Inferring network properties from topological and spectral analysis

Vittorio Rosato
ENEA Casaccia
Computing & Modeling Unit
P.O.Box 2400
00100 Roma A.D.
rosato@casaccia.enea.it

Sandro Bologna
ENEA Casaccia
Computing & Modeling Unit
P.O.Box 2400
00100 Roma A.D.
bologna@casaccia.enea.it

Fabio Tiriticco,
Università di Roma
"Tor Vergata"
Dip. Ing. Telecomunicazioni
Via O. Raimondo 8
00173 Roma (Italy)
tirix3@libero.it

Abstract

Self-organized complex systems can be analyzed in terms of their topological structure. This can unveil relevant properties related to robustness and efficiency. We show how a suitable combination of topological and spectral analysis may improve the understanding of the mechanisms and the driving force which lead to the spontaneous formation of the internet network at the Autonomous System level.

Introduction

In the last 20 years, communication and information technology have undergone an impressive speed-up which has driven to a communication-biased form of social progress. Computers, telephones, televisions, domestic appliances and cars contain a number of electronic devices for the exchange of information and data, enabling “intelligent” services designed for a maximum efficiency and user comfort. World is extremely interconnected; however, more than ever, interconnection relies on the integrity of complex networks, including the Internet, telecommunications, electric power systems, oil and gas networks, rail and road networks, etc. All these complex networks have some common properties and are being investigated by different scientific communities.

Complex systems are functional structures sharing the property of being self-generated (i.e. grown under any external supervision) under the action of some "functional" driving force. Although related to a variety of domains (biology, sociology, telecommunications, ethology etc.), they appear to share a number of properties (common topology, common functional assets, robustness, self-healing etc.) whose nature and "necessity" are key points it is worth investigating with the aim of: (a) understanding the nature of the generating driving forces and the details of the growth mechanisms; (b) control, duplicate and/or transpose into different domains, under the form of technological objects, naturally-occurring structures and functions aimed at producing and controlling "complex" man-made systems.

Results and advancements in fundamental and technological areas are seldom cross-fertilizing each other. Lack of common languages, decades of incommunicability and attention to different properties and functions of systems, have inhibited the flow of methods and results from one field to the others. It is now time to export the most recent acquisitions (methods, tools, paradigms), which have lead to a dramatic improvement of the knowledge in some domains of basic science (physics, biology) into technological domains where they might have an high functional impact.

The field of analysis and protection of "Large Complex Critical Infrastructures" (LCCI), such as an electric grid, a telecommunication network, a railway/air/road transportation network, an information network, a social network, etc. might highly benefit from recent theoretical advancements [1,2]. These are "complex systems"; as such, they might be thought, after all, as made up by the superposition of interacting sub-units whose internal complexity is intrinsically similar to that of the whole. If, on one side, the theory of system control has developed its own strategies and methods, on the other side new ideas and "solutions" are emerging in the area of statistical mechanics [1,2]; these ideas have been promptly applied to several other domains, such as Cell Biology, Sociology etc. [3-5]. In Cell Biology, for instance, cells are seen as multi-functional systems capable of producing different response to a practically infinite variety of inputs [3,4]. They have been analyzed in general terms and many interesting unveiled properties have been discovered [3]. Problems are thus reconducted to a unifying theoretical frame and analyzed therein with same methods and tools.

We have attempted to analyze, under this unifying perspective [2] and following an extensive seminal paper on the argument [6], data of the topological structure of the internet in US, at the Autonomous System (AS) level. The 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 [7]. This system represents an example of a complex infra-structure undergone a sizeable growth in the last few years, thus allowing the study of its growth process.

Methods and results

Let us indicate a generic network G=(N,L) as a set of nodes (N) and links (L) where |N|=*n* and |L|=*m*. A mathematical object allowing a complete definition of the network is the *Adjacency Matrix* **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. A further matrix which can be associated to the network is the so-called *Laplacian Matrix* **L**, defined as **L** = **D** - **A** (where **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 **A** and **L** matrices.

The AS-level routers network data [7] have been analyzed through the evaluation of several properties: the degree distribution, the clustering coefficient *c* [2], the network diameter *d* and the average path length <l> (average of the $n(n-1)/2$ node's distances). We have also evaluated some property deduced by the spectral analysis. There is a wide literature on the spectral analysis of the **A** and **L** matrices. The analysis of the defined quantities on the AS-level network data [7] (spanning a period from 17.3.1998 to 02.01.2000, configurations AS1 and AS4 respectively) has produced the results reported in Table 1.

date	*n*	*m*	*?*	*c*	*?*	k_{max}	*d*	*<l>*	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

Table 1: Relevant properties evaluated on the snapshots of the AS-level network at the different dates: n is the number of nodes, m the number of links, ? the ratio between existing number of links and the maximum possible number of links (n(n-1)/2); c the average clustering coefficient [8]; ? is the coefficient of the distribution of 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 [9]), <l> the average path lenght,* n_l *the number of links connecting the two subgraphs solution of the "min-cut" problem.*

As most of the analyzed self-generated networks [2,6], the AS-level router network shows a "scale-free" behaviour, i.e. its probability distribution of node's degree decays with a power law

$$P(k) \sim k^{?} \qquad (1)$$

where ? takes values of the order of 2.35 up to 2.51, at different times (see Table 1). A further intriguing property, which has been characterized for the first time in a recent work [10], concerns with the number of links showed by the network as solution of the min-cut theorem [11]. This theorem states that the Fiedler eigenvector, resulting from the complete diagonalization of the Laplacian matrix, allows to divide the network into two equivalent sub-nets linked by the minimum possible number of connections, called n_l. Interesting suggestions inferred from data in Table 1 are:

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

(b) the large value of the network clustering (0.19<*c*<0.25) gives evidence in favour of a high cliquishness of the network;

(c) low values of the average path length <*l*>;

(d) the rapidly-growing number of n_l with time; n_l begins with a very small value (n_l=11 at the first observation date) and grows at n_l =493 after less than two years, with a growth rate much higher than that exhibited by the whole network.

