SCALE FREE NETWORKS – A CHALLENGE IN MODELING COMPLEXITY

Radu Dobrescu
POLITEHNICA University of Bucharest
Splaiul Independentei 313, Faculty of Control and Computers
radud@isis.pub.ro

Abstract: The paper proposes a model that relieves the characteristics of several complex systems having a similar scale free network architecture. The properties of this kind of networks are compared with those of other methods which are specific for studying complex systems: nonlinear dynamics and statistical methods. We place particular emphasis on scale free network theory and its importance in augmenting the framework for the quantitative study of complex systems, by discussing three important applications: Internet topology and traffic characteristics, epidemics broadcast and cellular communication system in biological networks (in particular in Sepsis). Finally the new ways in modeling complex systems with scale-free networks are discussed.
.

Keywords: complexity, scale free networks, modeling; self similarity; cellular signalling

1. INTRODUCTION

Let start with some ideas for a definition of complexity, or more exactly of a complex system. We expect that its definition should be richer than that of algorithmic complexity, and should express the level of interconnectedness and interdependencies of a system, not just the instruction set for creating the system -the amount of effort it takes to use those instructions must addressed as well. In a complex system it is often the case that the utility of a structure or process is expressed at the next higher level of organization relative to the process itself. Unlike entropy and the related concept of information, complexity is not extensive, nor is it entirely intensive. What is clear though is that complexity concerns a specific description, which is of course dependent on the technology and subjective capabilities of the observer. Anyway, we can consider that a complex system is a system with a large number of elements, building blocks or agents, capable of interacting with each other and with their environment. The interaction between elements may occur only with immediate neighbors or with distant ones; the agents can be all identical or different; they may move in space or occupy fixed positions, and can be in one of two states or of multiple states. The common characteristic of all complex systems is that they display organization without any *external* organizing principle being applied.

Another problem is how to measure the complexity. The first and still classic measure of complexity is that introduced by Kolmogorov, which is (roughly) the shortest computer program capable of generating a given string. This quantity is in general uncomptable, in the sense that there is simply no algorithm which will compute it. Moreover, the Kolmogorov complexity is maximized by random strings, so it's really telling us what's random, not what's complicated, and it's gradually come to be called the "algorithmic information." Let also remind that Bennett proposed a measure for the computing complexity, which he called the logical depth of a system [1]. The basic idea is that a system should be called complex, or *logically deep*, if that system can be generated by a few simple rules, but those rules require a long time to run. So, for example, a human body is complex in that it is specified by a relatively

small amount of information encoded in DNA, but it takes a great deal of processing to get from that DNA to the human body.

So, if we can not define exactly the complexity and also we can not measure its dimensions, let try to model – if not complexity, at least the complex systems. Because we met more and more example of complex systems: physical, mechanical, biological, social. The stock market, cities, metabolic pathways, ecosystems, the Internet or the human brain, are all complex. We can ask what do have in common all these systems. In the last few years the answer that has emerged is that they all share similar network architectures. Network theory has become one of the most visible pieces of the body of knowledge that can be applied to the description, analysis, and understanding of complex systems and is now an essential ingredient their study. But it is not the only solution. Actually, as will be shown in the next section, network theory competes with other two complex systems modeling methods.

2. TOOLS FOR MODELING COMPLEX SYSTEMS

Remote In a rough sense, the current toolbox used in tackling complex systems involves three main areas: (i) nonlinear dynamics and chaos, (ii) statistical physics, including discrete models, and (iii) network theory.

2.1 Nonlinear dynamics and chaos

Nonlinear dynamics and chaos in deterministic systems are now an integral part of science and engineering. The theoretical foundations are well agreed upon mathematical definitions of chaos, many of them formally equivalent. However, because of its relative novelty and, in much case, counterintuitive nature, there are still many misconceptions about chaos and its implications. Extreme sensitivity to initial conditions does not mean that prediction is impossible. Memory of initial conditions is lost within attractors but the attractor itself may be extremely robust. In particular chaotic does not mean unstable. Chaos means that simple systems are capable of producing complex outputs. Many techniques have been developed to analyze signals and to determine if fluctuations stem from deterministic components. There are numerous applications in geophysics, physiology and neurophysiology [2].

