Information Theory of Complex Networks:
on evolution and architectural constraints

Ricard V. Solé[1,2] and Sergi Valverde[1]

[1]*ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Dr Aiguader 80, 08003 Barcelona, Spain*

[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA*

Complex networks are characterized by highly heterogeneous distributions of links, often pervading the presence of key properties such as robustness under node removal. Several correlation measures have been defined in order to characterize the structure of these nets. Here we show that mutual information, noise and joint entropies can be properly defined on a static graph. These measures are computed for a number of real networks and analytically estimated for some simple standard models. It is shown that real networks are clustered in a well-defined domain of the entropy-noise space. By using simulated annealing optimization, it is shown that optimally heterogeneous nets actually cluster around the same narrow domain, suggesting that strong constraints actually operate on the possible universe of complex networks. The evolutionary implications are discussed.

Keywords: Information theory, complex networks, evolution, selection, emergence, simulated annealing

I. INTRODUCTION

Many complex systems are to some extent describable by the network of interactions among its components. Beyond the specific features displayed by each net, it has been shown that a number of widespread properties are common to most of them. One is the presence of the small-world phenomenon and the second the observation that in many cases they are highly heterogeneous in their connectivity patterns (Albert and Barabási, 2002, Bornoldt and Schuster, 2002, Dorogovtsev and Mendes, 2002, 2003).

Heterogeneity can be easily identified by looking at the so called degree distribution P_k, which gives the probability of having a node with k links. Most complex networks (both natural and artificial) can be described by a degree distribution $P_k \sim k^{-\gamma}\phi(k/\xi)$ where $\phi(k/\xi)$ introduces a cut-off at some characteristic scale ξ (Amaral et al, 2000). An example of such scale-free networks is provided by the architecture of digital electronic circuits (figure 1). It has been shown (Ferrer, Janssen and Solé, 2001) that these systems exhibit long tail distributions of links, where the nodes are electronic components and the links are physical wires between units. Most elements are connected to a few others (for circuits this usually means nearest neighbors) but some are connected to many others. In figure 2 several examples of the observed distributions for both analogic (a-b) and digital (c-d) systems are shown. Although analogic systems are closer to an exponential distribution (i. e. ξ small) digital, large-scale systems[1] exhibit scaling behavior, with $\gamma \sim 3$.

Scale-free nets have been shown to be obtainable through a number of mechanisms, including preferential attachment (Barabási and Albert, 1999; Dogorovtsev and Mendes, 2002), optimization (Ferrer and Solé, 2003, Valverde, Ferrer and Solé, 2002), duplication and divergence (Solé et al, 2002, Vazquez et al. 2003) or fitness-dependent, rich-gets-richer mechanisms (Caldarelli et al., 2002). Beyond the common qualitative architecture shared by these systems, the dynamical patterns and their time scales that take place on top of these webs differs from system to system, although in a way or another deals with information propagation and/or processing. Besides, the response to node removal differ from system to system. Although genetic and metabolic networks seem to be fairly robust against perturbations of different types, a totally different situation arises in electronic circuits. In biological nets failure of highly connected components will tipically end in system's failure (for example, at the cellular level). But failure (by mutation or transient change) of a gene is often buffered by the rest of the system. This is not the case for electronic circuits

---

[1] Similar results have been obtained by looking at VLSI networks, see for example: http://citeseer.nj.nec.com/450707.html
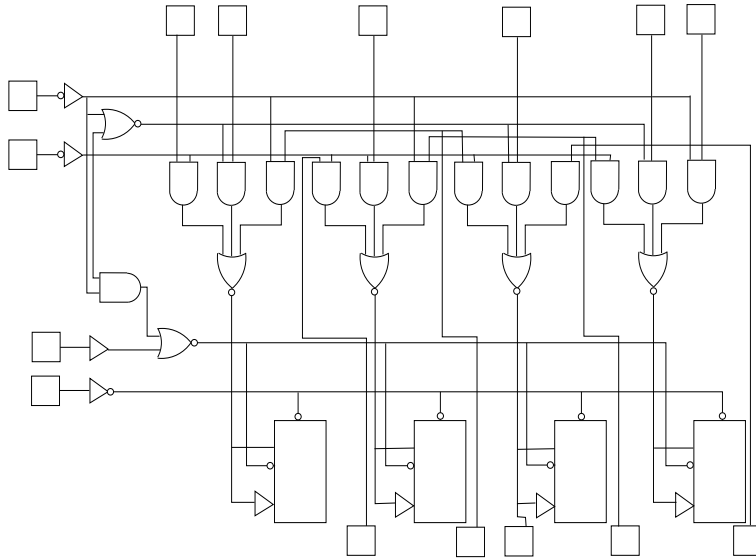
FIG. 1 Heterogeneity is a widespread feature of most (but not all) complex networks. An example from technology graphs are electronic circuits (upper plot) which have been shown to display scale-free distributions of links.

and, to a similar extent, by software networks. Failure of any component typically ens in system's failure, no matter how much linked is the given unit.

Several quantitative measures can be used in order to characterize a given network. The first step is to define an appropriate representation in terms of a graph Ω, defined by a pair $\Omega = (W, E)$, where $W = \{s_i\}, (i = 1, ..., N)$ is the set of N nodes (species, proteins, neurons, etc) and $E = \{\{s_i, s_j\}\}$ is the set of edges/connections between nodes. The *adjacency matrix* ξ_{ij} indicates that an interaction exists between two nodes $s_i, s_j \in \Omega_p$ ($\xi_{ij} = 1$) or that the interaction is absent ($\xi_{ij} = 0$). Several statistical properties, such as average degree, clustering or diameter can be defined from the adjacency matrix.

But the universe of possible networks (figure 3), although not arbitrarily diverse, displays a number of variations that cannot be compressed by the previous average quantities. Real networks are not only tipically heterogeneous, but they also involve other types of features, such as hierarchical organization (Ravasz et al., 2002).

In figure 3 we qualitatively summarize the basic types of network organization by using a generic parameter space. Here heterogeneity, modularity and randomness define three axes. Assuming that such three parameters can be properly defined, different real and model graphs can be located at different locations. The current knowledge of network architecture in many different systems strongly indicates that the domain of random networks with long tailed degree distributions and some amount of modular structure are rather densely occupied. In spite that the evolutionary process leading to these different systems are rather diverse, it is interesting to see that there is a strong convergence towards this type of architectures. Here we will explore this problem by using information-based statistical measures.

II. MEASURING CORRELATIONS

Beyond the degree distribution and average statistical measures, correlation measures offer considerable insight into the structural properties displayed by complex networks (Newman, 2002). One particularly interesting is network asortativeness (ref). Some networks show *assortative mixing* (AM): high degree vertices tend to attach to other high-degree vertices. At the other extreme there are graphs displaying *dissortative mixing* (DM), thus involving anticorrelation. The later are common in most biological nets, whereas the former are common in social and collaboration networks. It has been suggested that the presence and sign of assortativeness in this nets can have deep implications to their resilience under node removal or disease propagation.

