes and finite networks with size tending towards infinity. Comparison with simulations will be provided at the end of the section.

COROLLARY 3.3. *For large Poisson networks with N nodes and average degree z, the threshold p_c for random node failures is given by*

$$p_c = 1 - \frac{\sum_{k=0}^{K} z^k/(k-1)!}{\sum_{k=0}^{K} z^k/(k-2)!},$$

where K is the maximal degree or the network, related to p_k and N by Lemma 2.1.

COROLLARY 3.4 (COHEN ET AL. 2000). *For Poisson networks with size tending toward infinity and average degree z, the threshold p_c for random node failures is*

$$p_c = 1 - \frac{1}{z}.$$

COROLLARY 3.5 (COHEN ET AL. 2000). *For large continuous power-law networks with N nodes, exponent α and minimal degree m, the threshold p_c for random node failures is*

$$p_c = \begin{cases} 1 - \left[\frac{2-\alpha}{3-\alpha} m - 1\right]^{-1}, & \text{if } \alpha > 3 \\ 1 - \left[\frac{2-\alpha}{\alpha-3} m N^{\frac{3-\alpha}{\alpha-1}} - 1\right]^{-1}, & \text{if } 2 < \alpha < 3 \\ 1 - \left[\frac{2-\alpha}{3-\alpha} m N^{\frac{1}{\alpha-1}} - 1\right]^{-1}, & \text{if } 1 < \alpha < 2. \end{cases}$$

COROLLARY 3.6 (COHEN ET AL. 2000). *For continuous power-law networks with size tending toward infinity, exponent α and minimal degree m, the threshold p_c for random node failures is*

$$p_c = \begin{cases} 1 - \left[\frac{2-\alpha}{3-\alpha} m - 1\right]^{-1}, & \text{if } \alpha > 3 \\ 1, & \text{if } 1 < \alpha < 3. \end{cases}$$

COROLLARY 3.7. *For large discrete power-law networks with N nodes and exponent α, the threshold p_c for random node failures is given by*

$$p_c = 1 - \frac{H_K^{(\alpha-1)}}{H_K^{(\alpha-2)} - H_K^{(\alpha-1)}},$$

where $H_K^{(\alpha)} = \sum_{k=1}^{K} k^{-\alpha}$ is the Kth harmonic number for α, where K is the maximal degree or the network, related to p_k and N by Lemma 2.1.

COROLLARY 3.8 (COHEN ET AL. 2000). *For discrete power-law networks with size tending toward infinity and exponent α, the threshold p_c for random node failures is*

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

We plot numerical evaluations of these results in Figure 3, together with experimental results. We also give in Table II the thresholds for specific values of the exponent and the average degree.

The central point here is to notice that power-law and Poisson networks display a qualitatively different behavior in case of node failures. In theory, power-law networks have a threshold $p_c = 1$ as long as the exponent is lower than 3 (most real-world cases), which means that all nodes have to be removed to achieve a breakdown. On

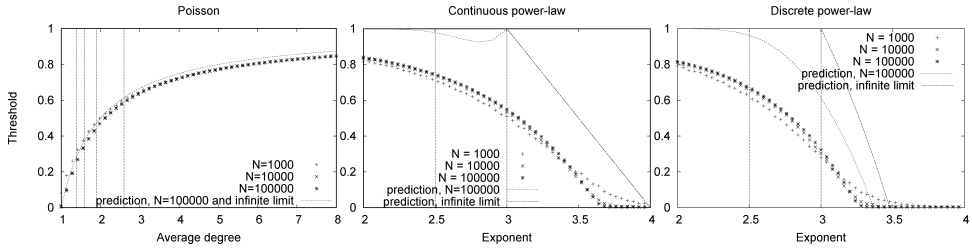

Fig. 3. Thresholds for random node failures. For technical details on our plots and on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 2.4.

Table II. Values of Threshold for Random Node Failures on Discrete and Continuous Power-law Networks of Exponents 2.5 and 3, and on Poisson Networks Having Same Average-Degree (See Table I); Values Are Analytic Previsions (Prev.) at Infinite Limit and Those Obtained for Experiments (Exp.) with Networks of $N = 100\,000$ Nodes

	Continuous power-law		Poisson		Discrete power-law		Poisson	
α	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.
2.5	1	0.74	0.62	0.59	1	0.67	0.47	0.45
3	1	0.55	0.38	0.34	1	0.32	0.29	0.26

the contrary, for Poisson networks only a finite (i.e., strictly lower than 1) fraction of the nodes has to be removed. This leads to the conclusion that power-law networks are significantly more resilient to node failures than Poisson networks, which confirms the experimental observations discussed in the Introduction.

However, this result is moderated by the two following observations. First, Poisson networks may have a quite large threshold when their average degree grows (which appears from both analytic previsions and experiments). Second, and more importantly, power-law networks of finite size N are much more sensitive to failures than what is predicted for the infinite limit. This is already true from the analytic previsions, and even more pronounced for experiments; see Table II.

We may therefore conclude that power-law networks are indeed more resilient to random node failures than Poisson ones, but that the difference in practice is not as striking as predicted by the infinite limit approximations.

3.2. Link Point of View of Random Node Failures

As discussed in the preliminaries, one may wonder what happens in networks during random *node* failures in terms of *the number of links removed*. The plots of the size of the largest component as a fraction of the number of links removed during random node failures are given in Figure 4 (notice that these plots are nothing but (nonlinear) rescalings of the plots in Figure 2). The question we address here therefore is: how many links have been removed when we reach the threshold for random node failures? This is not equivalent to random removals of links, which are studied in the next subsection.

One can evaluate the number of links removed during random node failures as follows.

PROPOSITION 3.9. *In large random networks, after the removal of a fraction p of the nodes during random nodes failures, the fraction of removed links is $m(p) = 2p - p^2$.*

PROOF. Let us consider a network in which we randomly remove a fraction p of the nodes. Since the nodes are chosen randomly, we can assume that the same fraction p

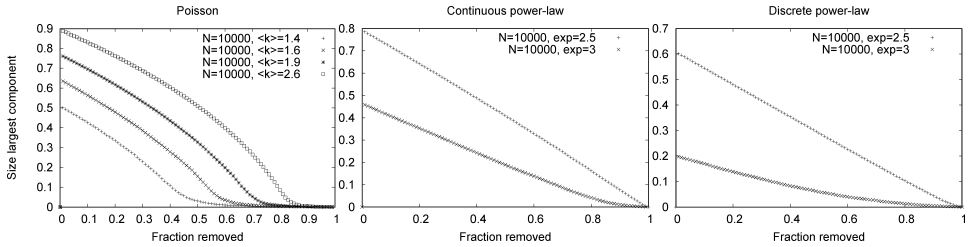

Fig. 4. Size of the largest connected component as a function of the fraction of removed *links*, during random *node* failures. For technical details on our plots, see Section 2.4.

