EUROPHYSICS LETTERS 15 May 2004

Europhys. Lett., **66** (4), pp. 471–477 (2004)

DOI: 10.1209/epl/i2003-10236-6

Growth mechanisms of the AS-level Internet network

V. Rosato[1,2] and F. Tiriticco[3]

[1] *ENEA, Ente per le Nuove Tecnologie, l'Energia e l'Ambiente*
Casaccia Research Center, Computing and Modelling Unit (CAMO)
P.O. Box 2400, 00100 Roma, Italy
[2] *Istituto Nazionale di Fisica della Materia (INFM)*
Unità di Ricerca di Roma 1 - Roma, Italy
[3] *Dipartimento di Ingegneria delle Telecomunicazioni*
Università di Roma "Tor Vergata" - Via O. Raimondo 8, 00173 Roma, Italy

(received 22 October 2003; accepted in final form 17 March 2004)

PACS. 05.65.+b – Self-organized systems.
PACS. 89.20.Hh – World Wide Web, Internet.
PACS. 89.75.Fb – Structures and organization in complex systems.

Abstract. – The growth of the AS (Autonomous System) level router network has been analyzed during a period of almost two years (1998–2000). The behavior of the clustering coefficient and that of the "min-cut" analysis on the network have been used as key features to simulate the network growth. The Triad Formation mechanism, which suitably incorporates also the Preferential Attachment mechanism, is an adequate recipe for reproducing relevant structural properties of the AS-level router network during its stages of growth.

Many "real-world" networks arising from *non-supervised* growth processes display common topological features [1–5]. Networks in biology (metabolic and protein-protein interaction networks), in sociology (scientific-publication co-authors network), in communication engineering (the Internet), in the information society (the web pages), although describing different systems and phenomena, show similarities in their structure [2]. The most striking among them is the distribution of the node's degree k, $P(k)$, characterized by a power law function of the type [1–3]

$$P(k) \sim k^{-\gamma} \tag{1}$$

with the value of γ restricted in a narrow interval ($2 < \gamma < 3$) [2]. The network structure originating such a distribution is "scale-free" [1, 2] and highly (*hubs*) and loosely (*leaves*) connected nodes coexist.

A major interest has been focussed on the growth mechanism able to produce these structures. The understanding of the growth mechanism, on the one hand, might be used to elucidate the hidden driving force which builds up such complex topological structures; on the other hand, the growth mechanism could be reproduced and inserted into network's design tools [6] (*i.e.* codes which attempt to reproduce the structure of the network used for research purpose, to design "virtual networks" with the same properties of the real ones).

Several mechanisms have been claimed to allow the growth of "scale-free" networks. The "preferential attachment" mechanism [2, 3] (PA hereafter) considers a growth process where new nodes stick to pre-existing ones by preferring those with a large degree. Other authors succeded in deriving the peculiar topology of the Internet from a complex multi-objective optimization, based on the minimization of the "last mile" connection costs and that of the transmission delays [7]. Although it has been proven [1–5] that, at least qualitatively, the PA mechanism is able to deliver networks with the scale-free properties, in some cases (such as that which we will describe), there are some quantitative differences between the measured and the simulated properties which cannot be reproduced by the PA mechanism alone [8]. This fact has been claimed by several authors, when dealing with the AS-level network [8]. Others [9], while focussing on different types of networks, have proposed a novel mechanism which improves the capability of the PA mechanism to properly reproduce the observed properties of the network. The mechanism we are referring to has been called "triad formation" (TF) [9]. This mechanism prescribes to use a PA mechanism to select the first node where a new node must be added: then, further links of the new node are connected either with the PA mechanism or to nearest neighbors of the first node, in a way to form "triangles". The TF mechanism allows the choice among the two options with a given probability, which is an adjustable parameter of the model.

In this work, we will analyze the behavior of the Autonomous System (AS) level routers network in the US. Tha analyzed data are snapshots of that network, resulting from a daily collection of BGP (Border Gateway Protocol) routing tables coming from a route server with BGP connections to multiple geographically distributed target operational routers [10]. This system represents an example of a complex infra-structure which has undergone a sizeable growth in the last few years, thus allowing the study of its growth process. Rather than presenting a complete analysis of this network, previously done by other authors [11,12], we wish to present further evidences of its "anomalous" behavior. To the purpose of explaining these anomalies, we first discuss the features which, we believe, contain the key properties for the understanding of the growth process; then we propose the use of the TF mechanism which, by incorporating the main features of the PA mechanisms, is able to quantitatively predict the behavior of the most important topological properties of the AS-level network during the different growth phases.