Following a previous analysis (Newman, 2002) we will be interested here not in the degree
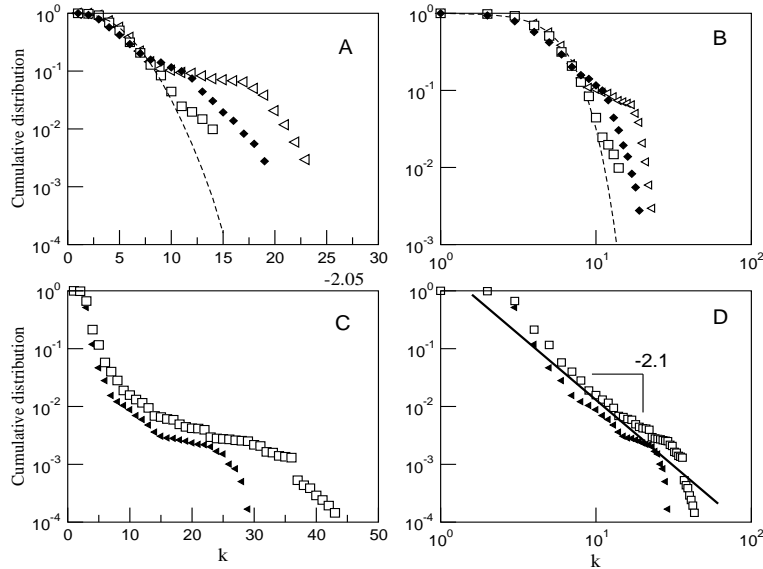
FIG. 2 Cumulative degree distributions for several examples of analogic (a-b) and digital (c-d). Although the analogic systems are less heterogeneous, analogic circuits (particularly large systems) display scaling in their degree distributions.

distribution P_k but instead in the *remaining degree*: the number of edges leaving the vertex other than the one we arrived along. This new distribution $q(k)$ is obtained from:

$$q(k) = \frac{(k+1)P_{k+1}}{<k>} \tag{1}$$

where $<k> = \sum_k kP_k$. In a network with no assortative (or disassortative) mixing $q_c(j,k)$ takes the value $q(j)q(k)$. If there is assortative mixing, $q_c(j,k)$ will differ from this value and the amount of assortative mixing can be quantified by the connected degree-degree correlation function

$$\langle jk \rangle - \langle j \rangle \langle k \rangle = \sum_{jk} jk q_c(j,k) - \left(\sum_j jq(j) \right)^2 \tag{2}$$

where $\langle \ldots \rangle$ indicates an average over edges.

The correlation function is zero for no assortative mixing and positive or negative for assortative or disassortative mixing respectively. In order to compare different networks, normalization is required. By dividing by its maximal value, which it achieves on a perfectly assortative network, i.e., one with $q_c(j,k) = q(k)\delta_{jk}$. This value is equal to the variance $\sigma_q^2 = \sum_k k^2 q(k) - \left[\sum_k kq(k) \right]^2$ of the distribution $q(k)$, and hence the normalized correlation function is

$$r = \frac{1}{\sigma_q^2} \left[\sum_{jk} jk(q_c(j,k) - \left(\sum_j jq(j) \right)^2 \right] \tag{3}$$

As defined from the previous equation, we have $-1 < r < 0$ for DM and $0 > r > 1$ for AM. Both biological ad technological nets tend to display DM, whereas social webs are clearly assortative.

Correlation functions have been widely used both in statistical physics (Stanley et al., 1996) and nonlinear dynamics (Abarbanel et al., 1993). A closely related, and more general approach involves the use of information-based measures (Ash 1965, Adami 1998, Li 1990, 1991). One specially important quantity is the so called *mutual information*, which is a general measure of dependence between two variables (Ash 1965, Li 1991). Correlation functions measure linear relations, whereas mutual information measures the general dependence and is thus a less biased statistic. The relevance of this difference is illustrated by the analysis of chaotic dynamical systems: the second
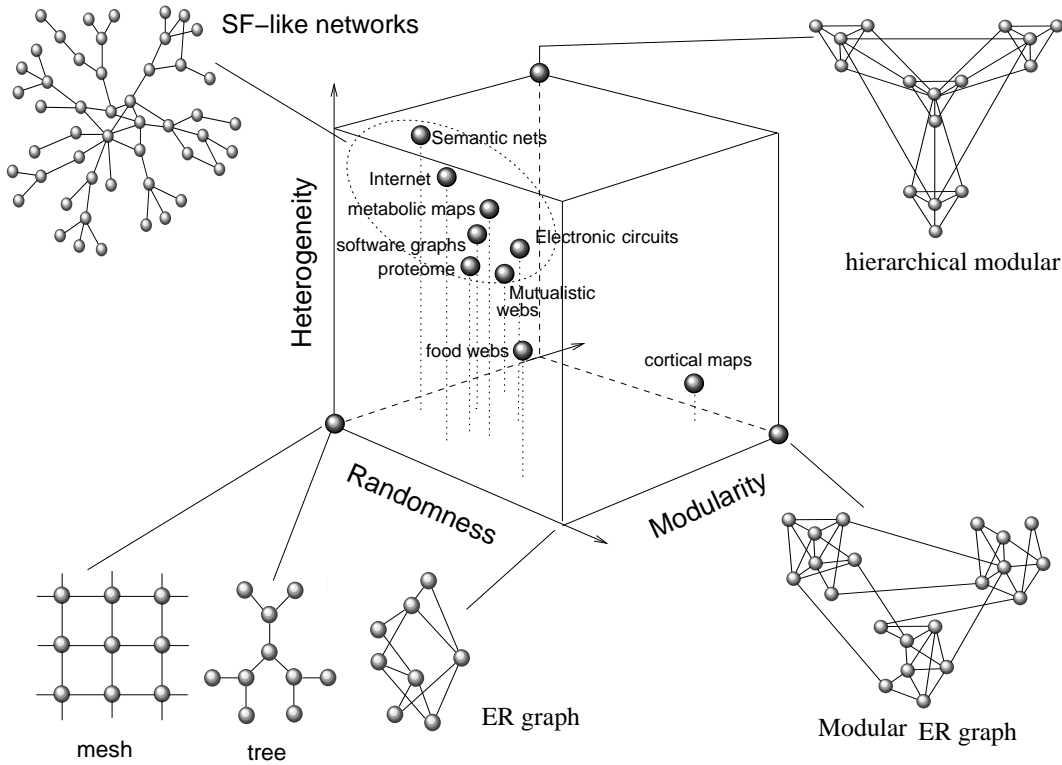
FIG. 3 A zoo of complex networks. In this qualitative space, three relevant characteristics are included: randomness, heterogeneity and modularity. The first introduces the amount of randomness involved in the process of network's building. The second measures how diverse is the link distribution and the third would measure how modular is the architecture. The position of different examples are only a visual guide. The domain of highly heterogeneous, random hierarchical networks appears much more occupied than others. Scale-free like networks belong to this domain.

allows to determine the independent variables for (re-) constructing phase trajectories (Fraser and Swinney, 1986). This cannot be done from linear correlation functions. Additionally, the definition of mutual information within the context of communication channels implies additional statistical quantities (such as channel entropy and noise) that provide a detailed characterization of system's complexity. Here we show how these quantities can be properly defined for complex networks, how they correlate with other statistical measures and what is their meaning and implications.