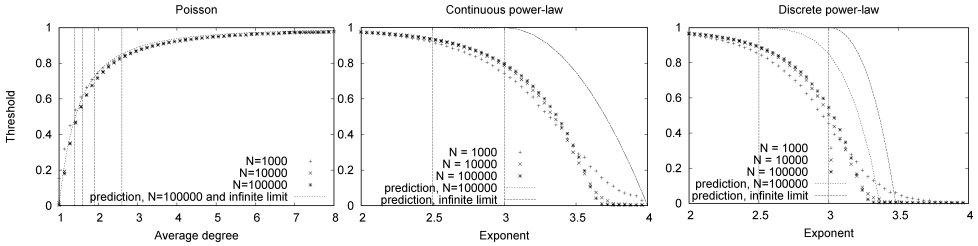

Fig. 5. Thresholds for the link point of view of random node failures. For technical details on our plots and on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 2.4.

Table III. Values of Threshold for Link Point of View of Random Node Failures on Discrete and Continuous Power-law Networks of Exponents 2.5 and 3, and on Poisson Networks Having Same Average-Degree (See Table I); Values Are Analytic Previsions (Prev.) at Infinite Limit and Those Obtained for Experiments (Exp.) with Networks of $N = 100\,000$ Nodes

	Continuous power-law		Poisson		Discrete power-law		Poisson	
α	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.
2.5	1	0.93	0.85	0.83	1	0.89	0.72	0.69
3	1	0.80	0.61	0.57	1	0.54	0.49	0.44

of the stubs in the network were attached to the removed nodes. Each stub is kept with probability $(1-p)$; the fraction of pairs of stubs linking nonremoved nodes is therefore $(1-p)^2$. This last quantity is the fraction of nonremoved links and $1-(1-p)^2 = 2p-p^2$ is finally the fraction of removed links. □

We can now use this result to study the threshold for random node failures in terms of the fraction of removed links.

COROLLARY 3.10. *The fraction of links removed at the threshold p_c for random node failures in large random networks with degree distribution p_k is*

$$m(p_c) = 2p_c - p_c^2 = 1 - \left(\frac{\langle k\rangle}{\langle k^2\rangle - \langle k\rangle}\right)^2.$$

The application of these results to the cases of interest is then straightforward. We plot numerical evaluations of the obtained results in Figure 5, together with experimental results. We also give in Table III the thresholds for specific values of the exponent and the average degree.

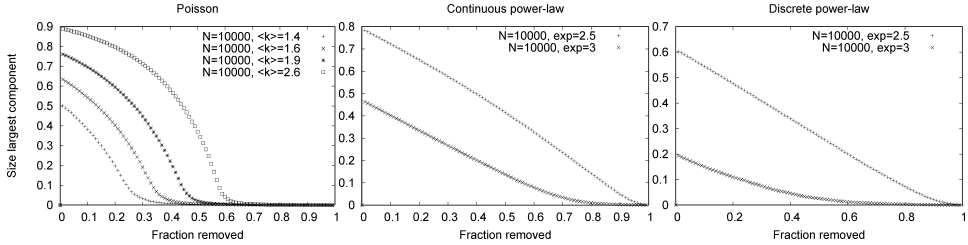

Fig. 6. Size of the largest connected component as a function of the fraction of randomly removed links. For technical details on our plots, see Section 2.4.

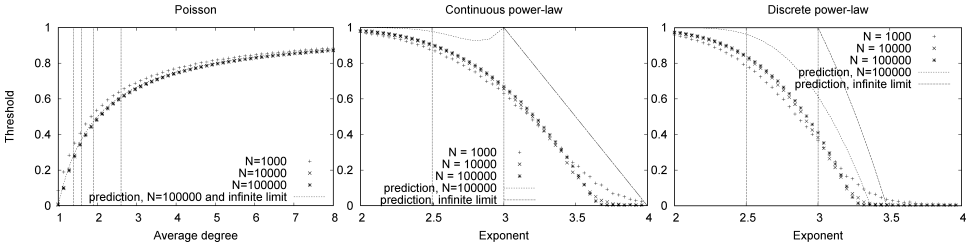

Fig. 7. Thresholds for random link failures. For technical details on our plots and on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 2.4.

As expected, these results are not qualitatively different from what is observed from the node point of view. Again, power-law networks are more resilient than Poisson ones, but the difference in practice is not as important as in the predictions.

Notice also that the fraction of removed links is significantly larger at the threshold than the fraction of removed nodes. This is a simple consequence of the fact that removing a node leads to the removal of both its stubs and some of its neighbors.

3.3. Random Link Failures

Until now we observed the behavior of random network when *nodes* are randomly removed. One may wonder what happens when we remove *links* at random. This may model link failures, just like the random removal of nodes models node failures.

Typical behaviors are plotted in Figure 6. Just like in the case of random node failures (see Figure 2), there is a qualitative difference between Poisson and power-law networks. Going further, the plots are very similar to the ones for node failures. We will see that the formal results for both cases are indeed identical.

We give here the threshold m_c for random link failures which is actually the same as the one for random *node* failures (see Theorem 3.1).

THEOREM 3.11 (CALLAWAY ET AL. 2000; COHEN ET AL. 2001a). *The threshold m_c for random link failures in large random networks with degree distribution p_k is*

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

Again there are two proofs for this result, similar to the ones sketched in Section 3.1. We plot numerical evaluations of the obtained results in Figure 7, together with experimental results. We also give in Table IV the thresholds for specific values of the exponent and the average degree.

Table IV. Values of Threshold for Random Link Failures on Discrete and Continuous Power-law Networks of Exponents 2.5 and 3, and on Poisson Networks Having Same Average-Degree (See Table I); Values Are Analytic Previsions (Prev.) at Infinite Limit and Those Obtained for Experiments (Exp.) with Networks of $N = 100\,000$ Nodes

	Continuous power-law		Poisson		Discrete power-law		Poisson	
α	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.
2.5	1	0.90	0.62	0.60	1	0.84	0.47	0.46
3	1	0.67	0.38	0.35	1	0.41	0.29	0.27

In principle, these plots and values should be exactly the same as the ones in Figure 3 and in Table II. This is true for the analytic previsions, but experiments differ significantly, which deserves more discussion.

When we consider the size of the largest connected component as a function of the fraction of removed nodes/links (see Figures 2 and 6) then it appears clearly that, though the plot seems to reach zero at the same fraction, they do not have the same shape. Since we chose to define the threshold as the value for which the largest connected component reaches 5 % of the total number of nodes, the different shapes give different experimental thresholds.

Finally, the same conclusions as the ones for random node failures hold: power-law networks are more resilient to random link failures than Poisson ones, but the difference in practice is not as striking as predicted by the results for the infinite limit.

3.4. Conclusion on Random Failures

Two main formal conclusions have been reached in this section concerning the case where the size of the network tends toward infinity. First, as expected from the empirical results discussed in the Introduction, Poisson and power-law networks behave qualitatively differently in case of (node or link) random failures: whereas Poisson networks display a clear threshold, in power-law ones all the nodes or links have to be removed to achieve a breakdown. Second, random link failures are very similar, if not identical, to random node failures. On the other hand, link point of view does not change the observations qualitatively but the fraction of removed links at the threshold is significantly larger than the fraction of removed nodes. This also means that the thresholds for the link point of view of random node failures are larger than the thresholds for random link failures.