All these data have contributed to propose the scenario of a network which grows under the combined action of the Preferential Attacchment (PA) [2] and the "Triad Formation" (TF) [12] mechanisms. The first is responsible of the scale-free structure of the network, the second is able to sustain a consistent growth of the clustering coefficient *c*, whose behaviour, if related to the only PA mechanism and depending on the resulting ? parameter, would have decreased with the network growth [13]. Table 2 reports the relevant network's properties evaluated on the network whose growth has been simulated by using the TF mechanisms [10].

model	*c*	*?*	k_{max}	*d*	*<l>*	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)

Table 2: Same properties as in Table 1 for networks generated by the use of the PA and TF mechanisms. First line refers to the real network at the first observation date. Other lines refer to the simulated network configurations. In parentheses, the same quantities as evaluated on the real AS-level network.

We have succeded to reproduce the growth data of the AS-level network of US router by using the TF growth mechanisms [10]. This fact supports the scenario where the onset of "local" strategies, aimed at increasing robustness and reducing the local diameter (the TF mechanism corresponds to the formation of local "consortia" of internet providers which tend to increase their connections) is able to produce a similar overall effect, by simultaneously strengthening the network against a "large scale" disconnection (increase of the solution of the min-cut problem.

A similar type of analysis is going to be carried out on the high-voltage electrical power transmission networks of several European countries [14].

Conclusions

Measuring network's properties is a first fundamental step for any activity aiming at improving the general conditions allowing network survivability and dependability.

Several are the potential benefits of the application of the results of theoretical studies at the technological level:

(1) the definition of an appropriate growth model for a given network can be used to predict its extent and structure at later times;

(2) changes to the network's structure can be designed to reduce its vulnerability and improve its functionality by operating on specific parameters;

(3) relevant mechanisms and parameters could be inserted into specific network's design tools, for building up new structures from scratch, in order to foster, into the systems, technical solutions able to prevent (and to cope with) abrupt components faults, human errors or malicious attacks;

Combining the deep understanding of the system properties with the biological programming paradigm [15-16] is a way to exploring how to design systems that can adequately operate even in the presence of catastrophic failures and large scale attacks. The approach should overcome the constraints introduced by limitations of the traditional fault-tolerant approaches, based on explicit programming of recovery mechanisms.

References

[1] D. J. Watts, S. H. Strogartz, "Collective dynamics of small-world networks", Nature **393** (1998) 440.

[2] R. Albert, A.-L. Barabàsi, "Statistical Mechanics of Complex Networks", Rev. Mod. Phys. **74** (2002) 47.

[3] A.-L. Barabàsi, R. Albert, "Emergence of scaling in random networks", Science **286** (1999) 509.

[4] H. Kitano, " A graphical notation for biochemical networks", Biosilico **1** (2003) 169.

[5] P. Holme, B. J. Kim, "Growing scale-free networks with tunable clustering", Phys. Rev. E **65** (2002) 026107.

[6] A. Vasquez, R. Paster-Satorras, A. Vespignani, "Large-scale topological and dynamical properties of the internet", Phys. Rev. E **65** (2002) 066130.

[7] Data have been taken from the raw data repository of the NLNAR Measurement and Networking Analysis Group at the University of California, S Diego (USA) (http://moat.nlanr.net/Routing/rawdata). Analyzed data span a period comprised from March 17, 1998 to January 2, 2000 .

[8] The clustering coefficient c is the average (over all nodes) of the ratio of the effective number of links and

its maximum possible value between each node's neighbors; if athe node *i* has k_i neighbors,

c_i= number of links between the k_i nodes/[k_i (k_i-1)/2]

and, thus, c=? $_i c_i$/n (where *n* is the total number of nodes).

[9] E. W. Dijkstra, "A Note on Two Problems in Connection with Graphs", Numerische Math. **1** (1959) 269.

[10] V. Rosato, F. Tiriticco, "Growth mechanism of the AS-level internet network", Europhys. Lett. in press.

[11] B. Mohar, "Laplace eigenvalues of a graph: a survey", Discrete Math. **109** (1992) 171.

[12] P. Holme, B. J. Kim, CN.Yoon, S.K. Han, "Attack vulnerabiliy of complex networks", Phys. Rev. E **65** (2002) 056109.

[13] In the PA growth mechanism, new nodes are added by linking them to pre-existing nodes by choosing them with a probability proportional to their degree (highly connected nodes are favoured for receiving links of new nodes). The TF mechanism acts when more links must be added from a new node; the first link is selected according to PA, the others in a way to connect the new node with neighbours of the node chosen for the first connection.

[14] V. Rosato, S. Bologna, F. Tiriticco, in preparation.

[15] BICAS: Biologically Inspired Complex Adaptive Systems, http://www.hpl.hp.com/research/bicas/

[16] D. Evans et al., "A Biological Programming Model for Self-Healing", First ACM Workshop on Survivable and Self-Regenerative Systems, October 31, 2003.