III. ENTROPY AND INFORMATION

By using the previous distribution $\mathbf{q} = (q(1), ..., q(i), ..., q(N))$, an entropy measure $H(\mathbf{q})$ can be defined:

$$H(\mathbf{q}) = -\sum_{k=1}^{N} q(k) \log(q(k)) \tag{4}$$

The entropy of a network will be a measure of uncertainty (Ash, 1965). Within the context of complex nets, it provides an average measure of network's heterogeneity, since it measures the diversity of the link distribution. The maximum is $H_{\max}(\mathbf{q}) = \log N$ is obtained for $q(i) = 1/N (\forall i = 1, ..., N)$ and $H_{\min}(\mathbf{q}) = 0$ which occurs when $\mathbf{q} = (1, 0., , , 0)$. In an information channel, there is a distinction between source and destination. Given the symmetric character of our system, no such distinction is made here. In figure 4 we can see the impact of heterogeneity on entropy. Specifically, we computed the entropy $H(\mathbf{q}; \gamma, \xi)$ for $\gamma \in (2, 3)$ and $\xi \in (0, 50)$ for a distribution $P_k \sim k^{-\gamma} \phi(k/\xi)$ using different scaling exponents γ and cut-offs ξ. The impact of diversity (long tails) is obvious, increasing the uncertainty. As the scaling exponent increases or
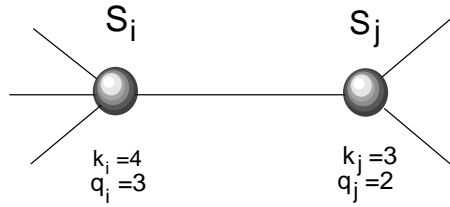
FIG. 4 Computing correlations in a network. Here two given, connected nodes s_i, s_j are shown, displaying different degrees k_i, k_j. Since we are interested in the remaining degrees, a different value needs to be considered (here indicated as q_i, q_j).

the cut-off decreases, the network becomes less heterogeneous and as a result a lower entropy is observed.

Similarly, the joint entropy can be computed by using the previous joint probabilities:

$$H(\mathbf{q}, \mathbf{q}') = -\sum_{k=1}^{N}\sum_{k=1}^{N} q_c(k, k') \log q_c(k, k') \tag{5}$$

Since it considers all possible pairs of events, this entropy provides a measure of the average uncertainty of the network as a whole. As before, it can be understood in terms of a measure of the diversity of linked pairs with given remaining degrees.

The information transfer $I(\{q_k\})$ of a given system is defined by means of the difference:

$$I(\mathbf{q}) = H(\mathbf{q}) - H_c(\mathbf{q}|\mathbf{q}') \tag{6}$$

where the last term $H_c(\mathbf{q}|\mathbf{q}')$ is the conditional entropy that involves a different set of *conditional* probabilities $\pi(k|k')$ (Ash, 1965). They give the probability of observing a vertex with k edges leaving it provided that the vertex at the other end of the chosen edge has k' leaving edges. This entropy is defined as:

$$H_c(\mathbf{q}|\mathbf{q}') = -\sum_{k=1}^{N}\sum_{k'=1}^{N} q(k)\pi(k|k') \log \pi(k|k') \tag{7}$$

Since the conditional and joint probabilities are related through:

$$\pi(k|k') = \frac{q_c(k, k')}{q(k')} \tag{8}$$

the conditional entropy can actually be computed in terms of the two previous distributions:

$$H_c(\mathbf{q}|\mathbf{q}') = -\sum_{k=1}^{N}\sum_{k'=1}^{N} q_c(k, k') \log \frac{q_c(k, k')}{q(k')} \tag{9}$$

we thus have, from the previous expressions,

$$I(\mathbf{q}) = H(\mathbf{q}) - H_c(\mathbf{q}|\mathbf{q}') \tag{10}$$

$$= -\sum_{k=1}^{N}\sum_{k'=1}^{N} q_c(k, k') \log q(k) + \sum_{k=1}^{N}\sum_{k'=1}^{N} q_c(k, k') \log \pi(k|k')$$

$$= -\sum_{k=1}^{N}\sum_{k'=1}^{N} q_c(k, k') \log \frac{q(k)}{\pi(k|k')}$$

which gives a final form for the information transfer function [2]:

$$I(\mathbf{q}) = \sum_{k=1}^{N}\sum_{k'=1}^{N} q_c(k, k') \log \frac{q_c(k, k')}{q(k)q(k')} \tag{11}$$

---

[2] The previous measures can be extended (with some care) into continuous distributions. In this case, we

Some limit cases are of interest here. The first corresponds to the maximum information transfer, which is obtained, for a given $\{q_k\}$, when $H_c(\mathbf{q}|\mathbf{q}') = 0$, i. e. when the conditional probabilities are such that $\pi(k|k') = 1$ or 0 for all $k, k' = 1, ..., N$. Another is given by $\pi(k|k') = \delta_{k,k'}$. This case corresponds to a deterministic channel in standard information theory (Ash, 1965). This implies that $q_c(k,k') = q(k')\delta_{k,k'}$ which is precisely the case of perfectly assortative network (Newman 2002).

In analogy with information channels, we can find a maximum value of the information, which we call the *network's capacity* $\mathcal{C} = \max_{\{q_k\}} I(\mathbf{q})$. There is no general method to compute $\mathcal{C}$ for an arbitrary channel. It can only be computed in some specific cases.

By using the previous functions, we will measure three key quantities: (a) the amount of correlation between nodes in the q-graph, as measured by the information transfer; (b) the noise level, which will provide a measure of assortativeness and (c) the average diversity associated to the $q(k)$ distribution. Since the total information involves the two last terms in a linear fashion, a noise-entropy space will be constructed and the distribution of real nets on this space will be analysed.

