Dynamical and Correlation Properties of the Internet

Romualdo Pastor-Satorras,[1] Alexei Vázquez,[2] and Alessandro Vespignani[3]
[1]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, Mòdul B4, 08034 Barcelona, Spain*
[2]*International School for Advanced Studies SISSA/ISAS, via Beirut 4, 34014 Trieste, Italy*
[3]*The Abdus Salam International Centre for Theoretical Physics (ICTP), P.O. Box 586, 34100 Trieste, Italy*
(Received 8 May 2001; published 28 November 2001)

The description of the Internet topology is an important open problem, recently tackled with the introduction of scale-free networks. We focus on the topological and dynamical properties of real Internet maps in a three-year time interval. We study higher order correlation functions as well as the dynamics of several quantities. We find that the Internet is characterized by nontrivial correlations among nodes and different dynamical regimes. We point out the importance of node hierarchy and aging in the Internet structure and growth. Our results provide hints towards the realistic modeling of the Internet evolution.

DOI: 10.1103/PhysRevLett.87.258701 PACS numbers: 89.20.Hh, 05.70.Ln, 87.23.Ge, 89.75.Hc

Complex networks play an important role in the understanding of many natural systems [1,2]. A network is a set of nodes and links, representing individuals and the interactions among them, respectively. Despite this simple definition, growing networks can exhibit a high degree of complexity, due to the inherent wiring entanglement occurring during their growth. The Internet is a capital example of growing network with technological and economical relevance; however, the recollection of router-level maps of the Internet has received the attention of the research community only very recently [3–5]. The statistical analysis performed so far has revealed that the Internet exhibits several nontrivial topological properties (wiring redundancy, clustering, etc.). Among them, the presence of a power-law connectivity distribution [6,7] makes the Internet an example of the recently identified class of scale-free networks [8].

In this Letter, we focus on the dynamical properties of the Internet. We shall consider the evolution of real Internet maps from 1997 to 2000, collected by the National Laboratory for Applied Network Research (NLANR) [3]. In particular, we will inspect the correlation properties of nodes' connectivity, as well as the time behavior of several quantities related to the growth dynamics of new nodes. Our analysis shows dynamical behavior with different growth regimes depending on the node's age and connectivity. The analysis points out two distinct wiring processes: the first one concerns newly added nodes, while the second is related to already existing nodes increasing their interconnections. A feature introduced in this paper refers to the Internet hierarchical structure, reflected in a nontrivial scale-free connectivity correlation function. Finally, we discuss recent models for the generation of scale-free networks in the light of the present analysis of real Internet maps. The results presented in this Letter could help develop more accurate models of the Internet.

Several Internet mapping projects are currently devoted to obtaining high-quality router-level maps of the Internet. In most cases, the map is constructed by using a hop-limited probe (such as the UNIX *trace-route* tool) from a single location in the network. In this case the result is a "directed" map as seen from a specific point on the Internet [5]. This approach does not correspond to a complete map of the Internet because cross-links and other technical problems (such as multiple internet provider aliases) are not considered. Heuristic methods to take into account these problems have been proposed [9]. However, it is not clear if they are reliable and if the corresponding completeness of maps can be constructed in this way. A different representation of the Internet is obtained by mapping the autonomous systems (AS) topology. Each AS number approximately maps to an internet service provider (ISP) and its links are inter-ISP connections. In this case it is possible to collect data from several probing stations to obtain complete interconnectivity maps [3,4]. In particular, the NLANR project has been collecting data since November 1997, and it provides topological as well as dynamical information on a consistent subset of the Internet. The first November 1997 map contains 3180 AS, and it has grown in time until the December 1999 measurement, consisting of 6374 AS. In the following we will consider the graph whose nodes represent AS and whose links represent the connections between AS.