The qualitative difference between Poisson and power-law networks leads to the conclusion that power-law networks are much more resilient to random failures.

These results, however, concern only the limit case where the size of the network tends toward infinity. For networks of a given size N, even for very large values of N, the difference between Poisson and power-law networks often is much less striking than predicted. This is even clearer for the link point of view.

4. RESILIENCE TO ATTACKS

The aim of this section is to study the resilience of random networks to targeted attacks. In our context, an attack consists of node or link removals which are *not* random anymore; instead, the nodes or links are chosen according to a *strategy*.

Obviously, one may define many strategies and we already presented one, defined in the initial paper [Albert et al. 2000]: it consists of the removal of nodes in decreasing order of their degree. We will call this strategy a *classical attack*.

We will first consider these classical attacks (Section 4.1) on general random networks, and then apply the obtained results to Poisson and power-law networks. In order to deepen our understanding, we will consider these attacks from the *link* point

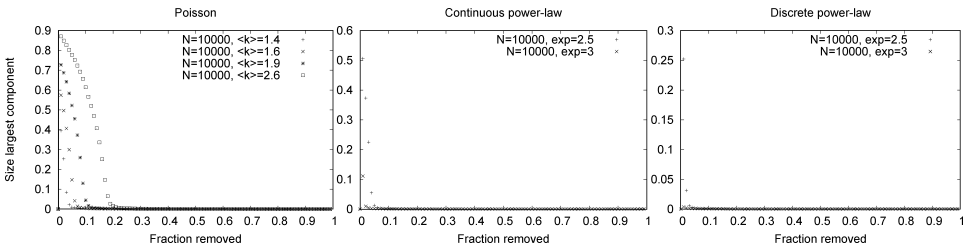

Fig. 8. Size of the largest connected component as a function of the fraction of nodes removed during classical attacks. For technical details on our plots, see Section 2.4.

of view (Section 4.2). We will also introduce new attack strategies (both on nodes and on links) to provide a deeper understanding of classical attacks (Section 4.3).

4.1. Classical Attacks

In this section, we first present a general result on classical attacks, independent of the type of underlying network, as long as it is a *random* network. We then apply this general result to the special cases under concern. Figure 8 displays the behaviors observed for these three types of networks.

As explained in the preliminaries, there is no fundamental difference in the behaviors of Poisson and power-law networks in case of classical attacks: in both cases the largest connected component is quickly destroyed. It is important, however, to notice that Poisson networks are significantly more resilient than power-law ones. The aim of this section is to formally confirm these observations, and give both formal and intuitive explanations.

THEOREM 4.1 (CALLAWAY ET AL. 2000; COHEN ET AL. 2001a). *The threshold p_c for classical attacks in random networks, with size tending toward infinity and degree distribution p_k, is given by*

$$\frac{\sum_{k=0}^{K(p_c)} k(k-1)p_k}{\langle k \rangle} = 1,$$

where $K(p_c)$ is the maximal degree in the network after the attack, related to p_c by Lemma 4.2.

As in the case of failures, there are two main ways to derive this result, proposed in Callaway et al. [2000] and Newman [2003a] and Cohen et al. [2001a, 2003b]. They both rely on the following result.

LEMMA 4.2 (COHEN ET AL. 2001a). *In a random network with size tending toward infinity and degree distribution p_k, after removal of a fraction p of the nodes during a classical attack, the maximal degree $K(p)$ is given by*

$$p = 1 - \sum_{k=0}^{K(p)} p_k.$$

In order to compute the threshold for random networks with a given degree distribution and size tending toward infinity, one therefore has to first compute the value of $K(p_c)$ using Theorem 4.1, then obtain the value of p_c using Lemma 4.2. Note that we will mainly use Lemma 4.2 to compute a fraction of removed nodes given a maximal degree and not the converse.

We now outline the two main proofs available for Theorem 4.1.

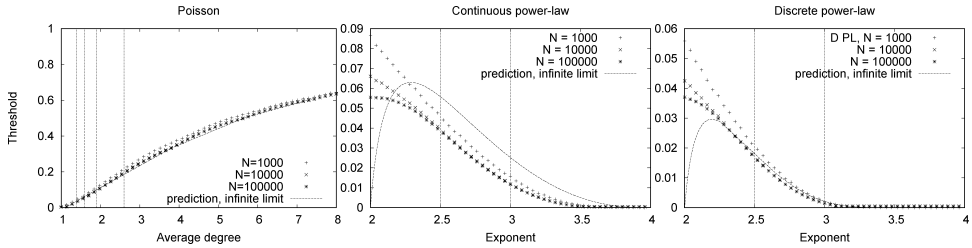

Fig. 9. Thresholds for classical attacks. For technical details on our plots and on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 2.4.

The first one, from Cohen et al. [2001a, 2003b], is based on the fact that the network obtained after the removal of a fraction p of the nodes during a classical attack is equivalent to a random network on which random link failures occurred.

Indeed, a classical attack has two kinds of effects: it reduces the maximal degree in the network by removing the nodes with highest degree, and it removes the links attached to these nodes. A classical attack removes all the stubs attached to the removed nodes and some *other* stubs, which were linked to removed stubs. Since pairs of stubs are linked randomly, this is equivalent to randomly removing the correct number of stubs from the subnetwork composed of the nodes which are not removed. If the classical attack removes a fraction p of the nodes, the fraction of stubs attached to removed nodes is $s(p) = \frac{1}{\langle k \rangle} \sum_{k=K(p)+1}^{\infty} k p_k$. The probability for any given stub of a remaining node to be linked to a stub of a removed node, and therefore its probability to be also removed, is $s(p)$.

Since links in this subnetwork are constructed by choosing random pairs of stubs, it is also a random network. Moreover, its degree distribution is nothing but the original one with a cutoff (which is the maximal degree after the attack).

Finally, a classical attack is equivalent to random link failures on a random network with known degree distribution. The value of the threshold can therefore be derived from Theorem 3.11 on random link failures. It is then possible to derive from this the result of Theorem 4.1.

The other proof [Callaway et al. 2000; Newman 2003a] relies on generating functions: the fraction p of nodes of highest degrees are marked as *absent*, and the others are marked as *present*.

Recall that $F_1(x)$ is the generating function for the probability of finding an unmarked (i.e., present) node with k other (marked or unmarked) neighbors at the end of a randomly chosen link. In our case, $F_1(x)$ therefore is

$$F_1(x) = \frac{1}{\langle k \rangle} \sum_{k=1}^{K(p)} k p_k x^{k-1}.$$

Theorem 4.1 is a direct consequence of Theorem 2.13 and its application to the cases of interest is straightforward. Numerical evaluations of the obtained results can be done using Lemma 4.2. We plot these evaluations in Figure 9, together with experimental results. We also give in Table V the thresholds for specific values of the exponent and the average degree.

It appears clearly that both types of networks are very sensitive to classical attacks: only a few percent of the nodes have to be removed to destroy them. Moreover, the thresholds for power-law networks are much lower than the ones for Poisson networks: they are almost one order of magnitude smaller than for comparable Poisson networks.