IV. MODEL NETWORKS

In the following sub-sections some simple, limit cases will be considered. Different types of architectures are represented by some standard networks exhibiting different degrees of heterogeneity and randomness. The list is far from exhaustive but provides an idea of what are the effects of each ingredient on information transfer and entropies.
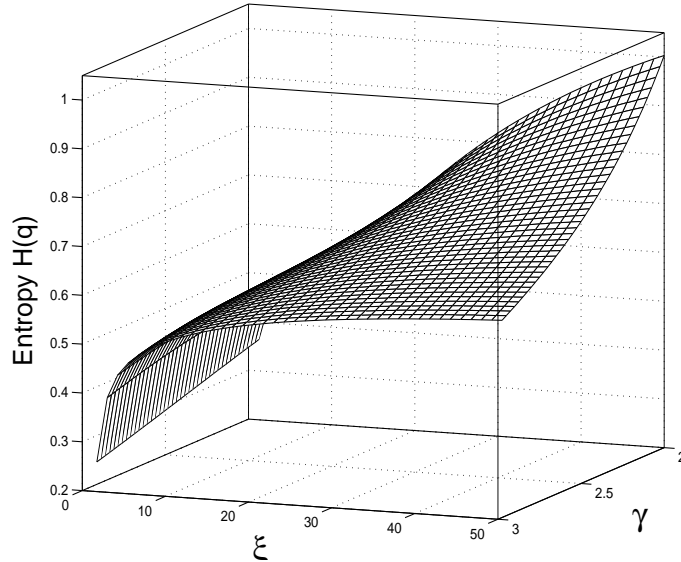

FIG. 5 Entropy of the remaining degree distribution obtained from a network with degree distribution $P_k \sim k^{-\gamma}\phi(k/\xi)$. Here $H(\mathbf{q}) = -\int P_k \log P(k)dk$ is shown against the scaling exponent γ and the cut off ξ. As expected, the entropy becomes larger for smaller γ and decreases as ξ is reduced.

———

must assume that the continuous counterparts of the previous degree distributions can be defined. The new distributions are such that the normalization conditions: $\int q(k)dk = 1$ and $\int\int q_c(k,k')dkdk' = 1$ are at work. Provided that the distributions are well behaved, the information transfer is now given by $I(\mathbf{q}) = \int\int q_c(k,k')\log\left(\frac{q_c(k,k')}{q(k)q(k')}\right)dkdk'$. Accordingly, entropy and noise would be obtained from: $H(\mathbf{q}) = -\int q(k)\log(q(k))dk$ and $H_c(\mathbf{q}|\mathbf{q}') = -\int\int q_c(k,k')\log\pi(k|k')dkdk'$

A. Lattices and trees

Lattice-like networks are common in some man-made architectures, particularly parallel computers (Germain-Renaud and Sansonnet 1991, Milutinovic 1988, Hillis 1985). These nets represent the highest degree of homogeneity and have no randomness. For a lattice, we have $P_k = \delta_{k,z}$, where z is a fixed number of links per node and δ_{ij} the Kronecker's delta function. For this ordered graph Ω_L, we have

$$q(k) = \delta_{k,z-1} \tag{12}$$

$$q_c(k, k') = \delta_{k,z-1}\delta_{k',z-1} \tag{13}$$

and thus

$$I(\mathbf{q}) = H(\mathbf{q}) = H_c(\mathbf{q}|\mathbf{q}') = 0 \tag{14}$$

This is a trivial case, since the homogeneous character of the degree distribution implies zero uncertainty. The same situation arises for a Cayley tree (Bethe lattice), where each node has exactly the same degree. Tree-like architectures are also common in designed systems, such as small-sized software graphs (Valverde, Ferrer and Solé, 2002) and communication networks.

B. Erdös-Renyi graphs

Erdös-Renyi graphs $\Omega_{N,p}$ are random graphs such that two nodes are joined with some probability p. These types of graphs have been widely used as the backbone of null models of genetic (Kauffman 1993, ecological (May 1976) and neural (Amari 1972) networks. It seems also appropriate in describing the topology of species-poor ecosystems (Montoya and Solé, 2003). The distributions are single-scaled and thus low uncertainty and high randomness are at work. The average degree will be $< k > \approx pN$, and it can be easily shown that the probability P_k that a vertex has a degree k follows a Poisson distribution $P_k = e^{-<k>} < k >^{-k} /k!$, and thus

$$q(k) = \frac{(k+1)e^{-<k>}}{< k > (k+1)!} < k >^{-(k+1)} = P_k \tag{15}$$

For this random graph, the independence associated to the link assignement implies (for N large) $q_c(k, k') = q(k)q(k')$ and thus information transfer is zero.

An interesting extension of the standard ER graph allows to introduce modularity into the graph structure (Ravasz et al., 2002). In general, the graph Ω is partitioned into m subgraphs $\{\Omega_i\}, (i = 1, ..., m)$ of relative size $\eta_i = |\Omega_i|$ such that

$$W = \bigcup_{i=1}^{m} W_i \quad (W_j \cap W_k = \emptyset) \tag{16}$$
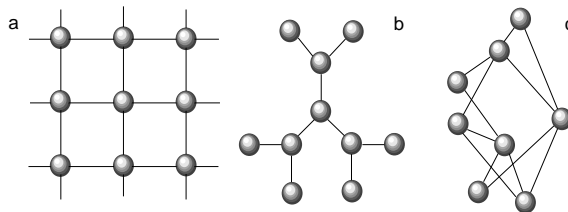

FIG. 6 Homogeneous networks: here two examples of a lattice (a) and a tree (b) are shown as examples of deterministic nets. In both cases each node has the same degree and thus both the entropy and the noise are zero. In (c) a random, Erdös-Renyi graph is shown. Here some amount of heterogeneity is at work, but the variance equals the mean and both noise and entropy are very close, giving as a result a small information i. e. no correlations.

and such that

$$\sum_i \eta_i = 1 \tag{17}$$

All nodes $s_j \in \Omega_i$ are connected with probability q and additionally we have a probability p of connecting two nodes belonging to different modules. The average degree of this system is $< k >= pN/m + (m-1)Nq/m$. Given the random wiring, it is not difficult to show that for large N information will be tipically very small.

C. Star Graph

Star graphs define another extreme within the universe of complex nets. Although no real network is likely to be described in terms of a pure star graph, it is certainly a common motif in many graphs. They are largely responsible for the short distances achieved in SF networks. Besides, a star graph can be shown to be optimal for low-cost communication (Ferrer and Solé, 2003).

