

Information Theory of Complex Networks: On Evolution and Architectural Constraints

Ricard V. Solé and Sergi Valverde

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

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

Abstract. 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.

1 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 [1-4].

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 [5]. An example of such scale-free networks is provided by the architecture of digital electronic circuits (Fig. 1). It has been shown [6] 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 Fig. 2 several examples of the observed distributions for both analog (a-b) and digital (c-d) systems are shown. Although analog systems are closer to an exponential distribution (i. e. ξ small) digital, large-scale systems³ exhibit scaling behavior, with $\gamma \sim 3$.

³ Similar results have been obtained by looking at large-scale systems, such as VLSI networks. See for example: <http://citeseer.nj.nec.com/450707.html>

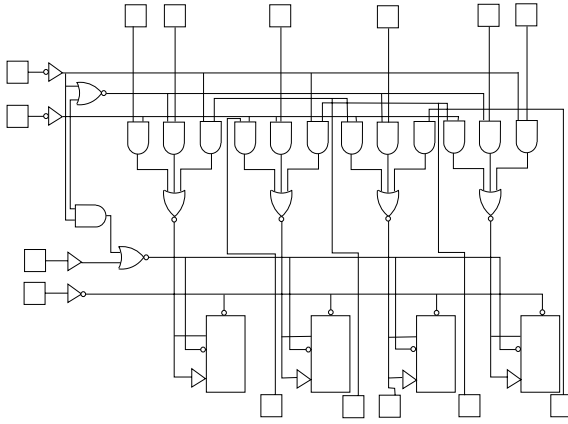


Fig. 1. Heterogeneity is a widespread feature of most (but not all) complex networks. An example from technology graphs are electronic circuits (upper plot) which have been shown to display scale-free distributions of links.

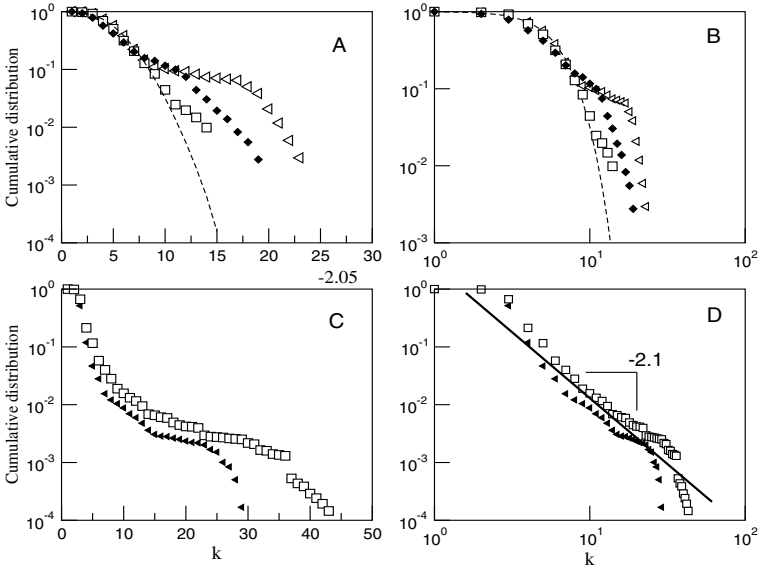


Fig. 2. Cumulative degree distributions for several examples of analogic (a-b) and digital (c-d). Although the analogic systems are less heterogeneous, analogic circuits (particularly large systems) display scaling in their degree distributions (Left plots are linear-log and right plots are in log-log scale).

Scale-free nets have been shown to be obtainable through a number of mechanisms, including preferential attachment [2,3,7,8], optimization [9,10], duplication and divergence [11,12] or fitness-dependent, rich-gets-richer mechanisms [13] or the “copying” model [14]. 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. Moreover, the response to node removal differs 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 typically 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 and, to a similar extent, by software networks. Failure of any component typically leads to 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 $W = \{s_i\}, (i = 1, \dots, N)$ 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 (Fig. 3), although not arbitrarily diverse, displays a number of structural variations that cannot be compressed by the previous average quantities. Real networks are not only typically heterogeneous, but they also involve other types of features, such as hierarchical organization [15].

In Fig. 3 we qualitatively summarize the basic types of network organization by using a generic, qualitative 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.

