Extremum Statistics in Scale-Free Network Models

André Auto Moreira,[1,2,*] José S. Andrade, Jr.,[2,†] and Luís A. Nunes Amaral[1,3,‡]
[1]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215*
[2]*Departamento de Física, Universidade Federal do Ceará, 60451-970 Fortaleza, Ceará, Brazil*
[3]*Department of Chemical Engineering, Northwestern University, Evanston, Illinois 60208*
(Received 29 April 2002; published 12 December 2002)

We investigate the statistics of the most connected node in scale-free networks. For a scale-free network model with homogeneous nodes, we show by means of extensive simulations that the exponential truncation, due to the finite size of the network, of the degree distribution governs the scaling of the extreme values and that the distribution of maxima follows the Gumbel statistics. For a scale-free network model with heterogeneous nodes, we show that scaling no longer holds and that the truncation of the degree distribution no longer controls the maxima distribution.

DOI: 10.1103/PhysRevLett.89.268703 PACS numbers: 89.75.Hc, 05.10.–a

The statistics of extrema is a classical subject of great interest in mathematics and physics [1]. In physics, extreme events have been studied in a number of fields, including self-organized fluctuations and critical phenomena [2], material fracture [3], disordered systems at low temperatures [4], and turbulence [5]. Knowledge of extreme event statistics is also of fundamental importance to predict and estimate risk in a variety of natural and man-made phenomena such as earthquakes, changes in climate conditions, floods [6], and large movements in financial markets [7]. A new field where extreme statistics is of interest is complex networks [8]. For one particular class of complex networks [9]—scale-free networks [10,11]—it is well known that the most connected nodes strongly influence the dynamics of the system, playing a fundamental role in many different phenomena such as Internet response to attacks [12], spreading of epidemics [13], or propagation of email virus [14]. Surprisingly, so far there has been no attempt to characterize the distribution of extreme connectivities in scale-free networks.

An important result in extreme statistics is that the distributions of maxima for independent identically distributed (iid) random variables fall onto a small number of universality classes [1]. Let $C \equiv \{k_1, k_2, \ldots, k_S\}$ be a set of iid variables drawn with probability density function $p(k)$. The distribution $\rho(K)$ of the maximum K in the set C is dictated by the asymptotic behavior of the tail of $p(k)$ [1]. Specifically, $\rho(K)$ converges to the Gumbel distribution,

$$\rho(K) = a \exp(-u - e^{-u}), \tag{1}$$

where $u = a(K - b)$, when $p(k)$ decays faster than a power law; and to the Frechet distribution,

$$\rho(K) = \alpha K^{-(\alpha+1)} \exp(-K^{-\alpha}), \tag{2}$$

when $p(k)$ decays as $k^{-(\alpha+1)}$ [15,16].

Unlike the case of iid random variables, little is known when correlations are present among the variables k_i [4,17]. Even though the universality classes of uncorrelated and correlated variables may not necessarily be the same, correlated systems have been generally studied under the framework of iid extreme statistics [18]. In the case of scale-free networks, which we consider in this Letter, correlations are present in the degrees k_i (i.e., the number of links) of the nodes [19].

To investigate the extreme statistics in scale-free networks, we consider here the fitness model of Ref. [20]. This model is a generalization of the scale-free model of Ref. [10], in which nodes have heterogeneous fitnesses $\{\eta_i; i = 1, \ldots, S\}$. The fitness η_i models an inherent quality of the node i that "weighs" its attractiveness to new links. In this model, the network starts with s_0 nodes, each with $s_0 - 1$ links. At time t, a new node is added to the network and establishes $s_0 - 1$ new links. A new link is established with a node i, from the set of the $t - 1 + s_0$ existing nodes, with a probability proportional to the node degree k_i and fitness η_i,

$$\Pi_i = \frac{k_i \eta_i}{\sum k_j \eta_j}. \tag{3}$$

This mechanism, typically denoted "preferential attachment," drives the network to a degree distribution that decays in the tail as a power law [10]. In the homogeneous case, $\eta_i = 1$ for all i, one recovers the original model of Ref. [10] and generates a network with cumulative degree distribution that decays as $P(k) \sim k^{-2}$ [10]. In the case where the fitnesses η_i are drawn from a uniform distribution $\eta_i \epsilon [0, 1]$, the network has a cumulative degree distribution of the form $P(k) \sim k^{-\alpha}/\log(k)$, with $\alpha = 1.225$ [20].