In dealing with the Internet as an evolving network, it is important to discern whether or not it has reached a stationary state whose average properties are time independent. As a first step, we analyzed the behavior in time of several average quantities, such as the connectivity $\langle k \rangle$, the clustering coefficient $\langle C \rangle$, and the average minimum path distance $\langle d \rangle$, of the network [10]. The first two quantities (see Table I) show a very slow tendency to increase in time, while the average minimum path distance slowly decreases with time. A more clear-cut characterization of the topological properties of the network is given by the connectivity distribution, $P(k)$. In Fig. 1 we show the probability $P(k)$ that a given node has k links to other nodes. We report the distribution of snapshots of the Internet at different times. In all cases, we found a clear power-law

 0031-9007/01/87(25)/258701(4)$15.00

TABLE I. Average properties for three different years. $\langle k \rangle$ is the average connectivity. $\langle d \rangle$ is the minimum path distance d_{ij} averaged over every pair of nodes (i, j). $\langle C \rangle$ is the clustering coefficient C_i averaged over all nodes i, where C_i is defined as the ratio between the number of links between the neighbors of i and its maximum possible value $k_i(k_i - 1)/2$. The numbers in parentheses indicate the statistical uncertainty from averaging the values of the corresponding months in each year.

Year	1997	1998	1999
$\langle k \rangle$	3.47(4)	3.62(5)	3.82(6)
$\langle C \rangle$	0.18(1)	0.21(2)	0.24(1)
$\langle d \rangle$	3.77(1)	3.76(2)	3.72(1)

behavior $P(k) \sim k^{-\gamma}$ with $\gamma = 2.2 \pm 0.1$. The distribution cutoff is fixed by the maximum connectivity of the system and is related to the overall size of the Internet map. On the other hand, the power-law exponent γ seems to be independent of time and in good agreement with previous measurements [6]. This evidence seems to point out that the Internet's topological properties have already settled into a rather well-defined stationary state.

Initially, the modeling of the Internet considered algorithms based on its static topological properties [11]. However, since the Internet is the natural outcome of a complex growth process, the understanding of the dynamical processes leading to its present structure must be considered a fundamental goal. From this perspective, the Barabási-Albert (BA) model (Refs. [8,12]) can be considered as a major step forward in the understanding of evolving networks. Underlying the BA model is the preferential attachment rule [8]; i.e., new nodes will link with higher probability to nodes with an already large connectivity. This feature is quantitatively accounted for by postulating that the probability of a new link attaching to an old

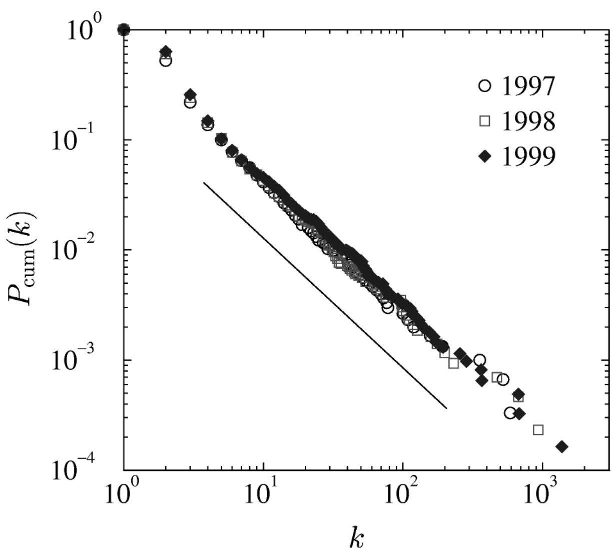

FIG. 1. The cumulated connectivity distribution for the 1997, 1998, and 1999 snapshots of the Internet. The power-law behavior is characterized by a slope -1.2, which yields a connectivity exponent $\gamma = 2.2$.

node with connectivity k_i, $\Pi(k_i)$, is linearly proportional to k_i, $\Pi(k_i) \sim k_i$. This is an intuitive feature of the Internet growth, where large provider hubs are more likely to establish connections than smaller providers. The BA model has been successively modified with the introduction of several ingredients in order to account for connectivity distribution with $2 < \gamma < 3$ [13,14], local geographical factors [15], wiring among existing nodes [16], and age effects [17]. While all these models reproduce the scale-free behavior of the connectivity distribution, it is interesting to inspect deeper the Internet's topology to eventually find a few discriminating features of the dynamical processes at the basis of the Internet growth.