2 Measuring Correlations

Beyond the degree distribution and average statistical measures, correlation measures offer considerable insight into the structural properties displayed by com-

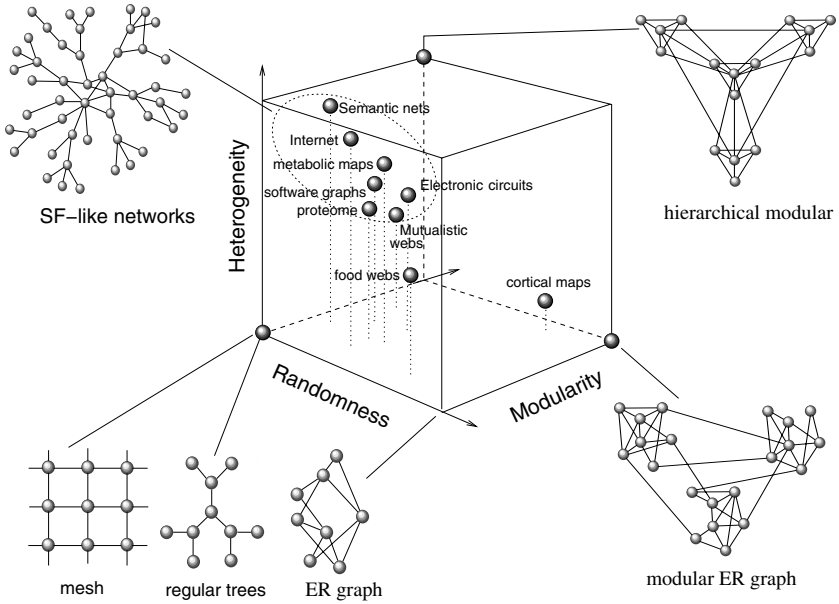


Fig. 3. A zoo of complex networks. In this qualitative space, three relevant characteristics are included: randomness, heterogeneity and modularity. The first introduces the amount of randomness involved in the process of network’s building. The second measures how diverse is the link distribution and 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.

plex networks. One particularly interesting is network assortativeness [16]. 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 *disassortative 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 [16] we will be interested here not in the degree distribution P_k but instead in the *remaining degree*: the number of edges leaving the vertex other than the one we arrived along (Fig. 4). This new distribution $q(k)$ is obtained from:

$$q(k) = \frac{(k + 1)P_{k+1}}{\langle k \rangle} \tag{1}$$

where $\langle k \rangle = \sum_k kP_k$. 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

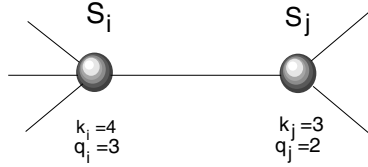


Fig. 4. Computing correlations in a network. Here two given, connected nodes s_i, s_j are shown, displaying different degrees k_i, k_j . Since we are interested in the remaining degrees, a different value needs to be considered (here indicated as q_i, q_j).

connected degree-degree correlation function

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

where $\langle \dots \rangle$ 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 obtained by dividing it with 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 $\sigma_q^2 = \sum_k k^2q(k) - [\sum_k kq(k)]^2$ of the distribution $q(k)$, and hence the normalized correlation function is

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

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

Correlation functions have been widely used both in statistical physics [17] and nonlinear dynamics [18]. A closely related, and more general approach involves the use of information-based measures [19-21]. One specially important quantity is the so called *mutual information*, which is a general measure of dependence between two variables [19,22]. 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 allows to determine the independent variables for (re-) constructing phase trajectories [23]. 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.

3 Entropy and Information

By using the previous distribution $\mathbf{q} = (q(1), \dots, q(i), \dots, q(N))$, 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 [19]. 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 $q(i) = 1/N (\forall i = 1, \dots, N)$ and $H_{\min}(\mathbf{q}) = 0$ which occurs when $\mathbf{q} = (1, 0, \dots, 0)$. 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 Fig. 5 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 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:

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

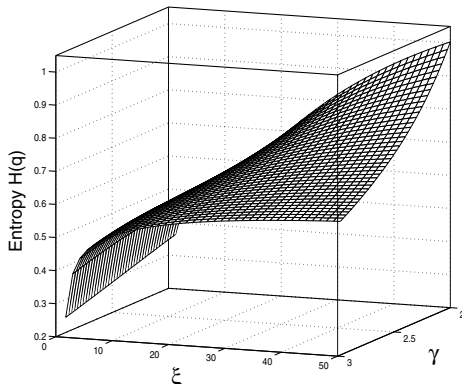


Fig. 5. Entropy of the remaining degree distribution obtained from a network with degree distribution $P_k \sim k^{-\gamma} \phi(k/\xi)$. Here $H(\mathbf{q}) = - \int P_k \log P_k dk$ is shown against the scaling exponent γ and the cut off ξ . Here we have used an exponential cut-off, i. e. $\phi(k/\xi)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

Here $q_c(k, k')$ is the joint probability, and it is normalized, i. e.:

$$\sum_{k=1}^N \sum_{k'=1}^N q_c(k, k') = 1 \quad (6)$$

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

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

$$I(\mathbf{q}) = H(\mathbf{q}) - H_c(\mathbf{q}|\mathbf{q}') \quad (7)$$

where the last term $H_c(\mathbf{q}|\mathbf{q}')$ is the conditional entropy that involves a different set of *conditional* probabilities $\pi(k|k')$ [19]. 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 (the “noise” in our graph) is defined as:

$$H_c(\mathbf{q}|\mathbf{q}') = - \sum_{k=1}^N \sum_{k'=1}^N q(k) \pi(k|k') \log \pi(k|k') \quad (8)$$

Since the conditional and joint probabilities are related through:

$$\pi(k|k') = \frac{q_c(k, k')}{q(k')} \quad (9)$$

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

$$H_c(\mathbf{q}|\mathbf{q}') = - \sum_{k=1}^N \sum_{k'=1}^N q_c(k, k') \log \frac{q_c(k, k')}{q(k')} \quad (10)$$

we thus have, from the previous expressions,

$$\begin{aligned} I(\mathbf{q}) &= H(\mathbf{q}) - H_c(\mathbf{q}|\mathbf{q}') \quad (11) \\ &= - \sum_{k=1}^N \sum_{k'=1}^N q_c(k, k') \log q(k) + \sum_{k=1}^N \sum_{k'=1}^N q_c(k, k') \log \pi(k|k') \\ &= - \sum_{k=1}^N \sum_{k'=1}^N q_c(k, k') \log \frac{q(k)}{\pi(k|k')} \end{aligned}$$

which gives a final form for the information transfer function⁴:

⁴ The previous measures can be extended (with some care) into continuous distributions. In this case, we must assume that the continuous counterparts of the pre-

$$I(\mathbf{q}) = \sum_{k=1}^N \sum_{k'=1}^N q_c(k, k') \log \frac{q_c(k, k')}{q(k)q(k')} \quad (12)$$

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 $k, k' = 1, \dots, N$. Another is given by $\pi(k|k') = \delta_{k,k'}$. This case corresponds to a deterministic channel in standard information theory [19]. This implies that $q_c(k, k') = q(k')\delta_{k,k'}$ which is precisely the case of perfectly assortative network [16].

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 graph, as measured by the information; (b) the noise level, as defined by the conditional entropy, which will provide a measure of assortativeness and (c) the entropy of 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.

4 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.

4.1 Lattices and Trees

Lattice-like networks are common in some man-made architectures, particularly parallel computers [24–26]. 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

vious 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 $H_c(\mathbf{q}|\mathbf{q}') = - \int \int q_c(k, k') \log \pi(k|k')dkdk'$

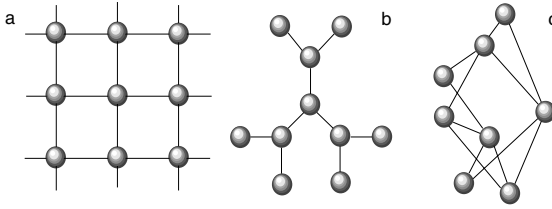


Fig. 6. Homogeneous networks: here two examples of a lattice (a) and a regular tree (b) are shown as examples of deterministic nets. In both cases each node has the same degree and thus both the entropy and the noise are zero. In (c) a random, Erdős-Renyi graph is shown. Here some amount of heterogeneity 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 (in the $N \rightarrow \infty$ limit).

$$q(k) = \delta_{k,z-1} \quad (13)$$

$$q_c(k, k') = \delta_{k,z-1} \delta_{k',z-1} \quad (14)$$

and thus

$$I(\mathbf{q}) = H(\mathbf{q}) = H_c(\mathbf{q}|\mathbf{q}') = 0 \quad (15)$$

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 [27] and communication networks.

