

Loops structure of the Internet at the Autonomous System Level

Ginestra Bianconi,^{1,2} Guido Caldarelli,^{3,4} and Andrea Capocci⁴

¹*The Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy*

²*INFN UdR Trieste, via Beirut 2-4, 34014, Trieste, Italy*

³*INFN UdR Roma1 and Dipartimento di Fisica, Università "La Sapienza," Piazzale A. Moro 2, 00185 Roma, Italy*

⁴*Centro Studi e Ricerche E. Fermi, Compendio Viminale, Roma, Italy*

(Received 14 February 2005; published 17 June 2005)

We present here a study of the clustering and loops in a graph of the Internet at the autonomous systems level. We show that, even if the whole structure is changing with time, the statistical distributions of loops of order 3, 4, and 5 remain stable during the evolution. Moreover, we will bring evidence that the Internet graphs show characteristic Markovian signatures, since the structure is very well described by two-point correlations between the degrees of the vertices. This indeed proves that the Internet belongs to a class of network in which the two-point correlation is sufficient to describe their whole local (and thus global) structure. Data are also compared to present Internet models.

DOI: 10.1103/PhysRevE.71.066116

PACS number(s): 89.75.Hc, 89.20.Hh, 89.75.Da, 89.75.Fb

In the last five years the physics community has started to look at the Internet [1] as a beautiful example of a complex system with many degrees of freedom resulting in global scaling properties. The Internet in fact can be described as a network, with vertices and edges representing, respectively, autonomous systems (ASs) and physical lines connecting them. Moreover it has been shown [2,3] that it belongs to the wide class of scale-free networks [4,5] emerging as the underlying structure of a variety of real complex systems. But, in addition to the common scale-free connectivity distribution, what distinguishes networks as different as the social networks of interactions and the technological networks such as, for example, the Internet? Researchers have started to characterize further the networks, introducing different topological quantities in addition to the degree distribution exponent. Among those, the clustering coefficient $C(k)$ [6] and the average nearest neighbor degree $k^{nn}(k)$ of a vertex as a function of its degree k [7,8]. In particular, measurements on the Internet yield $C(k) \sim k^{-\mu}$ with $\mu \approx 0.75$ [9] and $k^{nn} \sim k^{-\nu}$ with $\nu \approx 0.5$ [9]. A two-vertex degree anticorrelation has also been measured [10]. Accordingly, the Internet is said to display disassortative mixing [11], because nodes prefer to be linked to peers with different degree rather than similar. This situation is opposed to that in social networks where we observe so-called assortative mixing.

Moreover, the modularity of the Internet due to national patterns has been studied by measuring the slowly decaying modes of a diffusion process defined on it [12]. Recently, more attention has been devoted to network motifs [13,14], i.e., subgraphs appearing with a frequency larger than that observed in maximally random graphs with the same degree sequence. Among those, the most natural class includes loops [15–18], closed paths of various lengths that visit each node only once. Loops are interesting because they account for the multiplicity of paths between any two nodes. Therefore, they encode the redundant information in the network structure.

In this paper we will present data of the scaling of loops of length $h \leq 5$ in the Internet graph and we will show that

this scaling is very well reproduced by the two-point correlation matrix between the degrees of linked pairs of vertices. This allows us to suggest that the Internet is “Markovian,” i.e., correlations of order higher than 2 are negligible. In the paper we then study the structure of the graph in the two-point correlation assumption with the goal of characterizing the cycle structure of the Internet and defining an upper limit of the scaling of the number of loops with the system size valid for all possible lengths of the loops.

To measure the number of loops in an undirected network we consider its symmetrical adjacency matrix $\{a_{ij}\}$, with $a_{ij} = 1$ if i and j are connected and $a_{ij} = 0$ otherwise. If no loops (self-links in a vertex) are present, i.e., $a_{ii} = 0$ for all i , the number of loops of length h is given by a dominant term of the type $\text{Tr}(a^h)/h$ that counts the total number of paths of length h minus all the contributions coming from intersecting paths. For $h=3$ these terms are absent and the total number of loops N_3 of length $h=3$ is given by

$$N_3 = \frac{1}{6} \sum_i (a^3)_{ii}. \quad (1)$$

In the case of short loops $h \leq 5$ these terms can be easily evaluated and give the expressions for the total number of loops of size $h=4, 5$, N_4, N_5 [15],

$$N_4 = \frac{1}{8} \left[\sum_i (a^4)_{ii} - 2 \sum_i (a^2)_{ii} (a^2)_{ii} + \sum_i (a^2)_{ii} \right],$$

$$N_5 = \frac{1}{10} \left[\sum_i (a^5)_{ii} - 5 \sum_i (a^2)_{ii} (a^3)_{ii} + 5 \sum_i (a^3)_{ii} \right]. \quad (2)$$