The model. – Let us indicate a generic network $\mathrm{G} = (\mathrm{N}, \mathrm{L})$ as a set of nodes (N) and links (L), where $|\mathrm{N}| = n$ and $|\mathrm{L}| = m$. A mathematical object allowing a complete definition of the network is the *adjacency matrix* $\boldsymbol{A}$; if the network has undirected and unitary links, assumed to hold hereafter, it is defined as $A_{ij} = 1$ if nodes i and j are connected, 0 otherwise. All the relevant properties of the network can be deduced from the analysis of the adjacency matrix [2]. A further matrix which can be associated to the network is the so-called *Laplacian matrix* $\boldsymbol{L}$, defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$ (where $\boldsymbol{D}$ is the diagonal matrix having $D_{ii} = k_i$, with k_i defined as the degree of the i-th node). Further insights on the structure and the properties of the network can be gained by the spectrum analysis of the $\boldsymbol{A}$ and $\boldsymbol{L}$ matrices [2, 13].

The AS-level routers network data [10] have been analyzed through the evaluation of several properties: the degree distribution, the clustering coefficient c [2, 11], the network diameter d and the characteristic path length $\langle l \rangle$ (average of the $n(n-1)/2$ distances). We have also evaluated some property deduced by the spectral analysis. There is a wide literature on the spectral analysis of the $\boldsymbol{A}$ and $\boldsymbol{L}$ matrices [13–15]. Several authors have pointed out the scaling properties of the eigenvalues [13], while others were interested to extract information concerning the structure of the graph [15]. We focus, in turn, on a specific result derived by the spectral analysis of the $\boldsymbol{L}$-matrix and the so-called "min-cut" theorem [16–18]. This can be stated as follows. The lowest eigenvalue λ_1 of the $\boldsymbol{L}$-matrix is always vanishing ($\lambda_1 = 0$) and the orthonormalized components of its associated eigenvector $\boldsymbol{v}^{L}{}_1$ are all equal to $1/\sqrt{n}$ (n

TABLE I – *Relevant properties of the snapshots of the AS-level network at the different dates: n is the number of nodes, m the number of links, α the ratio between the existing number of links and the maximum possible number of links ($n(n-1)/2$); c the average clustering coefficient; γ is the coefficient of the distribution of the node's degree as in eq.* (1)*; $k_{\max}$ the degree of the network's hub; d is the network diameter (evaluated via the Dijkstra algorithm* [19]*), $\langle l \rangle$ the characteristic path length, n_{l} the number of links connecting the two–sub-networks solution of the "min-cut" problem.*

Date	n	m	α	c	γ	$k_{\max}$	d	$\langle l \rangle$	n_{l}
AS1 17.03.1998	3459	6137	$1.02 \cdot 10^{-3}$	0.1938	2.35	734	10	3.767	11
AS2 17.09.1998	4107	7571	$8.98 \cdot 10^{-4}$	0.2214	2.51	855	11	3.783	97
AS3 17.03.1999	4788	8990	$7.84 \cdot 10^{-4}$	0.2368	2.41	1083	11	3.719	378
AS4 02.01.2000	6474	12572	$6.00 \cdot 10^{-4}$	0.2522	2.46	1458	9	3.705	493

being the number of nodes). The components of the second eigenvector $\boldsymbol{v}^{L}{}_{2}$ of $\boldsymbol{L}$, associated to the second eigenvalue λ_2 of the Laplacian ($\lambda_2 \neq 0$) have, in turn, different signs. The eigenvector $\boldsymbol{v}^{L}{}_{2}$ of $\boldsymbol{L}$ provides a recipe allowing the partition of the network into two nearly equal sub-networks: the first formed by nodes with a positive component, the second with those with a negative component. The "min-cut" theorem ensures that these two sets of nodes are connected via the minimum number of connections n_{l}, *i.e.* the cut has a minimum "weight". We have thus defined n_{l} as the number of links joining nodes belonging to the different sub-networks (*i.e.* from the total number of links m, we count only those joining nodes belonging to different sub-networks). The links defined by the "min-cut" algorithm are those whose failure would induce the "most effective" perturbation to the network by producing the maximum number of "effective broken links". We have introduced n_{l} as a relevant quantity whose value and variation must be correctly predicted by the proposed growth mechanism.

The analysis of the defined quantities on the AS-level network data [10] has produced the following results, reported in table I.

The structure of the AS-level router network is scale-free (eq. (1)) in agreement with previous calculations [11]; the estimated γ-parameter is a slowly varying function of time, going from $\gamma = 2.35$, at the first observation, to $\gamma = 2.46$, at the final one. Our results are slightly different from those evaluated by other authors [11] who reported values of $\gamma \sim 2.2$. Concerning the clustering coefficient c, the AS-level network has a quite high c value which increases with the size of the network.