A first step in a more detailed characterization of the Internet concerns the exploration of the connectivity correlations. This factor is best represented by the conditional probability $P_c(k' \mid k)$ that a link belonging to a node with connectivity k points to a node with connectivity k'. If this conditional probability is independent of k, we are in the presence of a topology without any correlation among the nodes' connectivity. In this case, $P_c(k' \mid k) = P_c(k') \sim k'P(k')$, in view of the fact that any link points to nodes with a probability proportional to their connectivity. On the contrary, the explicit dependence on k is a signature of nontrivial correlations among the nodes' connectivity, and the possible presence of a hierarchical structure in the network topology. A direct measurement of the $P_c(k' \mid k)$ function is a rather complex task due to large statistical fluctuations. More clear indications can be extracted by studying the quantity $\langle k_{nn} \rangle = \sum_{k'} k'P_c(k' \mid k)$; i.e., the nearest neighbors average connectivity of nodes with connectivity k. In Fig. 2, we show the results obtained for the Internet map of 1998 which strikingly exhibit a clear power-law dependence on the connectivity degree $\langle k_{nn} \rangle \sim k^{-\nu}$, with $\nu \simeq 0.5$. This result clearly implies the existence of nontrivial correlation properties for the Internet. The primary known structural difference between Internet nodes is the distinction between *stub* and *transit* domains. Nodes in stub domains have links that go through only the domain itself. Stub domains, on the other hand, are connected via a gateway node to transit domains that, on the contrary, are fairly well interconnected via many paths. In other words, there is a hierarchy imposed on nodes that is very likely at the basis of the above correlation properties. As instructive examples, we report in Fig. 2 the average nearest-neighbor connectivity for the generalized BA model, with $\gamma = 2.2$ [13], and the fitness model described in Ref. [18], with $\gamma = 2.25$, for networks with the same size as the Internet snapshot considered. While in the first case we do not observe any noticeable structure with respect to the connectivity k, in the latter we obtain a power-law dependence similar to the experimental findings. The general analytic study of connectivity correlations in growing network models can be found in Ref. [19]. A detailed discussion of different

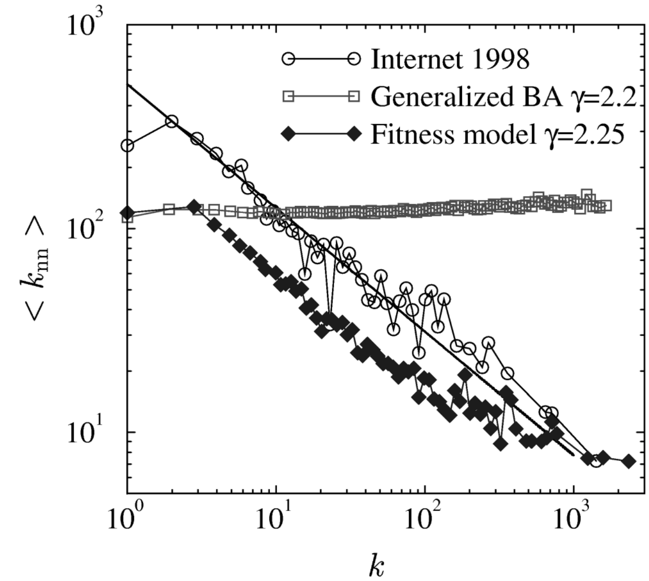

FIG. 2. The average connectivity $\langle k_{nn} \rangle$ of the nearest neighbors of a node depending on its connectivity k for the 1998 snapshot of the Internet, the generalized BA model with $\gamma = 2.2$ (Ref. [8]), and the fitness model (Ref. [18]). The solid line has a slope -0.5. The scattered results for very large k are due to statistical fluctuations.

models is beyond the scope of this paper; however, it is worth noticing that a k structure in correlation functions, as probed by the quantity $\langle k_{nn} \rangle$, does not arise in all growing network models.

In order to inspect the Internet dynamics, we focus our attention on the addition of new nodes and links into the maps. In the three-year range considered, we keep track of the number of links $\ell_{\rm new}$ appearing between a newly introduced node and an already existing node. We also monitor the rate of appearance of links $\ell_{\rm old}$ between already existing nodes. In Table II we see that the creation of new links is governed by these two processes at the same time. Specifically, the largest contribution to the growth is given by the appearance of links between already existing nodes. This clearly points out that the Internet growth is strongly driven by the need for redundancy wiring and an increased need for an available bandwidth for data transmission.