4.2 Erdős-Renyi Graphs

Erdős-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 [28] ecological [29] and neural [30] networks. It seems also appropriate in describing the topology of species-poor ecosystems [31]. The distributions are single-scaled and thus low uncertainty and high randomness are at work. The average degree will be $\langle k \rangle \approx pN$, and it can be easily shown that the probability P_k that a vertex has a degree k follows a Poisson distribution $P_k = e^{-\langle k \rangle} \langle k \rangle^k / k!$, and thus

$$q(k) = \frac{(k+1)e^{-\langle k \rangle}}{\langle k \rangle(k+1)!} \langle k \rangle^{(k+1)} = P_k \quad (16)$$

For this random graph, the independence associated to the link assignment 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 [15]. In general, the graph Ω is partitioned into m subgraphs $\{\Omega_i\}$, ($i = 1, \dots, m$) of relative size $\eta_i = |\Omega_i|$ such that

$$W = \bigcup_{i=1}^m W_i \quad (W_j \cap W_k = \emptyset) \quad (17)$$

and such that

$$\sum_i \eta_i = 1 \quad (18)$$

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 $\langle k \rangle = pN/m + (m-1)Nq/m$. Given the random wiring, it is not difficult to show that for large N information will be typically very small.

4.3 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 [32].

This graph Ω_* is characterized by a degree distribution:

$$p(k) = \frac{n-1}{n} \delta_{k,1} + \frac{1}{n} \delta_{k,n-1} \quad (19)$$

The corresponding distribution $q(k)$ is:

$$q(k) = \frac{1}{2} [\delta_{k,0} + \delta_{k,N-1}] \quad (20)$$

and the joint probabilities are reduced to:

$$q_c(k, k') = \delta_{k,N-2} \delta_{k',0} \quad (21)$$

The entropy is maximal, given by:

$$H(\mathbf{q}) = -q(0) \log q(0) - q(N-1) \log q(N-1) \quad (22)$$

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.

5 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:

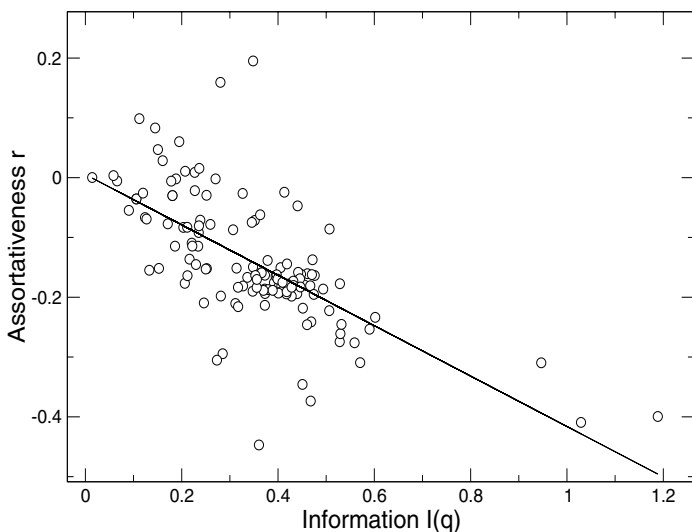


Fig. 7. Information transfer and assortativeness appear to be roughly correlated in a negative way. Here several systems have been used (all those analysed in this paper) and a linear interpolation has been displayed. Although the trend is clear, considerable variance can be appreciated, probably due to the underlying nonlinear mapping between both measures.

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 [33].
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 [34,35].
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 [6].