Further interesting features characterizing the structure of the network are related to the low values of the network's diameter d, $9 < d < 11$ in the different growth stages, and the low average point-to-point distance $\langle l \rangle$, whose lowest value is $\langle l \rangle = 3.705$ at AS4. An even most surprising result concerns the value of the number of links n_{l}. Although being characterized by a large number of nodes ($n = 3459$), the AS-level router network initially shows a quite small number of connections ($n_{\mathrm{l}} = 11$) between the two sub-sets resulting from the application of the "min-cut" algorithm.

This implies a large "usage" of these links to ensure the communications between the two sub-sets, giving origin to an intrinsic fragility of the system (failure, overloads). The value of n_{l}, however, increases very rapidly: at the end of the observation period, while nodes and links are increased by only a factor 2, n_{l} has increased of almost 45 times ($n_{\mathrm{l}} = 493$). This value is small with respect to the total number of connections m. The ratio between n_{l} and m, however, constantly increases with the network growth, rising from $1.79 \cdot 10^{-3}$ (at AS1) to $3.92 \cdot 10^{-2}$ (at AS4).

All quantities we have reported in table I might be thus considered as pieces of a unique

TABLE II – *Same properties as in table* I *for simulated networks generated by the use of PA and TF mechanisms.*

Model	n	m	c	γ	$k_{\max}$	d	$\langle l \rangle$	n_{l}
AS1-PA	3459	6137	$5.76 \cdot 10^{-3}$	2.74	99	10	4.92	702
AS1-TF			0.608	2.61	168	15	5.77	54
AS2-PA	4107	7571	$1.15 \cdot 10^{-2}$	2.62	195	9	4.76	924
AS2-TF			0.651	2.58	129	16	6.37	37
AS3-PA	4788	8990	$7.85 \cdot 10^{-3}$	2.79	146	9	4.89	1026
AS3-TF			0.666	2.56	131	13	6.04	13
AS4-PA	6474	12572	$7.52 \cdot 10^{-3}$	2.70	238	9	4.88	1309
AS4-TF			0.705	2.71	218	14	6.25	40

mosaic which reveals the peculiar structure of the AS-level network and its growth process. We have thus decided to finalize our efforts on the reproduction of these quantities and to build up a model to simulate the network growth.

We have firstly compared the values of the properties of the real systems with those evaluated on simulated network structures, of the same size (same n and same m), generated by using the PA and the TF mechanisms. Each network is the result of an independent growth process (*i.e.* large networks are produced by a new growth process and not as a further growth of smaller networks) performed by using the standard algorithms implementing the PA and the TF growth mechanisms. In both cases, however, the requirement of producing networks with given n and m (producing a fractional average degree) has been introduced in the algorithm by imposing, to the newly attached nodes, a probabilistic choice of further connections. In the adopted algorithm for the PA growth, *all* further links are drawn with the PA rules; in that adopted for the TF growth, *all* further links are selected to form new triangles. This choice is equivalent to setting the value of the parameter $P_t = 1$ in the original Holme and Kim algorithm [9]; this option could be regarded as an "extreme" case of the TF growth. In table II, we present the results obtained by applying the PA (ASx-PA) and the TF (ASx-TF) growth mechanisms ($x = 1, \ldots, 4$).

The PA- and TF-grown networks show a very different behavior from each other, providing different results of the relevant properties with respect to the real structure. The PA network shows a quite low value of c which hardly increases with the network growth. The TF network, in turn, exhibits a very high clustering which increases with the network size. This was expected on the basis of previous calculations made in [9] and the fact that the TF growth mechanism has been proposed for reconciling the evidence of large clustering coefficients in scale-free networks.

Concerning the diameter of the networks, the PA network has a similar d with respect to the real network, while TF networks seem to be much less efficient in the containment of the largest point-to-point distances. TF networks, moreover, display a quite larger value of the characteristic path length $\langle l \rangle$ which seems to saturate with the network growth.

The n_{l} values, as much as the clustering coefficient, show almost opposite behavior: PA-grown networks display very large values of n_{l}; TF-grown ones, in turn, tend to have very small n_{l} values.

