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

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

Contact: magnien@shs.polytechnique.fr

Abstract

It appeared recently that the underlying degree distribution of a network may play a crucial role concerning its robustness. Empiric and analytic results have been obtained, based on asymptotic and mean-field approximations. Previous work insisted on the fact that power-law degree distributions induce a high resilience to random failure but a high sensitivity to some attack strategies, while Poisson degree distributions are quite sensitive in both cases. Then, much work has been done to extend these results.

We focus here on these basic results, with the aim of deepening significantly our understanding of their origin and their limitations. We review in detail previous contributions and we give full proofs in a unified framework, in which the approximations on which these results rely are well identified. We then add to these known results a set of new ones aimed at enlightening some important aspects. We also provide extensive and rigorous experiments which make it possible to evaluate the relevance of the analytic results.

We reach the conclusion that, even if it is clear that the basic results of the field are true and 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.

Introduction.

It has been shown recently, see for instance [6, 44, 89, 105, 113], that most real-world complex networks have non-trivial properties in common. This is in particular true for their degree distribution (probability p_k that a randomly chosen node has k links, for each k): most real-world complex networks display heterogeneous degree distributions, well fitted by power-laws $p_k \sim k^{-\alpha}$ with an exponent α between 2 and 3 in general. This property has been observed on a variety of networks, including internet and web graphs, see for instance [48, 58, 106, 19, 95, 78, 25, 95, 107, 115, 26, 116, 21, 7, 3, 72, 18, 75], social networks, see for instance [77, 85, 86, 46, 68], and biological networks, see for instance [71, 109, 67, 49].

[1] CREA – CNRS – École Polytechnique – 1, rue Descartes, 75005 Paris, France,

[2] LIAFA – CNRS – Université Paris 7 – 2 place Jussieu, 75005 Paris, France.

In most of these networks, the existence of a path in the network from most nodes to most others, called *connectivity*, is a crucial feature. In the case of the internet, it means that computers can communicate, in the case of the web if 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, for instance. Notice that one may want to preserve connectivity (in the case of the internet for instance), or prevent it (in the case of virus propagation, for instance), depending on the application under focus.

These networks are subject to damages (either accidental or not) which may endanger this property. For instance, failures may occur on machines on the internet, causing removal of nodes in internet and web 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. Accidental failures may therefore be modeled by removals of random nodes and/or links, while attacks may be modeled by removals following a given strategy.

Networks of different natures may behave differently when one removes nodes and/or links. Likewise, the way removals are done may influence significantly the obtained behavior. It has been confirmed that this is indeed the case in the famous paper [8], in which the authors consider networks with Poisson and power-law degree distributions, and then remove nodes either randomly or by decreasing order of their degree. They measure the size of the largest connected component (*i.e.* set of nodes such there is a path in the network between any two of them) as a function of the fraction of removed nodes.

The authors of [18] had pursued the same kind of idea earlier. They tried to establish whether the connectivity of the web is mainly due to the very popular pages with a very large number of incoming links by studying the connectivity of the web graph from which the links going to these pages have been removed.

The authors of [8] obtained the results in Figure 1, showing that the two kinds of networks behave significantly differently, and that the removal strategies play an important role. In particular, it seems that power-law degree distributions make networks very resilient to random failures but very sensitive to attacks. This particularity has even been named the *Achille's heel of the internet* [1, 11]. In the case of a social network on which one want to design vaccination strategies, it means that one may expect better performance than with random vaccination [41, 97, 32, 64].

Since then, much work has been done to extend this initial result. Other kinds of failures and attacks, in particular cascade ones, as well as other kinds of topologies, have been studied. See for instance [76, 34, 35, 92, 20, 94, 118, 66, 51, 84, 83, 82, 101, 117]. Some studies introduced other criteria for measuring the state of the network, see for instance [74, 36, 20, 94]. Cases where the underlying networks have non-trivial degree correlations have also been studied, see for instance [15, 111]. Recent studies focus on the identification and design of robust topologies, or repair strategies, see for instance [110, 100, 99, 108, 37, 27, 64, 54, 55, 103, 102, 13, 14].

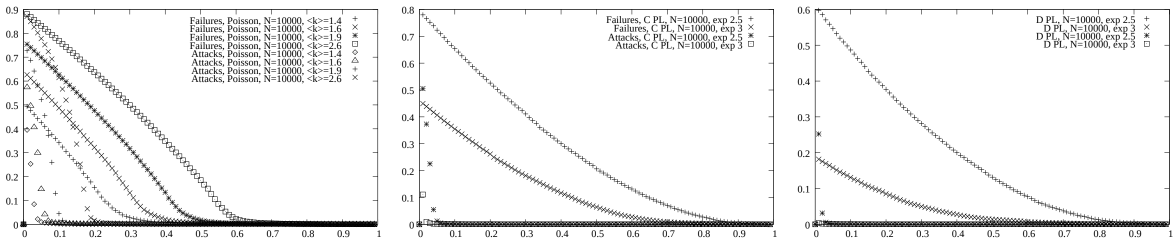

Figure 1: Size of the largest connected component as a function of the fraction of randomly removed nodes (failures) and nodes removed by decreasing order of node degrees (classical attacks). From left to right: Poisson, continuous power-law and discrete power-law networks. We will define properly these different kinds of networks in Section 1.2. For technical details on our plots see Section 1.4.

In several cases, an important effort has been made to give analytic results completing the experimental ones, based on mean-field approximations, see in particular [29, 30, 33, 24, 88, 6]. The aim of our contribution is to present in detail these results, to deepen them with new results, and to discuss their implications.

A significant part of this paper is therefore devoted to detailing the existing proofs of previously known results. Indeed, these proofs rely strongly on mean-field approximations which are classical in statistical mechanics, but quite unusual in computer science. They are therefore only sketched in the original papers, and many approximations are implicit. It however seems important to us to give proofs with full details and explicit approximations, which we do here.

Moreover, the original papers focus on special cases of interest. We give here a unified and completed set of results on the questions under concern, which will make it possible to deepen significantly our understanding of the field. In particular, we will give results concerning link failures and attacks, as well as results on finite cases, which have received little attention. We will also compare the two different approaches proposed in [30, 29, 33] and in [24, 88].

This paper may therefore be considered as an in-depth and didactic survey of the main current results of the field, with the aim of discussing further their implications using a few new results.

It is organized as follows. Section 1 is devoted to some important preliminaries, which consist both in some preliminary results, in definitions and models, and in methodological discussions. In particular, the approximations made in the proofs in this paper are presented and explained. Sections 2 and 3 deal with random failures and with attacks, respectively. We will present together classical results of the field and several new results aimed at improving our understanding of the phenomena under concern. We will discuss the behavior of several real-world complex networks in Section 4, and compare them to expected behaviors from theory. Finally, we give an in-depth discussion and synthesis of our understanding of the field in Section 5.

1 Preliminaries.

Before entering in the core of the paper, we need some important preliminaries. They consist both in preliminary results, mainly on distributions, definitions, and the models we will consider, and methodological aspects. Important notions are introduced and discussed here; therefore this section should be read carefully before the rest of the paper.

Before entering in more details, let us insist on the fact that most results in this paper are obtained using *approximations*, aimed at simplifying the computation. These approximations are valid in the limit of large sizes (of networks and in general of the considered samples). They typically consist in neglecting the distinctions between N and $N \pm n$ when n is small compared to N, or in supposing that random values are equal to the average. More subtle approximations are also done, and all belong to the *mean-field* approximation framework, classical in statistical mechanics and widely used in the context of complex networks [38, 2, 91].

It is important however to understand that the proofs we provide are valid only in this framework, and that effort should be made to provide exact results and proofs. Indeed, some details hidden in the approximation may play a significant role, which can be infirmed or confirmed only with exact proofs. This is why we will always explicitly point out the approximations we make, and we will always compare the analytic results to experiments. This makes this approach rigorous and relevant. Moreover, there is still much to do to obtain exact results and proofs in this context: most results presented here are currently beyond the areas to which exact methods have been applied with success.

1.1 Poisson and power-law distributions.

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 centered around this average.

A power-law distribution with exponent α is such that p_k is proportional $k^{-\alpha}$. In the whole paper, we will generally consider exponents between 2 and 4, which are the relevant cases for our purpose (see Section 1.2), but we will also state some results valid out of this range. In such distributions, the probability of occurrence of a value x decays only polynomially with x. This implies that, though most values are small, one may obtain very large values with such a distribution.

We will consider here two types of power-law distributions, 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.

The corresponding discrete power-law distribution is $p_k = \frac{1}{C}\ k^{-\alpha},\ k \geq m$, where $C = \sum_{k=m}^{\infty} k^{-\alpha}$ is the normalization constant necessary to ensure that each p_k is between 0 and 1 and that their sum is 1. In such a distribution, therefore, p_k is exactly proportional to $k^{-\alpha}$ for all k. In order to simplify the computation, we will always take $m = 1$ for discrete power-law distributions in this paper. 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, $p_k = \frac{1}{\zeta(\alpha)}\ k^{-\alpha}$.

The corresponding continuous power-law is obtained by taking p_k equal to $\int_k^{k+1} Cx^{-\alpha}\mathrm{d}x$, where C is the normalization constant, which is proportional to $k^{-\alpha}$ in the limit where k is large[3]. We must moreover ensure that the sum of the p_k is equal to 1: $\sum_{k=m}^{\infty} p_k = \int_m^{\infty} C\ x^{-\alpha}\mathrm{d}x = C\ \frac{m^{-\alpha+1}}{\alpha-1} = 1$. We then obtain $C = m^{\alpha-1}(\alpha-1)$, and finally $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1})$.

Discrete power-law distributions and continuous power-law distributions each have their own advantages and drawbacks. For instance, continuous power-laws are easier to use in experiments than discrete ones, which themselves are more rigorous than continuous ones for small values. For a more complete discussion on the advantages and drawbacks of discrete and continuous distributions, see for instance [43, 31]. We will use both of them in the sequel, and discuss their differences.

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, *i.e.* the fraction $p_k(N)$ of values equal to k for each k, is slightly different from the original distribution p_k. First, for all $k > K$, $p_k(N) = 0$ while in general $p_k \neq 0$. We will therefore call these distributions *bounded distributions.* The difference between bounded and unbounded distributions goes to zero when N tends towards infinity, but for any finite value of N it may be taken into account.

We detail below important properties of bounded distributions, starting with their expected maximal value K.

Lemma 1.1 *[29] 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 is given 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

[3] One can also consider that p_k is proportional to $\int_{k-1/2}^{k+1/2} x^{-\alpha}\mathrm{d}x$, see [31]. This has little impact on the obtained results.

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. □

Lemma 1.2 *For a Poisson distribution with average value z, the expected maximal value K among a sample of N values is given by*

$$\sum_{0}^{K-1} \frac{z^k}{k!} = e^z \left(1 - \frac{1}{N}\right).$$

Proof : Direct application of Lemma 1.1 with $p_k = e^{-z}\frac{z^k}{k!}$. □

Lemma 1.3 *[29] For a continuous power-law with exponent α and minimal value m, the expected maximal value K among a sample of N values is $K = mN^{\frac{1}{\alpha-1}}$.*

Proof : From Lemma 1.1, K satisfies $\sum_1^{K-1} p_k = 1 - \frac{1}{N}$. Therefore, $\frac{1}{N} = \sum_K^{\infty} p_k$. We have that $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1}) = (\alpha-1)m^{\alpha-1}\int_k^{k+1} x^{-\alpha}\mathrm{d}x$. Therefore, $\frac{1}{N} = (\alpha-1)m^{\alpha-1}\int_K^{\infty} x^{-\alpha}\mathrm{d}x = m^{\alpha-1}K^{-\alpha+1}$. The result follows directly. □

Lemma 1.4 *For a discrete power-law with exponent α, the expected maximal value K among a sample of N values is given by $\zeta(\alpha)(1-\frac{1}{N}) = H_{K-1}^{(\alpha)}$, with $H_K^{(\alpha)} = \sum_{k=1}^{K} k^{-\alpha}$ being the K-th harmonic number for α .*

Proof : Direct application of Lemma 1.1 with $p_k = \frac{k^{-\alpha}}{\zeta(\alpha)}$. □

These results may be used in practice to compute the expected maximal value K among a sample of N values. For Lemmas 1.2 and 1.4, it is obtained by putting $K = 0$ and then increase it until $\sum_{k=0}^{K} p_k \geq 1 - \frac{1}{N}$.

Figure 2 plots the estimates of the maximal value for samples of size $N = 100\,000$ as obtained from the results above, together with experimental values obtained by sampling $1\,000$ sets of N values and then by taking their average maximal value. In the case of power-law distributions, the theoretical evaluations underestimate slightly the experimental values.

For Poisson distributions, the evaluation fits experiments exactly, but for some precise values only. This is due to the fact that K can only take integer values in the evaluation: it is actually the first integer such that $\sum_{k=0}^{K} p_k > \frac{N-1}{N}$. Since K can only take integer values, this sum may sometimes be significantly larger than $\frac{N}{N-1}$, and then the evaluation of K is poor. To evaluate this bias, we have plotted the relative error $\left(\sum_{k=0}^{K} p_k\right) \frac{N-1}{N}$ in Figure 2 (left).

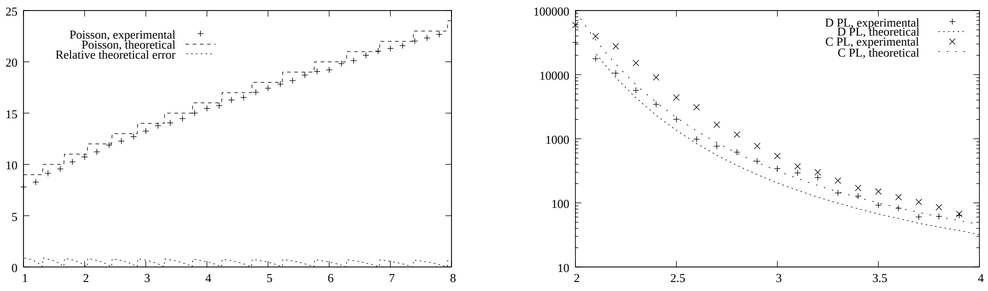

Figure 2: Analytic and experimental estimates of the expected maximal among $N = 100\,000$ sampled values. Left: for Poisson distributions, as a function of the average value; right: for discrete and continuous power-law distributions, as a function of the exponent.

An important point here is to notice that, for any of the three kinds of distributions we consider here, the expected maximal among N sampled values grows sublinearly with N. This is obvious for Poisson distributions, and it is also true for power-law distributions: Lemma 1.3 explicitly states it for continuous power-laws, and one may also check the discrete power-law case. As we will see, this is important for some approximations we will make in the following, and for some results in Section 3.3.

Until now, we discussed the fact that sampling N values induces an expected maximal one. But it also induces an expected distribution of the N values, denoted by $p_k(N)$, which is different from the original distribution p_k. We now study more precisely this expected distribution.

Lemma 1.5 *The expected distribution $p_k(N)$ of N values sampled from a given distribution p_k is, for all $k \leq K$,*

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

where K is the expected maximal value, related to p_k by Lemma 1.1.

Proof : We must have that $p_k(N)$ is proportional to p_k for all $k \leq K$: $p_k(N) = C\ p_k$, and that the sum of all $p_k(N)$ is 1: $\sum_{k=0}^{\infty} p_k(N) = 1$. Moreover, we know from Lemma 1.1 that $\sum_{K}^{\infty} p_k = \frac{1}{N}$. We obtain $1 = \sum_{k=0}^{\infty} p_k(N) = \sum_{k=0}^{K} p_k(N) = C \sum_{k=0}^{K} p_k = C(1 - \frac{1}{N})$, where we neglected the difference between $\sum_{K}^{\infty} p_k$ and $\sum_{K-1}^{\infty} p_k$. The claim follows. □

Lemma 1.6 *For a Poisson distribution with average value z, the expected distribution $p_k(N)$ of a sample of N values is, for all $k \leq K$,*

$$p_k(N) = \frac{N}{N-1}\ \frac{e^{-z} z^k}{k!},$$

where K is the expected maximal value, related to p_k by Lemma 1.2.

Proof : Direct application of Lemma 1.5 with $p_k = e^{-z}\frac{z^k}{k!}$. □

Lemma 1.7 *For a continuous power-law distribution with exponent α and minimal value m, the expected distribution $p_k(N)$ of a sample of N values is, for all $m \leq k \leq K$,*

$$p_k(N) = \frac{N}{N-1} m^{\alpha-1} \left(k^{-\alpha+1} - (k+1)^{-\alpha+1}\right),$$

where K is the expected maximal value, related to p_k by Lemma 1.3.