In Table 1 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 $\langle k \rangle$, information measures and the assortative mixing coefficient r are provided. We can see that most nets are disassortative, as predicted in [16]. Actually, information and r appear to be negatively correlated. This is shown in Fig. 7, where r is shown against $I(\mathbf{q})$ 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

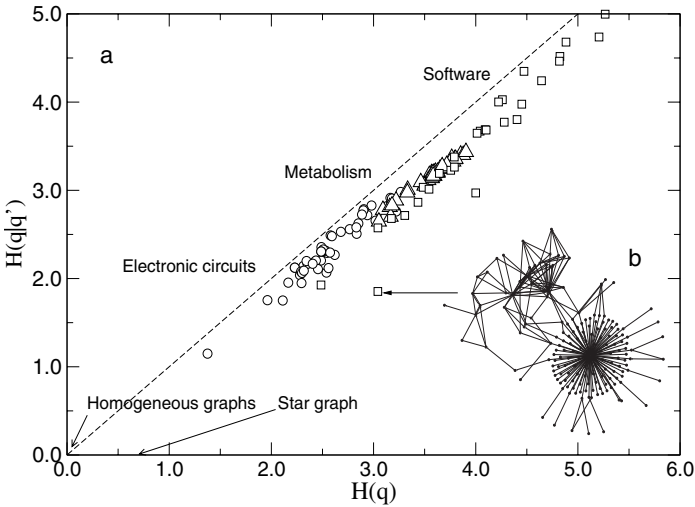


Fig. 8. Noise-entropy plot for different real networks, both natural and artificial. Here electronic circuits (open circles), metabolic (triangles) and software maps (squares) are shown to be close to the zero-information line, i. e. when entropy equals to noise. An example of a software graph that significantly 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.

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 typically uncorrelated structure. This is clear from the strongly linear dependence shown between noise and entropy (Fig. 8). If two given, randomly chosen nodes with remaining degrees k, k' are typically connected with some probability, roughly irrespective of their mutual degree (i. e. low assortativeness is present) we should expect:

$$q_c(k, k') \approx q(k)q(k') \quad (23)$$

and thus we would have

$$\pi(k|k') \approx q(k) \quad (24)$$

in this case, the noise will be given by:

$$H_c(\mathbf{q}|\mathbf{q}') = - \sum_{k=1}^N \sum_{k'=1}^N q(k, k') \log \pi(k|k') \quad (25)$$

$$= - \sum_{k=1}^N \sum_{k'=1}^N q(k, k') \log q(k) = H(\mathbf{q}) \quad (26)$$

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 conclu-

Table 1. 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	$\langle k \rangle$	$I(\mathbf{q})$	$H(\mathbf{q})$	$H_c(\mathbf{q} \mathbf{q}')$	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
Internet AS	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
Erdős-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

sion from this analysis is that network correlations in real graphs are small, being 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 $I(\mathbf{q}; N) \sim N^{-1}$.

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 Fig. 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 engineered, small-sized systems but unlikely to result from

a natural process or from artificial evolution when some complexity thresholds are reached.

One possible explanation for the previous result is that correlations simply do not play any particular role in shaping network architecture⁵. 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 structures are actually the only possible choices, at least when some complexity threshold is reached.

6 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 boundaries⁶:

$$\partial_1\Gamma = \{(H(\Omega), H_c(\Omega)) \mid H_c(\Omega) = 0\} \quad (27)$$

$$\partial_2\Gamma = \{(H(\Omega), H_c(\Omega)) \mid H(\Omega) = H_c(\Omega)\} \quad (28)$$

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 $U(\Omega)$:

$$U(\Omega) = \sqrt{(H - H(\Omega))^2 + (H_c - H_c(\Omega))^2} \quad (29)$$

Here, we use the Boltzmann strategy presented in [36,37]. 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)

⁵ This conclusion is reached under our specific, quasi-local definition of remaining degree. Other approaches, considering instead shortest paths among nodes might reveal important differences

⁶ Strictly speaking, we are considering the entropies associated to the remaining degree distribution of a graph sampled from some graph ensemble by a stochastic process

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:

$$\frac{dp_i(t)}{dt} = \sum_{i \neq j} A_{ij} p_i(t) - A_{ji} p_j(t) \quad (30)$$

where

$$A_{ij} = A_{ij}^0 * \begin{cases} 1 & ; U_i < U_j \\ \exp(-(U_i - U_j)/T(t)) & ; U_i \geq U_j \end{cases} \quad (31)$$

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 always fixed. Transitions to lower energy states are always accepted but local minima is avoided because thermal fluctuations like in simulated annealing. As the search progresses, the temperature $T(t)$ is decreased following a power law rule:

$$T(t) = \frac{T_0}{1 + at} \quad (32)$$

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 $\langle k \rangle$ and lasts a given number of simulation steps.