This graph Ω_* is characterized by a degree distribution:

$$p(k) = \frac{n-1}{n}\delta_{k,1} + \frac{1}{n}\delta_{k,n-1} \tag{18}$$

The corresponding distribution $q(k)$ is:

$$q(k) = \frac{1}{2}\left[\delta_{k,0} + \delta_{k,N-1}\right] \tag{19}$$

and the joint probabilities are reduced to:

$$q_c(k,k') = \delta_{k,N-2}\delta_{k',0} \tag{20}$$

The entropy is maximal, given by:

$$H(\mathbf{q}) = -q(0)\log q(0) - q(N-1)\log q(N-1) \tag{21}$$

which gives $H(\mathbf{q}) = \log 2$. The noise term is $H_c(\mathbf{q}|\mathbf{q}') = 0$, since $\pi(k|k') = \delta_{kk'}$. The information is thus maximal, with $I(\mathbf{q}) = H(\mathbf{q}) = \log 2$. The star graph displays maximum information, as expected given the deterministic character of the conditional probabilities.

V. REAL NETWORKS

In this section we present some analysis of the information measures as applied to real networks. A large set of both technological and biological graphs has been studied. Specifically, three groups of data sets were used in our analysis, all of them known to be highly heterogeneous displaying scale-free architecture:

1. Metabolic networks: a graph theoretic representation of the biochemical reactions taking place in a given metabolic network can be easily constructed. In this representation, a metabolic network is built up of nodes, the substrates, that are connected to one another through links, which are the actual metabolic reactions (Jeong et al., 2000).

2. Software class diagrams: Nodes are software components and links are relationships between software components. *Class diagrams* constitute a well-known example of such graphs (Valverde et al., 2002; Valverde and Solé, 2003; Myers, 2003).

3. Electronic circuits: they can be viewed as networks in which vertices (or nodes) are electronic components (e.g. logic gates in digital circuits and resistors, capacitors, diodes and so on in analogic circuits) and connections (or edges) are wires in a broad sense (Ferrer, Janssen and Solé, 2001).
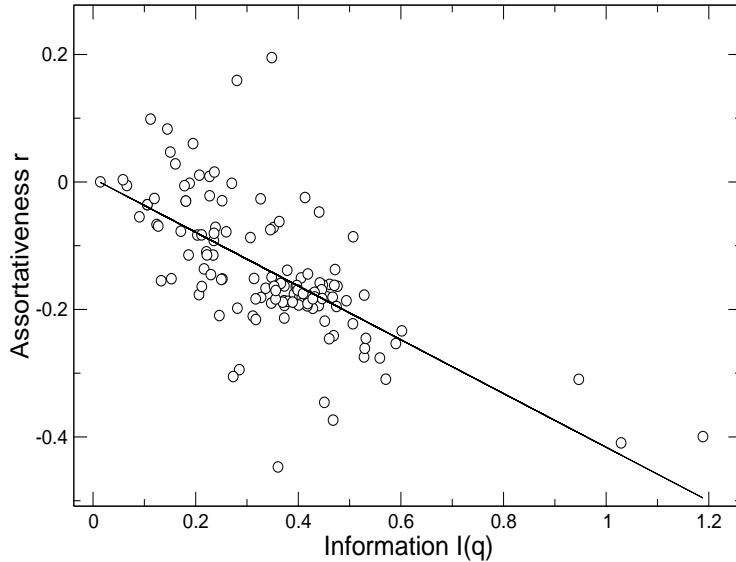
FIG. 7 Information transfer and assortativeness appear to be roughly correlated in a negative way. Here several systems have been used (all those analysed in this paper) and a linear interpolation has been displayed. Although the trend is clear, considerable variance can be appreciated, probably due to the underlying nonlinear mapping between both measures.

In table I we also show a list of selected networks obtained from very different systems and ordered from the higher to the lower information. The system's size N, average connectivity $< k >$, information measures and the assortative mixing coefficient r are provided. We can see that most nets are disassortative, as predicted in (Newman, 2002). Actually, information and r appear to be negatively correlated. This is shown in figure 7, where $I(\mathbf{q})$ is shown against r for different systems.

It is important to see that, in spite of the roughly negative correlation (a linear interpolation has been used) a large variance is observable, and a range of r values is associated to each information transfer. Such a variable plot is likely to be the result of the nonlinear character of the information transfer, not shared by the (linear) correlation defined by assortative mixing measures.

By displaying noise against entropy, the general picture that emerges is that the set of complex networks analysed here displays tipically uncorrelated (and hence non-assortative) structure. This is clear from the strongly linear dependence shown between noise and entropy (figure 8). If two given, randomly chosen nodes with remaining degrees k, k' are tipically connected with some probability, roughly irrespective of their mutual degree (i. e. low assortativeness is present) we should expect:

$$q_c(k, k') \approx q(k)q(k') \tag{22}$$

and thus we would have

$$\pi(k|k') \approx q(k) \tag{23}$$

in this case, the noise will be given by:

$$H_c(\mathbf{q}|\mathbf{q}') = -\sum_{k=1}^{N}\sum_{k'=1}^{N} q(k, k') \log \pi(k|k') \tag{24}$$

$$= -\sum_{k=1}^{N}\sum_{k'=1}^{N} q(k, k') \log q(k) = H(\mathbf{q}) \tag{25}$$
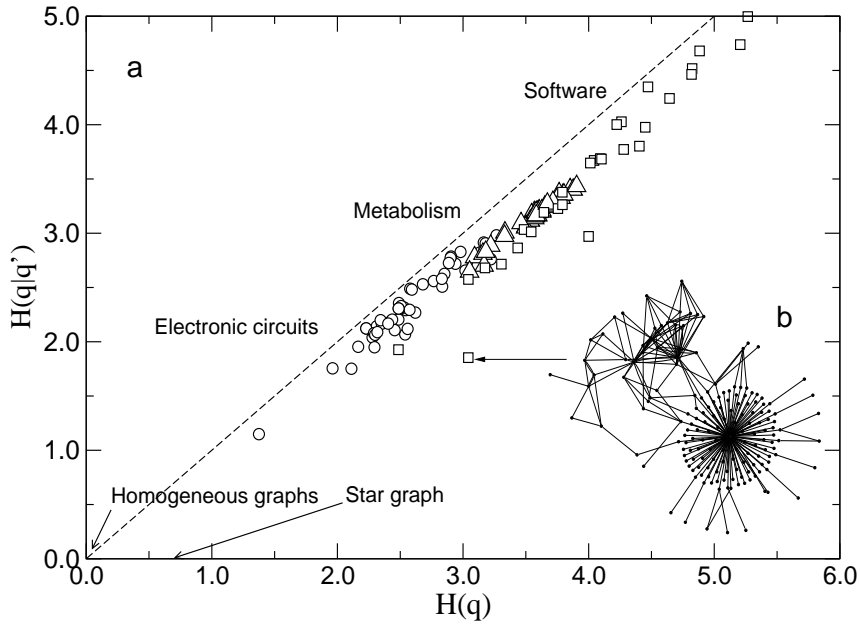
FIG. 8 Noise-entropy plot for different real networks, both natural and artificial. Here electronic circuits (open circles), metabolic (triangles) and software maps (squares) are shown to be close to the zero-information line, i. e. when entropy equals to noise. An example of a software graph that significantly deviates from the $H = H_c$. The network is small and has a rather particular shape, involving a large hub plus another cluster of connected classes.
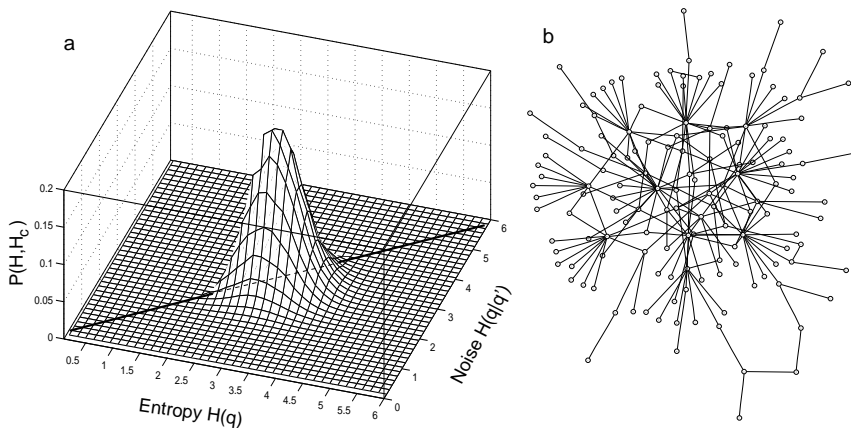