Proof : Direct application of Lemma 1.5 with $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1})$. □

Lemma 1.8 *For a discrete power-law distribution with exponent α, the expected distribution $p_k(N)$ of a sample of N values is, for all $k \leq K$,*

$$p_k(N) = \frac{N}{N-1}\,\frac{k^{-\alpha}}{\zeta(\alpha)} = \frac{k^{-\alpha}}{H_{K-1}^{(\alpha)}},$$

with $H_K^{(\alpha)} = \sum_{k=1}^{K} k^{-\alpha}$ being the K-th harmonic number for α , where K is the expected maximal value, related to p_k by Lemma 1.4.

Proof : Direct application of Lemma 1.5, with $p_k = \frac{k^{-\alpha}}{\zeta(\alpha)}$. □

The results above give a precise description of what one may expect from finite samples from Poisson and power-law distributions. They will be useful when dealing with finite networks below.

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 i-th moment being defined as $\langle k^i \rangle = \sum_{k=0}^{\infty} k^i\ p_k$. In the whole paper, the first and second moments will play a central role. We present here the results we will need about them. Namely, we give formulae for the Poisson and power-law cases, both in the infinite case and in the case of a sample of finite size N.

Lemma 1.9 *For a Poisson distribution with average value z, the first two moments of the expected distribution of a sample of N values are*

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

where K is the expected maximal value, related to N by Lemma 1.2.

Proof : Direct application of Lemma 1.6. □

Lemma 1.10 *For a Poisson distribution with average value z, the first two moments are*

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

Proof : Direct computation, with $p_k = e^{-z}\frac{z^k}{k!}$. □

Lemma 1.11 *[29] 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 are*

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

where K is related to N by Lemma 1.3.

Proof : If we approximate $\frac{N-1}{N}$ by 1, we obtain from Lemma 1.7 that $p_k(N) = m^{\alpha-1}\left(k^{-\alpha+1} - (k+1)^{-\alpha+1}\right) = (\alpha-1)m^{\alpha-1}\int_k^{k+1} x^{-\alpha}\mathrm{d}x$, hence $\langle k\rangle = \frac{(\alpha-1)m^{\alpha-1}}{-\alpha+2}(K^{-\alpha+2} - m^{-\alpha+2})$ and $\langle k^2\rangle = \frac{(\alpha-1)m^{\alpha-1}}{-\alpha+3}(K^{-\alpha+3} - m^{-\alpha+3})$.
Moreover, when N is large, we have $K \gg m$ and we can approximate $K^{-\alpha+2} - m^{-\alpha+2}$ by $K^{-\alpha+2}$ and $K^{-\alpha+3} - m^{-\alpha+3}$ by $K^{-\alpha+3}$ if $1<\alpha<2$; $K^{-\alpha+2} - m^{-\alpha+2}$ by $-m^{-\alpha+2}$ and $K^{-\alpha+3} - m^{-\alpha+3}$ by $K^{-\alpha+3}$ if $2<\alpha<3$; and $K^{-\alpha+2} - m^{-\alpha+2}$ by $-m^{-\alpha+2}$ and $K^{-\alpha+3} - m^{-\alpha+3}$ by $-m^{-\alpha+3}$ if $\alpha>3$. The result follows. □

Lemma 1.12 *[29] For a continuous power-law distribution with exponent α and minimal value m, the first two moments are*

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

and they diverge in all the other cases.

Proof : Direct application of Lemma 1.11 with K tending towards infinity. □

Lemma 1.13 *For a discrete power-law distribution with exponent α, the first two moments of the expected distribution of a sample of N values are*

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

with $H_K^{(\alpha)} = \sum_{k=1}^{K} k^{-\alpha}$ being the K-th harmonic number for α , where K is the expected maximal value, related to N by Lemma 1.4.

Proof : Direct application of Lemma 1.8. □

Lemma 1.14 *[88] For a discrete power-law distribution with exponent α, the first two moments are*

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

Proof : Direct computation, with $p_k = \frac{k^{-\alpha}}{\zeta(\alpha)}$. □

We would like to discuss here the differences between the moments of bounded and unbounded distributions. Although the difference between the distributions themselves is small, both for the Poisson and the power-law case (the difference between a bounded and an unbounded distribution is $N/(N-1)$), 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 $\sum_{k=K+1}^{\infty} kp_k$ 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, the values of p_k becomes neglectible when k becomes large, 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 even for values of k larger than K the probabilities p_k are much larger (though still very small), and therefore values of kp_k and $k^2 p_k$ for $k > K$ (but still close to 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.

We finally have all the preliminary results we need on distributions; we can now use them in the context of complex networks.

1.2 Modeling issues.

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

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 of the N nodes with respect to p_k, then to attach to each node as many *stubs* as its degree, and finally to construct links by choosing random pairs of stubs. 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. Then the procedure is guaranteed to terminate (if it is possible with the given distribution). This model is known as the *configuration model* [12] and is widely used in the literature, see for instance [16, 80, 81, 4]. We will call all networks obtained using it *random networks*[4].

If one chooses a Poisson distribution of average z then one obtains an equivalent of the Erdös-Rényi model [47] in which the network is constructed from N initially disconnected nodes by adding $M = \frac{z\,N}{2}$ links between pairs chosen at random. One then obtains a network taken uniformly at random among the ones having N nodes and M links.

As already discussed in the introduction, and as we will see all along this contribution, 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, which are natural representative of all the networks having the same degree distribution, is therefore a key issue. Much work has already been done to this regard, see for instance [88, 4, 96, 33]. 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. We will therefore use them 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, 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 1.12 and 1.14) and above 4 the network has only small connected components, as we will see below. Moreover, most real-world cases fit in these ranges.

Let us insist finally on the fact that real-world complex networks may have other properties that influence their robustness, like for instance correlations between degrees, clustering (local density), and others. Capturing these properties in formal models however remains a challenge, in particular for the clustering [113, 69, 45, 44, 70, 65, 87, 60, 59]. The case of degree correlations has been studied in [23, 70, 93]. This is however out of the scope of this contribution: we will only discuss this point when observing the behavior of real-world networks in Section 4.

[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 1.3.

Failures and attacks.

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

Random node failures are then 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 in series of removal of links chosen at random.

Attacks cannot be defined as easily: they follow a *strategy* which has to be defined. We then say that we observe an *attack following this strategy.* For instance, we presented in the introduction the most famous strategy, which consists in removing nodes in decreasing order of their degrees. We will call this the *classical attack*, and we will define other strategies in Section 3.

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 in which one observes the fraction of *links* removed during *node* failures or attacks.

In the sequel we will consider all these situations: random node or link failures, attacks following various strategies, and link point of view of node failures and attacks. More complex situations, like cascading failures, have been considered for instance in [34, 76, 35, 92, 83, 82, 101, 117].

We want to observe the resilience of networks in these various cases. To achieve this, there are again many ways to capture the current state of the network. We will here measure it by the size of its largest connected component, or more precisely the fraction of nodes in this component. It captures the ability of nodes to communicate, which is central in our context.
Notice however that one may use other criteria. See for instance [74, 36, 20, 94, 66, 84].

Largest connected component.

In many cases, the largest connected component of a random network contains most nodes of the network. More precisely, depending on the underlying degree distribution, the size of the largest connected component may scale linearly with the size of the network. The network is then said to have a *giant component*.

One can actually give a precise and simple criterion on the degree distribution to predict if a random network having this degree distribution 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 1.15 *[80, 4, 29, 88] A random network with size N tending towards 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 [80, 4] and has been proved in the mean-field approximation framework in [29, 88]. Detailing these proofs is out of the scope of this paper.

This result may be applied to the three kinds of networks we consider here (since their maximal degree is sublinear, as explained in Section 1.1), which gives the following results[5].

Lemma 1.16 *A Poisson network with size tending towards infinity and average degree z almost surely has a giant component if and only if $z > 1$.*

Proof : Direct application of Theorem 1.15 and Lemma 1.10. □

Lemma 1.17 *A continuous power-law network with size tending towards infinity, exponent α and minimal degree $m = 1$ almost surely has a giant component if and only if $\alpha < 4$.*

Proof : Direct application of Theorem 1.15 and Lemma 1.12. □

Lemma 1.18 *A discrete power-law network with size tending towards infinity and exponent α almost surely has a giant component if and only if α is such that $\frac{\zeta(\alpha-2)}{\zeta(\alpha-1)} > 2$.*

Proof : Direct application of Theorem 1.15 and Lemma 1.14. □

One may compute the numerical value from this last lemma. One then obtains the conditions $\alpha < 3.48\ldots$ for discrete power-law networks. Very simple conditions based only on the degree distributions are therefore obtained for random networks to have a giant component.

1.3 Mean-field framework and generating functions.

As already emphasized in the beginning of Section 1, most results in this paper are made using *approximations*, valid in the mean-field framework. Most of these approximations are very simple and classical ones, like neglecting small values when added to large ones, but some are specific to random networks and deserve more attention. We detail them below. We then present the generating function framework, which makes it possible to embed these approximations in a powerful formalism. We finally recall some results on generating functions which will be useful in the rest of the paper.

[5] Refer to Section 1.3 for the conditions under which the previous theorem is going to be applied.

Mean-field approximations in random networks.

The fact that stubs are linked together fully at random in a random network is a feature which has important consequences in our context. Notice for instance that when one removes a link chosen at random in such a network, then 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. These simple remarks will be essential in the following.

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

Consider for instance the neighbors of a source node in a large random network. Suppose that the network is sparse (the probability for two randomly chosen nodes to be linked together is almost 0) and that its maximal degree is small compared to its size, which will always be true in our context. Then the probability that two of these neighbors are directly linked together is negligible. Likewise, if we take all the nodes at distance 2 of the source node then the probability of having a link between two of them is very small and may be neglected. So does the probability to have 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 towards infinity.

The approximation above relies in the fact that we neglect very small probabilities, or equivalently that we consider the limit where the size of the network tends towards infinity.

In the same manner, it is known [28, 22] that any random network that has a maximal degree lower than $\sqrt{\langle k \rangle N}$ almost surely has no loops or multiple links. Though the networks we consider will often have maximal degrees greater than these values, we will neglect the probabilities that they possess loops and multiple links. Likewise, Theorem 1.15 is formally true only for networks with maximal degree less than $N^{1/4}$, but we will suppose it true for all random networks.

The mean-field framework allows another important approximation in our context, which we detail now. It consists in making no distinction between starting from one node and then follow one of its link, and choosing a stub at random. Indeed, since links are formed by pairs of randomly chosen stubs, it makes in principle no difference. It means in particular that we suppose that there is no correlation between the degree of a node and the degrees of its neighbors. This is true when the maximal degree is below $N^{\min(1/2, 1/(\alpha-1))}$ [22]. If the maximal degree is larger, however, which will often be the case in the sequel, this is not true anymore. We will however neglect the possible correlations, which is classical in the mean-field approach.

This approximation may be used to describe the degree distribution of neighbors of nodes, in other words the degree of a node reached from a randomly chosen node by following one of its links chosen at random. This is equivalent to choose a random stub, according to the mean-field approximation above. Since the probability that a random stub belongs to a given node is proportional to its degree, the degree distribution of such a node therefore is different from the degree distribution of a random node. Finally, the probability of reaching a node of degree k is proportional to $k\ p_k$, and the sum of these probabilities must be equal to 1. We finally 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:

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

We insist on the fact that, in the current state of our knowledge, such approximations are necessary to derive the results we seek. It is important however to pursue the development of exact methods in order to verify them and deepen our understanding. It is important, too, to know exactly the approximation we make and when we make them. We will carefully point out the uses of these approximations in the whole paper.

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, which often are much more powerful. See [114] for a general introduction.

The application of generating functions to the random network context is presented in details in [91]. Using them to encode series of probabilities (like for instance degree distributions), the authors show how mean-field approximations may be embedded with benefit 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, and we refer to [91] for a detailed and didactic presentation with illustrations. We follow the notations in this reference, and we will use them all along the paper.

Let us begin by encoding the degree distribution p_k by the following generating function:

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

This function is an encoding of the whole distribution since one may obtain p_k by derivating it k times, then evaluate it at $x = 0$ and divide the result by $k!$: $p_k = G_0^{(k)}(0)/k!$.

Moreover, we have $G_0(1) = \sum_{k=0}^{\infty} p_k = 1$, like for any generating function encoding distribution 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 link at random of a randomly chosen node. This number is distributed according to q_k, defined in Equation 1. We then have

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

This generating function will be useful in the sequel. For more details on how to use generating functions in the context of random networks, see [91].

Useful results.

We give now a few results on generating functions which will play an important role in the sequel. These results are rewritings of results in [24, 88].

Let us consider a random network with size tending towards infinity and with degree distribution p_k encoded in G_0. Let us suppose that some of its nodes (resp. links, *i.e.* pairs of stubs) are marked. We are interested in components composed of unmarked nodes, *i.e.* sets of unmarked nodes such that there exists a path composed only of unmarked nodes (resp. links) between any two of them.

Let us consider a node reached by following a random link, *i.e.* a node obtained by picking a random stub. We will first compute the number of unmarked nodes that can be reached from this node by following links between unmarked nodes (resp. unmarked links) only. We will call such sets of nodes *clusters*.

Two cases may occur: either the chosen node (resp. stub) is marked, in which case the cluster is of size 0, or it is unmarked, in which case the node has a number k of other stubs. Let us denote by $F_1(x)$ the generating function for the probability that such a node is unmarked and has k other stubs. Notice that the case where the chosen node (resp. stub) is marked plays no role in $F_1(x)$. Notice also that $F_1(1)$ is the fraction of unmarked nodes (resp. links) in the network.

When the size of the network tends towards infinity, the clusters have a limit distribution of sizes. We will call *finite* clusters the ones with a finite size in this limit distribution, while we call *infinite* clusters the others. We denote by $H_1(x)$ the generating function for the distribution of the size of such *finite* clusters. Notice that $H_1(x)$ does not take into account infinite clusters, if they exist.

Lemma 1.19 *[24, 88] The generating function $H_1(x)$ satisfies the following self-consistency condition:*

$$H_1(x) = 1 - F_1(1) + xF_1(H_1(x)).$$

Proof : The cluster is of size 0 if the chosen node (resp. stub) is marked, which happens with probability $1 - F_1(1)$ since $F_1(1)$ is the fraction of unmarked nodes (resp. links).

In the other case, let us denote by r_k the probability that the initial node has k other links, *i.e.* $F_1(x) = \sum_{k=0}^{\infty} r_k x^k$. Since we consider networks whose size tends towards infinity, according to the mean-field framework we can neglect cycles (*i.e.* multiple paths between two nodes) in finite clusters. Then, the size of the cluster is 1 plus the sum of the sizes of the clusters at the end of these k links. The distribution for the sum of the sizes of k independent clusters is given by $H_1^k(x)$, see [91]. Moreover, the distribution of 1 plus a value is obtained by multiplying the corresponding generating function of this value by x. We therefore obtain $H_1(x) = 1 - F_1(1) + x \sum_{k=0}^{\infty} r_k H_1^k(x) = 1 - F_1(1) + xF_1(H_1(x))$. □

Theorem 1.20 *[24, 88] If τ is the fraction of marked nodes (resp. links) such that removing all the marked nodes (resp. links) gives a network with no giant component, then τ is such that $F_1'(1) = 1$.*

Before proving this result, we need a new approximation, made implicitly in [24, 88]. It consists in assuming that the average size of components in a random network is finite if and only if there is no giant component. Notice that this is not always true: one can construct graphs such that all the components are of infinite but sub-linear size (thus there is no giant component), in which case the average is infinite. Conversely, there may be a giant component but a finite average size[6]. This approximation is however necessary for the following proof of Theorem 1.20.

Proof : Suppose we removed enough nodes (resp. links) to ensure that there is no giant component in the remaining network. According to the assumption above, we then have that the average size of components is finite (which does not mean that there is no infinite component). This average size is given by $H_1'(1)$. From Lemma 1.19, $H_1'(x) = F_1(H_1(x)) + xF_1'(H_1(x))H_1'(x)$, and since $H_1(1) = 1$, we obtain

$$H_1'(1) = \frac{F_1(1)}{1 - F_1'(1)}.$$

This reasoning is valid only if we removed enough nodes (resp. links). If there is still a giant component in the network, $H_1(1)$ is no longer equal to 1, and the above calculations do not hold. They are valid only for fractions of removed nodes (resp. links) in the interval $]\tau, 1]$ for a given τ which is the threshold below which there is still a giant component.

Notice now that the expression above for $H_1'(1)$ diverges at the point $F_1'(1) = 1$, which defines τ: it is the fraction of marked (thus removed) nodes (resp. links) above which

[6] Computing the distribution of component sizes is a difficult task [80, 24].

there is no giant component anymore. If we choose to remove a fraction of nodes (resp. links) closer and closer to τ, but still larger than it, the size of remaining components grows. It keeps growing until the point where the fraction of removed nodes (resp. links) is not large enough to destroy the giant component. At this point, the average size of finite components tends towards infinity. □

The result we have just described is very powerful and general. We will see that it can be applied to many cases and give simple results with direct proofs: to compute the fraction of nodes (resp. links) to remove from a network in order to ensure that there is no giant component, it is sufficient to give an expression for $F_1(x)$ and then to determine the fractions which leads to $F_1'(1) = 1$.