A customarily measured quantity in the case of growing networks is the average connectivity $\langle k_i(t) \rangle$ of new nodes as a function of their age t. In Refs. [8,19] it is shown that $\langle k_i(t) \rangle$ is a scaling function of both t and the absolute time of birth of the node t_0. We thus consider the total number of nodes born within a small observation window Δt_0, such that $t_0 \simeq$ const with respect to the absolute time scale that is the Internet lifetime. For these nodes, we measure the average connectivity as a function of the time t elapsed since their birth. The data for two different time windows are reported in Fig. 3, where it is possible to distinguish two different dynamical regimes: At early times, the connectivity is nearly constant with a very slow increase $[\langle k_i(t) \rangle \sim t^{0.1}]$. Later, the behavior approaches a power-law growth $\langle k_i(t) \rangle \sim t^{0.5}$. While exponent estimates are affected by noise and limited time window effects, the crossover between two distinct dynamical regimes is compatible with the general aging form obtained in Ref. [19]. In particular both the generalized BA model [13] and the fitness model [18] present aging effects similar to those obtained in real data. A more detailed comparison would require quantitative knowledge of the parameters to be used in the models and will be reported elsewhere.

A basic issue in the modeling of growing networks concerns the preferential attachment hypothesis [8]. By generalizing the BA model algorithm it is possible to define models in which the rate $\Pi(k)$, with which a node with k links receives new nodes, is proportional to k^α. The inspection of the exact value of α in real networks is an important issue since the connectivity properties strongly depend on this exponent [14,20]. Here we use a simple recipe that allows us to extract the value of α by studying the appearance of new links. We focus on links emanating from newly appeared nodes in different time windows ranging from one to three years. We consider the frequency $\mu(k)$ of links that connect to nodes with connectivity k. By using the preferential attachment hypothesis, this effective probability is $\mu(k) \sim k^\alpha P(k)$. Since we know

TABLE II. Monthly rate of new links connecting existing nodes to new ($\ell_{\rm new}$) and old ($\ell_{\rm old}$) nodes.

Year	1997	1998	1999
ℓ_{new}	183(9)	170(8)	231(11)
ℓ_{old}	546(35)	350(9)	450(29)
ℓ_{new}/ℓ_{old}	0.34(2)	0.48(2)	0.53(3)

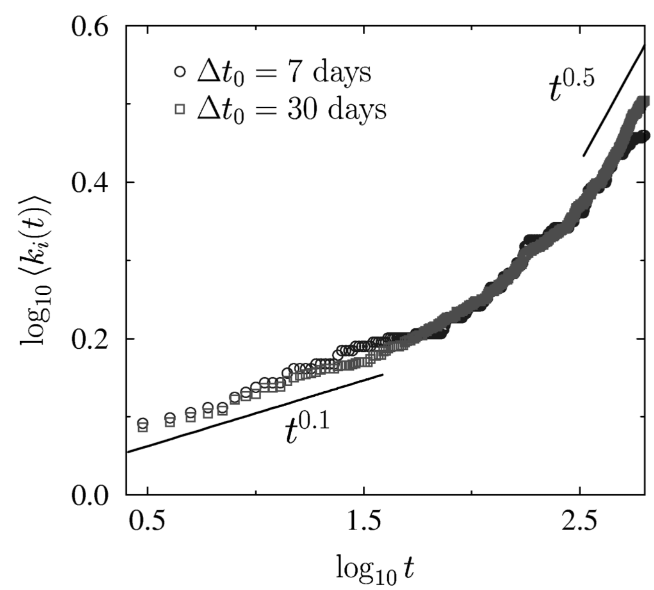

FIG. 3. The average connectivity of nodes borne within a small time window Δt_0, after a time t elapsed since their appearance. Time t is measured in days. As a comparison we report the lines corresponding to $t^{0.1}$ and $t^{0.5}$.

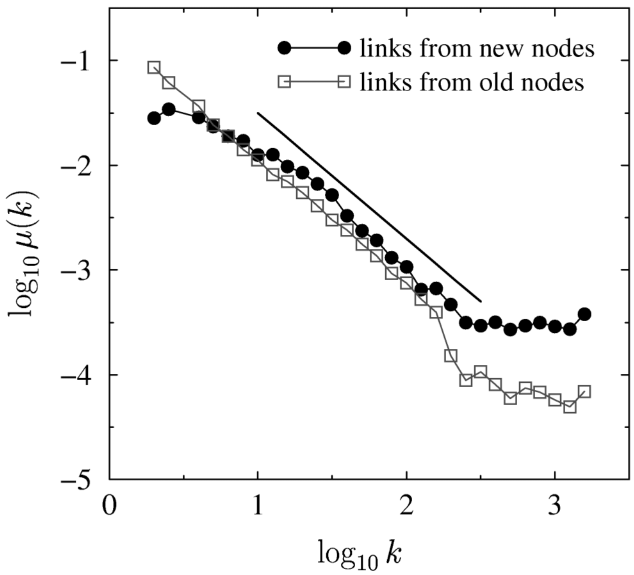

FIG. 4. The cumulative frequency of links emanating from new and existing nodes that attach to nodes with connectivity k. The straight line corresponds to a slope -0.2. The flat tail originates from the poor statistics at very high k values.