Table V. Values of Threshold for Classical Attacks on Discrete and Continuous Power-law Networks of Exponents 2.5 and 3, and on Poisson Networks Having Same Average-Degree (See Table I); Values Are Analytic Previsions (Prev.) at Infinite Limit and Those Obtained for Experiments (Exp.) with Networks of $N = 100\,000$ Nodes

	Continuous power-law		Poisson		Discrete power-law		Poisson	
α	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.
2.5	0.056	0.038	0.18	0.19	0.018	0.017	0.08	0.09
3	0.025	0.012	0.05	0.05	0.002	0.0015	0.03	0.035

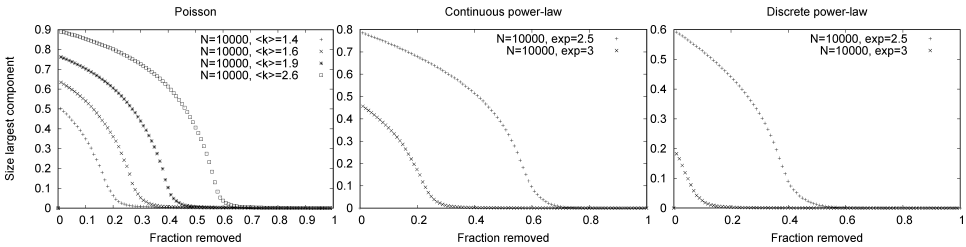

Fig. 10. Size of the largest connected component as a function of the fraction of *links* removed during classical attacks. For technical details on our plots, see Section 2.4.

These are certainly the main results on the topic and we will deepen them in the rest of the section.

4.2. Link Point of View of Classical Attacks

The classical attack strategy removes highest-degree nodes first. Since in a power-law network there are some very-high-degree nodes, this leads in this case to the removal of a huge number of links. One may then wonder if its efficiency on power-law networks is due to the fact that the number of removed links is much larger than in the case of random failures. Likewise, one may wonder if the fact that a classical attack removes many more links in a power-law network than in a Poisson one is the cause of its difference of efficiency on these two types of networks. These explanations actually have been proposed by some authors as an intuitive explanation of the results presented above.

Figure 10 displays the behaviors observed for these three types of networks. One can see there that the thresholds for Poisson and power-law networks are much closer than from the node point of view (see Figure 8). One may also observe that, though there are significant differences, when one removes from power-law networks as many links as needed to destroy a Poisson network with the same average degree, then the size of the largest connected component becomes very small.

It is possible to obtain the following general result.

THEOREM 4.3. *In a large random network, the fraction $m(p)$ of links removed when a fraction p of the nodes have been removed during a classical attack is*

$$m(p) = 2s(p) - s(p)^2,$$

where $s(p) = \frac{1}{\langle k \rangle} \sum_{k=K(p)+1}^{\infty} kp_k$ is the fraction of stubs attached to removed nodes.

Theorem 4.3 is valid for any random network, whatever its degree distribution. It makes it possible to compute the fraction of links removed at the threshold for classical attacks. To study the behavior of Poisson and power-law networks, we therefore only have to apply it to these cases. In each case, one first has to compute the (node)

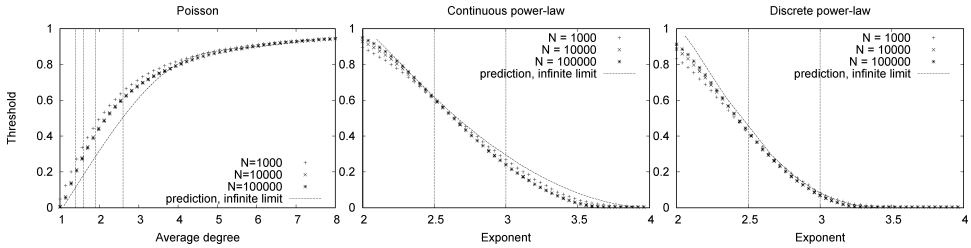

Fig. 11. Thresholds for the link point of view of classical attacks. For technical details on our plots and on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 2.4.

Table VI. Values of Threshold for Link Point of View of Classical Attacks on Discrete and Continuous Power-law Networks of Exponents 2.5 and 3, and on Poisson Networks Having Same Average-Degree (See Table I); Values Are Analytic Previsions (Prev.) at Infinite Limit and Those Obtained for Experiments (Exp.) with Networks of $N = 100\,000$ Nodes

	Continuous power-law		Poisson		Discrete power-law		Poisson	
α	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.	Prev.	Exp.
2.5	0.62	0.6	0.5	0.6	0.45	0.42	0.28	0.4
3	0.3	0.24	0.15	0.28	0.08	0.07	0.1	0.2

threshold p_c for classical attacks, then compute $s(p_c) = \frac{1}{\langle k \rangle} \sum_{k=K(p_c)+1}^{\infty} k p_k$, before applying Theorem 4.3.

We plot numerical evaluations of these results in Figure 11, together with experimental results. We also give in Table VI the thresholds for specific values of the exponent and the average degree.

The results are striking: the thresholds are much larger from the link point of view than from the node point of view (see Table V for comparison). More importantly, while the number of nodes to be removed is much lower for power-law networks than for Poisson ones, the corresponding number of links is similar for both kinds of networks: the links thresholds are similar.

The conclusion from these observations is that the fact that power-law networks are rapidly destroyed during classical attacks may be viewed as a consequence of the fact that many links are removed. It is, however, important to notice that the obtained behavior for power-law networks is not the same as the one obtained if we remove the same number of links at random (see Figure 7 and Table IV for comparison). Therefore, although the number of removed links is huge and this plays a role in the behavior of power-law networks, this is not sufficient to explain the observed behavior. This means that the links attached to highest-degree nodes play a more important role regarding the network connectivity than random links.

4.3. New Attack Strategies

In this section we introduce two very simple new attack strategies, one targeting nodes (Section 4.3.1) and the other targeting links (Section 4.3.2). These strategies are close to random failures. Our aim is not to provide efficient attack strategies, but rather to deepen our understanding of previous results.

These two new attack strategies rely on the following observation. We have seen (Theorem 2.9) that a random network with size tending toward infinity has a giant component if $\langle k^2 \rangle - 2\langle k \rangle > 0$. This is equivalent to the condition $p_1 < \sum_{k=3}^{\infty} k(k-2)p_k$. The fraction of nodes of degree 1 in the network therefore plays a key role. The two

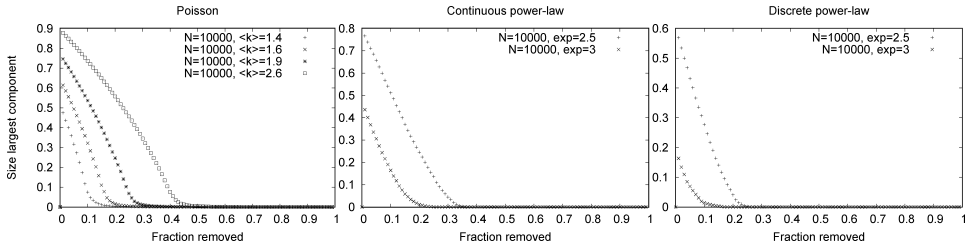

Fig. 12. Size of the largest connected component as a function of the fraction of nodes removed during almost-random node attacks. For technical details on our plots, see Section 2.4.