One must however keep in mind that they rely on mean-field approximations, and that the formalism sometimes makes it difficult to see exactly when approximations are performed.

1.4 Plots and thresholds.

In all plots of this paper, *Poisson*, *C PL* and *D PL* stand for Poisson networks, continuous power-law networks, and discrete power-law networks, respectively.

The first main kind of plots we will consider in the sequel represents the fraction of nodes in the largest connected component of a network as a function of the fraction of removed nodes or links. Figure 1 provides an example. In these plots, we actually sampled a large number of networks (typically 1 000) on which we repeated the experiment, and then plotted the average behavior. 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 plotted results for both continuous and discrete power-law networks with these exponents, and for Poisson networks with the average degrees corresponding to these four cases. These values are summarized in Table 1. The figures of this kind are Figures 3, 5, 7, 9, 11, 13 and 15.

	average degree	
exponent	continuous power-law	discrete power-law
2.5	2.6	1.9
3	1.6	1.4

Table 1: The exponents we will consider in our experiments on power-law networks, and the average degrees they induce (the given values are the ones obtained in practice with minimal value $m = 1$ and $N = 100\,000$ nodes; they are slightly lower than previsions from Lemma 1.11 for continuous power-laws, but fit very well the previsions from Lemma 1.13 for discrete power-law). We will therefore consider Poisson networks with these average degrees all along the paper.

In our context, it is usual to witness a *threshold* phenomenon (typical of statistical mechanics and more precisely percolation theory, see for instance [104]): 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). These thresholds play a central role in the phenomenon under consideration; in the sequel we will often study such thresholds.

Notice that, 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. In these cases, there are several ways to define a threshold. One may notice that, when we reach the threshold, the slope of the plots of the fraction of nodes in the largest connected component in function of the number of removed nodes goes to infinity. In finite-size computation, we may therefore consider that we reach the threshold when this slope is maximal [24]. Notice that this does not always make sense: it may happen, like in Figure 3 (right), that the slope is maximal at 0 (while the expected value of the threshold would rather be 1). One can then adopt the convention that in such cases there is no threshold, but this reduces our ability to discuss practical cases.

Another way, described in [98], consists in computing the degree distribution of the network after each removal of a node or a link, and see if it satisfies the criterion of Theorem 1.15 for it to have a giant component. The threshold is then the fraction of nodes or links to remove so that the network does not satisfy this criterion anymore.

The solution we have chosen is to consider that the threshold is reached when the largest connected component contains less than a given (small) fraction of all the nodes. We adopted this last definition in this paper, using a fraction which makes both definitions quite equivalent in our cases, namely 0.05. In other words, we consider that a network does not have any giant component whenever the size of its largest connected component is less than 5 % of the whole. Notice that this may have a impact on numerical results. Changing this value, to 1 % for instance, would change the results. However, similar observation would be made.

This leads us to the second main kind of plots encountered in this paper. In each context, these plots represent the threshold as a function of the main character of each kind of networks: the average degree for Poisson networks, and the exponent of the power-law for power-law networks. Again, we will plot experimental results obtained by averaging results on large number of networks (typically 1 000), and this for networks of different sizes (typically 1 000, 10 000 and 100 000). To help in the comparison between different kinds of networks, we will add on these plots vertical lines at the values quoted in Table 1. We will also plot the theoretical predictions we will obtain, on the same plots as the experimental results. This will therefore make it possible to compare them. The figures of this kind are Figures 4, 6, 8, 10, 12, 14 and 16.

We will see that the experimental results do not always fit analytic predictions very well. This is influenced in part by the choice to consider 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 large ones. This is why we will present results for both finite cases and the infinite limit in Section 2. 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.

For Poisson networks, for instance, we are faced with the same problem as the one concerning the evaluation of the maximal degree of finite networks, see Section 1.1 and Figure 2: since some parameters can only take integer values, their analytic evaluation may lead to values quite different from their true values. Therefore, we have chosen to use, both for the plots and for numerical evaluations, only the analytic values of the threshold for those values such that the error due to this effect is minimal.

Notice finally that the plot for a particular instance may vary significantly from the average behavior, in particular for power-law and/or small networks. We do not enter in these considerations here.

2 Resilience to random failures.

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

We will first consider random node failures (Section 2.1) on general random networks, and then apply the obtained results to Poisson and power-law networks. We will see that the empirical observations cited in the preliminaries concerning the different behaviors of Poisson and power-law networks are formally confirmed. In order to deepen our understanding of random node failures, we will consider in Section 2.2 these failures from the *link* point of view: what fractions of the *links* are removed during random node failures? Finally, we will consider random *link* failures (Section 2.3).

2.1 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 detail the two main proofs proposed for this result [24, 88, 29, 33]. We then apply this general result to the special cases under concern: Poisson and power-law networks (both discrete and continuous versions).

Figure 3 displays the behaviors observed for the three types of networks we consider, namely Poisson, continuous power-law, and discrete power-law networks.

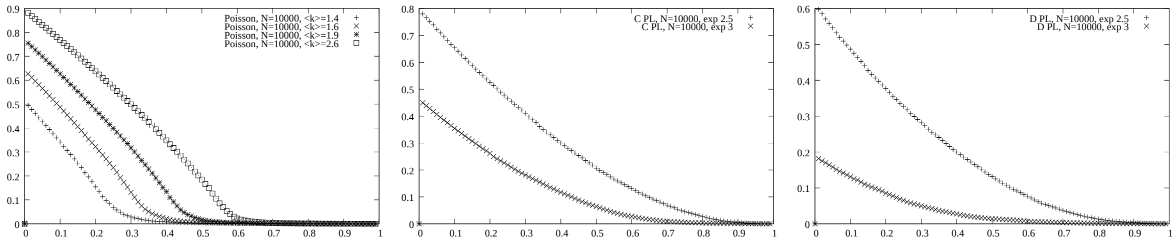

Figure 3: Size of the largest connected component as a function of the fraction of randomly removed nodes. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots see Section 1.4.

As explained in the preliminaries, 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.

2.1.1 General results.

Our aim here is to prove the following general result, stating the value of the threshold for random node failures.

Theorem 2.1 *[24, 29] 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 0. 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 1.15, this implies that the network almost surely has no giant component. In this case, the notion of threshold therefore has no meaning, and the theorem is irrelevant.

Theorem 2.1 has been derived in different ways in the literature. The two main methods were proposed in [29] and in [24]. We detail both approaches below.

Let us begin with the proof in [29, 33]. It relies on the fact that, as explained in the preliminaries, 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, one can use the criterion in Theorem 1.15 to decide if there is still a giant component or not.

Lemma 2.2 *[29] 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*

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

Proof : If a given node had degree k_0 before the removal, then the probability that it has degree $k' \leq k_0$ after the removal is $\binom{k_0}{k'}(1-p)^{k'} p^{k_0-k'}$. Indeed, $k_0 - k'$ of its neighbors have been removed with probability $p^{k_0-k'}$, and k' of its neighbors have not been removed with probability $(1-p)^{k'}$. □

In order to apply Theorem 1.15, we now have to compute the first and second moments of the new degree distribution:

Proposition 2.3 *[29] With the notations of Lemma 2.2, the average and the second moment of the degree distribution $p_k(p)$ are*

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

In order to prove this proposition, we need the following technical lemma.

Lemma 2.4 *For any integer k and k_0, and any real p, we have*

$$\sum_{k=0}^{k_0} k \binom{k_0}{k} (1-p)^k p^{k_0-k} = (1-p)k_0,$$

and

$$\sum_{k=0}^{k_0} k^2 \binom{k_0}{k} (1-p)^k p^{k_0-k} = (1-p)^2 k_0^2 + p(1-p)k_0.$$

Proof : Let us start with:

$$(x+y)^{k_0} = \sum_{k=0}^{k_0} \binom{k_0}{k} x^k y^{k_0-k}.$$

If we derivate this equality with respect to x and then multiply the resulting equality by x, we obtain:

$$x k_0 (x+y)^{k_0-1} = \sum_{k=0}^{k_0} \binom{k_0}{k} k x^k y^{k_0-k}.$$

We obtain the first claim by setting $x = 1 - p$ and $y = p$ in this equation.
If again we derivate the last equation with respect to x and multiply the resulting equality by x, we obtain:

$$x(k_0(x+y)^{k_0-1} + xk_0(k_0-1)(x+y)^{k_0-2}) = \sum_{k=0}^{k_0} k^2 \binom{k_0}{k} x^k y^{k_0-k}.$$

By setting $x = 1 - p$ and $y = p$ we obtain:

$$\begin{array}{rcl} \sum_{k=0}^{k_0} k^2 \binom{k_0}{k} (1-p)^k p^{k_0-k} & = & (1-p)k_0 + (1-p)^2 k_0(k_0-1) \\ & = & (1-p)^2 k_0^2 + p(1-p)k_0, \end{array}$$

which ends the proof. □

We can now prove Proposition 2.3:
Proof : The claims follow from the following series of equations.

$$\begin{array}{rcl} \langle k(p) \rangle & = & \sum_{k=0}^{\infty} k p_k(p) \\ & = & \sum_{k=0}^{\infty} k \sum_{k_0=k}^{\infty} p_{k_0} \binom{k_0}{k} (1-p)^k p^{k_0-k} \\ & = & \sum_{k_0=0}^{\infty} \sum_{k=0}^{k_0} k p_{k_0} \binom{k_0}{k} (1-p)^k p^{k_0-k} \\ & = & \sum_{k_0=0}^{\infty} p_{k_0} \sum_{k=0}^{k_0} k \binom{k_0}{k} (1-p)^k p^{k_0-k} \\ & = & \sum_{k_0=0}^{\infty} p_{k_0} (1-p) k_0 \\ & = & (1-p)\langle k \rangle \\ \langle k^2(p) \rangle & = & \sum_{k=0}^{\infty} k^2 p_k(p) \\ & = & \sum_{k=0}^{\infty} k^2 \sum_{k_0=k}^{\infty} p_{k_0} \binom{k_0}{k} (1-p)^k p^{k_0-k} \\ & = & \sum_{k_0=0}^{\infty} \sum_{k=0}^{k_0} k^2 p_{k_0} \binom{k_0}{k} (1-p)^k p^{k_0-k} \\ & = & \sum_{k_0=0}^{\infty} p_{k_0} \sum_{k=0}^{k_0} k^2 \binom{k_0}{k} (1-p)^k p^{k_0-k} \\ & = & \sum_{k_0=0}^{\infty} p_{k_0} [(1-p)^2 k_0^2 + p(1-p)k_0] \\ & = & (1-p)^2 \langle k^2 \rangle + p(1-p)\langle k \rangle \end{array}$$

□

Finally, this yields the following proof for Theorem 2.1:
Proof : The threshold p_c is reached when the network does not have a giant component anymore. From Theorem 1.15, this happens when $\langle k^2(p_c) \rangle - 2\langle k(p_c) \rangle = 0$. From Proposition 2.3, this is equivalent to $(1-p_c)[(1-p_c)\langle k^2 \rangle - (2-p_c)\langle k \rangle] = 0$, which gives the result. □

Let us now describe the method developed in [24, 88], to obtain Theorem 2.1. It relies on the use of generating functions (see Section 1.3), each node being marked as *absent* with probability p, or 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 1.3. We can then prove Theorem 2.1 as a direct consequence of Theorem 1.20:
Proof : From Theorem 1.20, the threshold p_c is reached when $F_1'(1) = 1$, which is equivalent here to $(1-p_c)G_1'(1) = 1$. Therefore p_c satisfies

$$p_c = 1 - \frac{1}{G_1'(1)}.$$

We know that $G_1(x) = \sum_{k=0}^{\infty} q_k x^k = \sum_{k=1}^{\infty} k p_k x^{k-1}/\langle k\rangle$. Therefore $G_1'(x) = \sum_{k=2}^{\infty} k(k-1)p_k x^{k-2}/\langle k\rangle = \sum_{k=0}^{\infty} k(k-1)p_k x^{k-2}/\langle k\rangle$, and $G_1'(1) = \frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle}$. This ends the proof. □

The two proofs have different advantages and drawbacks. The first one is self contained and relies only on classical probabilistic notions, but it is quite long and technical. The second one is very concise and simple, but it relies on the generating function formalism, which has to be first introduced and understood. These differences do not only have an impact on the aspect of the proofs: they also imply that one has to think carefully about each approximation in the first approach, while they are hidden in the generating function formalism in the second one. As a counterpart, the first approach makes it easier to tune and locate approximations precisely.

2.1.2 The cases of Poisson and power-law networks.

Theorem 2.1 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. More precisely, we will consider Poisson, continuous power-law and discrete power-law networks, and, for each of these classes, both finite networks with N nodes and finite networks with size tending towards infinity. Comparison with simulations will be provided at the end of the subsection.

Notice that we will derive all the results for finite size networks as corollaries of results presented in previous sections. The results for networks with size tending towards infinity can then be derived either from results of the previous sections, or as limits of the corresponding finite cases.

Corollary 2.5 *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 N by Lemma 1.2.

Proof : Direct application of Theorem 2.1 using Lemma 1.9. □

Corollary 2.6 *[29] For Poisson networks with size tending towards infinity and average degree z, the threshold p_c for random node failures is*

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

Proof : Direct application of Corollary 2.5 when the size tends towards infinity. □

Corollary 2.7 *[29] For large continuous power-law networks with N nodes, exponent α and minimal degree m, the threshold p_c for random node failures is*

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

Proof : We can rewrite Theorem 2.1 into $p_c = 1 - 1/(\langle k^2\rangle/\langle k\rangle - 1)$. From the approximations of $\langle k\rangle$ and $\langle k^2\rangle$ in Lemma 1.11, we then obtain:

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

Using the evaluation of K in Lemma 1.3, we obtain the result. □

Corollary 2.8 *[29] For continuous power-law networks with size tending towards infinity, exponent α and minimal degree m, the threshold p_c for random node failures is*

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

Proof : Direct application of Corollary 2.7 when the size tends towards infinity. □

Corollary 2.9 *For large discrete power-law networks with N nodes and exponent α, the threshold p_c for random node failures is given by*

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

with $H_K^{(\alpha)} = \sum_{k=1}^{K} k^{-\alpha}$ being the K-th harmonic number for α , where K is the maximal degree or the network, related to N by Lemma 1.4.

Proof : Direct application of Theorem 2.1 using Lemma 1.13. □

Corollary 2.10 *[24] For discrete power-law networks with size tending towards infinity and exponent α, the threshold p_c for random node failures is*

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

Proof : Direct application of Corollary 2.9 when the size tends towards infinity. □

We plot numerical evaluations of these results in Figure 4, together with experimental results. We also give in Table 2 the thresholds for specific values of the exponent and of the average degree.

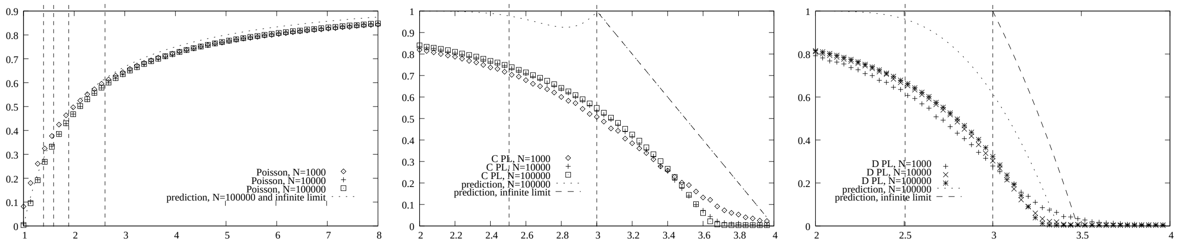

Figure 4: Thresholds for random node failures. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots, on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 1.4.

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

	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

Table 2: Values of the threshold for random node failures on discrete and continuous power-law networks of exponents 2.5 and 3, and on Poisson networks having the same average degree (see Table 1). The values are the analytic previsions at the infinite limit and the ones obtained for experiments with networks of $N = 100\ 000$ nodes.

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 the contrary, for Poisson network 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 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 the predictions at the infinite limit, including cases where the exponent is lower than 3. This is already true from the analytic previsions, and even more pronounced for experiments.

This is particularly clear when one compares the behavior of networks of various kinds but with the same average degree, see Table 2. When the exponent is 3, for instance, the experimental threshold for continuous (resp. discrete) power-law networks is approximately 0.55 (resp. 0.32). For Poisson networks with the same average degree, it is larger than 0.25. Likewise, for exponent 2.5 the difference is not huge.

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.