By measuring the final error $\varepsilon(\Omega) = U(\Omega)$ for a large number of Monte Carlo samples it is possible to approximate the likelihood of a particular candidate network. Here we have used $\epsilon = 0.003$ and the optimization parameters are: $N = 500$, $\langle k \rangle = 3$, $T_0 = 0.01$, $a = 0.002$ and 350000 steps. 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 Fig. 9a 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 (compare with Fig. 8). 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 Fig. 9b. 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 $\langle \gamma \rangle \sim 2.5$. 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

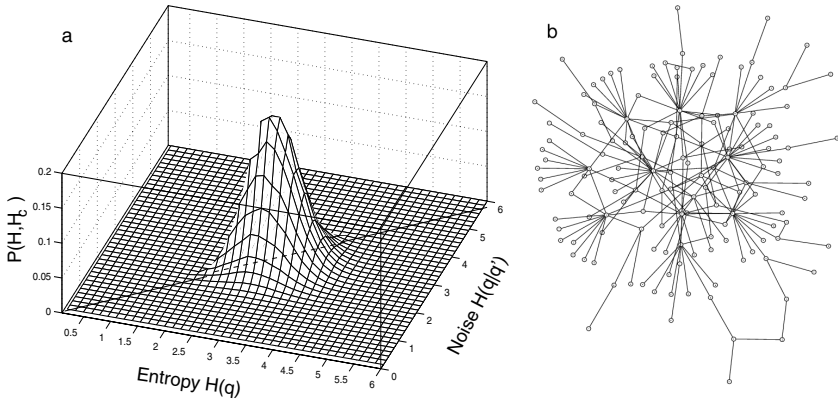


Fig. 9. (a) Noise-entropy probability plot obtained by exploring the Γ space using a Monte Carlo sampling. Different pairs of noise and entropy are generated and a simulated annealing search is performed looking for candidate networks. Here the smoothed probability distribution obtained from this algorithm 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 ($N = 142, \langle k \rangle = 2.64$) 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 $H(\mathbf{q}) = 3.69, H(\mathbf{q}|\mathbf{q}') = 3.09$).

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 frustrated optimization in software design processes [10].

7 Discussion

Complex networks display heterogeneous structures that result from different mechanisms of evolution [38]. Some are created through multiplicative processes (such as preferential attachment) while others seem to be well described in terms of optimization mechanisms [9]. 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 algorithm 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 [38]. The need of a sparse graph can be a consequence of different requirements. In an electronic circuit, 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 [39-41] (see also [42] 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.

Acknowledgments

The authors thank the members of the Complex Systems Lab for useful discussions and to an anonymous referee for valuable comments. This work was supported by a grant BFM2001-2154, FET Open Project IST DELIS and by the Santa Fe Institute.

References

1. R. Albert and A.-L. Barabási. Statistical Mechanics of Complex Networks. *Rev. Mod. Phys.* **74**, 47-97 (2002a).
2. S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Adv. Phys.* **51**, 1079-1187 (2002).
3. S. N. Dorogovtsev and J. F. F. Mendes. *Evolution of Networks: from biological nets to the Internet and WWW* Oxford U. Press, Oxford (2003).
4. S. Bornholdt and H. G. Schuster, eds. *Handbook of Graphs and Networks: From the Genome to the Internet*. Springer, Berlin (2002).
5. L. A. N. Amaral, A. Scala, M. Barthélemy and H. E. Stanley. Classes of behavior of small-world networks. *Proc. Nat. Acad. Sci. USA* **97**, 11149-11152 (2000).
6. R. Ferrer, C. Janssen and R. V. Solé. Topology of Technology Graphs: Small World Patterns in Electronic Circuits *Physical Review E* **64**, 32767 (2001).
7. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science* **286**, 509-512 (1999).