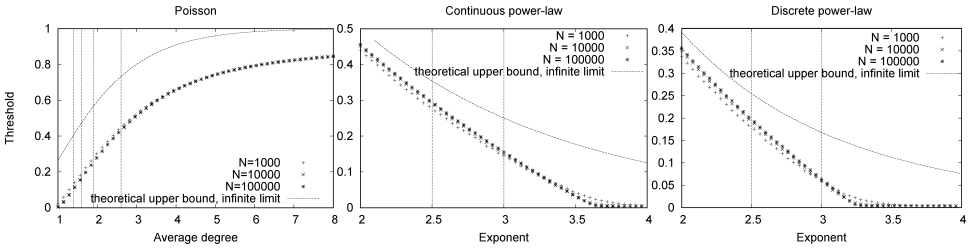

Fig. 13. Thresholds and upper bounds for almost-random node attacks. For technical details on our plots and on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 2.4.

attack strategies are based on the idea that increasing this fraction should quickly break the network.

Since our aim here is not to compute the exact value of the threshold, but rather to understand a general behavior, we will only consider in the sequel the case of networks with size tending toward infinity.

4.3.1. Almost-Random Node Attacks. The first attack strategy simply consists of randomly removing nodes of degree at least 2. We call it the *almost-random node attack* strategy. Figure 12 displays the behaviors observed for the three types of networks under consideration.

Although this strategy is barely different from random node failures, it is actually much more efficient than random node failures (see Figure 2 for comparison). In particular, it has a finite threshold for all the types of networks we consider.

THEOREM 4.4. *The threshold p_c for almost-random node attacks for large random networks with degree distribution p_k is bounded by*

$$p_c < 1 - p_1 - p_0.$$

The extension of this theorem to the three types of networks we consider is straightforward. We plot experimental results for the value of the threshold in Figure 13, as well as the upper bounds given above. We also give in Table VII the thresholds for specific values of the exponent and the average degree.

We recall that our aim here is not to obtain an efficient attack strategy, but to study the ability of a strategy very similar to random failures to have the same qualitative behavior as classical attacks, namely, to display a finite threshold for power-law networks.

In this regard, the values of the thresholds displayed in Table VII are quite large (one has to remove a large fraction of the nodes to destroy the networks), but remain significantly lower than 1 and much lower than the thresholds for node failures (see

Table VII. Values of Threshold for Almost-Random Node Attacks on Discrete and Continuous Power-Law Networks of Exponents 2.5 and 3, and on Poisson Networks Having Same Average Degree (see Table I); Values Are Ones Obtained for Experiments (Exp.) with Networks of $N = 100\,000$ Nodes, and Theoretical Upper Bounds

	Continuous power-law		Poisson		Discrete power-law		Poisson	
α	Bound	Exp.	Bound	Exp.	Bound	Exp.	Bound	Exp.
2.5	0.35	0.29	0.73	0.43	0.25	0.2	0.57	0.25
3	0.25	0.16	0.48	0.16	0.17	0.06	0.41	0.10

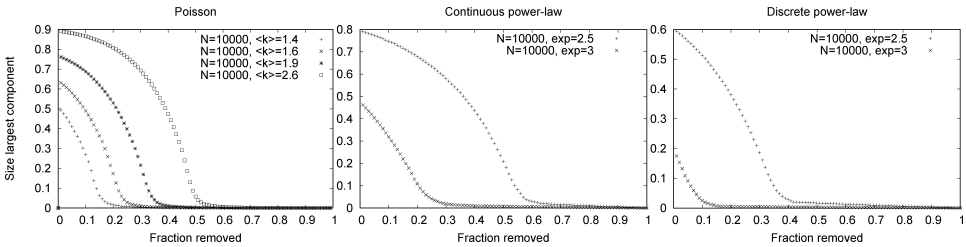

Fig. 14. Size of the largest connected component as a function of the fraction of links removed during almost-random link attacks. For technical details on our plots, see Section 2.4.

Table VIII. Values of Threshold for almost-random link attacks on Discrete and Continuous Power-Law Networks of Exponents 2.5 and 3, and on Poisson Networks Having Same Average Degree (see Table I); Values Are Ones Obtained for Experiments (Exp.) with Networks of $N = 100\,000$ Nodes, and Theoretical Upper Bounds

	Continuous power-law		Poisson		Discrete power-law		Poisson	
α	Bound	Exp.	Bound	Exp.	Bound	Exp.	Bound	Exp.
2.5	0.62	0.55	0.86	0.51	0.37	0.35	0.72	0.32
3	0.39	0.22	0.64	0.23	0.15	0.07	0.57	0.15

Table II). This shows that the efficiency of classical attacks relies in part on simple properties like removing nodes of degree larger than 1.

4.3.2. Almost-Random Link Attacks. The other attack strategy consists of randomly removing links between nodes of degree at least 2. We call it the *almost-random link attack* strategy. Figure 14 displays the behaviors observed for the three types of networks under concern.

Although this strategy is barely different from random link failures, it is actually much more efficient than random link failures (see Figure 6 for comparison). In particular, it has a finite threshold for all the types of networks we consider, which makes it much more efficient on power-law networks:

THEOREM 4.5. *The threshold m_c for the almost-random link attack strategy for large random networks with maximal degree sublinear in the number of nodes and degree distribution p_k is bounded by*

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

The extension of this theorem to the three types of networks we consider is straightforward. We plot experimental results for the value of the threshold in Figure 15, as well as the upper bounds given above. We also give in Table VIII the thresholds for specific values of the exponent and the average degree.

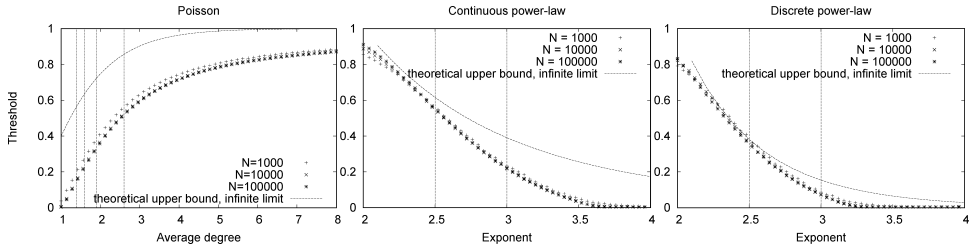

Fig. 15. Thresholds and upper bounds for almost-random link attacks. For technical details on our plots and on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 2.4.

As in the case of almost-random node attacks, the values of the thresholds are quite large but remain significantly lower than 1. Since our aim is still to study the ability of a strategy very similar to random failures to display a finite threshold, this result is satisfactory. This shows that the efficiency of classical attacks relies in part on simple properties like removing links between nodes of degree at least 2.

Going further, one may notice that almost-random link attacks perform better than classical attacks in terms of the number of removed links (see Table VI). This shows that classical attacks, although they focus on high-degree nodes, actually remove many links connected to nodes of degree 1, which play little role in the connectivity of the network. The simple almost-random strategy, on the other hand, focuses on those links which really disconnect the network.