2.2 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 question we address here therefore is: how many links have been removed when we reach the threshold for random node failures? The corresponding plots are given in Figure 5. Notice that these plots are nothing but (nonlinear) rescalings of the plots in Figure 3. However, they are not equivalent to random removals of links, which are studied in the next subsection.

Like in the previous subsection, we will begin with general results and then apply

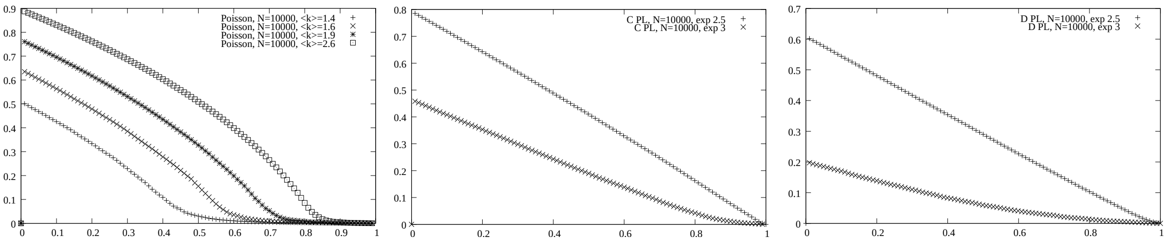

Figure 5: Size of the largest connected component as a function of the fraction of removed *links*, during random *node* failures. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots see Section 1.4.

them to the cases under concern.

2.2.1 General results.

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

Proposition 2.11 *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. This induces the removal of both the links between two removed nodes, and the ones between a removed node and a remaining one.

Since the nodes are chosen randomly, we can assume that the same fraction p of the stubs in the network were attached to the removed nodes. Since pairs of stubs are linked at random, any stub then has a probability p to be linked to a stub attached to a removed node.

Therefore the fraction $m(p)$ of removed stubs may be decomposed into a fraction p of stubs attached to removed nodes, and a fraction $p(1-p)$ of stubs attached to remaining nodes that are linked to stubs attached to a removed node.

Finally, we remove a fraction $p + p(1-p) = 2p - p^2$ stubs, which corresponds to an equal fraction of links. □

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

Corollary 2.12 *The fraction of links removed at the threshold p_c for random node failures in large random networks with degree distribution p_k is*

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

Proof : Immediate from Theorem 2.1 and Proposition 2.11. □

2.2.2 The cases of Poisson and power-law networks.

We now apply the general result above to the cases of interest, which gives a corollary in each case.

Corollary 2.13 *For large Poisson networks with N nodes and average degree z, the fraction of links removed at the threshold p_c for random node failures is*

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

where K is the maximal degree or the network, related to N by Lemma 1.2.

Proof : Direct application of Corollaries 2.5 and 2.12. □

Corollary 2.14 *For Poisson networks with size tending towards infinity and average degree z, the fraction of links removed at the threshold p_c for random node failures is*

$$m(p_c) = 1 - \frac{1}{z^2}$$

Proof : Direct application of Corollaries 2.6 and 2.12. □

Corollary 2.15 *For large discrete power-law networks with N nodes and exponent α, the fraction of links removed at the threshold p_c for random node failures is given by*

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

with $H_K^{(\alpha)} = \sum_{k=1}^{K} k^{-\alpha}$ being the K-th harmonic number for α , where K is the maximal degree or the network, related to N by Lemma 1.4.

Proof : Direct application of Corollaries 2.9 and 2.12. □

Corollary 2.16 *For discrete power-law networks with size tending towards infinity and exponent α, the fraction of links removed at the threshold p_c for random node failures is*

$$m(p_c) = 1 - \left(\frac{\zeta(\alpha-1)}{\zeta(\alpha-2) - \zeta(\alpha-1)} \right)^2 .$$

Proof : Direct application of Corollaries 2.10 and 2.12. □

Corollary 2.17 *For large continuous power-law networks with N nodes, exponent α and minimal degree m, the fraction of links removed at the threshold p_c for random node failures is*

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

Proof : Direct application of Corollaries 2.7 and 2.12. □

Corollary 2.18 *For continuous power-law networks with size tending towards infinity, exponent α and minimal degree m, the fraction of links removed at the threshold p_c for random node failures is*

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

Proof : Direct application of Corollaries 2.8 and 2.12. □

We plot numerical evaluations of these results in Figure 6, together with experimental results. We also give in Table 3 the thresholds for specific values of the exponent and of the average degree.

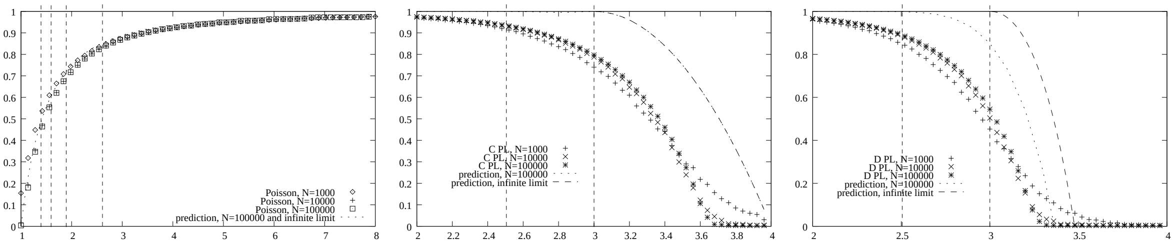

Figure 6: Thresholds for the link point of view of random node failures. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots, on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 1.4.

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.

	continuous		Poisson		discrete		Poisson	
	power-law				power-law			
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

Table 3: Values of the threshold for the 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 the same average degree (see Table 1). The values are the analytic previsions at the infinite limit and the ones obtained for experiments with networks of $N = 100\ 000$ nodes.

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, as explained in the proof of Proposition 2.11.

2.3 Random link failures.

Until now we observed the behavior of random network when *nodes* are randomly removed, both from the nodes and from the link points of view. One may then wonder what happens when we remove *links* rather than nodes, still at random. This may model link failures, just like random removal of nodes models node failures.

Typical behaviors for each type of random networks under concern, when one randomly removes links, are plotted in Figure 7. Just like in the case of random node failures (see Figure 3), 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 are indeed identical.

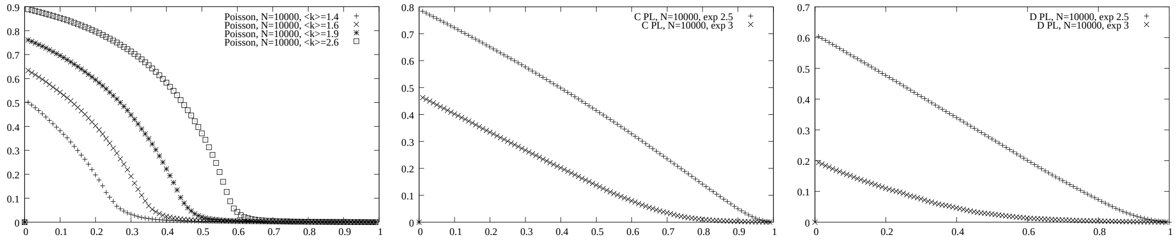

Figure 7: Size of the largest connected component as a function of the fraction of randomly removed links. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots see Section 1.4.

Again, in this section we will first prove a general result which we apply to the three cases under concern. We then compare formal results to experiments, and discuss them.

2.3.1 General results.

The goal of this section is to prove that the the threshold m_c for random link failures actually is the same as the one for random *node* failures (see Theorem 2.1):

Theorem 2.19 *[24, 30] The threshold m_c for random link failures in large random networks with degree distribution p_k is*

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

Just as we did for Theorem 2.1, we will present the main ways to derive this result, as described in [24, 88, 30, 33]. They are very similar to the ones for Theorem 2.1 therefore we will present them in less detail.

The first proof is based on the fact that, as explained in the preliminaries, a random network in which one randomly removes links may still be viewed as a random network, with another degree distribution. Therefore, one can use the criterion given in Theorem 1.15 with this new degree distribution to decide if the network still has a giant component. We therefore begin with the computation of the new degree distribution.

Lemma 2.20 *[30] In a large random network with degree distribution p_k, after the removal of a fraction m of the links during random link failures the degree distribution $p_k(m)$ is given by*

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

Proof : Removing randomly a fraction m of the links corresponds to removing randomly a fraction m of the stubs. If a given node has degree k_0 before the removals, then the probability that its degree becomes $k' \leq k_0$ is $\binom{k_0}{k'}(1-m)^{k'} m^{k_0-k'}$. Indeed, $k_0 - k'$ of its stubs have been removed, with probability $m^{k_0-k'}$, and k' of its stubs are still present, with probability $(1-m)^{k'}$. The result follows. □

This leads to the first proof of Theorem 2.19:

Proof : Notice that Lemma 2.20 actually is nothing but a direct rewriting of Lemma 2.2 on random node failures. Therefore Theorem 2.19 is derived from Lemma 2.20 and Theorem 1.15 in exactly the same way as Theorem 2.1 is derived from Lemma 2.2 and Theorem 1.15. □

The other method used to obtain this result [24] relies on generating functions. As explained in the preliminaries, each link is marked as *present* with probability $1-m$, and *absent* with probability m.

Recall that $F_1(x)$ is the generating function for the probability that, when following a random link, this link is unmarked (*i.e.* present) and leads to a node with k other (marked or unmarked) links emanating from it. In our case, $F_1(x)$ therefore is

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

This leads us to the second proof of Theorem 2.19:
Proof : Again, the generating function $F_1(x)$ obtained here is exactly the same as the one obtained in Section 2.1.1, page 24, for random node failures. The proof therefore is the same as the proof for Theorem 2.1, page 24. □

2.3.2 The cases of Poisson and power-law networks.

Since theoretical results for random link failures are the same as those for random node failures, we focus in this section on experimental results. See Figure 8 and Table 4.

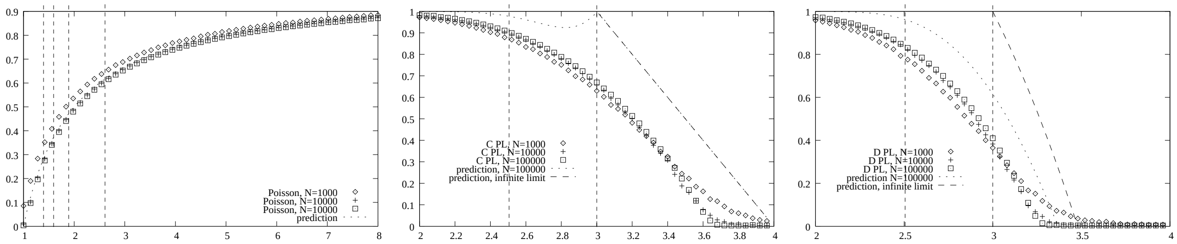

Figure 8: Thresholds for random link failures. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots, on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 1.4.

In principle, these plots and values should be exactly the same as the ones in Figure 4 and in Table 2. This is true for the analytic prevision, 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 3 and 7, then it appears clearly that, though the plot seems to reach 0 at the same fraction, they do not have the same shape. This is responsible for the fact that the size of the largest connected component reaches 5 % of the whole network for different fractions in these two cases. Recall that this is the way we

	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

Table 4: Values of the threshold for random link failures on discrete and continuous power-law networks of exponents 2.5 and 3, and on Poisson networks having the same average degree (see Table 1). The values are the analytic previsions at the infinite limit and the ones obtained for experiments with networks of $N = 100\ 000$ nodes.

compute the threshold in our experiments. As explained in Section 1.4, this is somewhat arbitrary, but we insist on the fact that, in the case of power-law networks considered in these experiments, the other main method for computing the threshold cannot be applied: in several cases, the slope of the plots for these networks is always decreasing.

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

2.4 Conclusion on random failures.

Two mains formal conclusions have been reached in this section. 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 have a clear threshold, in power-law ones all the nodes or links have to be removed to achieve a breakdown. Second, what happens in case of random link failures is very similar, if not identical, to what happens in case of 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.

The qualitative difference between Poisson and power-law networks leads to the conclusion that power-law networks are much more resilient to random failures. This may be used in the design of large scale networks, and it may also be seen as an explanation of the fact that real-world networks like the internet or biological networks seem very resilient to random errors. This has been widely argued in the literature, see for instance [42, 10].

These results however concern only the limit case where the size of the network tends towards infinity. When one consider networks of a given size N, even for very large values of N, then the difference between Poisson and power-law networks often is much less striking. This is even clearer when one considers the link point of view. Moreover, experiments show that the networks we consider, in particular power-law ones, are more resilient to random link failures than to random node failures.

3 Resilience to attacks.

The aim of this section is to study the resilience of random networks to targeted attacks. In our context, an attack on a network consists in series of node and/or link removals, just like in failures. The difference lies in the fact that the removals are *not* random anymore; instead, the nodes and/or links to remove are chosen according to a *strategy*.

Obviously, one may define many different strategies, and failures themselves could be considered as attacks where the strategy consists in random choices. Defining more subtle strategies may however lead to much more efficient methods to destroy a network. As announced in the introduction, such a strategy has been defined in the initial paper on the topic [8] and received since then much attention. It consists in the removal of nodes by decreasing value of their degree, which we will call a *classical attack*.

We will first consider these classical attacks (Section 3.1) on general random networks, and then apply the obtained results to Poisson and power-law networks. In order to deepen our understanding of classical attacks, we will consider in Section 3.2 these attacks from the *link* point of view: what fractions of the *links* are removed during classical attacks? Our effort towards in-depth understanding of classical attacks will also lead us to the introduction of new attack strategies (both on nodes and on links) in Section 3.3. Finally we will conclude this section with a detailed discussion on the efficiency of classical attacks, as well as on the difference of the behaviors of different networks under various attack strategies (Section 3.4).

3.1 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 detail the two main proofs proposed for this result [24, 88, 30, 33] We then apply this general result to the special cases under concern: Poisson and power-law networks (both discrete and continuous versions).

Figure 9 displays the behaviors observed for the three types of networks we consider, namely Poisson, continuous power-law, and discrete power-law networks.

As explained in the preliminaries, there is no fundamental difference between the way Poisson networks and power-law networks behave in case of classical attacks: both are quite sensitive. 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 of them.

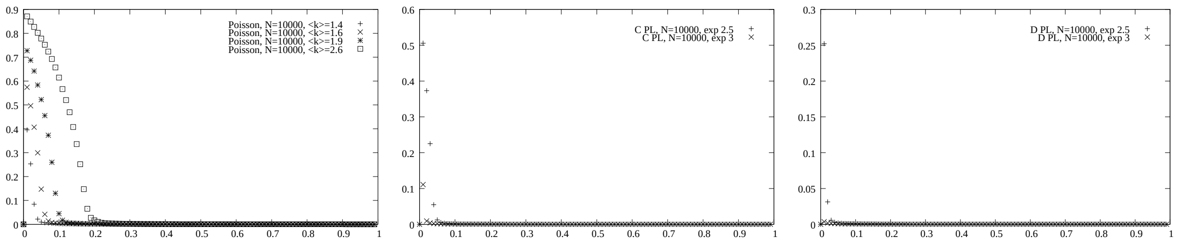

Figure 9: Size of the largest connected component as a function of the fraction of nodes removed during classical attacks. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots see Section 1.4.

3.1.1 General results.

Our aim here is to prove the following general result, which gives the value of the threshold for classical attacks.

Theorem 3.1 *[24, 30] The threshold p_c for classical attacks in random networks, with size tending towards 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 3.2.

As it was the case for failures, there are two main ways to derive this result, provided in [24, 88, 30, 33]. They both rely on the following result, which we therefore prove before entering in the core of each proof. It concerns the maximal degree of random networks after removal of a fraction p of the nodes during a classical attack, denoted by $K(p)$.

Lemma 3.2 *[30] In a random network with size tending towards 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*

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

Proof : Before the removals, the network has a maximal degree K. The new maximal degree $K(p)$ can then be evaluated using $\sum_{k=K(p)+1}^{K} p_k = p$. From Lemma 1.1, this is equivalent to $\sum_{k=K(p)}^{\infty} p_k = p + \frac{1}{N}$ (neglecting the difference between $\sum_{k=K}^{\infty} p_k$ and $\sum_{k=K+1}^{\infty} p_k$). Since N tends to infinity, we can neglect $1/N$, which lead to $p = \sum_{k=K(p)+1}^{\infty} p_k$, hence the result. □

Therefore, in order to compute the threshold at the infinite limit for random networks with a given degree distribution, one has to follow a two step computation: first compute the value of $K(p_c)$ using Theorem 3.1 and then obtain the value of p_c using Lemma 3.2.