FIG. 9 Noise-entropy plot obtained by exploring the Γ space using a Monte Carlo sampling. Different pairs of noise and entropy are generated and a simulated annealing search is performed looking for candidate networks. Here the smoothed probability distribution obtained from this algorithm is shown in (a). The highest density of observed networks appears to be close to the same domain observed for real networks. In (b) an example of a small sized network ($N = 142, < k >= 2.64$).is shown, together with its degree distribution (c). The cumulative degree distribution follows power law with exponent -1.26 (i. e. $\gamma = 2.26$). The graph has been obtained close to the boundary $H = H_c$ (with $H(\mathbf{q}) = 3.69, H(\mathbf{q}|\mathbf{q}') = 3.09$.

Statistical independence among node degree thus gives a predicted straight line $H_c(\mathbf{q}|\mathbf{q}') = H(\mathbf{q})$ which seems to be suggested by our data. One first conclusion from this analysis is that network correlations in real graphs are small, begin the diversity of pairs of linked nodes a direct consequence of the heterogeneous character of the degree distribution and nothing else. In spite that the cloud of points deviates from the straight line, these deviations might result from finite-size effects. Actually, if we plot information measures $I(\mathbf{q}; N)$ against system's size N, it can be shown that they follow a scaling $I(\mathbf{q}; N) \sim N^{-1}$.

TABLE I Information-based measures computed for different real and theoretical systems. For each subset the list is ordered from higher to lower information transfer.

Network type	N	$<k>$	$I(\mathbf{q})$	$H(\mathbf{q})$	$H_c(\mathbf{q}\|\mathbf{q}')$	r
Technological networks						
Software 1	168	2.81	1.19	3.04	1.85	-0.39
Software 2	159	4.19	1.03	3.99	2.97	-0.41
Circuit AS97a	3200	3.56	0.50	4.77	4.27	-0.22
Software 3	1993	5.00	0.30	4.82	4.51	-0.08
Circuit TV	320	3.17	0.23	1.37	1.14	0.010
Circuit EC05	899	4.14	0.15	2.98	2.82	-0.15
Software linux	5285	4.29	0.12	4.47	4.35	-0.06
Powergrid	4941	2.67	0.06	3.01	2.95	0.003
Biological networks						
Silwood park	154	4.75	0.94	4.09	3.14	-0.31
Ythan estuary	134	8.67	0.53	4.74	4.21	-0.24
p53 subnetwork	139	5.09	0.46	4.00	3.54	-0.24
Metabolic map	1173	4.84	0.39	3.58	3.19	-0.17
Neural net (C.elegans)	297	14.5	0.37	5.12	4.74	-0.16
Metabolic map	821	4.76	0.37	3.46	3.09	-0.18
Romanian syntax	5916	5.65	0.31	5.45	5.14	-0.18
Proteome map	1458	2.67	0.24	3.85	3.61	-0.21
Theoretical systems						
Star graph	17	1.88	1.00	1.00	0.00	-1.00
Barabási-Albert	3000	3.98	0.25	4.12	3.85	-0.078
Erdös-Renyi	300	6.82	0.06	3.31	3.25	-0.005
modular E-R	500	10.3	0.04	3.67	3.62	-0.001

Two points clearly deviate from the general pattern displayed by the majority of networks analysed here. Both are small systems and correspond to software graphs, and one of them is shown in figure 8b. As we can see this is a rather peculiar system, involving a large hub connected to a small module. It is thus a small structure dominated by the star graph component together with a homogeneous component. Such a nonuniform structure is likely to result from an process dealing with a small structure but unlikely to result from a natural process or from artifical evolution when some complexity thresholds are reached (Valverde, Ferrer and Solé, 2002).

One possible explanation for the previous result is that correlations simply do not play any particular role in shaping network architecture. However, it could be also argued that such a lack of correlation has been either chosen or selected for some underlying reason. But there's also another (more likely) scenario: that the observed stuctures are actually the only possible choices, at least when some complexity threshold is reached.

VI. SIMULATED ANNEALING SEARCH

The spread of real networks close to the zero-information boundary suggests that the possible structures allowed to occur (with a given heterogeneity and a given correlation) is rather constrained. This might be a consequence of the irrelevance of correlations for these systems but it

would also be the case that some selective pressure is made towards heterogeneous networks with small correlations (i. e. no assortativeness).

In order to test the previous idea we can perform a Monte Carlo search in network space. Specifically, we explore the space of possible pairs entropy-noise available to candidate graphs Ω, i. e. $\Gamma = \{H(\mathbf{q}), H_c(\mathbf{q}|\mathbf{q}')\}$, which is constrained by two well-defined boundaries:

$$\partial_1\Gamma = \{(H(\Omega), H_c(\Omega)) \mid H_c(\Omega) = 0\} \tag{26}$$

$$\partial_2\Gamma = \{(H(\Omega), H_c(\Omega)) \mid H(\Omega) = H_c(\Omega)\} \tag{27}$$

where $H(\Omega)$ and $H_c(\Omega)$ indicate the entropy and noise associated to a given graph Ω. It is not difficult to show that only two points occupy the lower boundary, i. e. $\partial_1\Gamma = \{(0,0), (\log 2, 0)\}$. These correspond to purely homogeneous graphs and the star graph. The second boundary has already been studied.

