

# Selection, Tinkering and Emergence in Complex Networks

Ricard V. Solé<sup>1,2</sup>, Ramon Ferrer i Cancho<sup>1</sup>, Jose M. Montoya<sup>1,3</sup>, and Sergi Valverde<sup>2</sup>

<sup>1</sup> *ICREA-Complex Systems Lab, IMIM-UPF, Dr Aiguader 08003 Barcelona, Spain*

<sup>2</sup> *Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA*

<sup>3</sup> *Department of Ecology, University of Alcalá, 28871 Alcalá de Henares, Madrid, Spain*

Complex biological networks have a very different origin from technologic ones. The second involve a large amount of design and, as engineered structures, include a large amount of optimization. The former involve (in principle) contingency and structural constraints, with new structures being incorporated through tinkering with previously evolved modules or units. However, the observation of the topological features of different biological nets suggests that nature can have a limited repertoire of “attractors” that essentially optimize communication under some basic constraints of cost and architecture or that allow to reach a high degree of homeostasis. Conversely, the topological features exhibited by some technology graphs indicate that tinkering and internal constraints play a key role, in spite of the “designed” nature of these structures. Previous scenarios suggested to explain the overall trends of evolution are re-analyzed under the light of topological patterns.

## I. INTRODUCTION

The emergence of the telegraph marked the appearance of totally new social and economic exchanges. As a technological innovation, it defined a new scenario of communication and information processing within human societies that lead to the creation of new, previously inexistent structures (Standage, 1999). The resulting telegraph network rapidly increased in size (after a delay in its acceptance as a real, useful innovation) and at its climax involved a whole network with millions of users, fully developed codes, encrypted messages and code crackers, chats and congestion problems.

In many ways, the telegraph network was very similar to Internet. At some point, the emergence of a new innovation (such as the telephone, figure 1) triggered the fall of the rich telegraph network. A whole culture of human communication and the networks that covered most urban centers around the globe vanished in a few decades. The telegraph went extinct, as many species and innovations through biological evolution, once a new, highly competitive novelty emerges. However, as it happens with most extinct life forms, the underlying innovations introduced by the telegraph still persist in modern communication networks.

The previous example is interesting for two reasons. First, because there is a common pattern between different types of communication networks that suggest common principles of organization. Second, because it also illustrates the presence of underlying, subtle connections between technologic and biologic evolution. This observation is not new: different sources of evidence and theoretical arguments indicate that technologic innovation shares some basic traits with the patterns displayed by biological novelty (Kauffman, 1993). The rise and fall of technological creations also resembles the origination and extinction patterns observable in some groups of organisms and Jacques Monod actually suggested that the evolution of technology is sometimes closer to Darwinian

selection than biology itself (Monod, 1970). But are the patterns of organization displayed by complex biosystems deeply related to those displayed by technological structures?

The previous question was raised by a number of authors within the context of evolution. As discussed by Francois Jacob in his influential 1977 paper “Evolution as Tinkering”, one important source of divergence between engineering (technology) and evolution is that the engineer works according to a preconceived plan (in that he foresees the product of his efforts) and second that in order to build a new system a completely new design and units can be used without to resort to previous designs (Jacob, 1977).

Jacob also mentions the point that the engineer will tend to approach the highest level of perfection (measured in some way) compatible with the technology available. Evolution, is argued, is far from perfection, a point already made by Darwin in the *Origin of species* (Darwin, 1859). Jacob’s conclusion is that natural selection does not work as an engineer, but as a tinkerer, who knows what he is going to produce but is limited by the constraints present at all levels of biological organization as well as by historical circumstances (Jacob, 1977).

Although the presence of historical contingencies plays certainly a role in evolution (Gould, 1989, Conway Morris, 1998, Gould, 2002) recent studies on fractal transport networks in biology seem to provide strong support for the presence of effective optimization processes (Brown and West, 2000). Specifically, when looking at the general principles of biological scaling, the assumption of a minimization of the energy required to transport materials through the network (assuming that it has hierarchical, space-filling structure) leads to a remarkable agreement with the diversity of biological structures and functions observed in nature through many orders of magnitude in size. Optimization would then be able to operate in a successful manner at least when the constraints are easily avoided due to the flexibility allowed by the un-

derlying rules of network construction. When looking at some artificial networks, such as the local telephone network displayed in figure 1, we can often appreciate a hierarchical organization in the tree-shaped structure. However, once some complexity threshold is reached, the final structure turns to strongly deviate from a tree structure.