Before entering in the core of the proofs, notice that the above results hold for *random networks with size tending towards infinity.* It would be possible to write equivalent results for large networks of a given size N, using the (original) maximal degree given by Lemma 1.1. We however restrict ourselves to the infinite limit in this section for two main reasons. First, and most important, the results for large networks of a given size N would be very similar to the ones for the infinite limit. Indeed, the equivalent of Lemma 3.2 for finite networks is:

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

(where we neglected the difference between $\sum_{K+1}^{\infty} p_k$ and $\sum_{K}^{\infty} p_k$). For large networks, $1/N$ is very small compared to p when p is the threshold for classical attacks (which we will prove later in this section). Therefore the maximal degree of a large network of a given size N after a classical attack is very close to the one at the infinite limit. Second, considering large networks of a given size N would make the following computations much more intricate. One must however keep in mind that the case of finite networks is tractable, and can be derived in a very similar way.

We will now give the two main proofs available for Theorem 3.1. The first one, from [30, 33], uses the following preliminary result.

Lemma 3.3 *[30] In a random network with size tending towards infinity and degree distribution p_k, the fraction $s(p)$ of stubs attached to nodes removed during a classical attack where a fraction p of the nodes is removed is given by*

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

where $K(p)$ is the maximal degree of the network after the attack, related to p by Lemma 3.2.

Proof : Each node of degree k has k stubs attached to it. Therefore the fraction of stubs attached to all nodes of degree k is equal to $kp_k/\langle k \rangle$. Therefore, the total number of stubs attached to removed nodes is $s(p) = \frac{1}{\langle k \rangle} \sum_{K(p)+1}^{K} kp_k$. At the infinite limit, this is equivalent to $s(p) = \frac{1}{\langle k \rangle} \sum_{K(p)+1}^{\infty} kp_k$, hence the result. □

We can now give the first proof of Theorem 3.1:

Proof : The central point here is to understand 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. First, it reduces the maximal degree in the network by removing the nodes with highest degrees, and second it removes links attached to these nodes. At this point, it is better to see links as pairs of randomly chosen stubs, as explained in the preliminaries: a classical attack removes all the stubs attached to the removed nodes *and* some stubs attached to other nodes. These last are the stubs that were linked to stubs attached to removed nodes. Since pairs of stubs are linked randomly, this is equivalent to removing the correct number of stubs randomly from the remaining nodes. This is again equivalent to the random removal of half as many links in the subnetwork composed of the nodes which will not be removed.

Since the 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, but with a cutoff (the maximal degree after the attack): $\left(p'_k = \frac{p_k}{\sum_{k=0}^{K(p)} p_k}\right)_{k \leq K(p)}$, where $K(p)$ is the maximal degree after the attack.

We finally obtain that a classical attack is equivalent to random link failures on a random network of which we know the degree distribution. The value of the threshold can then be derived from Theorem 2.19 on random link failures. We then have to relate the number of links that are removed in the subnetwork consisting in the remaining nodes to the number of nodes removed during classical attacks.

During a classical attack which removes a fraction p of the nodes, the fraction of stubs attached to a removed node is $s(p)$, given by Lemma 3.3. Since the links are nothing but pairs of randomly chosen stubs, the probability for a given stub of a remaining node to be linked to a stub of a removed node therefore is $s(p)$. Finally, each stub attached to a remaining node is removed with probability $s(p)$, which is equivalent to the removal of the same fraction of links.

We can now apply Theorem 2.19 to find the fraction p_c of nodes to remove in a classical attack to destroy a random network. Indeed, this theorem makes it possible to compute the fraction $s(p_c)$ of links to remove randomly to destroy the random network described above, which is equivalent. This gives

$$1 - s(p_c) = \frac{\langle k(p_c) \rangle}{\langle k^2(p_c) \rangle - \langle k(p_c) \rangle},$$

where $\langle k(p_c) \rangle$ and $\langle k^2(p_c) \rangle$ are the first and second moment of the degree distribution $\left(p'_k = \frac{p_k}{\sum_{k=0}^{K(p_c)} p_k} = \frac{p_k}{1-p_c}\right)_{0 \leq k \leq K(p_c)}$, which is the degree distribution of the remaining nodes before some of their stubs are removed. The first two moments are given by $\langle k(p_c) \rangle = \sum_{k=0}^{K(p_c)} k p_k / (1 - p_c)$ and $\langle k^2(p_c) \rangle = \sum_{k=0}^{K(p_c)} k^2 p_k / (1 - p_c)$.

We can finally transform the above relation into the claim:

$$\begin{aligned}
1 - s(p_c) &= \frac{\langle k(p_c)\rangle}{\langle k^2(p_c)\rangle - \langle k(p_c)\rangle} \\
\frac{1}{\langle k\rangle}\sum_{k=0}^{K(p_c)} kp_k &= \frac{\sum_{k=0}^{K(p_c)} kp_k}{\sum_{k=0}^{K(p_c)} k^2p_k - \sum_{k=0}^{K(p_c)} kp_k} \\
\frac{1}{\langle k\rangle}\sum_{k=0}^{K(p_c)} kp_k &= \frac{\sum_{k=0}^{K(p_c)} kp_k}{\sum_{k=0}^{K(p_c)} k(k-1)p_k} \\
\langle k\rangle &= \sum_{k=0}^{K(p_c)} k(k-1)p_k.
\end{aligned}$$

□

The other method used to obtain this result [24, 88] relies on generating functions. As explained in the preliminaries, 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:

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

We can then prove Theorem 3.1 as a direct consequence of Theorem 1.20:
Proof : From Theorem 1.20, the threshold p_c is reached when $F_1'(1) = 1$. Derivating F_1 gives the result: $F_1'(x) = \frac{1}{\langle k\rangle}\sum_{k=2}^{K(p)} k(k-1)p_kx^{k-2}$ and $F_1'(1) = \frac{1}{\langle k\rangle}\sum_{k=2}^{K(p)} k(k-1)p_k$. □

Again, the two proofs of Theorem 3.1 have different advantages and drawbacks. See the comments at the end of Section 2.1.1.

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 classical attacks, we therefore only have to apply it to these cases. More precisely, we will consider Poisson, discrete power-law and continuous power-law networks, with size tending towards infinity. Comparison with simulations will be provided at the end of the section, see Figure 10.

Corollary 3.4 *For Poisson networks with size tending towards infinity and average degree z, the threshold p_c for classical attacks is given by*

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

where $K(p_c)$ is the maximal degree of the network after the attack, related to p_c by Lemma 3.2.

Proof : Direct application of Theorem 3.1, with $p_k = e^{-z} z^k / k!$. □

Corollary 3.5 *[24] For discrete power-law networks with size tending towards infinity and exponent α, the threshold p_c for classical attacks is given by*

$$H_{K(p_c)}^{(\alpha-2)} - H_{K(p_c)}^{(\alpha-1)} = \zeta(\alpha - 1),$$

with $H_K^{(\alpha)} = \sum_{k=1}^{K} k^{-\alpha}$ being the K-th harmonic number for α , where $K(p_c)$ is the maximal degree of the network after the attack, related to p_c by Lemma 3.2.

Proof : Direct application of Theorem 3.1, with $p_k = k^{-\alpha}/\zeta(\alpha)$. □

Corollary 3.6 *[30] For continuous power-law networks with size tending towards infinity, exponent α and minimal degree m, the threshold p_c for classical attacks is given by*

$$\left(\frac{K(p_c)}{m}\right)^{2-\alpha} - 2 \quad = \quad \frac{2-\alpha}{3-\alpha}\, m \left(\left(\frac{K(p_c)}{m}\right)^{3-\alpha} - 1\right),$$

where $K(p_c)$ is the maximal degree of the network after the attack, related to p_c by Lemma 3.2.

Proof : From Theorem 3.1, we have: $\left(\sum_{k=m}^{K(p_c)} k(k-1)p_k\right) / \langle k \rangle = 1$, thus $\langle k \rangle = \sum_{k=m}^{K(p_c)} k^2 p_k - \sum_{k=m}^{K(p_c)} k p_k$. We have $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1})$, and $\langle k \rangle = m(\alpha-1)/(\alpha-2)$, from Lemma 1.12. From this we obtain $\sum_{k=m}^{K(p_c)} k p_k = \frac{\alpha-1}{-\alpha+2} m^{\alpha-1} \left(K(p_c)^{-\alpha+2} - m^{-\alpha+2}\right)$ and $\sum_{k=m}^{K(p_c)} k^2 p_k = \frac{\alpha-1}{-\alpha+3} m^{\alpha-1} \left(K(p_c)^{-\alpha+3} - m^{-\alpha+3}\right)$.
The above equality then becomes:

$$(\alpha-1)m^{\alpha-1}\left[\frac{K(p_c)^{-\alpha+3} - m^{-\alpha+3}}{-\alpha+3} - \frac{K(p_c)^{-\alpha+2} - m^{-\alpha+2}}{-\alpha+2}\right] = m\frac{\alpha-1}{\alpha-2}.$$

We can finally transform this equation to obtain the result:

$$\begin{aligned}
\frac{m^{-\alpha+3}}{-\alpha+3}\left[\left(\frac{K(p_c)}{m}\right)^{-\alpha+3}-1\right]-\frac{m^{-\alpha+2}}{-\alpha+2}\left[\left(\frac{K(p_c)}{m}\right)^{-\alpha+2}-1\right] &= \frac{m^{-\alpha+2}}{\alpha-2}\\
\frac{m}{-\alpha+3}\left[\left(\frac{K(p_c)}{m}\right)^{-\alpha+3}-1\right]-\frac{1}{-\alpha+2}\left[\left(\frac{K(p_c)}{m}\right)^{-\alpha+2}-1\right] &= \frac{1}{\alpha-2}\\
\frac{m}{-\alpha+3}\left[\left(\frac{K(p_c)}{m}\right)^{-\alpha+3}-1\right]-\frac{1}{-\alpha+2}\left[\left(\frac{K(p_c)}{m}\right)^{-\alpha+2}-2\right] &= 0
\end{aligned}$$

□

Numerical evaluations of these results can be done by computing the maximal degree $K(p_c)$ using the appropriate corollary and then using it to evaluate p_c with Lemma 3.2. The equations in Corollaries 3.4 and 3.5 can be solved in a similar way as what we did to compute the maximal degree of a random network, see Section 1.1. The equation in Corollary 3.6 can be solved using a computer algebra system [79]. Notice that this equation is not defined for $\alpha = 3$; one may obtain the threshold for this value as the limit of its values when α tends to it.

Moreover, still in the case of continuous power-law networks, one may use the following results, instead of Lemma 3.2, which is simpler and more precise (since it allows non-integer values for the degree).

Lemma 3.7 *[30] For continuous power-law networks with size tending towards infinity, exponent α and minimal degree m, the maximal degree $K(p)$ after the removal of a fraction p of the nodes during a classical attacks is related to p by*

$$p = m^{\alpha-1}K(p)^{-\alpha+1}.$$

Proof : From Lemma 3.2, we have $p = \sum_{K(p)+1}^{\infty} p_k dk$, which we can approximate to be equal to $\sum_{K(p)}^{\infty} p_k \mathrm{d}k$. We have that $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1}) = (\alpha-1)m^{\alpha-1}\int_k^{k+1} x^{-\alpha}\mathrm{d}x$. We then have $p = (\alpha-1)m^{\alpha-1}\left[\frac{k^{-\alpha+1}}{-\alpha+1}\right]_K^{\infty} = (\alpha-1)m^{\alpha-1}K^{-\alpha+1}/(\alpha-1) = m^{\alpha-1}K^{-\alpha+1}$, hence the result. □

Finally, we plot numerical evaluations of these results in Figure 10, together with experimental results. We also give in Table 5 the thresholds for specific values of the exponent and of the average degree.

It appears clearly that the thresholds for power-law networks are much lower than the ones for Poisson networks: they are are almost one order of magnitude larger for Poisson

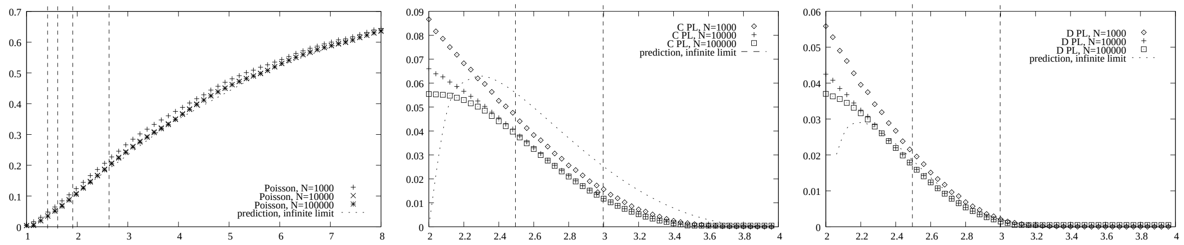

Figure 10: Thresholds for classical attacks. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots, on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 1.4.

	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

Table 5: Values of the threshold for classical attacks on discrete and continuous power-law networks of exponents 2.5 and 3, and on Poisson networks having the same average degree (see Table 1). The values are the analytic previsions at the infinite limit and the ones obtained for experiments with networks of $N = 100\ 000$ nodes.

networks than for comparable power-law networks. Moreover, both types of networks are very sensitive to classical attacks: only a few percents of the nodes have to be removed to destroy the network.

This leads to the conclusion that power-law networks are much more sensitive to classical attacks than Poisson ones, which certainly is the main result on the topic. We will deepen this in the rest of the section.

3.2 Link point of view of classical attacks.

The classical attack strategy removes highest degree nodes first. Since in a power-law network there is a high heterogeneity between node degrees, this leads in this case to the removal of huge numbers of links. One may then wonder if its efficiency on power-law networks is a consequence of 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 results in the removal of much more links in a power-law network than in a Poisson one is the cause of the difference between the two. These explanations actually have been proposed by some authors to give an intuitive explanation of the results presented above.

The aim of this section is to evaluate these ideas by the study of classical attacks under the link point of view. The questions we address here therefore are: how many links have been removed when we reach the threshold for classical attacks? Is this number similar for power-law and Poisson networks? And is it similar to the amount of links which have to be removed at random to disconnect the network?

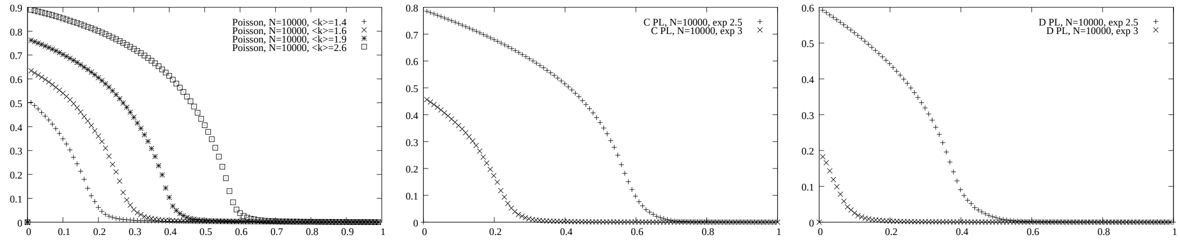

Figure 11: Size of the largest connected component as a function of the fraction of *links* removed during classical attacks. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots see Section 1.4.

Experiments are presented in Figure 11. 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 9. One may also observe that, though there are significant differences, when one removes from power-law networks as many links as what is needed to destroy a Poisson network with the same average degree, then the size of the largest connected component becomes very small. We will investigate all this more precisely in this section and the next one.

3.2.1 General results.

Our aim here is to prove the following general result, which allows to compute the fraction $m(p)$ of links removed when a fraction p of the nodes are removed during classical attacks.

Theorem 3.8 *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 given by*

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

where $s(p)$ is the fraction of stubs attached to removed nodes, and is related to p by Lemma 3.3.

Proof : Let us consider a network in which we remove a fraction p of the nodes during a classical attack. This induces the removal of the stubs attached to removed nodes, as well as those which formed links with such stubs. Since pairs of stubs are linked at random, any stub attached to a remaining node has a probability $s(p)$ of being linked to a stub attached to a removed node.

Therefore the fraction $m(p)$ of removed stubs may be decomposed into a fraction $s(p)$ of stubs attached to removed nodes, and a fraction $(1 - s(p))s(p)$ of stubs attached to remaining nodes that are linked to stubs attached to removed nodes.

Therefore the total fraction of removed stubs is equal to $s(p) + s(p)(1 - s(p))$. This fraction corresponds to an equal fraction of removed *links*. □

3.2.2 The cases of Poisson and power-law networks.

The general result above makes it possible to convert thresholds for classical attacks from the node point of view into thresholds from the link point of view, for any kind of random network. We now apply it to the cases of interest.

In each case, one first has to compute the threshold p_c for classical attacks, in terms of nodes, using the appropriate corollary in Section 3.1.2. Then, one has to apply Lemma 3.3 to obtain $s(p_c)$, and finally Theorem 3.8.

In the case of continuous power-law networks, it is possible to obtain $s(p_c)$ from p_c more easily as follows.