4.4. Conclusion on Attacks

There are two main formal conclusions for this section. First, as expected from the empirical results discussed in the Introduction, power-law networks are very sensitive to classical attacks, much more than Poisson networks. Second, the link point of view shows that many links are actually removed when the thresholds for classical attacks are reached. Moreover, very simple attack strategies close to random node or link failures also lead to finite (and reasonably small) thresholds.

Altogether, these results make it possible to discuss precisely the efficiency of classical attacks. First, although the number of links removed during such attacks is huge, this is not sufficient to explain the collapse of the network. Indeed, the removal of the same number of links at random does not collapse the network. Second, the number of removed links during classical attacks in a Poisson network and in a power-law network are very similar. This moderates the conclusion that power-law networks are particularly sensitive to classical attacks, since in terms of links both are equally robust.

Finally, the almost random attack strategies we introduced show that the efficiency of classical attacks relies strongly on simple properties like removing nodes of degree larger than 1 and links between nodes of degree at least 2.

5. CONCLUSION AND DISCUSSION

In this contribution, we focused on a set of previously known results which received much attention in the last few years [Albert et al. 2000; Broder et al. 2000; Cohen et al. 2000, 2001a, 2003b; Callaway et al. 2000; Newman 2003a]. These analytical and empirical results state that, although power-law networks are very resilient to random (node or link) failures and Poisson ones are not, they are very sensitive to a special type of attack (which we call *classical attacks*) consisting of removing the highest-degree nodes first, while Poisson networks are not. This had led to the conclusion that

its power-law degree distribution may be seen as an *Achille's heel of the Internet* [Nature 2000; Barabási 2003].

Our first contribution is to give a unified and complete presentation of these results. Our second contribution is to introduce some new results for cases which received less attention, maybe because these results are less striking. They are, however, essential for deepening one's understanding of this topic. We focused in particular on two aspects: studying the finite case, and studying the link point of view of random node failures, and classical attacks. We also introduced new attacks, very similar to random failures, which allowed us to deepen our understanding of the phenomena at play. Finally, we conducted extensive simulations on random graphs of different types in order to confront them to theoretical results.

All this showed that many of the classical conclusions of the field should be discussed further. We may now put all these results and their relations together to derive global conclusions.

Concerning random node and link failures (i.e., random node and link removals), the striking point is that, although analysis predict completely different behaviors for Poisson and power-law networks, in practice the differences, though important, are not huge (see Tables II to IV). This is even more pronounced for link failures. This overestimation of the difference was due to the study of the infinite limit and to the approximations made. It may also be a consequence of our choice to consider that a network breakdown occurs when the size of the the giant component reaches 5% of all nodes, but other conventions lead to similar conclusions.

Concerning classical attacks (i.e., removal of nodes in decreasing order of their degree), we have shown that, although the thresholds for power-law networks indeed are very low, and much smaller than the ones for Poisson networks, our other observations tend to moderate this conclusion. Indeed, the number of links removed during a classical attack is huge. When one considers the number of removed links, power-law networks are not more fragile than Poisson ones.

The large number of removed links, though it clearly plays a role, is, however, not sufficient to explain the efficiency of classical attacks: if one removes the same fraction of links randomly, then there is no breakdown. This invalidates the often claimed explanation that classical attacks are very efficient on power-law networks because they remove many links.

Going further, if one removes the same, or even a smaller, fraction of links, but *almost* randomly (i.e., randomly among the ones which are linked to nodes of degree at least 2) then a breakdown occurs. In terms of the fraction of removed links, classical attacks therefore lie between random link failures and almost-random link attacks, which makes them not so efficient.

Finally, the efficiency of classical attacks resides mainly in the fact that they remove many links, which are mostly attached to nodes of degree larger than 1. Conversely, this explains the robustness of power-law networks to random node failures: such failures often remove nodes of degree 1 and links attached to such nodes.

Another conclusion of interest comes from the study of classical attacks on Poisson networks (which was not done in depth until now). Although these networks behave similarly in the cases of random node failures and classical attacks, it must be noted that their threshold is significantly lower in the second case. This goes against the often claimed assumption that, because all nodes have almost the same degree in a Poisson network, there is little difference between random node failures and classical attacks. This is worth noting, since it reduces the difference, often emphasized, in the behavior of Poisson and power-law networks.

All these results led us to the conclusion that, although random node failures and classical attacks clearly behave differently and though the Poisson or power-law nature

of the network has a strong influence in this, one should be careful in deriving conclusions. This is confirmed by our experiments on real-world networks. The sensitivity of networks to attacks relies less on the presence of high-degree nodes than on the fact that they have many low-degree nodes. Conversely, their robustness to failures relies strongly on the fact that, when we choose a node at random, we choose such a node with high probability, and not so much on the fact that high-degree nodes hold the network together. Moreover, the fact that a classical attack on a power-law network removes many links may be considered as partly, but not fully, responsible for the network's rapid breakdown.

Although this article is already quite long, we had to make some choices in the results presented, and there are of course many omissions. We did not mention the various contributions considering other attack strategies and other definitions of the robustness than the size of the giant component. We also ignored random networks with degree correlations, or other types of modeling. For instance, the fact that nodes of real-world complex networks are organized in communities (groups of densely connected nodes) plays a key role. This is also captured in part by the notion of *clustering coefficient* [Blondel et al. 2008; Gibson et al. 1998; Kumar et al. 1999; Flake et al. 2000, 2002; Girvan and Newman 2002; Newman 2004; Latapy and Pons 2006]. Studying robustness of networks with more subtle properties than degree distributions would therefore be highly relevant, but most of the work on that remains to be done.

Finally, let us insist once more on the necessity of developing formal results to enhance our understanding of empirical results. There is no doubt that experiments provide much understanding and intuition about phenomena of interest. The need for rigor and for a deeper understanding of what happens during these experiments is, however, strong. It led to several approaches to analyzing them. The main ones in our context were developed in Cohen et al. [2000, 2001a, 2003b], Callaway et al. [2000], and Newman [2003a]. They all rely on mean-field approximations, and we gave here the details of the underlying approximations and assumptions. Such an approach is definitively rigorous, but is not *formal*. Obtaining exact results, or even approximate results, with formal methods would be another improvement. Some results in that direction have begun to appear [Bollobás and Riordan 2003], but much remains to be done and the task is challenging.

ACKNOWLEDGMENTS

We thank the anonymous referees for taking the time to read this article in depth and making very valuable comments for improving it. We thank Fabien Viger for valuable comments on degree distributions.

REFERENCES

Adamic, L. and Huberman, B. 2000. Power-law distribution of the world wide web. *Science 287*, 2115.

Aiello, W., Chung, F., and Lu, L. 2000. A random graph model for massive graphs. In *Proceedings of the ACM Symposium on Theory of Computing*. 171–180.

Albert, R. and Barabási, A.-L. 1999. Emergence of scaling in random networks. *Science 286*, 509–512.

Albert, R. and Barabási, A.-L. 2002. Statistical mechanics of complex networks. *Rev. Mod. Phys. 74*, 47–97.

Albert, R., Jeong, H., and Barabási, A.-L. 1999. Diameter of the World Wide Web. *Nature 401*, 130–131.