Here, we demonstrate that for the homogeneous case—i.e., all η_i are equal—the distribution of maxima obeys Gumbel statistics. This is a surprising result for two reasons: (i) the degrees k_i are not iid variables and are not equally distributed—hence, there is not an *a priori* justification to expect that one of the two universality classes (1) or (2) will hold—and (ii) the distribution $P(k)$ decays as a power law—hence, one would more likely expect to find Frechet statistics. For the case of nodes with heterogeneous fitness, we find that the distribution of maximum, for finite network sizes, is not consistent with either of the universality classes represented by Eqs. (1) and (2). However, our results suggest that the distribution of maxima converges to the Frechet distribution in the thermodynamic limit.

First we consider the case $\eta_i = 1$ for all i. The distribution of the maximum degree K is nontrivial because (i) the degrees k_i display a constraint on the total number of links and (ii) the variables k_i are not identically distributed. Indeed, recent studies have shown that each node has a different probability distribution for its degree $p_i(k_i)$ obeying an exponential form with a characteristic scale that depends on the square root of the node index i [21,22].

In Fig. 1(a) we show the cumulative degree distribution, $P(k) = \sum_{k'>k} p(k')$, for different network sizes S and $\eta_i = 1$ (homogeneous case). These results have been obtained numerically through iteration of the rate equation proposed in Ref. [22]. As expected, the curves display a power-law decay, $P(k) \propto k^{-\alpha}$, with $\alpha = 2$, followed by an exponential truncation. The data collapse is obtained by scaling all curves according to [21,22]

$$SP(k) \propto \left(\frac{k}{S^{1/\alpha}}\right)^{-\alpha}. \qquad (4)$$

Also shown in Fig. 1(a) are the distributions of maximum degree for different network sizes S and averaged over 10^5 network realizations [23]. After scaling with the same exponent that governs the degree distribution, $K' \equiv S^{-1/\alpha}K$, we observe that the maxima overlap the exponentially decaying region of the degree distribution. As a consequence, the scaling of the degree distribution's truncation controls the extremum statistics. Figure 1(b) shows the distribution of maximum degree—scaled in the fashion of Fig. 1(a)—compared with the best fittings to the data of the Frechet and Gumbel distributions. It is apparent that the Gumbel distribution describes the maximum statistics quite well for $K' < 5$ [24], suggesting that the truncation of the degree distribution dictates the exponential decay of the distribution of maxima.

The surprising result that the statistics of maxima for homogeneous scale-free networks is described by the Gumbel distribution has its origin in the non-iid character of the degrees in the network. Indeed, one would have expected Frechet statistics to hold in this case due to the power-law form of the degree distribution. However, the constraint in the total number of links limits the maximum degree of a node and the maximum distribution assumes the exponentially decaying Gumbel form [25].

Next, we consider the case of a growing network with nodes having a uniform distribution of fitnesses. In this case, fitter nodes will enter the network and compete for new links with the less fit nodes in the network, eventually becoming the most connected nodes. This is clearly shown in Fig. 2, where we plot the fitness distribution of the most connected node for different network sizes. As the network grows, the distribution converges logarithmically to a delta function at $\eta = 1$. In the thermodynamic limit, the fittest node, with fitness one, becomes the most connected. This is to be expected since the growth of the degree of a node increases over time as a power law with an exponent proportional to its fitness [20].

In Fig. 3(a) we compare the degree distribution and the maximum statistics for the heterogeneous case [26]. The cumulative degree distribution follows the expected scaling and displays, as in the homogeneous case, an