Lemma 3.9 *[30] For continuous power-law networks with size tending towards infinity, exponent α and minimal degree m, the fraction $s(p)$ of stubs attached to the fraction p of nodes removed during a classical attack is given by*

$$s(p) = p^{(2-\alpha)/(1-\alpha)}.$$

Proof : From Lemma 3.3, we know that $s(p) = (\sum_{k=K(p)+1}^{\infty} kp_k)/\langle k\rangle$. We have $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1}) = (\alpha-1)m^{\alpha-1}\int_k^{k+1} x^{-\alpha}\mathrm{d}x$, and $\langle k\rangle = \frac{\alpha-1}{\alpha-2}m$ from Lemma 1.12.

Therefore, $s(p) = \frac{1}{\langle k\rangle}(\alpha-1)m^{\alpha-1}\left[\frac{k^{-\alpha+2}}{-\alpha+2}\right]_{K(p)}^{\infty} = m^{\alpha-2}K(p)^{-\alpha+2}$.

From Lemma 3.7, we finally obtain $K(p) = mp^{1/(-\alpha+1)}$, hence the result. □

We plot numerical evaluations of these results in Figure 12, together with experimental results. We also give in Table 6 the thresholds for specific values of the exponent and of the average degree.

The results are striking: the thresholds are much larger for power-law networks from the link point of view than from the node point of view, see Table 5. More importantly, though the fraction of removed nodes during a classical attack is much lower in power-law networks than in Poisson ones, the fractions of removed links during the very same attack are similar.

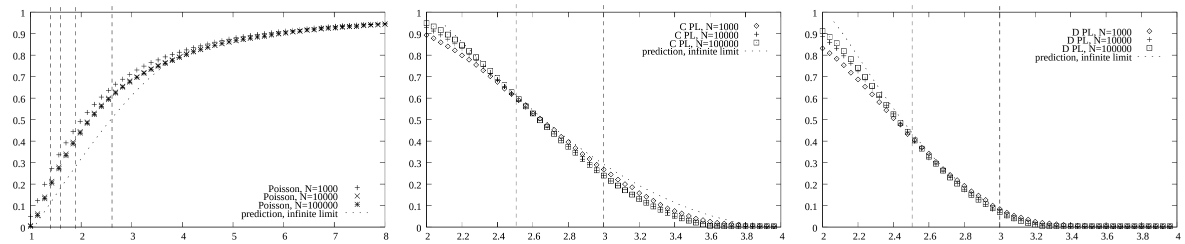

Figure 12: Thresholds for the link point of view of classical attacks. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots, on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 1.4.

	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

Table 6: Values of the threshold for the 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 the same average degree (see Table 1). The values are the analytic previsions at the infinite limit and the ones obtained for experiments with networks of $N = 100\ 000$ nodes.

These observations lead to the conclusion 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 amount of links at random, see Figure 8 and Table 4. Therefore, despite the fact that the amount of removed links is huge and that 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 an important role regarding the network connectivity.

3.3 New attack strategies.

In this section we introduce two very simple new attack strategies, one targeting nodes (Section 3.3.1) and the other targeting links (Section 3.3.2). Let us insist on the fact that these strategies are *not* designed to be efficient; our aim here rather is to deepen our understanding of previous results by considering attack strategies which are very close to random failures.

Our new attack strategies rely on the following observation. We have seen (Theorem 1.15) that a random network with size tending towards 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 key point therefore is the fraction of nodes of degree 1 in the network. It seems that any strategy aimed at increasing this fraction should quickly break the network. The two attack strategies we propose below are based on this.

3.3.1 Almost-random node attacks.

Our first attack strategy simply consists in randomly removing nodes of degree at least 2. We call it the almost-random node attack strategy. Figure 13 displays the behaviors observed for the three types of networks we consider.

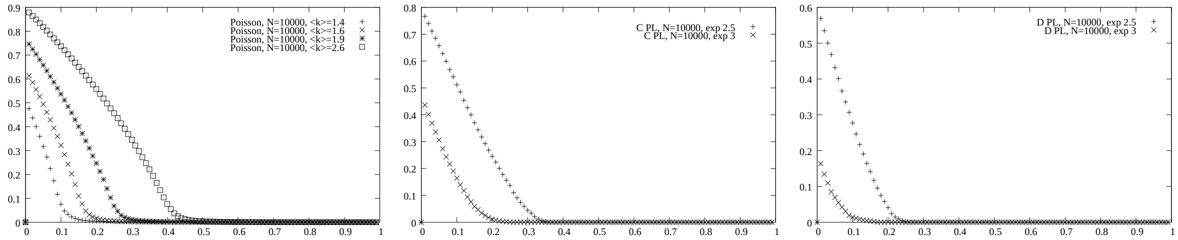

Figure 13: Size of the largest connected component as a function of the fraction of nodes removed during almost-random node attacks. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots see Section 1.4.

Although this strategy is barely different from random node failures, it is actually much more efficient than failures, see Figure 3. In particular, it has a finite threshold for all random networks with a non 0 fraction of nodes of degree 0 or 1, which includes Poisson and power-law networks.

We will begin by proving a general result valid for all random networks, and then apply it to the three cases of interest.

Theorem 3.10 *The threshold for the almost-random node attack strategy for large random networks with degree distribution p_k is bounded by $1 - p_1 - p_0$.*

Proof : When all nodes that had initially a degree higher that one have been removed, then the network surely has no giant component anymore since all nodes have degree at most 1. All nodes of degree higher than one represent a fraction $1 - p_1 - p_0$ of all the nodes, and this is therefore an upper bound for the threshold for this attack strategy. □

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

Corollary 3.11 *For Poisson networks with size tending towards infinity and average degree z, the threshold for almost-random node attacks is bounded by $1 - e^{-z}(z+1)$.*

Proof : Direct application of Theorem 3.10 with $p_k = e^{-z}z^k/k!$. □

Corollary 3.12 *For discrete power-law networks with size tending towards infinity and exponent α, the threshold for almost-random node attacks is bounded by $1 - 1/\zeta(\alpha)$.*

Proof : Direct application of Theorem 3.10 with $p_k = k^{-\alpha}/\zeta(\alpha)$. □

Corollary 3.13 *For continuous power-law networks with size tending towards infinity, exponent α and minimal degree m, the threshold for almost-random node attacks is bounded by $1 - m^{\alpha-1}(1 - 2^{-\alpha+1})$.*

Proof : Direct application of Theorem 3.10, with $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1})$. □

We plot experimental results for the value of the threshold in Figure 14, as well as the upper bounds given above. We also give in Table 7 the thresholds for specific values of the exponent and of the average degree.

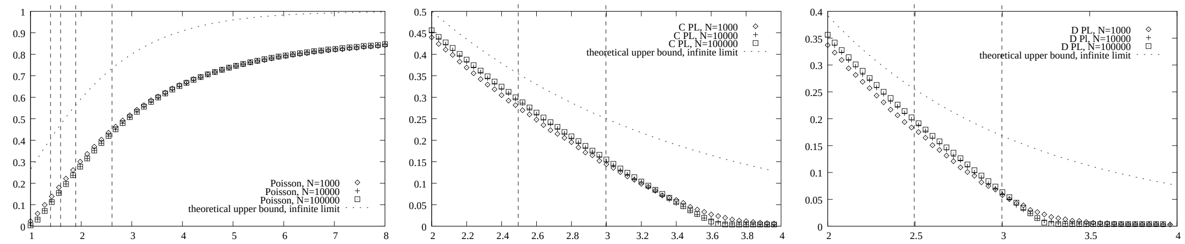

Figure 14: Thresholds and upper bounds for almost-random node attacks. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots, on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 1.4.

Notice that the values of the thresholds are quite large (one has to remove a large fraction of the nodes do destroy the networks), but remain significantly lower than 1. We recall that our aim here is not to obtain an efficient attack strategy, but to study the ability of an attack strategy very similar to random failures to have the same qualitative behavior as classical attack, namely to display a finite threshold for power-law networks.

This is indeed the case of almost-random node attacks. This shows that the efficiency of classical attacks relies in part on simple properties like the fact that it removes nodes of degree larger than 1.

	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

Table 7: Values of the threshold for almost-random node attacks on discrete and continuous power-law networks of exponents 2.5 and 3, and on Poisson networks having the same average degree (see Table 1). The values are the ones obtained for experiments with networks of $N = 100\ 000$ nodes, and the theoretical bounds.

3.3.2 Almost-random link attacks.

We now propose the following attack strategy on links: we randomly remove links between two nodes of degree at least 2. We call it the almost-random link attack strategy. Figure 15 displays the behaviors observed for the three types of networks we consider.

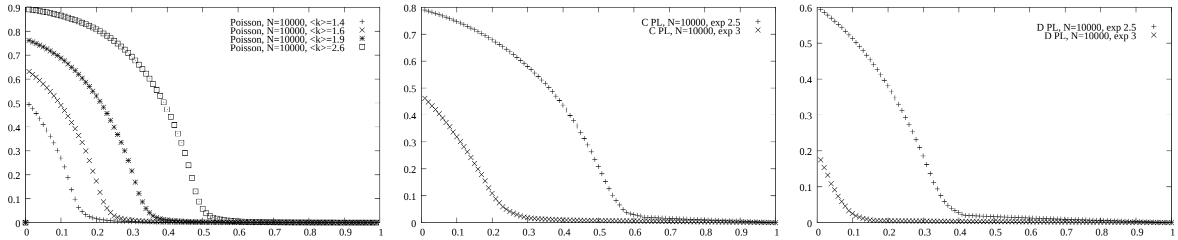

Figure 15: Size of the largest connected component as a function of the fraction of links removed during almost-failure link attacks. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots see Section 1.4.

Although this strategy is barely different from random link failures, it is actually much more efficient than failures, see Figure 7. In particular, it has a finite threshold for all random networks, including power-law ones.

We will begin by proving a general result valid for all random networks, and then apply it to the three cases of interest.

Theorem 3.14 *The threshold 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 $1 - \frac{2p_1}{\langle k \rangle} + \frac{p_1^2}{\langle k \rangle^2}$.*

Proof : The upper bound is the fraction of links between two nodes of degree at least 2. Indeed, when all such links have been removed, the network is nothing but a set of disjoint stars (each central node being connected to nodes of degree 1). The size of the largest component therefore is less than $K + 1$, where K is the maximal degree in the original network. Hence, if K is sublinear with respect to the number of nodes, so is the size of the largest component after the attack.

Let us now evaluate the number of links between nodes of degree at least 2. This quantity is $|E|$ minus the number of links incident to at least one node of degree 1. The number of such links is given by the number of nodes of degree 1, minus the number of links between two nodes of degree 1.

The number of links between two nodes of degree 1 can be evaluated as follows. There are Np_1 nodes of degree 1, each of them having a probability $Np_1/2|E|$ of being connected to another node of degree 1^7. Therefore the number of *nodes* of degree 1 adjacent to another node of degree 1 is $N^2p_1^2/2|E| = Np_1^2/\langle k\rangle$. Finally, the number of links between two such nodes is $Np_1^2/2\langle k\rangle$.

From this we have that the number of links adjacent to at least one node of degree 1 is: $Np_1 - Np_1^2/2\langle k\rangle$, and the number of links *not* adjacent to any node of degree 1 is: $|E| - Np_1 + Np_1^2/2\langle k\rangle$. The fraction of such links therefore is $1 - \frac{2p_1}{\langle k\rangle} + \frac{p_1^2}{\langle k\rangle^2}$, hence the result. □

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

Corollary 3.15 *For Poisson networks with size tending towards infinity and average degree z, the threshold for almost-random link attacks is bounded by $1 - 2e^{-z} + e^{-2z}$.*

Proof : Direct application of Theorem 3.14, the maximal degree of the network being sublinear in the size of the network (see Lemma 1.2), with $p_k = e^{-z}z^k/k!$. □

Corollary 3.16 *For discrete power-law networks with size tending towards infinity and exponent α, the threshold for almost-random link attacks is bounded by $1 - \frac{2\zeta(\alpha-1)-1}{\zeta^2(\alpha-1)}$.*

Proof : Direct application of Theorem 3.14, the maximal degree of the network being sublinear in the size of the network (see Lemma 1.4), with $p_k = k^{-\alpha}/\zeta(\alpha)$. □

Corollary 3.17 *For continuous power-law networks with size tending towards infinity, exponent α and minimal degree m, the threshold for almost-random link attacks is bounded by*

$$1 - \frac{2(\alpha-2)m^{\alpha-2}(1-2^{-\alpha+1})}{(\alpha-1)} + \left(\frac{(\alpha-2)m^{\alpha-2}(1-2^{-\alpha+1})}{(\alpha-1)}\right)^2.$$

[7] This is an approximation of the real value $(N-1)p_1/2|E|$

Proof : Direct application of Theorem 3.14, the maximal degree of the network being sublinear in the size of the network (see Lemma 1.3), with $p_k = m^{\alpha-1}(k^{-\alpha+1} - (k+1)^{-\alpha+1})$. □

We plot experimental results for the value of the threshold in Figure 16, as well as the upper bounds given above. We also give in Table 8 the thresholds for specific values of the exponent and of the average degree.

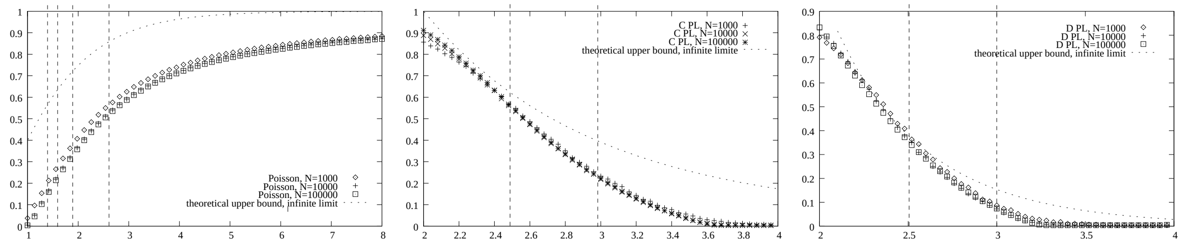

Figure 16: Thresholds and upper bounds for almost-random link attacks. From left to right: Poisson, continuous power-law and discrete power-law networks. For technical details on our plots, on the computation of thresholds, and for discussions on the origins of differences between experiments and predictions, see Section 1.4.

	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 8: Values of the threshold for almost-random link attacks on discrete and continuous power-law networks of exponents 2.5 and 3, and on Poisson networks having the same average degree (see Table 1). The values are the ones obtained for experiments with networks of $N = 100\ 000$ nodes, and the theoretical bounds.

Like in the case of almost-random node attacks, the values of the thresholds are quite large but remain significantly lower than 1. Since out aim is still to study the ability of an attack 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 the fact that it removes 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. This shows that classical attacks, although they focus on high degree nodes, actually remove many links connected to nodes of degree one, which play little role in the connectivity of the network. The simple almost-random strategy, on the opposite, focuses on those links which really disconnect the network.

3.4 Conclusion on attacks.

There are two main formal conclusion for this section. First, as expected from the empirical results discussed in 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, despite the number of links removed during such attacks is huge, it is not sufficient to explain that the network collapses: if the same number of links is randomly removed, then the network does not collapse. However, 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: in terms of links, they are as robust to classical attacks as Poisson networks.

Finally, the attack strategies we introduced, which are very close to random failures, show that the efficiency of classical attacks relies strongly on simple properties like the fact that it removes nodes of degree larger than 1 and links between nodes of degree at least 2.

4 Resilience of real-world networks.

We have seen in previous sections that, apart from the general shape of their degree distributions, precise properties of the networks under concern, like for instance their fraction of nodes of degree 1, may play a crucial role in their behavior in case of failures or attacks. Other properties not captured by the models, like clustering for instance, may also play an important role. In order to observe this and give some insight on the practical incidence of the results above, we present in this section empirical results on real-world complex networks.

We will consider the following real-world cases, which constitute a set of complex networks often taken as illustration in studies of the field. The *actor* network is obtained from the *Internet Movie DataBase* [40]. It consists in links between movie actors, where two actors are linked together if they played in the same movie. See [113, 10] for results on this network. The *co-authoring* network is obtained from the *arXiv* site [9] and consists in links between scientific authors, two authors being linked together if they signed a paper (present in the archive) together. See [85, 86] for results on this kind of networks. The *cooccurrence* network is obtained from the *Bible* [112]. The nodes are the words in this text, with a link between two words if they belong to the same sentence. See [50] for results on this kind of networks. The *internet1* and the *internet-core* networks are two internet maps where the nodes are routers in the internet, two routers being linked if they are at one hop at the IP level. See [58, 61] for details on these complex networks. The *protein* network represents interactions between proteins in a cell. It is provided at [39], and studied in [67]. The *www* network is a set of web pages, two of them being linked together if there is a hyperlink from one of them to the other. This sample is provided at [39] and studied in [7]. The *p2p* network is a set of exchanges between peers, captured in a running peer-to-peer system, two peers being linked if they exchanged a file during the capture. See [62, 63] for precise description and results on this network. We do not detail more the description of these data, which is not our purpose here; details are provided in the references.