2.2 Statistical physics: Universality and scaling

Statistical physics brought three very important conceptual and technical advances: 1. It lead to a new conception of prediction; 2. It circumvented classical mechanics and it casted solutions in terms of ensembles; 3. It introduced the concept of discrete models, ranging from the cellular automata to agent-based models. In the 1970s, fundamental advances occurred in our understanding of phase transitions and critical phenomena leading to the development of two important new concepts: *universality* and *scaling* [3]. The finding, in physical systems, of universal properties that are independent of the specific form of the interactions gives rise to the intriguing hypothesis that universal laws or results may also be present in complex social, economic and biological systems. The *scaling hypothesis* which arose in the context of the study of critical phenomena led to two categories of predictions, both of which have been remarkably well verified by a wealth of experimental data on diverse systems. The first category is a set of relations, called *scaling laws* that serve to relate the various critical-point exponents characterizing the singular behavior of the order parameter and of response functions. The second category is *data collapsing*. Another fundamental concept arising from the study of critical phenomena is *universality*. For systems in the same universality class, exponents and scaling functions are the same in the vicinity of the critical point. This fact suggests than when studying a given problem, one may pick the most tractable system to study and the results one obtains will hold for all other systems in the same universality class. Fractal analysis seems to be one of the most promising tools.

In what concern *discrete models* the main assumption in that some phenomena can and should be modeled directly in terms of computer programs (algorithms) rather than in terms of equations. Cellular automata are the simplest example of discrete time and space models that were developed with the computer in mind. Examples of the application of cellular automata exist in physical, chemical, biological and social sciences; they can be as simple as elementary predator-prey models between a handful of species and as complex as the evolution of artificial societies. Discrete, or agent-based, modeling has been extremely successful because of the intuition-building capabilities it provides and the speed with which it permits the investigation of multiple scenarios. For this reason discrete modeling has led in some cases to a replacement of equation based approaches in disciplines such as ecology, traffic optimization, supply networks, and behavior-based economics.

2.3 Networks

The third element in the toolbox is networks. A network is a system of nodes with connecting links. Once one adopts this viewpoint, networks appear everywhere. Consider some examples from two main fields: a) biological networks: autonomous nervous systems of complex organisms, a network of neurons connected by synapses, gene regulation networks, a network of genes connected by cross-regulation interactions or metabolic networks, a network of metabolites connected by chemical reactions, b)

social networks, like e-mails services, Internet and the World Wide Web. The structure of such social networks was formalized exactly by using *random graphs*, in which the existence of a link between any pair of nodes has probability *p*. Erdos, in collaboration with Renyi, pursued the theoretical analysis of the properties of random graphs obtaining a number of important results, including the identification of the percolation threshold, that is, the average number of links per node necessary in order for a random graph to be fully connected, or the typical number of intermediate links in the shortest path between any two nodes in the graph. Another important class of network is represented by the *Small-world networks,* that have as main characteristic the so-called small-world phenomenon, which is defined by the co-existence of two apparently incompatible conditions, (i) the number of intermediaries between any pair of nodes in the network is quite small - typically referred to as the six-degrees of separation phenomenon and (ii) the large local redundancy of the network, i.e., the large overlap of the circles of neighbors of two network neighbors. The latter property is typical of ordered lattices, while the former is typical of random graphs. Recently, Watts and Strogatz [4] proposed a minimal model for the emergence of the small-world phenomenon in simple networks. In their model, small-world networks emerge as the result of randomly rewiring a fraction *p* of the links in a *d*-dimensional lattice (Fig. 1). The parameter *p* enables one to continuously interpolate between the two limiting cases of a regular lattice ($p = 0$) and a random graph ($p = 1$).