To measure the actual scaling in Internet at the AS level, we used Eqs. (1) and (2). The data of the Internet at the autonomous system level are collected by the University of Oregon Route Views Project and made available by the NLANR (National Laboratory of Applied Network Research). The subset we used in this manuscript is given in [30]. We considered 13 snapshots of the Internet network at the AS level

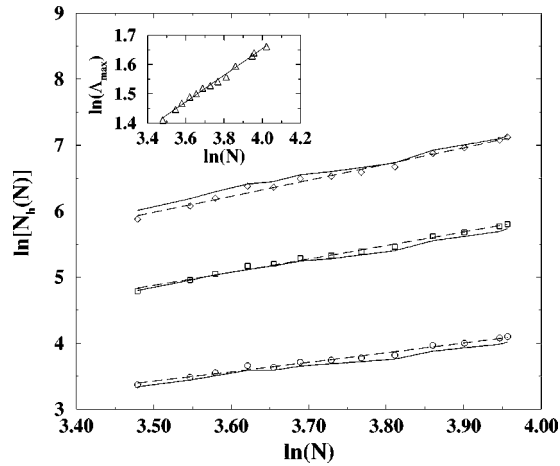


FIG. 1. Number of h loops N_h as a function of the system size N shown with empty symbols for loops of length 3,4,5 (circles, squares, and diamonds). In the solid line we report the first order approximation and in the dashed line the power-law fit to the data. In the inset we report the logarithm of the largest eigenvalue of the matrix C as a function of the system size.

at different times starting from November 1997 (when $N = 3015$) to January 2001 ($N = 9048$). Throughout this period, the degree distribution is a power law with a nearly constant exponent $\gamma \approx 2.22(1)$. Using relations (1) and (2), we measure $N_h(t)$ for $h=3, 4, 5$ in the Internet at different times, corresponding to different network sizes. We observe in Fig. 1 that the data follow a scaling of the type

$$N_h(N) \sim N^{\xi(h)} \quad (3)$$

with the $\xi(h)$ exponents reported in Table I.

To model the Internet means to find a class of networks defined by a stochastic algorithm that share the main characteristics of the Internet graph. Consequently, we suppose that the real Internet graphs belong to a certain ensemble of graphs and it is actually a realization of the Internet. Supposing one knows this ensemble, in order to evaluate the number of loops one theoretically would need to know the entire probability distribution for each element of the adjacency matrix, i.e., the probability distribution

TABLE I. The exponent $\xi(n)$ for $n=3, 4, 5$ as defined in Eq. (3) for real data, in the zero order approximation (ZOA) and in the first order approximation (FOA), and for network models.

System	$\xi(3)$	$\xi(4)$	$\xi(5)$
AS	1.45 ± 0.07	2.07 ± 0.01	2.45 ± 0.08
ZOA	2.26 ± 0.06	3.15 ± 0.07	3.94 ± 0.09
FOA	1.34 ± 0.03	1.86 ± 0.04	2.25 ± 0.05
Fitness	0.59 ± 0.02	0.86 ± 0.02	1.10 ± 0.02
GNG ($p=0.5$)	0.53 ± 0.03	0.72 ± 0.03	0.96 ± 0.02
GNG ($p=0.6$)	0.53 ± 0.03	0.74 ± 0.03	0.99 ± 0.02
D	1.60 ± 0.01	2.20 ± 0.03	2.70 ± 0.03
ND	1.59 ± 0.03	2.11 ± 0.03	2.64 ± 0.03

$P(a_{11}, \dots, a_{1,N}, \dots, a_{N,1}, \dots, a_{N,N})$. Let us make the assumption that the probability for a set of h nodes to be connected depends only on the connectivities. The zero-order approximation to Eqs. (1) and (2) would be then to assume that the connectivity of the nodes are completely uncorrelated and then the formula for calculation of the loops of size h would be [19]

$$N_h^{(1)} = \frac{1}{2h} \left[\frac{\sum_k k(k-1)P(k)}{\langle k \rangle} \right]^h, \quad (4)$$