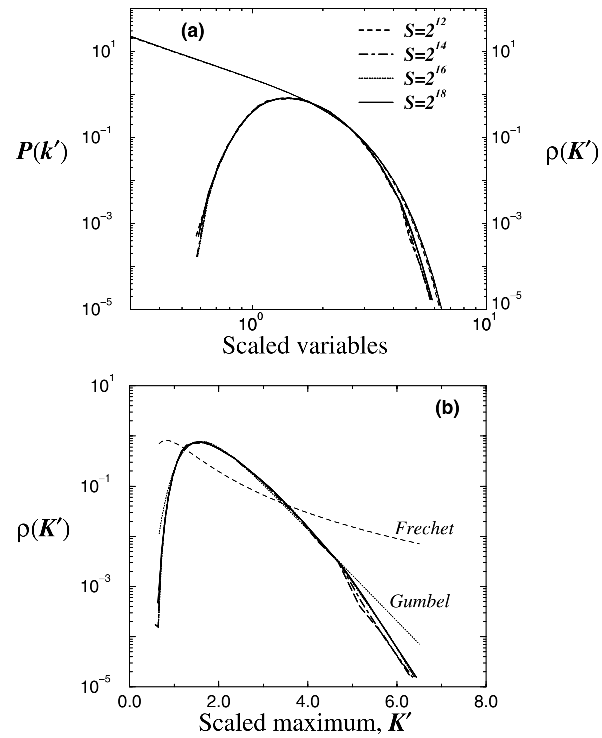

FIG. 1. (a) Log-log plot of the cumulative degree distribution (thin lines) and distribution of maximum degree (thick lines) for the homogeneous case [10]. The cumulative distribution decays as a power law with an exponent $\alpha = 2$, followed by an exponential truncation. Using the rescaling relation $K' \equiv S^{-1/\alpha}K$, all curves collapse after the onset of the exponential truncation [21,22]. The distribution of maxima displays the same scaling of the degree distribution. It can also be seen that the maxima are mainly drawn from the exponential region of the degree distribution. (b) Semilog plot of the distribution of maximum degree. For comparison, we also show the best fits to the data of the Gumbel statistics with $u = -2.1(K' - 1.6)$ (dotted line), and Frechet distribution with $\alpha = 2$ (dashed line). The maximum statistics agree well with the Gumbel distribution for $K' < 5$. The exponential truncation of the degree distribution determines the form of the distribution of maximum degree.

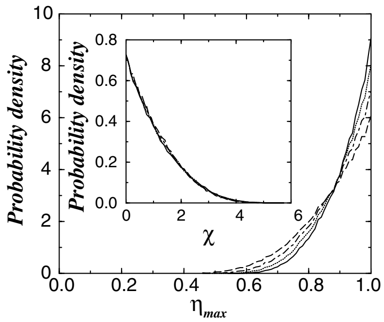

FIG. 2. Fitness distribution of the node with the maximum degree. As expected, the node with the largest fitness eventually becomes the most connected during the growth of the network. The inset shows the collapse of the data obtained under the transformation $\chi \equiv (1 - \eta_{\max}) \log S$, implying that the typical maxima $\eta_{\rm typ}$ converges to one as $1/\log S$.

exponential truncation that scales as $S^{1/\alpha}$, with $\alpha = 1.255$. However, in this case the scaling of the exponential truncation $S^{1/\alpha}$ does not lead to the data collapse of the distributions of maximum degree; cf. Fig. 3(a). Indeed, as the network grows, the distribution of maxima gradually shifts into the region of the degree distribution before the exponential truncation.

This fact is explained by the fact that the larger values of K are due to nodes with η very close to 1, while the typical maxima $K_{\rm typ}$ correspond to nodes with smaller fitnesses. The results of Fig. 2 indicate that the fitness of the typical maxima, $\eta_{\rm typ}$, scales as $1 - 1/\ln S$. Since the degree of a site i grows as $S^{\eta_i/\alpha}$ [20], we obtain

$$\frac{d\ln K_{\rm typ}}{d\ln S} = \frac{\eta_{\rm typ}}{\alpha} \propto \frac{1}{\alpha}\left(1 - \frac{1}{\ln S}\right), \tag{5}$$

where $K_{\rm typ}$ is the typical degree of the maxima. The integration of Eq. (5) results in

$$K_{\rm typ} \sim \left(\frac{S}{\ln S}\right)^{1/\alpha}. \tag{6}$$

We test this result in Fig. 3(b), where we plot the distributions of maximum degree for different network sizes scaled according to (6). The scaling succeeds in collapsing the peaks of the distributions. The scaling also shows that $\rho(K)$ does not follow either of the two classical forms, but we note that the curves approach the Frechet distribution as the system size increases. This is consistent with the results shown in Fig. 3(a). Since as the network grows the distribution of maxima shifts away from the region of the truncation of the degree distribution, we expect that the power-law region of the degree distribution will dictate the statistics of the maxima, which, according to Eq. (6), will converge logarithmically to the Frechet statistics.