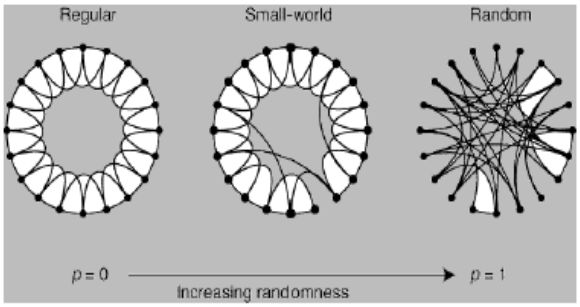

Fig.1. Small-world networks model generation

Watts and Strogatz probed the structure of their small-world network model via two quantities: (i) the mean shortest distance *L* between all pairs of nodes in the network, and (ii) the mean clustering coefficient *C* of the nodes in the network. For a *d*-dimensional lattice one has $L \sim N^{1/d}$ and $C = O(1)$, where *N* is the number of nodes in the network; for a random graph one has $L \sim \ln N$ and $C \sim 1/N$.

3. SCALE-FREE NETWORKS

An important characteristic of a graph that is not taken into consideration in the small-world model of Watts and Strogatz is the degree distribution, i.e., the distribution of number of connections of the nodes in the network. The Erdos-Renyi class of random graphs has a Poisson degree distribution, while lattice-like networks have even more strongly peaked distributions - a perfectly ordered lattice has a delta-Dirac degree distribution. Similarly, the small-world networks generated by the Watts and Strogatz model also have peaked, single-scale, degree distributions, i.e., one can clearly identify a typical degree of the nodes comprising the network. Against this theoretical background, Barabasi and coworkers found that a number of real-world networks have a scale-free degree distribution with tails that decay as a power law [5]. These networks were called Scale Free Networks (SFN).

3.1. General SFN properties.

Barabasi and Albert suggested that scale-free networks emerge in the context of growing network in which new nodes connect preferentially to the most connected nodes already in the network. Specifically,

$$p_i(n+1) = \frac{k_i(n)}{\sum_{-n_0+1}^{n} k_i(n)}$$

where *n* is the time and number of nodes added to the network, n_0 is the number of initial nodes in the network at time zero, k_i is the degree of node *i* and $p_i(n + 1)$ is the probability of a new node, added at time $n+1$ linking to node *i*.

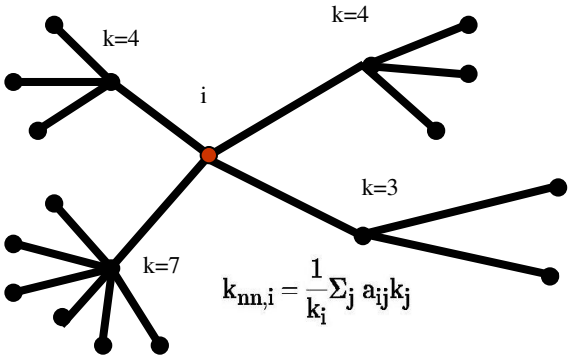

Fig.2. A scale free network graph

As illustrated in Figure 2, as time ticks by the degree distribution of the nodes becomes more and more heterogeneous since the nodes with higher degree are the most likely to be the ones new nodes link to. Significantly, scale-free networks provide extremely efficient communication and navigability as one can easily reach any other node in the network by sending information through the "hubs", the highly-connected nodes. The efficiency of the scale-free topology and the existence of a simple mechanism leading to the emergence of this topology led many researchers to believe in the *complete* ubiquity of scale-free network. Note that scale-free networks are a subset of all small-world networks because (i) the

mean distance between the nodes in the network increases extremely slowly with the size of the network and (ii) the clustering coefficient is larger than for random networks.

3.2 Diameter of scale-free networks

It was shown that scale-free networks with degree exponent $2<\lambda<3$ possess a diameter $D \sim \ln \ln N$, smaller even than that of random and small world networks [6]. If the network is fragmented, we will only be interested in the diameter of the largest cluster (assuming there is one). In this study we consider the diameter of a Molloy-Reed scale-free network definite as the *average* distance between any two sites on the graph. Actually, it easier still to focus on the radius of a graph, $L \equiv \langle l \rangle$ as the average distance of all sites from the site of highest degree in the network. The diameter of the graph D is restricted to $L \leq D \leq 2L$ and thus scales like L .