Rather, we will discuss their resilience to failures and attacks and will explain it using our knowledge of their structure and the results presented in this paper. Figures 17 to 24 show the obtained plots, together with the actual degree distribution of the network. Many points are worth noticing in these experiments.

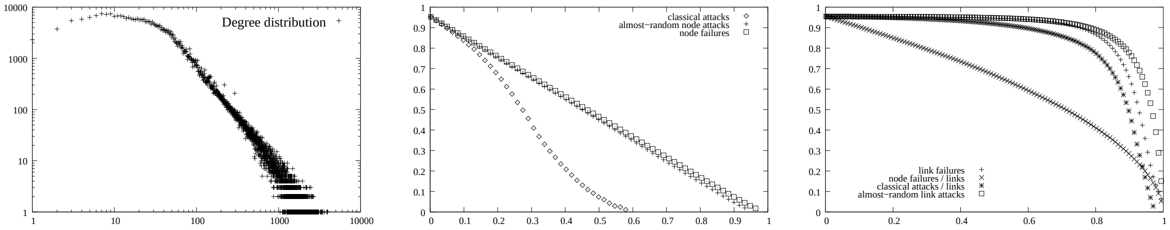

Figure 17: Behavior of the *actor* network in case of failures and attacks. From left to right: the degree distribution, node removals and link removals.

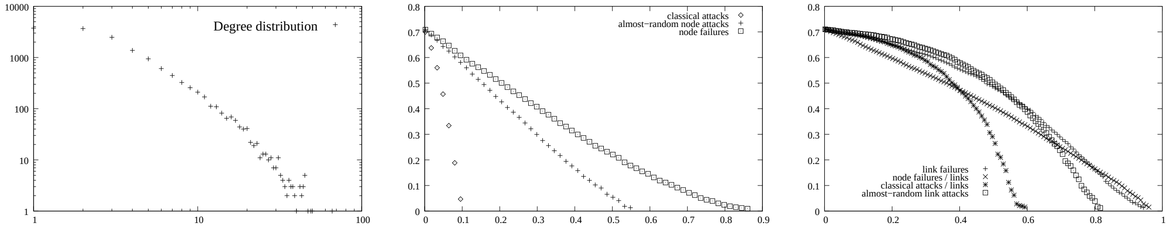

Figure 18: Behavior of the *coauthoring* network in case of failures and attacks. From left to right: the degree distribution, node removals and link removals.

In several cases, namely *actor*, *cooccurrence*, *internet-core* and *p2p*, the behaviors in the case of random node failures and almost-random node attacks are very similar. This is due to the fact that, in these networks, there are few nodes of degree 1 with respect to the total number of nodes.

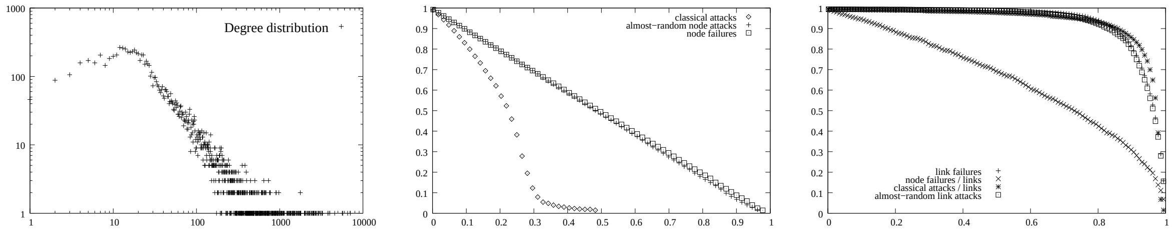

Figure 19: Behavior of the *cooccurrence* network in case of failures and attacks. From left to right: the degree distribution, node removals and link removals.

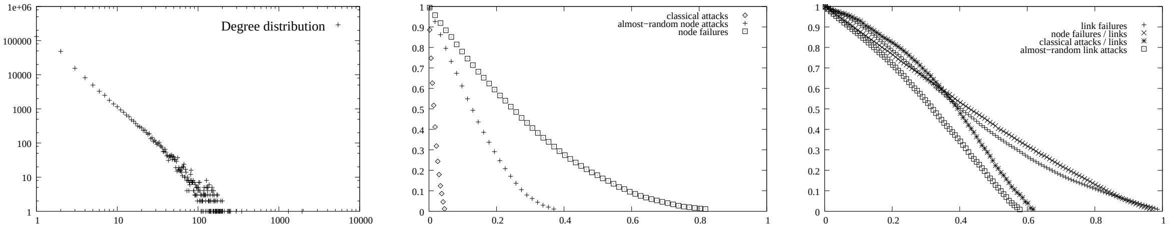

Figure 20: Behavior of the *internet1* network in case of failures and attacks. From left to right: the degree distribution, node removals and link removals.

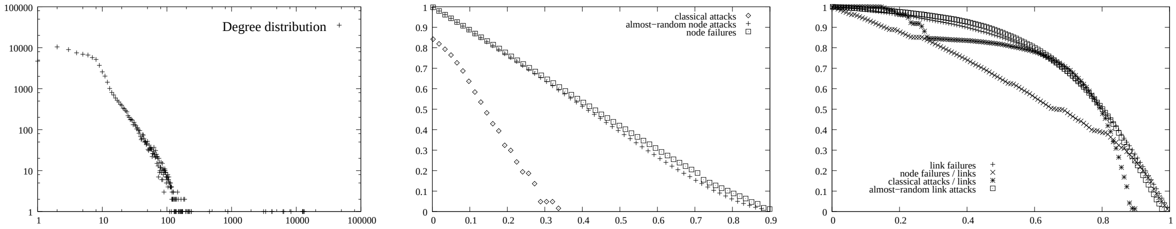

Figure 21: Behavior of the *internet-core* network in case of failures and attacks. From left to right: the degree distribution, node removals and link removals.

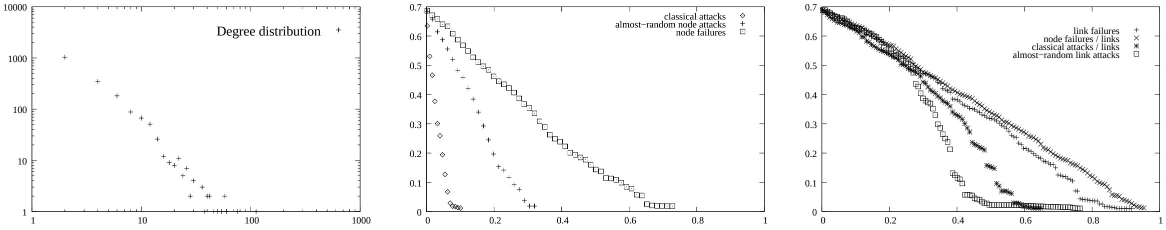

Figure 22: Behavior of the *protein* network in case of failures and attacks. From left to right: the degree distribution, node removals and link removals.

The plots for the *actor*, *cooccurrence*, and *p2p* networks show that, for all link removal strategies except node failures seen from the link point of view, almost all links have to be

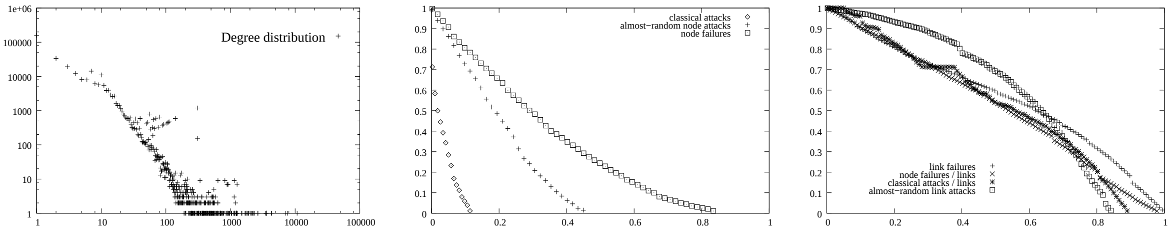

Figure 23: Behavior of the *www* network in case of failures and attacks. From left to right: the degree distribution, node removals and link removals.

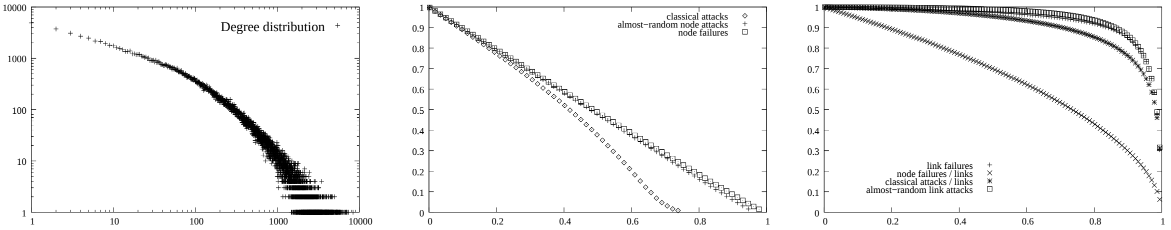

Figure 24: Behavior of the *p2p* network in case of failures and attacks. From left to right: the degree distribution, node removals and link removals.

removed to simply *reduce significantly* the size of the largest connected component. This can be understood as follows. First, notice that for these three networks, the beginning of the plots for random node failures, almost-random node attacks and classical attacks are almost identical. This means that these three strategies have the same impact on the connectivity of the network when few nodes have been removed (up to 20 % in the case of *p2p*). Intuitively, this means that there are few nodes of degree 1 (as already noted), and that the nodes of highest degree are redundant: they are all linked to each other, and linked mostly to the same lower degree nodes, which explains why their removal does not damage the network more than the random removal of nodes. However, a very large fraction of the links is attached to the nodes of highest degree, which explains why classical attacks are less efficient than random node failures when seen from the link point of view. The fact that such a large fraction of links is attached to the highest degree nodes also explains why the behavior observed for random link failures and almost-random link attacks are very similar to classical attacks seen from the link point of view.

The behaviors of the two maps of the internet are very different, which shows that one must be very careful when driving conclusions about such networks. Indeed, the measurement procedure only gives a partial and biased view, see [61]. This moderates the often claimed assertion that the internet is very robust to failures and very sensitive to attacks, which has been derived from such experiments, typically conducted on maps of the kind of the *internet1* one.

In this last case, one may notice that almost-random link failures destroy the network more efficiently than classical attacks, when viewed from the link point of view. This indicates that the robustness of this network in case of failures is strongly due to the fact that the amount of nodes of degree 1 is huge, which is indeed confirmed by the plots.

In the case of *www*, the behaviors concerning node failures or attacks are exactly what one would intuitively expect. It makes the behaviors concerning link failures and attacks even more striking: whatever the strategy used, it seems that almost all links must be removed in order to destroy the network. However, to remove all links, one needs to remove fewer nodes by classical attacks than by random node failures.

More generally, it appears that in all the cases there is a significant difference between random node failures and classical attacks, even if this difference is quite small in the cases of *actor* and *p2p* networks. Instead, when we observe link removals, there are clearly two different classes of networks: the ones in which one has to remove almost all the links to break the network, in all the strategies (namely *actor*, *cooccurrence*, *internet-core*, *www* and *p2p* networks), and the ones in which some strategies break the network after removal of a fraction of the links significantly lower than 1 (namely *coauthoring*, *internet1*, and *protein* networks). There is no simple *a priori* explanation for this, but this certainly points out fundamental differences between these networks.

These experiments show that all the aspects we have discussed in this paper must be taken into account when dealing with practical cases. They also show that much remains to be done, as we will discuss further in the next section. On the other hand, one may see the study of the resilience of a given network as a way to deepen the study of its structure and point out some non-trivial features which should be explored. This is the case, for instance, of the remarks we made on the *www* network above.

Notice however that none of these networks has a Poisson structure, though their degree distribution sometimes is quite far from a power-law. In all the cases the degree distribution is highly heterogeneous, which makes the discussed attack strategies relevant. In most cases, classical attacks are much more efficient than random failures. As illustrated above, considering them from the link point of view, and studying almost-random attacks, however, helps much in understanding what happens.

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. These results state that, despite 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 attacks (which we named the *classical attacks*), while Poisson networks are not. As already quoted in the introduction, this had lead to the conclusion that its power-law degree distribution may be seen as an *Achille's heel of the internet* [1, 11].

These results were first obtained empirically [8, 18], but an important analytic effort has been done to prove them with mean-field and asymptotic approximations. This has been done with success in [29, 30, 33, 24, 88].

Our first notable contribution here is to give a unified and complete presentation of these results (both empiric and analytic ones). Since some of the involved techniques (in particular mean-field approximations) are unusual in computer science, we emphasized on the methods and gave proofs much more detailed than in the original papers. In particular, we pointed out the approximations where they occur, we discussed them in the light of the experiments, and we tried to give a didactic presentation.

This complete and unified presentation of the field was an excellent opportunity to present some new results on cases which received less attention (like the case of finite networks for instance), maybe because these results are less striking. They are however essential if one wants to deepen one's understanding of the more studied cases. We focused in particular on two aspects: the link point of views, and attacks very similar to random failures.

We explained in the core of the paper what happens in each case. We also gave at the end of the section of random failures, Section 2, a summary of what this section in the end teaches us on random failures. We proceeded likely for attacks. In each case, we saw 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 failures, the striking point is that, although analysis predicts completely different behaviors between Poisson and power-law networks, in practice the differences, though important, are not huge (see Tables 2 to 4). This overestimation of the difference was due to the study of the infinite limit, as shown by our analysis of the finite cases, and to the approximations. It may also be a consequence of our choice to consider that the giant component must contain at least 5% of all the nodes, but taking another convention leads to similar conclusions. These conclusions hold for random node failures, and are even more pronounced for link failures.

Concerning classical attacks, we have shown that, although the thresholds for power-law networks indeed are very low, and much smaller than for Poisson ones, the other cases we studied tend to moderate this conclusion. Indeed, as one may have guessed, 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, however is not sufficient to explain the efficiency of classical attacks: if one removes the same fraction of links but randomly, then there is no breakdown. This invalidates the often claimed explanation of the efficiency of classical attacks on power-law networks by the fact that 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 is reached. In terms of the fraction of removed links, therefore, classical attacks 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 it removes many links, and that these links are mostly attached to nodes of degree larger than 1. Conversely, this explains the robustness of power-law networks to random node failures by the fact that such failures often remove nodes of degree 1 and/or links attached to such nodes.

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

The observation of practical cases in Section 4 also provided interesting insights: in several cases, some observed behaviors may be explained using the results in this paper and our knowledge of the properties of the underlying network. It appears clearly however that other properties than degree distributions are at work concerning network resilience. In particular, we identified two distinct classes of behaviors, which points out interesting directions for further analysis. Conversely, one may see the study of a particular network's resilience as a way to obtain some insight on its structure.

All these results lead 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 driving conclusions from this. This is confirmed by our experiments on real-world networks. The sensitivity of networks to attacks relies on the fact that they have many low-degree nodes. Their robustness relies on the fact that when we choose a node at random, we choose such a node with high probability. Moreover, the fact that a classical attack on a power-law network removes many links may be considered as partly, but not fully, responsible for its rapid breakdown.

Despite the fact that this paper is already quite long, there are of course many omissions, and we had to make some choices in the presented results. For instance, we did not mention random networks with degree correlations, on which interesting results exist [15, 111]. We also ignored the various contributions considering other attack strategies [76, 34, 35, 92, 20, 94, 66, 51, 84, 83, 82, 101, 117]. Likewise, we could have compared real-world networks in Section 4 with random networks having exactly the same degree distribution, which would certainly be enlightening. It would be interesting also to compare the results obtained when we consider that a component of less than 5 % of the nodes is still a giant component.

Presenting and discussing all these aspects is impossible in a reasonable space; instead, we chose to deepen the basic results of the field, which we hope leads to a significantly improvement of our understanding of them. This work could serve as a basis for further deepening of some aspects we have ommitted, like the ones pointed out above.

Going further, it must be clear that other properties than the degree distribution are at work in real-world complex networks, and that they certainly play a role in their robustness. This appears clearly in Section 4, where several classes of behaviors appear. It seems obvious, for instance, that the fact that most nodes of real-world complex networks are spread between pieces such that most links of the network are between nodes of the same piece (captured in part by the notion of clustering coefficient) [56, 72, 52, 57, 53, 90, 73] plays a key role. Random networks of the kind we considered here (as in most of the literature) do not have these kinds of properties.

Studying robustness of networks with more subtle properties than degree distributions would therefore be highly relevant, but most remains to be done. In particular, while there is a consensus on the modeling of degree distributions of networks in the community (through the use of the configuration model [12], like here, or the preferential attachment principle [5]), there is no consensus for more subtle properties like the clustering. Many models have been proposed, but each has its own advantages and drawbacks. Some of them seem however well suited for analysis, as they are simple extensions of the configuration model [60, 59] or of the preferential attachment one [44, 45].