that $P(k) \sim k^{-\gamma}$, we expect to find a power-law behavior $\mu(k) \sim k^{\alpha-\gamma}$ for the frequency. In Fig. 4, we report the obtained results which show a clear algebraic dependence $\mu(k) \sim k^{-1.2}$. By using the independently obtained value $\gamma = 2.2$, we find a preferential attachment exponent $\alpha \simeq 1.0$, in good agreement with the result obtained with a different analysis in Ref. [20]. We also performed a similar analysis for links emanated by existing nodes, recovering the same form of preferential attachment (see Fig. 4).

In summary, we have shown that the Internet map exhibits a stationary scale-free topology, characterized by nontrivial connectivity correlations. An investigation of the Internet's dynamics confirms the presence of a preferential attachment behaving linearly with the nodes' connectivity and identifies two different dynamical regimes during the nodes' evolution. We point out that very likely several other factors, such as the nodes' hierarchy, resource constraints, and the real geographical location of nodes, can influence the Internet evolution. The results reported here could be relevant for a more realistic modeling of the Internet growth and evolution, and could have implications in the study of the resilience to attacks and spreading phenomena in this network [21,22].

This work has been partially supported by the European Network Contract No. ERBFMRXCT980183. R. P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnologia (Spain) and from the Abdus Salam International Centre for Theoretical Physics (ICTP), where part of this work was done. We thank M.-C. Miguel, Y. Moreno-Vega, and R. V. Solé for helpful comments and discussions.

[1] S. H. Strogatz, Nature (London) **410**, 268 (2001).
[2] L. A. N. Amaral, A. Scala, and M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11 149 (2000).
[3] The National Laboratory for Applied Network Research, sponsored by the National Science Foundation, provides Internet routing related information based on border gateway protocol data (see http://moat.nlanr.net/).
[4] The Cooperative Association for Internet Data Analysis, located at the San Diego Supercomputer Center, provides measurements of Internet traffic metrics (see http://www.caida.org/home/).
[5] B. Cheswick and H. Burch, Intenet mapping project at Lucent Bell Labs (see http://www.cs.bell-labs.com/who/ches/map/).
[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Comput. Commun. Rev. **29**, 251 (1999).
[7] G. Caldarelli, R. Marchetti, and L. Pietronero, Europhys. Lett. **52**, 386 (2000).
[8] A.-L. Barabási and R. Albert, Science **286**, 509 (1999); A.-L. Barabási, R. Albert, and H. Jeong, Physica (Amsterdam) **272A**, 173 (1999).
[9] Mapping the Internet within the SCAN project at the Information Sciences Institute (see http://www.isi.edu/div7/scan/).
[10] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[11] E. W. Zegura, K. Calvert, and M. J. Donahoo, IEEE/ACM Trans. Netw. **5**, 770 (1997).
[12] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[13] R. Albert and A.-L. Barabási, Phys. Rev. Lett. **85**, 5234 (2000).
[14] P. L. Krapivsky, S. Redner, and F. Leyvraz, Phys. Rev. Lett. **85**, 4629 (2000).
[15] A. Medina, I. Matt, and J. Byers, Comput. Commun. Rev. **30**, 18 (2000).
[16] S. N. Dorogovtsev and J. F. F. Mendes, Europhys. Lett. **52**, 33 (2000).
[17] S. N. Dorogovtsev and J. F. F. Mendes, Phys. Rev. E **63**, 025101(R) (2001).
[18] G. Bianconi and A.-L. Barabási, Europhys Lett. **54**, 436 (2001).
[19] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[20] H. Jeong and Z. Néda, and A.-L. Barabási, cond-mat/0104131.
[21] R. A. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **406**, 378 (2000); D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000); R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).
[22] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).