Given a distribution $P(k)k^{-\gamma}$ with a cutoff at $k_c = N^{1/\chi}$ we get the scaling prediction Eq. (3) with $\xi(h) = h(3-\gamma)/\chi$, in the relevant case $2 < \gamma < 3$. In the special case of an uncorrelated graph with $\gamma=3$ we obtain the scaling behavior $N_h(N) \sim [\log(N)]^{\psi(h)}$, with $\psi(h)=h$. Interestingly enough, the same calculation is exactly valid also in a Barabási-Albert [20] network which is an off-equilibrium network but with zero correlations [15]. We need to observe that the fact itself that in the Internet data the exponent χ follows

$$\frac{1}{\chi} = \frac{1}{\gamma-1} \quad (5)$$

indicates that the network is strongly correlated, in fact for uncorrelated networks we would expect $1/\psi=1/2$ [21,22].

The real exponents $\xi(h)$ as expected depend on h , but unfortunately they significantly differ from the zero-order approximation values $\xi(h) = h(3-\gamma)/\chi$ with χ given by Eq. (5) for and $\gamma \approx 2.22$ (see Table I). So, the correlation nature of the Internet cannot be neglected when one looks at the scaling of the loops in the network.

The first order approximation for Eqs. (1) and (2) consists in taking into account that the connectivity of the nodes is correlated. In order to calculate the number of small loops in the network one can approximate $N_h \sim \text{Tr}(a^h/2h)$. Fixed a direction on the loops, each node is reached by one link connected to the previous node. The probability that a node of degree k_1 , already part of the loop, is connected to a successive node of degree k_2 is given by $(k_1-1)P(k_2|k_1)$ since we can decide to follow one of its remaining k_1-1 nodes. [In our notation $P(k|k')$ indicates the probability that, following one link starting at node k' , one reaches a node with connectivity k .] Consequently, the number of loops of size h in this first order approximation is given by

$$N_h^{(2)} = \frac{1}{2h} \text{Tr}(C^h) \quad (6)$$

where the matrix C is defined as

$$C_{k,k'} = (k'-1)P(k|k'). \quad (7)$$

Of course for higher order loops it will not be possible to neglect the contributions of intersecting paths, but still Eq. (6) would provide an upper limit to the behavior of $N_h(N)$. In Fig. 1 we compare the real data with the first order approximation given by Eq (6). It is clear that this approximation captures most of the cycle structure, at least for small values of h . Since we observe this peculiar characteristic of the

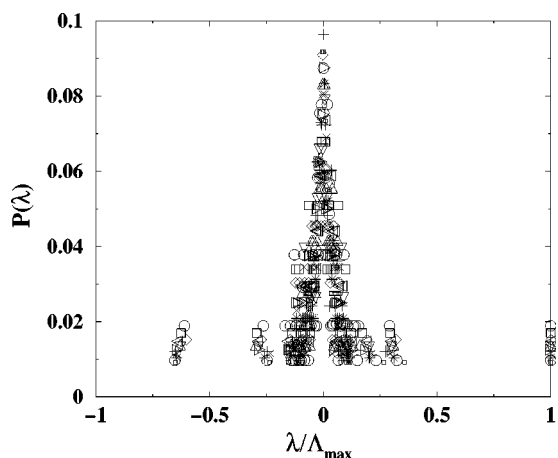


FIG. 2. The rescaled spectra of the matrix C calculated over the 13 snapshots of the Internet under study.

Internet graphs it is worth looking at the structure of the matrix C . Indeed the matrix C is characterized by a spectrum in which there are eigenvalues λ which scale as

$$\lambda(N) \sim N^\theta \quad (8)$$

where $\theta=0.47\pm 0.01$. In Fig. 2 we show how this spectrum scales for the different snapshots of the Internet at the autonomous system level. The largest eigenvalue $\Lambda_{max}(N)$ is the one of much interest to us in this paper since it is responsible for the behavior of N_h at large N . Indeed we can estimate an upper limit for the scaling of the loops of generic length h with the system size, i.e., $N_h^{(2)} \leq O(\Lambda_{max}^h/2h)$ where the scaling is supposed to be valid until $h \ll h^*$ where some arguments support the scaling $h^* \sim N^{(3-\gamma)/2}$ for random scale-free graphs [23] and $h^* \sim N^{1/(\gamma-1)}$ for correlated graphs [18] (see for the behavior of the number of loops at large h in regular random graphs [24]).