8. S. N. Dorogovtsev and J. F. F. Mendes. 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. R. Ferrer and R. V. Solé. Optimization in Complex Networks, Lect. Notes Phys. **625**, 114–125 (2003).
10. S. Valverde, R. Ferrer and R. V. Solé. Scale free networks from optimal design Europhys. Lett. **60**, 512-517 (2002).
11. R. V. Solé, R. Pastor-Satorras, R., Smith, E.D. and Kepler, T. A model of large-scale proteome evolution. *Adv. Complex Systems* **5**, 43-54 (2002).
12. A. Vazquez, A. Flammini, A. Maritan and A. Vespignani. Modeling of protein interaction networks. *Complexus*, **1**, 38-44 (2002).
13. G. Caldarelli, A. Capocci, P. De Los Rios and M. A. Muñoz. Scale-free networks from varying vertex intrinsic fitness. Phys. Rev. Lett. **89**, 258702 (2002).
14. F. Menczer. Growing and navigating the small world web by local content. Proc. Nat. Acad. Sci. USA **99**, 14014-14019 (2002).
15. E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai and A.-L. Barabási. Hierarchical Organization of Modularity in Metabolic Networks. Science 297, 1551-1555 (2002).
16. M. E. J. Newman. Assortative mixing in networks. Phys. Rev. Lett. **89**, 208701 (2002).
17. E. H. Stanley, S. V. Buldyrev, A. L. Goldberger, Z. D. Goldberger, S. Havlin, R. N. Mantegna, S. M. Ossadnik, C. K. Peng and M. Simon, Statistical mechanics in biology: how ubiquitous are long-range correlations? Physica A **205**, 214-253 (1996).
18. H. D. Abarbanel, R. Brown, J. L. Sidorowich and L. S. Tsimring. The analysis of observed chaotic data in physical systems. Rev. Mod. Phys. **65**, 1331-1392 (1993).
19. R. B. Ash. *Information Theory*, Dover, London (1965).
20. C. Adami. *Introduction to Artificial Life*. Springer, New York (1998).
21. W. Li. Mutual information versus correlation functions. J. Stat. Phys. **60**, 823-837 (1990).
22. W. Li. On the relationship between complexity and entropy for Markov chains and regular languages. Complex Syst. **5**, 381-399 (1991).
23. A. Fraser and H. Swinney. Independent coordinates for strange attractors from mutual information. Phys. Rev. **A33**, 1134-1140 (1986).
24. C. Germain-Renaud and J. P. Sansonnet. *Ordinateurs massivement paralleles*, Armand Colin, Paris (1991).
25. V. M. Milutinovic. *Computer Architecture*, North Holland, Elsevier (1988).
26. W. D. Hillis. *The Connection Machine*, MIT Press (Cambridge, MA, 1985).
27. S. Valverde, R. Ferrer and R. V. Solé, Scale-free networks from optimal design. Europhys. Lett. **60**, 512-517 (2002).
28. S. A. Kauffman. *Origins of Order*. Oxford U. Press, New York (1993).
29. R. M. May. *Stability and complexity in model ecosystems*. Princeton U. Press, New York (1973).
30. S. Amari. Characteristics of random nets of analog neuron-like elements. IEEE Trans. Man and Cybernetics **2**, 643-657 (1972).
31. J. M. Montoya and R. V. Solé. Topological properties of food webs: from real data to community assembly models Oikos **102**, 614-622 (2003).
32. R. Ferrer and R. V. Solé, Optimization in Complex Networks, Lect. Notes Phys. **625**, 114–125 (2003).
33. H. Jeong, S. Mason, A.-L. Barabási and Z. N. Oltvai. Lethality and centrality in protein networks Nature **411**, 41 (2001).

34. S. Valverde and R. V. Solé. Hierarchical small worlds in software architecture. Santa Fe Institute Working Paper 03-07-044.
35. C. R. Myers. Software systems as complex networks: structure, function, and evolvability of software collaboration graphs, *Phys. Rev. E* **68**, 046116 (2003).
36. F. Schweitzer, W. Ebeling, H. Rose and O. Weiss. 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.
37. F. Schweitzer. *Brownian Agents and Active Particles*. Springer, Berlin (2002).
38. R. V. Solé, R. Ferrer-Cancho, J. M. Montoya and S. Valverde. Selection, tinkering and emergence in complex networks. *Complexity* **8**(1), 20-33 (2002).
39. F. Jacob. Evolution as tinkering. *Science* **196**, 1161-1166 (1976).
40. P. Alberch. The logic of monsters: evidence for internal constraint in development and evolution. *Geobios* **19**, 21-57 (1989).
41. B. C. Goodwin. *How the Leopard Changed Its Spots: the Evolution of Complexity*. Charles Scribner's Sons, New York (1994).
42. S. J. Gould. *The structure of evolutionary theory*. Belknap, Harvard (2003).