3.3 Minimal graphs and lower bound

Cohen, *et al*., show that the radius of any scale-free graph with $\lambda>2$ has a rigorous lower bound that scales as $\ln \ln N$. It is easy to convince oneself that the smallest diameter of a graph, of a given degree distribution, is achieved by the following construction: Start with the highest degree site, then connect to each successive layer the extant sites of highest degree, until the layer is full. By construction loops will occur only in the last layer [7]. To bound the radius L of the graph, we will assume that the low degree sites are connected randomly to the giant cluster. On the other hand, if we start uncovering the graph from any site - provided it belongs to the giant component – then within a distance l_2 from this site there are at least l_2 bonds. Since $l= l_1+l_2$, all sites are at a distance of order $\ln \ln N$ from the highest degree site, and $L= \ln \ln N$ is a rigorous lower bound for the diameter of scale-free networks with $\lambda>2$. In a similar way one can demonstrate that the scaling of $D \sim \ln \ln N$ is actually realized in the general case of *random* scale-free graphs with $2<\lambda<3$. For $\lambda>3$ and $N>>1$, k is independent of N , and the radius of the network is $L \sim \ln N$.

4. USING SFN IN COMPLEX SYSTEMS MODELLING

4.1 The topology of the Internet and the dynamics of Internet traffic

The Internet is a prime example of a self-organizing complex system, having grown mostly in the absence of centralized control or direction. In this network, information is transferred in the form of packets from the sender to the receiver via routers, computers which are specialized to transfer packets to another router “closer” to the receiver. A router decides the route of the packet using only local information obtained from its interaction with neighboring routers, not by following instructions from a centralized server. A router stores packets in its finite queue and processes them sequentially. However, if the queue overflows due to excess demand, the router will discard incoming packets, a situation corresponding to congestion.. A number of studies have probed the topology of the Internet and its implications for traffic dynamics. It has been reported that Internet traffic fluctuations are statistically self-similar [8] and that the traffic displays two separate phases, congested and non-congested. It was also shown that time series of number of connections are nonstationary and are characterized by different mean values depending on the observation period. Barthelemy, *et al.* analyzed data from the French national network which comprises 30 interconnected routers and is used by approximately 2 million individuals [9]. They found that the Internet flow is strongly localized: most of the traffic takes place on a spanning network connecting a small number of routers which can be classified either as “active centers,” which are gathering information, or “databases,” which provide information. A number of groups have also demonstrated that the Internet displays a number of properties that distinguishes it from random graphs: wiring redundancy and clustering, non-trivial eigenvalue spectra of the connectivity matrix and a scale-free degree distribution.
Experimental evidence for self-similarity in various types of data network traffic is already overwhelming and continues to grow. So far, simulations and analytical studies have shown that it may have a considerable impact on network performance that could not be predicted by the traditional short-range-dependent models. The most serious consequence of self-similar traffic concerns the size of bursts. Within a wide range of time-scales, the burst size is unpredictable, at least with traditional modeling methods.

4.2 Spread of epidemics in complex networks

The propagation of errors occurring on routers and servers that are physically linked in a large network is a typical example of epidemic process, in which the corruption (virus) is transmitted from infected to healthy individuals. Computer viruses are usually referred to as little programs that can reproduce themselves by infecting other programs [10]. The basic mechanism of infection is as follows: When the virus is active inside the computer, it is able to copy itself, by different ways, into the code of other, clean, programs. When the newly infected program is run into another computer, the code of the virus is executed first, becoming active and being able to infect other programs. Apart from reproducing themselves, computer viruses perform threatening tasks that range from flashing innocuous messages on the screen to seriously corrupt data stored in the