Albert, R., Jeong, H., and Barabási., A.-L. 2000. Error and attack tolerance in complex networks. *Nature 406*, 378–382.

Barabási, A.-L. 2003. Emergence of scaling in complex networks. In *Handbook of Graphs and Networks: From the Genome to the Internet*, S. Bornholdt and H. G. Schuster, Eds. Wiley-VCH, Weinheim, Germany.

Bender, E. A. and Canfield, E. R. 1978. The asymptotic number of labelled graphs with given degree sequences. *J. Combin. Theor. (A) 24*, 357–367.

Blondel, V., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. 2008. Fast unfolding of communities in large networks. *J. Stat. Mech.* 10, 10008+.

Boguná, M., Pastor-Satorras, R., and Vespignani, A. 2003. Epidemic spreading in complex networks with degree correlations. In *Statistical Mechanies of Complex Networks*, R. Pastor-Satorras, M. Rubi, and A. Diaz-Guelera, Eds. Lecture Notes in Physics, vol. 625. Springer, Berlin, Germany, 127–147.

Bollobás, B. 1985. *Random Graphs*. Academic Press, New York, NY.

Bollobás, B. and Riordan, O. 2003. Robustness and vulnerability of scale-free random graphs. *Internet Math. 1*, 1, 1–35.

Broder, A., Kumar, S., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. 2000. Graph structure in the web. *Comput. Netw. 33,* 1-6, 309–320.

Broido, A. and Claffy, K. 2001. Internet topology: Connectivity of IP graphs. In *Proceedings of the SPIE International symposium on Convergence of IT and Communication.*

Bu, T. and Towsley, D. 2002. On distinguishing between Internet power law topology generators. In *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Socities.*

Burda, Z. and Krzywicki, A. 2003. Uncorrelated random networks. *Phys. Rev. E 67,* 046118.

Callaway, D., Newman, M., Strogatz, S., and Watts, D. 2000. Network robustness and fragility: Percolation on random graphs. *Phys. Rev. Lett. 85*, 5468–5471.

Chang, H., Jamin, S., and Willinger, W. 2001. Inferring AS-level Internet topology from router-level path traces. In *Proceedings of SPIE ITCom*.

Chen, Q., Chang, H., Govindan, R., Jamin, S., Shenker, S., and Willinger, W. 2002. The origin of power laws in Internet topologies revisited. In *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies*.

Chung, F. and Lu, L. 2002. Connected compnents in a random graph with given degree sequences. *Ann. Combinator. 6*, 125–145.

Cohen, R., Erez, K., ben Avraham, D., and Havlin, S. 2000. Resilience of the Internet to random breakdown. *Phys. Rev. Lett. 85*, 4626.

Cohen, R., Erez, K., ben Avraham, D., and Havlin, S. 2001a. Breakdown of the Internet under intentional attack. *Phys. Rev. Lett. 86*, 3682–3685.

Cohen, R., Erez, R., ben Avraham, D., and Havlin, S. 2001b. Reply to the comment on 'Breakdown of the Internet under intentional attack.' *Phys. Rev. Lett. 87*, 219802.

Cohen, R., Havlin, S., and ben Avraham, D. 2003a. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett. 91,* 247901.

Cohen, R., Havlin, S., and ben Avraham, D. 2003b. Structural properties of scale-free networks. In *Hankbook of Graphs and Networks: From the Genome to the Internet*, S. Bornholdt and H. G. Schuster, Eds. Wiley-VCH, Weinheim, Germany.

Crucitti, P., Latora, V., and Marchiori, M. 2004a. Error and attack tolerance of complex networks. *Physica A 340*, 388–394.

Crucitti, P., Latora, V., and Marchiori, M. 2004b. A model for cascading failures in complex networks. *Phys. Rev.* E *69,* 045104.

Crucitti, P., Latora, V., Marchiori, M., and Rapisarda, A. 2003. Efficiency or scale-free networks: Error and attack tolerance. *Physica A 320*, 622–642.

Csete, M. and Doyle, J. 2002. Reverse engineering of biological complexity. *Science 295*, 5560, 1664–1669.

Dezsö, Z. and Barabási, A.-L. 2002. Halting viruses in scale-free networks. *Phys. Rev. E 65,* 055103.

Dorogovtsev, S. and Mendes, J. 2001. Comment on 'Breakdown of the Internet under intentional attack.' *Phys. Rev. Lett. 87*.

Dorogovtsev, S. and Mendes, J. 2002. Evolution of networks. *Adv. Phys. 51*, 1079–1187.

Dorogovtsev, S., Mendes, J., and Samukhin, A. 2000. Structure of growing networks with preferential linking. *Phys. Rev. Lett. 85*, 4633–4636.

Doyle, J., Alderson, D., Li, L., Low, S., Roughan, M., Shalunov, S., Tanaka, R., and Willinger, W. 2005. The "Robust Yet Fragile" nature of the Internet. *Proc. Nat. Acad. Sci. 102*, 40, 14123–14475.

Ebel, H., Mielsch, L.-I., and Bornholdt, S. 2002. Scale-free topology of e-mail networks. *Phys. Rev. E 66,* 035103.

Erdös, P. and Rényi, A. 1959. On random graphs I. *Publ. Math. Debrecen 6*, 290–297.

Faloutsos, M., Faloutsos, P., and Faloutsos, C. 1999. On power-law relationships of the Internet topology. In *Proceedings of the ACM SIGCOMM Data Communications Festival*. 251–262.

Farkas, I., Derényi, I., Jeong, H., Neda, Z., Oltvai, Z., Ravasz, E., Schrubert, A., and Barabasi, A. 2003. The topology of the transcription regulatory network in the yeast *saccharomyces cerevisiae*. *Physica A 318*, 601–612.

FLAKE, G., LAWRENCE, S., AND GILES, C. L. 2000. Efficient identification of Web communities. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 150–160.

FLAKE, G., LAWRENCE, S., GILES, C. L., AND COETZEE, F. 2002. Self-organization of the Web and identification of communities. *IEEE Compt. 35,* 3, 66–71.

GIBSON, D., KLEINBERG, J., AND RAGHAVAN, P. 1998. Inferring Web communities from link topology. In *Proceedings of the U.K. Conference on Hypertext*. 225–234.

GIRVAN, M. AND NEWMAN, M. 2002. Community structure in social and biological networks. *Proc. Nat. Acad. Sci. 99*, 7821–7826.

GOVINDAN, R. AND TANGMUNARUNKIT, H. 2000. Heuristics for Internet map discovery. In *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies*. 1371–1380.

HOLME, P. 2004. Efficient local strategies for vaccination and network attack. *Europhys. Lett. 68,* 6, 908–914.

HOLME, P. AND KIM, B. J. 2002. Growing scale-free networks with tunable clustering. *Phys. Rev. E 65,* 026107.

HOLME, P., KIM, B. J., YOON, C. N., AND HAN, S. K. 2002. Attack vulnerability of complex networks. *Phys. Rev. E 65,* 056109.

JEONG, H., TOMBOR, B., ALBERT, R., OLTVAI, Z., AND BARABÁSI, A.-L. 2000. The large-scale organization of metabolic networks. *Nature, 407, 651*.