For every random sample point (H, H_c), an optimizing searching process looks for candidate networks that minimize the error term or potential function $U(\Omega)$:

$$U(\Omega) = \sqrt{(H - H(\Omega))^2 + (H_c - H_c(\Omega))^2} \tag{28}$$

Here, we use the Boltzmann strategy presented in (Schweitzer et al. 1996, Schweitzer, 2002). The algorithm explores the search space defined by all possible networks of N nodes. We assume that every possible state visited by the search process can be properly characterized by the scalar U_i. In the stationary limit (for a large number of searchers) define the occupation probability $p_i(t)$ of certain state i at time t. We require the optimization process to increase the occupation probability for the state of minimal potential. In general, many local minima exist and the search could be trapped in one of these states, which is undesirable.

A dynamics that finds the minimum is given by:

$$\frac{dp_i(t)}{dt} = \sum_{i \neq j} A_{ij}p_i(t) - A_{ji}p_j(t) \tag{29}$$

where

$$A_{ij} = A_{ij}^0 * \begin{cases} 1 & ; U_i < U_j \\ \exp\left(-(U_i - U_j)/T(t)\right) & ; U_i \geq U_j \end{cases} \tag{30}$$

is the transition probability for the searcher to move from state i to state j. The term A_{ij}^0 is 1 if and only if the state j can be reached by a little change or mutation and 0 otherwise. Here, the valid changes involve edge addition, edge removal and edge rewire, which are all equally selected with the same probability. The number of nodes of the network is allways fixed. Transitions to lower energy states are allways accepted but local minima is avoided because termal fluctuations like in simulated annealing . As the search progresses, the temperature $T(t)$ is decreased following a power law rule:

$$T(t) = \frac{T_0}{1 + at} \tag{31}$$

where T_0 is the initial temperature (or starting degree of disorder) and a is the cooling rate. This allows the optimization process to perform a smooth transition from coarse to detailed search. The process starts from a random graph of N nodes with a given connectivity $< k >$ and lasts a given number of simulation steps.

By measuring the final error $\varepsilon(\Omega) = U(\Omega)$ for a large number of Montecarlo samples it is possible to approximate the likely of a particular candidate network. Here we have used $\epsilon = 0.003$ and the optimization parameters are: $N = 500, < k >= 3, 350000$ steps, $T0 = 0.01, a = 0.002$. Our results indicate that the potential is effectively minimized only for a domain of pairs (H, H_c) along the second boundary $\partial_2\Gamma$. In figure 10a the (smoothed) probability density $\mathcal{P}(H, H_c)$ of optimized networks is shown (for the upper part of the parameter space, $\mathcal{P}(H, H_c) = 0$). The distribution is peaked around a domain of Γ that fits very well the range of values satisfied by most real networks
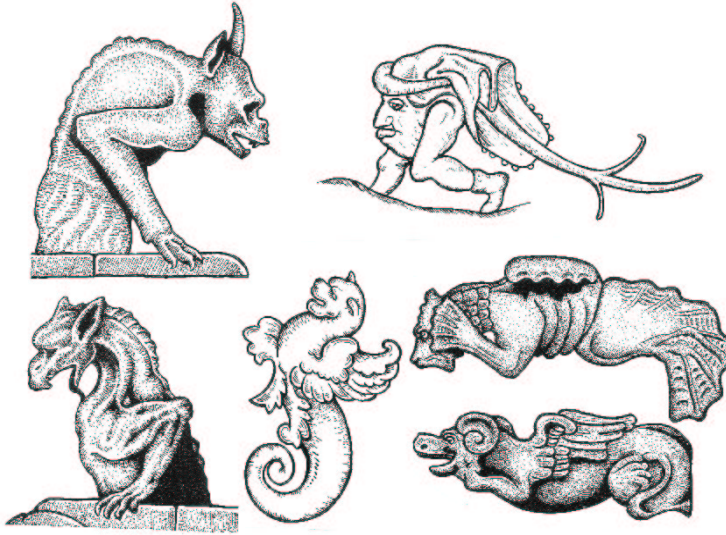
FIG. 10 The logic of monsters. Mythological creatures, gargoles and other imaginary creatures define a parallel universe of structures that are often mixtures of real creatures. Although possible in principle, they are not observed in nature.

(compare with figure 9). There is also a clearly empty zone outside this domain, indicating that networks are difficult or simply impossible to find. An example of the optimized graphs is shown in figure 10b. This particular graph is scale-free, with an exponent $\gamma \approx 2.26$ and a cut-off at $\xi \sim 50$. The other networks in this domain are also SF, with an average scaling exponent $< \gamma > \sim 2.5$. By searching candidate networks that simultaneously fit the two requirements of given entropy and noise, the only possible solutions to be found are scale-free graphs with small levels of correlations. Interestingly, software networks deviate from this rule and are to be found along the upper region of the boundary ($H > 4$), where potential is not minimum. This might be a signature of frustated optimization in software design processes (Valverde, Ferrer and Solé, 2002).

VII. DISCUSSION

Complex networks display heterogeneous structures that result from different mechanisms of evolution (Solé et al., 2002). Some are created through multiplicative processes while others seem to be well described in terms of optimization mechanisms. Our study indicates that the possible universe of complex networks is actually rather constrained. Networks display scale-free architecture but also small assortativeness. The search algorithm, instead of assuming the presence of a given predefined mechanism of network growth, simply searches for candidate solutions to an optimization algoritm trying to approach simultaneously some amount of network heterogeneity and correlations. The result is that indeed networks are scale-free and involve low degree of correlations, but such situation is constrained to a well-defined domain. This domain is remarkably similar to the one inhabited by real graphs. Outside this domain, it is not feasible to find graphs simultaneously satisfying the two requirements.

The impact of SF architecture on biological and artificial networks is clearly different. Although the first can take advantage of the high homeostasis provided by scaling laws the second are completely dependent on the correct functioning of all units. Failure of a single diode in a circuit or of a single component in a software system leads to system collapse. Thus, homeostasis can not be a general explanation for scaling. We have conjectured that the leading force here is an optimization process where reliable communication at low cost shapes network architecture in first place (Solé et al., 2002). The need of a sparse graph can be a consequence of different requirements. In an electronic circuits, saving wire is a strong constraint. In metabolic or genetic networks, it might be important in order to reduce the impact of unstable positive feedbacks. This can be satisfied by means of sparse graphs displaying scale-free architecture. What is the role of correlations? For

the systems analysed here correlations don't seem to be of relevance to network performance. But what is more important: the lack of networks outside the densely populated domain is not due to some relevant, perhaps adaptive trait. It is actually a consequence of higher-level limitations imposed to network architecture.