computer. These deleterious effects render most computer viruses as dangerous as their biological homonyms, and explain the interest, both commercial and scientific, arisen around their study. Computer viruses can be classified into three main classes, or *strains*. The first strain includes *file viruses* that infect application programs. A second and more harming family contains the *boot-sector viruses* that infect the boot sector of floppy disks and hard drives, a portion of the disk containing a small program in charge of loading the operating system of the computer. A third and nowadays prevailing strain is formed by the *macro viruses*. These viruses are independent of the platform's hardware and infect data files, such as documents produced with spreadsheets or word processors. They are coded using the *macro* instructions that are appended in the document, instructions used to perform a set of automatic actions, such as formatting the documents or typing long sequences of characters. In addition, with the ever more efficient deployment of antivirus software, more harmful viruses combining together the properties of the main strains have been developed. Noticeably, however, the nowadays dominant and most aggressive type of cyber organisms is represented by the *worms* family. Worms are actually viruses infecting the computer with mechanisms similar to usual viruses and making a particularly effective use of the e-mail for infecting new computers. In fact, by using the instructions of some commercial mail software applications, worms are capable of sending themselves to all the e-addresses found in the address-book of the person receiving the infected mail. This possibility renders worms the most effective viruses, especially in terms of the velocity at which they can propagate starting from a single infection.

The spreading of computer viruses has been studied for long years, in close analogy with the models developed for the study of the transmission of biological diseases. In this biological framework, the key point is the description of the epidemic process in terms of *individuals* and their *interactions*. In this simplified formalism, individuals can only exist in a discrete set of states, such as susceptible (or healthy), infected (and ready to spread the disease), immune, dead (or removed), etc. On the other hand, the interactions among individuals are schematized in the structure of the contacts along which the epidemics can propagate. Within this formalism, the system can be described as a *network* or graph, in which the nodes represent the individuals and the links are the connections along which the epidemics propagates. Standard epidemiological models usually consider *homogeneous* networks, which are those that have a connectivity distribution peaked at an average connectivity $\langle k \rangle$, and decaying exponentially fast for $k << \langle k \rangle$ and $k >> \langle k \rangle$. A typical example of deterministic homogeneous network is the standard hypercubic lattice, while among the random homogeneous network we can count the Erdos-Renyi model and the Watts-Strogatz model. On the other computer viruses and worms spread in environments characterized by scale-free connectivities. This will lead to the failure of the standard epidemic picture and will naturally introduce the scale-free connectivity as an essential ingredient for the understanding the spread of a computer virus. In a recent paper [11] was studied the effect of the special nature of scale-free distribution on the properties of random network models, including methods for the study of the layer structure of the graph, the percolation threshold and the critical exponents. The cited paper analyses also an epidemiological framework obtained in population networks characterized by a scale-free connectivity pattern. It was shown that SFN are very weak in face of infections, and its susceptibility to epidemic spreading is reflected also in an intrinsic difficulty in protecting them with uniform immunization policies. But targeted or selective immunization procedures achieve the desired lowering of epidemic outbreaks and prevalence. The special properties of scale-free networks might prove useful for applications such as the design of more robust networks, the improvement of routing algorithms and the prevention of an epidemic broadcast of computer or human viruses.

4.3 Biological networks in Sepsis modeling

The evolution in Sepsis can be considered as a result of information transfer in a complex cellular and even molecular communication system. Although molecular biology is mainly focused on identification of genes and functions of their products, which are components of the system, the major challenge in analysing Sepsis is to understand at the system level the biological system within a consistent framework of knowledge built up from the molecular level to the functional system level – not only gene networks, but also protein networks, signaling networks, metabolic networks and specific systems such as the immune system. At a very abstract level, a cell can be divided into two general subnetworks, a regulatory network and a metabolic network. These networks possess very different characteristics. The metabolic network is mainly occupied with substance transformation to provide metabolites and cellular structures. The regulatory network's main task is information processing for the adjustment of enzyme concentrations to the requirements of variable internal and external conditions. This network involves the use of genetic information.