JIN, E., GIRVAN, M., AND NEWMAN, M. 2001. The structure of growing social networks. *Phys. Rev. E 64,* 046132.

KLEINBERG, J. 2000. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the Annual ACM Symposium on Theory of Computing*.

KLEMM, K. AND EGUILUZ, V. 2002. Highly clustered scale-free networks. *Phys. Rev. E 65,* 036123.

KOHN, K. 1999. Molecular interaction map of the mammalian cell cycle control and DNA repair system. *Mol. Biol. Cell 10*, 2703–2734.

KUMAR, S. R., RAGHAVAN, P., RAJAGOPALAN, S., AND TOMKINS, A. 1999. Trawling the Web for emerging cyber-communities. *WWW8/Comp. Netw. 31,* 11-16, 1481–1493.

LATAPY, M. AND PONS, P. 2006. Computing communities in large networks using random walks. *J. Graph Alg. Appl. 10*, 2, 191–218.

LATORA, V. AND MARCHIORI, M. 2001. Efficient behavior of small-world networks. *Phys. Rev. Lett. 87,* 198701.

LAURA, L., LEONARDI, S., MILLOZZI, S., MEYER, U., AND SIBEYN, J. 2003. Algorithms and experiments for the Webgraph. In *Proceedings of the European Symposium on Algorithms*.

LEE, E., GOH, K.-I., KAHNG, B., AND KIM, D. 2005. Robustness of the avalanche dynamics in data packet transport on scale-free network. *Phys. Rev. E 71,* 256108.

LI, L., ALDERSON, D., DOYLE, J., AND WILLINGER, W. 2006. Towards a theory of scale-free graphs: Definition, properties, and implications. *Internet Math. 2*, 4, 431–523.

LI, L., ALDERSON, D., WILLINGER, W., AND DOYLE, J. 2004. A first-principles approach to understanding the Internet's router-level topology. In *Proceedings of the ACM SIGCOMM Data Communications Festival*.

LILJEROS, F., EDLING, C., AMARAL, L. N., STANLEY, H. E., AND ABERG, Y. 2001. The web of human sexual contacts. *Nature 411*, 907–908.

MAGONI, D. AND PANSIOT, J.-J. 2001. Analysis of the autonomous system network topology. *ACM SIGCOMM Comp. Commun. Rev. 31,* 3, 26–37.

MOLLOY, M. AND REED, B. 1995. A critical point for random graphs with a given degree sequence. *Rand. Struct. Alg. 6*, 161–179.

MOLLOY, M. AND REED, B. 1998. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput. 7*, 3, 295–305.

MOTTER, A. 2004. Cascade control and defense in complex networks. *Phys. Rev. Lett. 93,* 098701.

MOTTER, A. AND LAI., Y.-C. 2002. Cascade-based attacks on complex networks. *Phys. Rev. E 66,* 065102.

Nature. 2000. Nature *406*, 6794, Cover. (July 27).

NEWMAN, M. 2001a. Scientific collaboration networks: I. Network construction and fundamental results. *Phys. Rev. E 64,* 016131.

NEWMAN, M. 2001b. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E 64,* 016132.

NEWMAN, M. 2002. Assortative mixing in networks. *Phys. Rev. Lett. 89,* 208701.

NEWMAN, M. 2003a. Random graphs as models of networks. In *Hankbook of Graphs and Networks: From the Genome to the Internet*, S. Bornholdt and H. G. Schuster, Eds. Wiley-VCH, Weinheim, Germany.

NEWMAN, M. 2003b. The structure and function of complex networks. *SIAM Rev. 45,* 2, 167–256.

NEWMAN, M. 2004. Fast algorithm for detecting community structure in networks. *Phys. Rev. E 69,* 066133.

NEWMAN, M., STROGATZ, S., AND WATTS, D. 2001. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E 64,* 026118.

NEWTH, D. AND ASH, J. 2004. Evolving cascading failure resilience in complex networks. In *Proceedings of the Asia Pacific Symposium on Intelligent and Evolutionary Systems*. 125–136.

PARK, S.-T., KHRABROV, A., PENNOCK, D., LAWRENCE, S., GILES, C. L., AND UNGAR, L. 2003. Static and dynamic analysis of the Internet's susceptibility to faults and attacks. In *Proceedings of the Annual Joint Conference of the IEEE Computer and Communications Societies*.

PASTOR-SATORRAS, R., VAZQUEZ, A., AND VESPIGNANI, A. 2001. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett. 87,* 258701.

PASTOR-SATORRAS, R. AND VESPIGNANI, A. 2002. Immunization of complex networks. *Phys. Rev. E 65,* 036104.

PERTET, S. AND NARASIMHAN, P. 2005. Handling cascading failures: The case for topology-aware fault tolerance. In *Proceedings of the IEEE First Workshop on Hot Topics in System Dependability*.

STROGATZ, S. 2001. Exploring complex networks. *Nature 410*, 268–276.

TANGMUNARUNKIT, H., DOYLE, J., GOVINDAN, R., JAMIN, S., SHENKER, S., AND WILLINGER, W. 2001. Does AS size determine degree in AS topology? *ACM Comp. Communu. Rev. 31*, 7–10.

TANGMUNARUNKIT, H., GOVINDAN, R., JAMIN, S., SHENKER, S., AND WILLINGER, W. 2002. Network topologies, power laws, and hierarchy. *ACM Comp. Commun. Rev. 32*, 76.

UETZ, P., GIOT, L., CAGNEY, G., MANSFIELD, T., JUDSON, R., KNIGHT, J., LOCKSHON, D., NARAYAN, V., SRINIVASAN, M., POCHART, P., QURESHI-EMILI, A., LI, Y., GODWIN, B., CONOVER, D., KALBFLEISCH, T., VIJAYADAMODAR, G., YANG, M., JOHNSTON, M., FIELDS, S., AND ROTHBERG, J. 2000. A comprehensive analysis of protein-protein interactions in *saccharomyces cerevisiae*. *Nature 403*, 623–627.

VÁZQUEZ, A. AND MORENO, Y. 2003. Resilience to damage of graphs with degree correlations. *Phys. Rev. E 67,* 015101.

WATTS, D. AND STROGATZ, S. 1998. Collective dynamics of small-world networks. *Nature 393*, 440–442.

WILF, H. 1994. *Generating Functionology*. Academic Press/Harcourt Brace, New York, NY.

WILLINGER, W., GOVINDAN, R., JAMIN, S., PAXSON, V., AND SHENKER, S. 2002. Scaling phenomena in the Internet: Critically examining criticality. *Proc. Nat. Acad. Sci. 99*, 2573–2580.

YOOK, S.-H., JEONG, H., AND BARABÁSI, A.-L. 2002. Modeling the Internet's large-scale topology. *Proc. Nat. Acad. Sci. 99*, 13382–13386.

ZHAO, L., PARK, K. AND LAI, Y.-C. 2004. Attack vulnerability of scale-free networks due to cascading breakdown. *Phys. Rev. E 70,* 035101.

Received September 2007; revised August 2009; accepted August 2009