Such a constrained set of possibilities fits very well the view of evolution as strongly dominated by intrinsic constraints (Jacob, 1976; Alberch, 1989; Kauffman, 1993; Goodwin, 1994; see also Gould, 2003 for a critical discussion). Under this view, the outcome of evolutionary searches would be not *any* possible architecture from the set of possible patterns but a choice from a narrow subset of attainable structures. In this context, in spite of the contingencies intrinsic to evolutionary dynamics and history, the same basic repertoire of architectural motifs would be observable if the tape of evolution were rewound and played again (and this includes the evolution of technology). Although monsters are in principle possible (figure 10) they are unlikely to occur (the software graph shown in figure 8b would be an example). The surprising convergence of complex networks towards heterogeneous, scale-free graphs might become a good example of the role of structural constraints to evolutionary dynamics.

acknowledgments

VIII. REFERENCES

1. Albert, R. and Barabási, A.-L. Statistical Mechanics of Complex Networks. Rev. Mod. Phys. **74**, 47-97 (2002).

2. Dorogovtsev, S.N. and Mendes, J.F.F. Evolution of networks. *Adv. Phys.* **51**, 1079-1187 (2002).

3. Dorogovtsev, S.N. and Mendes, J.F.F. *Evolution of Networks: from biological nets to the Internet and WWW* Oxford U. Press, Oxford (2003).

4. Bornholdt, S. and Schuster, H. G.,eds. *Handbook of Graphs and Networks: From the Genome to the Internet.* Springer, Berlin (2002).

5. Amaral, L. A. N., Scala, A., Barthélemy, M. and Stanley, H. E. Classes of behavior of small-world networks. Proc. Nat. Acad. Sci. USA **97**, 11149-11152 (2000).

6. Ferrer, R., Janssen, C. and Solé, R. V. Topology of Technology Graphs: Small World Patterns in Electronic Circuits Physical Review E 64, 32767 (2001).

7. Barabási, A.-L. and Albert, R. Emergence of scaling in random networks. Science **286**, 509-512 (1999).

8. Dorogovtsev, S. N. and Mendes J.F.F. Accelerated growth of networks, in: *Handbook of Graphs and Networks: From the Genome to the Internet*, eds. S. Bornholdt and H.G. Schuster. pp. 320-343 Wiley-VCH, Berlin (2002).

9. Ferrer, R. and Solé, R. V. Optimization in Complex Networks In: *Statistical Physics of Complex Networks*, Lecture Notes in Physics, Springer (Berlin) to appear.

10. Valverde, S., Ferrer, R. and Solé, R. V. Scale free networks from optimal design Europhys. Lett. **60**, 512-517 (2002).

11. Solé, R.V., Pastor-Satorras, R., Smith, E.D. and Kepler, T. A model of large-scale proteome evolution. *Adv. Complex Systems* **5**, 43-54 (2002).

12. Vazquez, A., Flammini, A., Maritan, A. and Vespignani, A. Modeling of protein interaction networks. *Complexus*, **1**, 38-44 (2002).

13. Caldarelli, G., Capocci, A., De Los Rios, P. and Muñoz, M. A. Scale-free networks from varying vertex intrinsic fitness. Phys. Rev. Lett. **89**, 258702 (2002).

14. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. and Barabási, A.-L. Hierarchical Organization of Modularity in Metabolic Networks. Science 297, 1551-1555 (2002).

15. Newman, M. E. J. Assortative mixing in networks. Phys. Rev. Lett. **89**, 208701 (2002).

16. Stanley, E. H. et al., Statistical mechanics in biology: how ubiquitous are long-range correlations? Physica A**205**, 214-253 (1996).

17. Abarbanel H., D., I., Brown R., Sidorowich J., L. and Tsimring L., S. The analysis of observed chaotic data in physical systems. Rev. Mod. Phys. **65**, 1331-1392 (1993).

18. Ash, R. B. *Information Theory*, Dover, London (1965).

19. Adami, C. *Introduction to Artificial Life.* Springer, New York (1998).

20. Li, W. Mutual onformation versus correlation functions. J. Stat. Phys. **60**, 823-837 (1990).

21. Li, W. On th erelationship between complexity and entropy for Markov chains and regular languages. Complex Syst. **5**, 381-399 (1991).

22. Fraser, A. and Swinney, H. Independent coordinates for strange attractors from mutual information. Phys. Rev. A**33**, 1134-1140 (1986).

23. Germain-Renaud, C. and Sansonnet, J. P. *Ordinateurs massivement paralleles*, Armand Colin, Paris (1991).

24. Milutinovic, V. M. *Computer Architecture*, North Holland, Elsevier (1988).

25. Hillis, W. D. *The Connection Machine*, MIT Press (Cambridge, MA, 1985).

26. Kauffman, S. A. *Origins of Order.* Oxford U. Press, New York (1993).

27. May, R.M.*Stability and complexity in model ecosystems.* Princeton U. Press, New York (1973).

28. Amari, S. Characteristics of random nets of analog neuron-like elements. IEEE Trans. Man and Cybernetics **2**, 643-657 (1972).

29. Montoya, J. M. and Solé, R. V. Topological properties of food webs: from real data to community assembly models Oikos 102, 614-622 (2003).

30. Jeong, H., Mason, S., Barabási, A. L. and Oltvai, Z. N. (2001) Lethality and centrality in protein networks Nature **411**, 41 (2001).

31. Valverde, S. and Solé, R. V. Hierarchical small worlds in sotfware architecture. Santa Fe Institute Working Paper 03-07-044

32. Myers, C. R.Software systems as complex networks: structure, function, and evolvability of software collaboration graphs", cond-mat/0305575 (2003).

33. Schweitzer, F., Ebeling, W., Rose, H. and Weiss, O. Network Optimization Using Evolutionary Strategies, in: *Parallel Problem Solving from Nature - PPSN IV*, (Eds. H.-M. Voigt, W. Ebeling, I. Rechenberg, H.-P. Schwefel), Lecture Notes in Computer Science, vol. 1141, Springer, Berlin (1996) pp. 940-949

34. Schweitzer, F. *Brownian Agents and Active Particles.* Springer, Berlin (2002).

35. Solé, R. V., Ferrer-Cancho, R., Montoya, J, M. and Valverde, S. Selection, tinkering and emergence in complex networks. Complexity **8**(1), 20-33.

36. Jacob, F. Evolution as tinkering. Science **196**, 1161-1166 (1976) .

37. Alberch, P. The logic of monsters: evidence for internal constraint in development and evolution. Geobios **19**, 21-57 (1989).

38. Goodwin, B.C. *How the Leopard Changed Its Spots: the Evolution of Complexity. .* Charles Scribner's Sons, New York (1994) .

39. Gould, S. J. *The structure of evolutionary theory.* Belknap, Harvard (2003)