In order to further test the effect of new, fitter, nodes becoming the new maxima, we consider an additional network growth model in which the nodes have uniformly

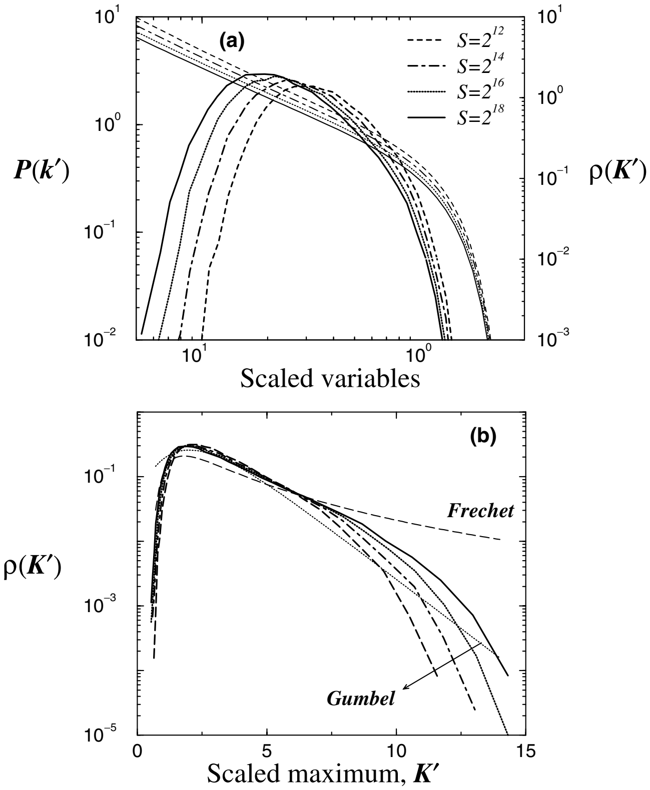

FIG. 3. (a) Log-log plot of the cumulative degree distribution (thin lines) and distribution of maximum (thick lines) for the model with uniform fitness distribution [20]. In this case, the distribution decay as $P(k) \sim k^{-\alpha}/\log k$, with $\alpha = 1.255$ [20]. Note that this result is different from the results of Fig. 2(a) in Ref. [20], which shows a plateau instead of an exponential truncation. Again, we use the rescaling relation $K' \equiv S^{-1/\alpha} K$ to determine the onset of the exponential truncation for large values of k. This scaling fails to collapse the distribution of maximum degree. In this case the maxima are mostly drawn from the power-law part of the degree distribution. (b) Data collapse for the maximum degree distribution using the rescaling relation $K'' \equiv S^{-1/\alpha}\ln(S)K$. The thin dotted and dashed lines correspond to the best fit to the data of the Gumbel [with $u = -0.7(K'' - 2)$] and Frechet (with $\alpha = 1.255$) distributions, respectively. For this case the curves do not collapse well. The distributions became broader as the network grows, appearing to converge to the Frechet distribution as S increases.

distributed fitnesses, but the two initial nodes in the network ($i = 1$ and 2) have $\eta = 1$. In this case, we expect that one of the two initial nodes will be the most connected; hence the distribution of $\eta_{\max}$ is a delta function at one. We calculate the cumulative degree distribution for this case and find that the distribution displays a power-law decay followed by an exponential truncation [27]. Moreover, in this case, the typical maxima fall in the exponentially decaying region of the degree distribution, and the maxima scale as $S^{1/\alpha}$, as the truncation of the degree distribution.

The results for this test case further suggest that in the case of heterogeneous fitness it is the slow, progressive entry into the system of nodes with larger fitnesses—which eventually overcome older nodes as the maxima—that is the reason for the lack of scaling for the distribution of maxima and for the fact the typical maxima is drawn from the power-law decaying region of the degree distribution.

The major finding of this study is the possibility to describe in a parsimonious way the statistics of the maximum degree for growing networks. This is of importance due to the role that the largest degree has in scale-free networks [11]. For the case of homogeneous nodes—i.e., nodes with identical fitnesses—we find that the distribution of maxima follows Gumbel statistics with parameters related to the exponent α characterizing the degree distribution. We explain this finding by the exponential truncation of $P(k)$ due to the finite size of the network. In contrast, for scale-free models with heterogeneous nodes having fitnesses drawn from a uniform distribution, we do *not* find scaling of the distribution of maxima. Nonetheless, our results suggest that the asymptotic distribution of maxima approaches the Frechet statistics as the network size increases, even though $P(k)$ is exponentially truncated due to the network's finite size. We explain this puzzling fact in terms of the progressive entry of nodes with larger fitness which over time will establish more links than nodes with lower fitness that entered the system earlier.