From these data, a possible growth scenario of the AS-level router network emerges where new links are formed between existing nodes; these links are thus not formed under the only pressure of connecting new nodes (as in the PA mechanism) but as a further result of a global strategy aimed at "tighting up" the network, in some sense. These considerations support the

TABLE III – *Same properties as in table* I *for networks generated by the use of the modified mechanisms of eqs.* (2) *and* (3)*, with* $q = 0.3$*. In parentheses, the same quantities as evaluated on the real AS-level network.*

Model	c	γ	$k_{\max}$	d	$\langle l \rangle$	n_l
AS1	0.194	2.35	734	10	3.77	11
AS2-(TF + PA)	0.210 (0.221)	2.52 (2.51)	814 (855)	10 (11)	3.75 (3.78)	111 (97)
AS3-(TF + PA)	0.214 (0.237)	2.71 (2.41)	884 (1083)	10 (11)	3.80 (3.72)	181 (378)
AS4-(TF + PA)	0.208 (0.252)	2.63 (2.46)	1056 (1458)	10 (9)	3.85 (3.70)	602 (493)

idea that the growth process increases the number of nodes and links in a way to maximize their impact on, at least, two important properties: a) the average clustering and b) the number of links between the sub-sets into which the system can be divided, with the further constraint of keeping the average point-to-point distance as low as possible.

These requirements should be, indeed, fulfilled by the application of the growth mechanisms proposed by Holme and Kim [9]: the TF mechanism, in fact, is able to fulfill the a) requirement. Moreover, when the creation of further links with the PA mechanism is allowed, this permits the reproduction also of the b) requirement.

In order to understand how these two constraints could be simoultaneously met, we have performed the following simulation of the network growth: starting from the network at the first observation date (AS1), we have grown the network by adding new nodes on the top of the AS1 configuration, up to obtaining a final network having the same nodes and the same links of the AS-level networks at later growth stages. The following growth mechanism has been used:

1) the first link of the new node connects to a pre-existing node i with a probability $p_i = k_i/\Sigma_j k_j$ (PA mechanism);

2) further links of the new nodes (if any) connect to a further node j with a probability $p_j = k_j/\Sigma_l k_l$ (PA mechanism) or to a randomly selected neighbor of the formerly chosen node i (TF mechanism).

The choice of the TF or the PA mechanism for connecting further links of the new nodes (stage 2) above) is triggered by a probability value q ($0 < q < 1$), in a way to compose a growth mechanism G of the following type:

$$G(1) = \mathrm{PA}\,, \tag{2}$$

$$G(2, \ldots k) = (1-q)\mathrm{PA} + q\mathrm{TF}, \tag{3}$$

where the index $1, 2, \ldots, k$ refers to the number of connections of the generic node added to the net. The parameter q measures the probability of drawing a new link, *if any*, with the TF mechanism. The proposed growth mechanism is thus equal to that proposed by Holme and Kim [9] (it has been called (TF + PA) to stress the use of both TF and PA links as types for further connections of new attached nodes), the only difference being in the variable number of links added for each new node, for the delivery of a network with the required average-degree connectivity. Results of the relevant structural quantities measured on the networks grown with the new recipes of (2) and (3) (with $q = 0.3$) are reported in table III.

The TF growth mechanism, used with the constraint of delivering a network with a given number of nodes and links, n and m, respectively, reconciles the results taken on the simulated networks with those measured on the real ones: the obtained values (see table III) are in good agreement with those measured on the experimental networks. Although being particularly

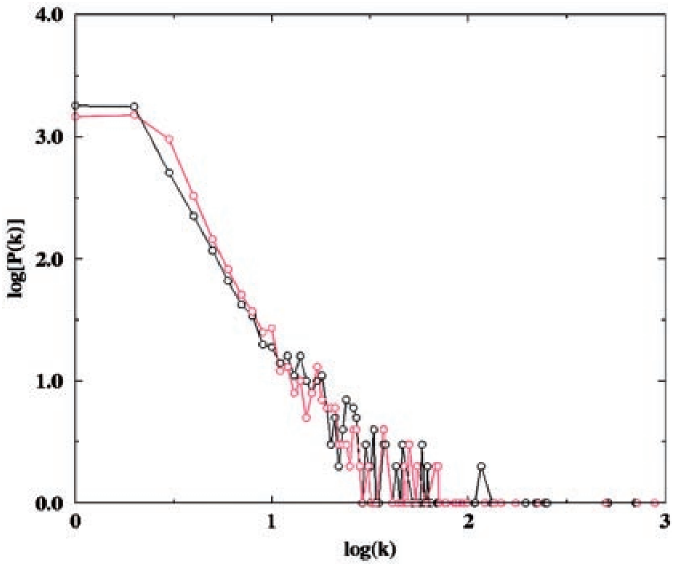

Fig. 1 – Degree distribution of the AS-level network (AS2 configuration) (black) compared to the simulated one (AS2-(TF + PA) with $q = 0.3$) (red).

effective for increasing the network's clustering, the TF mechanism alone would bring the network to a stage where very few connections exist among network's hubs (see results in table II). This fact implies a constantly low value of n_l which does not grow with the network growth. In turn, the PA mechanism alone, although producing high values of n_l, would have not allowed to reach sufficiently large clustering values.