To make a comparison between the real data and the model present in the literature at the moment we consider the fitness model [25] and the generalized network growth (GNG) model [26] and the competition and adaptation model [29] with (D) and without (ND) distance constraints. The fitness model has indeed $\gamma=2.255$ and the GNG model has a power-law exponent that depends on the intrinsic parameter p , $\gamma(p)=2+p/(2-p)$. In order to compare networks with a similar mean degree [$\langle k \rangle \in (3.4-4.0)$ [27,28] for the Internet], we consider the fitness model with $m=2$ ($\langle k \rangle=2m=4$) and the GNG model with parameter $p=0.5$ ($\langle k \rangle=2/p=4$) and $p=0.6$ ($\langle k \rangle=2/p=3.33$). All models present nontrivial correlations of the nodes as can be seen by observing the $C(k)$ and $k^m(k)$ functions.

In Table I we compare the $\xi(h)$ exponents of the real data with the exponents numerically calculated for the considered models. While $\xi(h)$ grows almost linearly with h as expected we observe that the D and ND models seem to best reproduce the data.

Following [16], we also measured the clustering coefficients $c_{3,i}$ and $c_{4,i}$ as a function of the connectivity k_i of node i for all i 's. In particular, $c_{3,i}$ is the usual clustering coefficient

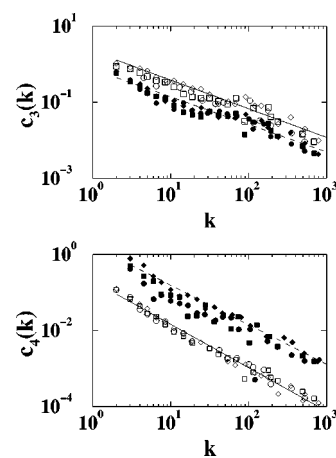


FIG. 3. The clustering coefficients $c_3(k)$ and $c_4(k)$ in the Internet for the data of November 1997 (circles), January 1999 (squares), and January 2001 (triangles). Filled symbols, the same results obtained in the first approximation assumption. Solid and dashed lines indicate the power-law fit to the data and to the first order approximation results, respectively.

cient C , i.e., the number of triangles including node i divided by the number of possible triangles $k_i(k_i-1)/2$.

Similarly, $c_{4,i}$ measures the number of quadrilaterals passing through node i divided by the number of possible quadrilaterals Z_i . This last quantity is the sum of all possible primary quadrilaterals Z_i^p (where all vertices are nearest neighbors of node i) and all possible secondary quadrilaterals Z_i^s (where one of the vertices is a second neighbor of node i). If node i has k_i^m second neighbors, $Z_i^p = k_i(k_i-1)(k_i-2)/2$ and $Z_i^s = k_i^m k_i(k_i-1)/2$. In Fig. 3(a) we plot $c_3(k), c_4(k)$ for the Internet data at three different times (November 1997, January 1999, and January 2001) showing that the behavior of $c_3(k)$ and $c_4(k)$ is invariant with time and scales as

$$c_h(k) \sim k^{-\delta(h)} \quad (9)$$

with $\delta(3)=0.7(1)$ and $\delta(4)=1.1(1)$.

In Fig. 3, we compare the behavior of $c_3(k)$ and $c_4(k)$ in real Internet data with the first order approximation results. Again we observe that the first order approximation results are quite satisfactory, reinforcing our thesis that to explain the loop structure of the Internet it is sufficient to stop at this order. However, the behavior of $c_3(k)$ and $c_4(k)$ cannot be explained by just looking at the largest eigenvalues of the C matrix but one has to consider the entire spectra. For completeness we also considered the behavior of the clustering coefficients $c_3(k)$ and $c_4(k)$ in Internet models (Table II). We observe that while in the D and ND models there are large deviations from the scaling (9) these models seem in general to capture better the cycle structure of the Internet with respect to the other non *ad hoc* models we have considered here.

In conclusion, we computed the number $N_h(t)$ of h loops of size $h=3, 4, 5$ in the Internet at the autonomous system level and the generalized clustering coefficients around individual nodes as a function of node degrees. We have

TABLE II. The exponent of the clustering coefficient $c_3(k)$ and $c_4(k)$ as measured from Internet data as a result of the first order approximation (FOA) and from simulations of Internet models.