We thank S. Mossa and H. E. Stanley for stimulating discussions and helpful suggestions. A. A. M. and J. S. A. thank the Brazilian Agencies CNPq and FUNCAP for support. L. A. N. A. thanks NIH/NCRR (P41 RR13622) and NSF for support.

*Electronic address: auto@fisica.ufc.br
†Electronic address: soares@fisica.ufc.br
‡Electronic addresses: amaral@northwestern.edu
URL: http://amaral.chem-eng.northwestern.edu

[1] E. J. Gumbel, *Statistics of Extremes* (Columbia University Press, New York, 1958); J. Galambos, *The Asymptotic Theory of Extreme Order Statistics* (R. E. Krieger, Malabar, FL, 1987), 2nd ed.; R. A. Fisher and L. H. C. Tippett, Proc. Cambridge Philos. Soc. **28**, 180 (1928).
[2] S. T. Bramwell *et al.*, Nature (London) **396**, 552 (1998).
[3] M. Sahimi and S. Arbabi, Phys. Rev. B **47**, 713 (1993).
[4] J.-P. Bouchaud and M. Mézard, J. Phys. A **10**, 7997 (1997).
[5] V. S. L'vov *et al.*, Phys. Rev. E **63**, 056118 (2001).
[6] T. N. Palmer and J. Raisanem, Nature (London) **415**, 512 (2002); R. Schnur, Nature (London) **415**, 483 (2002); P. C. D. Milly *et al.*, Nature (London) **415**, 514 (2002).
[7] P. Jefferies *et al.*, cond-mat/0201540.
[8] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[9] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[10] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[11] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002); S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).
[12] R. Albert *et al.*, Nature (London) **406**, 378 (2000).
[13] F. Liljeros *et al.*, Nature (London) **411**, 907 (2001).
[14] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001); Phys. Rev. E **63**, 066117 (2001).
[15] A third case, known as Weibull distribution, occurs when the distribution follows a positive power law. In this case the variables have to be bounded [1].
[16] In a recent study, these universality classes have been confirmed by means of the replica symmetry breaking method [4].
[17] D. S. Dean and S. N. Majumdar, Phys. Rev. E **64**, 046121 (2001).
[18] M. Z. Bazant, Phys. Rev. E **62**, 1660 (2000).
[19] The total number of links is constrained to the size of the network $\sum k_i = 2S$.
[20] G. Bianconi and A.-L. Barabási, Europhys. Lett. **54**, 436 (2001).
[21] R. Cohen *et al.*, Phys. Rev. Lett. **85**, 4626 (2000); **86**, 3682 (2001); S. N. Dorogovtsev *et al.*, Phys. Rev. Lett. **85**, 4633 (2000); S. N. Dorogovtsev *et al.*, Phys. Rev. E **63**, 062101 (2001).
[22] P. L. Krapivsky *et al.*, Phys. Rev. Lett. **85**, 4629 (2000).
[23] The results are indistinguishable from those obtained for 10^4 network realizations.
[24] Our results are consistent with the possibility that, as S increases, there is a very slow convergence to the Gumbel distribution even for $K' > 5$.
[25] Note that the network growth process generates finite size networks. This contrasts with the procedure when drawing a set of S variables from a given distribution. Specifically, if one draws ten samples of size 1000, this is no difference from drawing a single sample of size 10 000. In contrast, in the case of a scale-free network, there is a limit for the maximum degree possible that is controlled by the exponent $1/\alpha$ [21,22]. Hence, for scale-free networks, generating ten networks of size 1000 leads to a statistically different set of k_i's than generating one network of size 10 000. In the former case $K \sim 1000^{0.5} \approx 30$, while in the latter $K \sim 10\,000^{0.5} \approx 100$. To reduce this effect, we calculated the maximum statistics of a small *subset* of the nodes for a network of a size $S = 2^{18}$. We expect this case to be free of the constraint of the total number of links and indeed we find Frechet statistics for the maximum degree of the subset.
[26] In order to calculate the cumulative degree distribution in this model, we discretize the distribution of fitness in the range $[0, 1]$ to $n = 100$ equally spaced values. We then iterate consistently the rate equation through this sequence and sum over the fitness values. We obtain identical results both for a more fine discretization ($n = 1000$) and Monte Carlo simulations.
[27] The distributions obtained for different network sizes can be collapsed according to Eq. (4) with $\alpha = 1.27 \pm 0.05$. This exponent value is calculated as the average rate at which the quantity $\sum \eta_j k_j$ grows with S [20].