Finally, let us insist once more on the necessity to develop formal results to enhance our understanding of empiric results. It makes no doubt that experiments bring much understanding and intuition on phenomena of interest. In our case, these first steps were presented in [8, 18]. The need for a deeper understanding of what happens during these experiments, and the need of rigor of course, did lead to several approaches to analyze them. The main ones in our context are developed in [29, 30, 33, 24, 88]. They both rely on mean-field approximations, but with slightly different points of view. We extended them here, and we gave the details of the underlying approximations and assumptions. Such an approach is definitively rigorous, but is is not *formal.* Obtaining exact results with formal methods, or even approximate results with formal methods, would be another improvement. Some results begin to appear in this direction [17], but much remains to be done and the task is challenging.

Acknowledgments

We thank Fabien Viger for valuable comments on degree distributions. This work was supported in part by the MetroSec (*Metrology of the internet for Security and quality of services*, `http://www.laas/fr/~METROSEC`) project and by the GAP (*Graphs, Algorithms and Probabilities*) project.

References

[1] Cover of nature, volume 406, 2000.

[2] D. Achlioptas, A. Clauset, D. Kempe, and C. Moore. On the bias of traceroute sampling; or, power-law degree distributions in regular graphs. 2005. to appear.

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

[4] W. Aiello, F.R.K. Chung, and L. Lu. A random graph model for massive graphs. In *ACM Symposium on Theory of Computing*, pages 171–180, 2000.

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

[6] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics 74, 47*, 2002.

[7] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world wide web. *Nature*, 401:130–131, 1999.

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

[9] arXiv.org e Print archive. http://arxiv.org/.

[10] A.-L. Barabási. *Linked: The New Science of Networks.* Perseus Publishing, 2002.

[11] A.-L. Barabási. Emergence of scaling in complex networks. In Stefan Bornholdt and Heinz Georg Schuster, editors, *Hankbook of Graphs and Networks: From the Genome to the Internet*. Wiley-vch, 2003.

[12] E. A. Bender and E. R. Canfield. The asymptotic number of labelled graphs with given degree sequences. *Journal of Combinatorial Theory (A)*, 24:357–367, 1978.

[13] A. Beygelzimer, G. Grinstein, R. Linsker, and I. Rish. Improving network robustness. In *First International Conference on Autonomic Computing (ICAC'04)*, pages 322–323, 2004.

[14] A. Beygelzimer, R. Linsker, G. Grinstein, and I. Rish. Improving network robustness by edge modification. *to appear in Physica A*, 2005.

[15] M. Boguná, R. Pastor-Satorras, and A. Vespignani. Epidemic spreading in complex networks with degree correlations. *Lecture Notes in Physics*, 625:127–147, 2003.

[16] B. Bollobás. *Random Graphs.* Academic Press, 1985.

[17] B. Bollobás and O. Riordan. Robustness and vulnerability of scale-free random graphs. *Internet Mathematics*, 1 (1), 2003. available at http://www.internetmathematics.org/.

[18] A.Z. Broder, S.R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J.L. Wiener. Graph structure in the web. *WWW9 / Computer Networks*, 33(1-6):309–320, 2000.

[19] A. Broido and K. Claffy. Internet topology: connectivity of IP graphs. In *SPIE International symposium on Convergence of IT and Communication,Denver, CO.*, 2001.

[20] A. Broido and K. Claffy. Topological resilience in ip and as graphs. 2002. `http://www.caida.org/analysis/topology/resilience/`

[21] T. Bu and D. Towsley. On distinguishing between Internet power law topology generators, 2002. In Proceedings of INFOCOM, 2002.

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

[23] D.S. Callaway, J.E. Hopcroft, J.M. Kleinberg, M.E.J. Newman, and S.H. Strogatz. Are randomly grown graphs really random. Santa Fe Institute, Working Papers List, 2001.

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

[25] H. Chang, S. Jamin, and W. Willinger. Inferring AS-level internet topology from router-level path traces, 2001. in Proceeding of SPIE ITCom 2001, Denver, CO, August 2001.

[26] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. The origin of power laws in internet topologies revisited. In *INFOCOM*, 2002.

[27] L.P. Chi, C.B. Yang, and X. Cai. Stability of complex networks under the evolution of attack and repair. cond-mat/0505197, 2005.

[28] F. Chung and L. Lu. Connected compnents in a random graph with given degree sequences. *Annals of Combinatorics*, 6:125–145, 2002.

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

[30] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin. Breakdown of the internet under intentional attack. *Phys. Rev. Lett.*, 86:3682–3685, 2001.

[31] R. Cohen, R. Erez, D. ben Avraham, and S. Havlin. Reply to the comment on 'breakdown of the internet under intentional attack'. *Phys. Rev. Lett*, 87, 2001.

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

[33] R. Cohen, S. Havlin, and D. ben Avraham. Structural properties of scale-free networks. In Stefan Bornholdt and Heinz Georg Schuster, editors, *Hankbook of Graphs and Networks: From the Genome to the Internet*. Wiley-vch, 2003.

[34] P. Crucitti, V. Latora, and M. Marchiori. Error and attack tolerance of complex networks. *Physica A*, 340, 2004.

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

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

[37] L. da Fontoura Costa. Reinforcing the resilience of complex networks. *Phys. Rev. E*, 69(066127), 2004.

[38] L. Dall'Asta, I. Alavrez-Hamelin, A. Barrat, A. Vazquez, and A. Vespignani. Exploring networks with traceroute-like probes: theory and simulations. *To appear in Theoretical Computer Science*, 2005.

[39] Self-Organized Networks Database. http://www.nd.edu/˜networks/database/index.html.

[40] The Internet Movie Database. http://www.imdb.com/.

[41] Z. Dezsö and A.-L. Barabási. Halting viruses in scale-free networks. *Phys. Rev. E*, 65, 2002. cond-mat/0107420.

[42] S.N. Dorogovtsev and J.F.F. Mendes. *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, 2000.

[43] S.N. Dorogovtsev and J.F.F. Mendes. Comment on 'breakdown of the internet under intentional attack'. *phys. Rev. Lett*, 87, 2001.

[44] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. *Adv. Phys. 51, 1079-1187*, 2002.

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

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

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

[48] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–262, 1999.

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

[50] R. Ferrer and R.V. Solé. The small-world of human language. In *Proceedings of the Royal Society of London*, volume B268, pages 2261–2265, 2001.

[51] P. Flajolet, K. Hatzis, S. Nikoletseas, and P. Spirakis. On the robustness of interconnections in random graphs: a symbolic approach. *Theor. Comput. Sci.*, 287(2):515–534, 2002.

[52] G. Flake, S. Lawrence, and C. Lee Giles. Efficient identification of web communities. In *Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 150–160, Boston, MA, August 20–23 2000.

[53] G.W. Flake, S. Lawrence, C. Lee Giles, and F. Coetzee. Self-organization of the web and identification of communities. *IEEE Computer*, 35(3):66–71, 2002.

[54] L.K. Gallos, P. Argyrakis, A. Bunde, R. Cohen, and S. Havlin. Tolerance of scale - free networks : from friendly to intentional attack strategies. *Physica A: Statistical Mechanics and its Applications*, 344 (3-4):504–509, 2004.

[55] L.K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin. Stability and topology of scale-free networks under attack and defense strategies. *Phys. Rev. Lett.*, 94(188701), 2005.

[56] D. Gibson, J.M. Kleinberg, and P. Raghavan. Inferring web communities from link topology. In *UK Conference on Hypertext*, pages 225–234, 1998.

[57] M. Girvan and M.E.J. Newman. Community structure in social and biological networks. submitted to Proc. Natl. Acad. Sci., 2001.

[58] R. Govindan and H. Tangmunarunkit. Heuristics for internet map discovery. In *IEEE INFOCOM 2000*, pages 1371–1380, Tel Aviv, Israel, March 2000. IEEE.

[59] J.-L. Guillaume and M. Latapy. Bipartite graphs as models of complex networks. In *Lecture Notes in Computer Science, proceedings of the International Workshop on Combinatorial and Algorithmic Aspects of Networking, Banff, Canada*, 2004.

[60] J.-L. Guillaume and M. Latapy. Bipartite structure of *all* complex networks. *Information Processing Letters*, 90:5:215–221, 2004.

[61] J.-L. Guillaume and M. Latapy. Relevance of massively distributed explorations of the internet topology: Simulation results. In *Proceedings of the IEEE 24-th INFOCOM'05, Miami, USA*, 2005.

[62] J.-L. Guillaume, M. Latapy, and S. Le-Blond. Statistical analysis of a p2p query graph based on degrees and their time-evolution. In *Lecture Notes in Computer Science, proceedings of the 6-th International Workshop on Distributed Computing IWDC'04, Calcutta, Inde*, 2004.

[63] J.-L. Guillaume, M. Latapy, and S. Le-Blond. Clustering in p2p exchanges and consequences on performances. In *Lecture Notes in Computer Science, proceedings of the 4-th International Workshop on Peer to Peer Systems IPTPS'05, Cornell, USA*, 2005.

[64] P. Holme. Efficient local strategies for vaccination and network attack. *Europhys. Lett.*, 68(6):908–914, 2004.

[65] P. Holme and B. Jun Kim. Growing scale-free networks with tunable clustering. *Phys. Rev. E*, 65, 2002.

[66] P. Holme, B. Jun Kim, C. No Yoon, and S. Kee Han. Attack vulnerability of complex networks. *Phys. Rev. E*, 65, 2002.

[67] H. Jeong, B. Tombor, R. Albert, Z. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature, 407, 651*, 2000.

[68] E.M. Jin, M. Girvan, and M.E.J. Newman. The structure of growing social networks. *Phys. Rev. E 64*, (046132), 2001.

[69] J.M. Kleinberg. The small-world phenomenon: An algorithmic perspective, 1999.

[70] K. Klemm and V.M. Eguiluz. Highly clustered scale-free networks. *Phys. Rev. E*, 65, 2002.

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

[72] S. Ravi Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Trawling the web for emerging cyber-communities. *WWW8 / Computer Networks*, 31(11-16):1481–1493, 1999.

[73] M. Latapy and P. Pons. Efficient computation of communities in very large graphs. 2005. Submitted.

[74] V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Phys. Rev. Lett.*, 87, 2001.

[75] L. Laura, S. Leonardi, S. Millozzi, U. Meyer, and J. F. Sibeyn. Algorithms and experiments for the Webgraph, 2003. European Symposium on Algorithms.

[76] E.J. Lee, K.-I. Goh, B. Kahng, and D. Kim. Robustness of the avalanche dynamics in data packet transport on scale-free network. *to appear in Phys. Rev. E*, 2005.

[77] F. Liljeros, C.R. Edling, L.A. Nunes Amaral, H. Eugene Stanley, and Y. Aberg. The web of human sexual contacts. *Nature*, 411:907–908, 2001.

[78] D. Magoni and J.-J. Pansiot. Analysis of the autonomous system network topology. *ACM SIGCOMM Computer Communication Review*, 31(3):26 – 37, July 2001.

[79] Maple. http://www.maplesoft.com/.

[80] M. Molloy and B. Reed. A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6:161–179, 1995.

[81] M. Molloy and B. Reed. The size of the giant component of a random graph with a given degree sequence. *Combin. Probab. Comput.*, pages 295–305, 1998.

[82] A.E. Motter. Cascade control and defense in complex networks. *Phys. Rev. Lett.*, 93, 2004.

[83] A.E. Motter and Y.-C. Lai. Cascade-based attacks on complex networks. *Physical Review E 66*, 2002.

[84] A.E. Motter, T. Nishikawa, and Y.-C. Lai. Range-based attack on links in scale-free networks: are long-range links responsible for the small-world phenomenon? *Phys. Rev. E*, 66, 2002.

[85] M.E.J. Newman. Scientific collaboration networks: I. Network construction and fundamental results. *Phys. Rev. E*, 64, 2001.

[86] M.E.J. Newman. Scientific collaboration networks: II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E*, 64, 2001.

[87] M.E.J. Newman. Properties of highly clustered networks. *Phys. Rev. E*, 68, 2003.

[88] M.E.J. Newman. Random graphs as models of networks. In Stefan Bornholdt and Heinz Georg Schuster, editors, *Hankbook of Graphs and Networks: From the Genome to the Internet*. Wiley-vch, 2003.

[89] M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[90] M.E.J. Newman. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 69, 2004.

[91] M.E.J. Newman, S.H. Strogatz, and D.J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 2001.

[92] D. Newth and J. Ash. Evolving cascading failure resilience in complex networks. In *In proceedings of the Eighth Asia Pacific Symposium on Intelligent and Evolutionary Systems*, pages 125–136, 2004.

[93] J. Park and M.E.J. Newman. The origin of degree correlations in the internet and other networks. *Phys. Rev. E*, 68, 2003.

[94] S.-T. Park, A. Khrabrov, D.M. Pennock, S. Lawrence, C. Lee Giles, and L.H. Ungar. Static and dynamic analysis of the internet's susceptibility to faults and attacks. In *IEEE Infocom 2003*, San Francisco, CA, April 1–3 2003.

[95] R. Pastor-Satorras, A. Vazquez, and A. Vespignani. Dynamical and correlation properties of the internet. *Phys. Rev. Lett. 87, 258701*, 2001.

[96] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, 86:3200–3203, 2001.

[97] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Phys. Rev. E*, 65, 2002.

[98] G. Paul, S. Sreenivasan, and H. Eugene Stanley. Resilience of complex networks to random breakdown, 2005. cond-mat/0507202.

[99] G. Paul, T. Tanizawa, S. Havlin, and H.E. Stanley. Optimization of robustness of complex networks. In *2003 International Conference on Growing Networks and Graphs*, 2003.

[100] G. Paul, T. Tanizawa, S. Havlin, and H.E. Stanley. Optimization of robustness of complex networks. *Euro. Phys. J. B*, 38:187–191, 2004.

[101] S. Pertet and P. Narasimhan. Handling cascading failures: The case for topology-aware fault tolerance. In *Proceedings of the IEEE First Workshop on Hot Topics in System Dependability*, 2005.

[102] G. R. Raidl and I. Ljubic. Evolutionary local search for the edge-biconnectivity augmentation problem. *Information Processing Letters*, 82(1):39–45, 2002.

[103] B.A. Rezaei, N. Sarshar, P. Oscar Boykin, and V.P. Roychowdhury. Disaster management in scale-free networks: Recovery from and protection against intentional attacks. cond-mat/0504185.

[104] D. Stauffer and A. Aharony. *Introduction to Percolation Theory.* Taylor & Francis, London, 2nd edition, 1994.

[105] S.H. Strogatz. Exploring complex networks. *Nature 410*, March 2001.

[106] H. Tangmunarunkit, J. Doyle, R Govindan, S. Jamin, S. Shenker, and W. Willinger. Does AS size determine degree in AS topology? *Comput. Commun. Rev.*, 31:7–10, 2001.

[107] H. Tangmunarunkit, R Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topologies, power laws, and hierarchy. *Comput. Commun. Rev.*, 32:76–76, 2002.

[108] T. Tanizawa, G. Paul, R. Cohen, S. Havlin, and H.E. Stanley. Optimization of network robustness to waves of targeted and random attacks. *Phys. Rev. E (Rapid Communications)*, 71(047101), 2005.

[109] P. Uetz, L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J.M. Rothberg. A comprehensive analysis of protein-protein interactions in *saccharomyces cerevisiae. Nature*, 403:623–627, 2000.

[110] A.X.C.N. Valente, A. Sarkar, and H.A. Stone. 2-peak and 3-peak optimal complex networks. *Physical Review Letters*, 92(118702), 2004.

[111] A. Vázquez and Y. Moreno. Resilience to damage of graphs with degree correlations. *Phys. Rev. E*, 67, 2003.

[112] Bible Today New International Version. http://www.tniv.info/bible/.

[113] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[114] H. Wilf. *Generating functionology.* Academic Press/Harcourt Brace, 1994.

[115] W. Willinger, R Govindan, S. Jamin, V. Paxson, and S. Shenker. Scaling phenomena in the Internet: Critically examining criticality. *Proc. Natl. Acad. Sci. USA*, 99:2573–2580, 2002.

[116] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the Internet's large-scale topology. *Proc. Nat. Acad. Sci. USA*, 99:13382–13386, 2002.

[117] L. Zhao, K. Park, , and Y.-C. Lai. Attack vulnerability of scale-free networks due to cascading breakdown. *Phys. Rev. E*, 70, 2004.

[118] S. Zhou and R.J. Mondragon. Redundancy and robustness of the as-level internet topology and its models. *IEE Electronic Letters*, 40 (2):151, 2004.