System	$\delta(3)$	$\delta(4)$
AS	0.75 ± 0.05	1.13 ± 0.05
FOA	0.70 ± 0.05	1.00 ± 0.05
Fitness	0.67 ± 0.01	0.99 ± 0.01
GNG ($p=0.5$)	0.32 ± 0.02	1.68 ± 0.03
GNG ($p=0.6$)	0.27 ± 0.02	0.93 ± 0.01
D	0.3 ± 0.2	0.8 ± 0.2
ND	0.6 ± 0.2	1.0 ± 0.2

observed that this evolving network has a structure of the loops that is well captured by the two-point correlation matrix. Indeed it seems that the Internet is Markovian in the sense that it is not necessary to study a correlation function

of more than two points, at least to explain the cycle structure. For this reason we have characterized the correlation matrix $C_{k,k'} = (k'-1)P(k|k')$, studying its spectrum. Finally, we have compared these results with the behavior of the same quantities $N_h(N)$ and $c_h(k)$ in the fitness model, in the GNG model, and in the D and ND models, a chosen subset of the available Internet models present in the literature, finding that the *ad hoc* D and ND models seem to capture better the cycle structure of the Internet.

The authors are grateful to Uri Alon, Shalev Itzkovitz Matteo Marsili, and Yi-Cheng Zhang for useful comments and discussions and to the authors of Ref. [29] for suggesting measuring the cycle structure of their model. This paper has been financially supported by the Swiss National Foundation, under Grant No. 2051-067733.02/1, and by the European Commission FET Open Project No. COSIN IST-2001-33555, and IP Projects DELIS and EVERGROW.

-
- [1] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet* (Cambridge University Press, Cambridge, U.K., 2004).
- [2] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Proceedings, ACM SIGCOMM '99* [Comput. Commun. Rev. **29**, 251 (1999)].
- [3] G. Caldarelli, R. Marchetti, and L. Pietronero, *Europhys. Lett.* **52**, 386 (2000).
- [4] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [5] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press, Oxford, 2003).
- [6] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, *Science* **297**, 1551 (2002).
- [7] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
- [8] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. E* **65**, 066130 (2002).
- [9] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, e-print cond-mat/0206084.
- [10] S. Maslov, K. Sneppen, and A. Zaliznyak, e-print cond-mat/0205379.
- [11] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [12] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen, *Phys. Rev. Lett.* **90**, 148701 (2003).
- [13] S. Itzkovitz, R. Milo, N. Kashtan, G. Ziv, and U. Alon, *Phys. Rev. E* **68**, 026127 (2003).
- [14] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* **298**, 824 (2002).
- [15] G. Bianconi and A. Capocci, *Phys. Rev. Lett.* **90**, 078701 (2003).
- [16] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, e-print cond-mat/0212026.
- [17] G. Bianconi, *Eur. Phys. J. B* **38**, 223 (2004).
- [18] H. D. Rozenfeld, J. E. Kirk, E. M. Bollt, and D. ben-Avraham, e-print cond-mat/0403536.
- [19] Z. Burda, J. Jurkiewicz, and A. Krzywicki, e-print cond-mat/0401310.
- [20] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [21] Z. Burda and A. Krzywicki, *Phys. Rev. E* **67**, 046118 (2003).
- [22] M. Boguña, R. Pastor-Satorras, and A. Vespignani, *Eur. Phys. J. B* **38**, 205 (2004).
- [23] G. Bianconi and M. Marsili, e-print cond-mat/0502552.
- [24] E. Marinari and R. Monasson, e-print cond-mat/0407253.
- [25] G. Bianconi and A.-L. Barabási, *Europhys. Lett.* **54**, 436 (2001).
- [26] G. Caldarelli, P. De Los Rios, and L. Pietronero, e-print cond-mat/0307610.
- [27] Indeed the Internet shows an accelerated growth [2,28] that is not taken into account in this model. Indeed the acceleration is rather small $\langle k \rangle \sim N^\alpha$ with $\alpha=0.01$ and has been neglected in these models.
- [28] S. N. Dorogovtsev and J. F. F. Mendes, *Phys. Rev. E* **63**, 025101 (2001).
- [29] M.-A. Serrano, M. Boguña, and A. Díaz-Guilera, e-print cond-mat/0406765.
- [30] COSIN web page <http://www.cosin.org>