In fig. 1 we report the degree distribution of the real AS2 and the simulated AS2-(TF+PA) networks.

Conclusions. – We have described and characterized the growth process undergone by the AS-level router network during the period from March 17, 1998 to January 2, 2000. During this period, the network has almost doubled its size (n increases from $n = 3459$ to 6474, m from 6137 to 12572). The overall network structure exhibits a power law distribution of degree, in agreement with previous estimates [11].

Several interesting correlations have been further extracted from the data:

a) the growth of the network during the period of observation is faster than linear (with an average growth rate $\mathrm{d}n/\mathrm{d}t = 4.6$ nodes/day); the average number of links per node grows from 1.77 to 1.94, while the ratio α between the number of links and the maximum possible number of links decreases from 10^{-3} to $6 \cdot 10^{-4}$;

b) the average clustering coefficient increases from 0.19 to 0.25; this implies that, apart from new nodes, also new links are formed between pre-existing, non-connected nodes;

c) the number of links n_l, solution of the "min-cut" problem, grows dramatically from $n_\mathrm{l} = 11$ to $n_\mathrm{l} = 493$.

All these data have contributed to support the hypothesis that the TF growth mechanism, as proposed in [9], with a suitable choice of the parameter which allows the creation of PA-type links, is able to mimic a local strategy aimed at increasing robustness and reducing the local diameter. The incorporation of such a growth mechanism has allowed to simulate a growth behavior similar to that assumed by the real network.

* * *

We acknowledge the National Science Foundation Cooperative Agreement No. ANI-9807479, and the National Laboratory for Applied Network Research for the avaliability of raw data at the NLAR repository site. The authors have greatly benefited from several comments of the referees.

REFERENCES

[1] Watts D. J. and Strogartz S. H., *Nature*, **393** (1998) 440.
[2] Albert R. and Barabàsi A.-L., *Rev. Mod. Phys.*, **74** (2002) 47.
[3] Barabàsi A.-L. and Albert R., *Science*, **286** (1999) 509.
[4] Jeong H., Mason S. P., Barabàsi A.-L. and Oltvai Z. N., *Nature*, **411** (2001) 41.
[5] Eisenberg E. and Levanon E. Y., *Phys. Rev. Lett.*, **91** (2003) 138701.
[6] Vukadinovic D., Huang P. and Erlebach T., ETH Zürich, TIK Report No. 118, July 2001.
[7] Fabrikant A., Koutsoupias E. and Papadimitriou C. H., *Lect. Notes Comput. Sci.*, **2380** (2002) 110.
[8] Faloutsos M., Faloutsos P. and Faloutsos C., in *SIGCOMM'99*, 1999; Siganos G., Faloutsos M., Faloutsos P. and Faloutsos C., `http://ucr.edu/~michalis/PAPERS/SFFF.ps`.
[9] Holme P. and Kim B. J., *Phys. Rev. E*, **65** (2002) 026107.
[10] Data have been taken from the raw data repository of the NLNAR Measurement and Networking Analysis Group at the University of California, San Diego (USA) (`http://moat.nlanr.net/Routing/rawdata`). Analyzed data span a period comprised from March 17, 1998 to January 2, 2000.
[11] Vasquez A., Paster-Satorras R. and Vespignani A., *Phys. Rev. E*, **65** (2002) 066130.
[12] Maslov S., Sneppen K. and Zaliznyak A., con-mat/0205379, 6 Nov. 2002.
[13] Farkas I. J., Derényi I., Baràbasi A.-L. and Vicsek T., *Phys. Rev. E*, **64** (2001) 026704.
[14] Dorogovtsev S. N. and Mendes J. F. F., *Europhys. Lett.*, **52** (2000) 33.
[15] Goh K.-I., Kahng B. and Kim D., *Phys. Rev. E*, **64** (2001) 051903.
[16] Mohar B., *Discrete Math.*, **109** (1992) 171.
[17] Pothen A., Simon H. D. and Liou K. P., *SIAM J. Matrix Anal. Appl.*, **11** (1990) 430.
[18] Seary A. J. and Richards W. D., in *Proceedings of the International Conference on Social Networks*, edited by Everett M. G. and Rennolls K., Vol. **1**: *Methodology*, 1996, pp. 47-58.
[19] Dijkstra E. W., *Numer. Math.*, **1** (1959) 269.