In our approach to model intercellular communication in Sepsis [12], the basic network model consists of cell types as nodes and of *intercellular* signaling species (first messengers) connecting the nodes. Because communication between cell types occurs in an explicit direction and various kinds of communication might exist, the resulting graph is directed. The resulted model is a

highly inhomogeneous scale-free network in which a few highly connected cells play a central role in mediating interactions among numerous, less connected cells. There will be a lot of future work to make this model efficient, especially by using its self-similarity property in order to decide only of a few numbers of connections. One possible function of this model is to activate output only if the input signal is persistent and to allow a rapid deactivation when the input goes off. Between each node pair multiple edges (in both directions) are possible. Also edge weights (at least for the name/type of the connection) are necessary to reflect biological communication in a realistic manner. Thus, we do not model each individual cell, but the principal connections between cell types. In contrast to most other network models investigated recently this *intercellular* network possesses connectivity complexity rather than node complexity. The number of cell types in the human body is small (approximately 200) and fixed. The number of edges in contrast is principally orders of magnitude higher and varies over time. So, one of the challenges for future work will be to deal analytically and explanatory with this kind of complexity.

6. CONCLUSIONS AND FUTURE WORK

The need for enhanced computational ability is most evident when one attempts to couple large numbers of individual units into highly interactive and largely parallel networks. The proliferation of information transferred in such networks introduces the need for these systems that provide a framework for classifying information, spatial statistics for analyzing patterns, and dynamic simulation models that allow the integration of information across multiple spatial, temporal, and organizational scales. It is impossible to ignore the apparent universality of pair interactions among the various elements of a complex system. Instead of chance and randomness, one must consider a high degree of internal order that governs the system organization. Each node selected in order to be discussed as an element in a network of interacting constituents, ensures to spot and quantify the interplay between behavior, structure and function. It can be approached from the bottom up, moving from cells to modules, or from the top to the bottom, starting from the network's scale-free and hierarchical nature and moving to the specific modules. In either case, it must be acknowledged that structure, topology, network usage, robustness and function are deeply interlinked. The edge complexity could be reduced in different respects.. For instance a clustering of a network derived from the connectivity distribution of the nodes might show sub-networks of intense communication or the impact of distinct nodes for the whole system. Also the validation of the biological or physical plausibility of scale free networks reconstructed from databases is of major importance. Network modeling, quantitative analysis and laboratory experiments have to be combined in various ways to gain new insights.

REFERENCES

.

[1] C. H. Bennett, How to define complexity in physics, and why, in W. Zurek, ed., *Complexity, entropy and the physics of information*, Addison-Wesley, 1990, pp.137-148

[2] L. Glass, M.C. MacKey, *From Clocks to Chaos: The Rhythms of Life,* Princeton University Press, Princeton, NJ, 1988

[3] H.E. Stanley, *Introduction to Phase Transitions and Critical Phenomena,* Oxford University Press, Oxford, 1971

[4] D. J.Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature,* 393, 1998, pp. 440-442.

[5] A-L. Barabasi, R. Albert, Emergence of scaling in random networks, *Science*, 286, 1999, pp. 509-512

[6] S. N. Dorogovtsev, J. F. F. Mendes, Natural Scale of Scale-free Networks. *Phys. Rev. E* **63**, 2001, p. 62101

[7] Cohen, R., K. Erez, D. ben-Avraham and S. Havlin (2000). Resilience of the Internet to Random Breakdown. *Phys. Rev. Lett.* **85**, pp. 4626–4628

[8] Dobrescu, R., M. Dobrescu and St. Mocanu (2004). Using Self Similarity To Model Network Traffic, WSEAS Transactions on Computers, Issue 6, **3**, pp. 1752-1757

[9] M. Barthelemy, B. Gondran, E. Guichard, Spatial structure of the internet traffic. *Physica A*, 319, 2002, .pp.633-642

[10] Harley, C. D., R. Slade, D. Harley, E. H. Spafford and U E. Gattiker (2001) *Viruses Revealed*, McGraw-Hill, New York

[11] R.Dobrescu, C. Vasilescu, C. Ivan, Using scale-free networks to model epidemic broadcast, Proc. of the 2-nd Int. Symp. IAFA, 2005, vol.3, pp. 57-62

[12] R. Dobrescu, An informational approach in modeling molecular interactivity in sepsis, Proc. of the Int. Humboldt Workshop on Surgical Research, 2005, pg. 172-177