A different view of evolution implies the existence of constraints derived from the fundamental limitations exhibited by dynamical systems (Alberch, 1982; Kauffman, 1993; Goodwin, 1994). When looking at the macroscopic level (such as the organism level) strong regularities are perceived that indicate the presence of a limited (though diverse and tunable) range of basic structural plans of organization. Under this view, in spite of the historical contingency intrinsic to the evolutionary process, life forms would be nevertheless predictable, at least to some extent. In this context, it has been suggested that *emergent phenomena* might play a leading role in shaping biological evolution. An example that might illustrate this idea is provided by the presence of phase transitions in random graphs (Erdős and Renyi, 1959; Bollobas, 1985; Kauffman, 1993; Newman, 2002).

Let us consider a graph  $\Omega_{n,p}$  that consists of  $n$  nodes (or vertices) joined by links (or edges) with some probability  $p$ . Specifically, each possible edge between two given nodes occurs with a probability  $p$ . The average number of links (also called the average *degree*) of a given node will be  $z \approx np$ , and it can be easily shown that the probability  $p(k)$  that a vertex has a degree  $k$  follows a Poisson distribution. This so called Erdős-Renyi (ER) random graph will be fairly well characterized by an average degree  $z$  (where the distribution  $p(k)$  shows a peak).

This model displays a phase transition at a given critical average degree  $z_c = 1$  (figure 2). At this critical point, a *giant component* forms (Bollobas, 1985; Newman, 2002): for  $z > z_c$  a large fraction of nodes are connected in a web, whereas for  $z < z_c$  the system is fragmented into small subwebs. This type of random model has been used in different contexts, including ecological (May, 1976; Cohen 1978; Solé *et al.*, 2002), genetic and metabolic (Kauffman, 1962; Kauffman, 1993) and neural (Hertz *et al.*, 1991; Sompolinsky *et al.*, 1988) networks. The importance of this phenomenon is obvious in terms of the collective properties that arise at the critical point: communication among the whole system becomes available (thus information can flow from the units to the whole system and back). Besides, the transition occurs suddenly and implies an innovation. No less important, it takes place at a low cost in terms of the number of required links. Since the only requirement in order to reach whole communication is to have a given (small) number of links per node, once the threshold is reached, order can emerge *for free* (Kauffman, 1993).

A new theoretical framework, provided by the study of

complex networks, might help to answer some key questions raised by the previous views. In particular, the discovery that both natural and artificial systems display a high heterogeneity in their wiring maps challenges the early approaches based on purely random graphs (May, 1976; Kauffman, 1962) and provide a new picture of how complexity (defined in terms of the interactions among system's parts) might emerge. In this paper the different features exhibited by four types of natural and artificial networks are reviewed, after a brief account of the basic quantitative characterizations that allow to measure network complexity. Some key questions that will be explored are:

1. What mechanisms have originated observed topological regularities in complex networks?
2. To what extent does optimization shape network topology?
3. What is the origin of homeostasis in complex networks?
4. Is homeostasis a driving force or a side-effect in network topology?
5. Is tinkering an inevitable component of network evolution?
6. Are engineered systems free of tinkering?

Comparison between the mechanisms that drive the building process of different graphs reveals that optimization might be a driving force, canalized in biological systems by both tinkering and the presence of conflicting constraints common to any hard multidimensional optimization process. Conversely, the presence of global features in technology graphs that closely resemble those observed in biological webs indicates that, in spite of the engineered design that should lead to hierarchical structures (such as the one shown in figure 1) the tinkerer seems to be at work.

## II. MEASURING NETWORK COMPLEXITY

Since we are interested in comparing the global features of both biological and artificial (engineered) networks, we need to consider a number of quantitative measures in order to characterize them properly. In order to do so, the network structure is represented by a graph  $\Omega$ , as before. Some of these measures (minimal distance, clustering coefficient) are usually applied to topological (i. e. static) descriptors of the graph structure, but others (entropy, redundancy, degeneracy) also apply to states that average dynamic variables. Most of these measures are unable to explicitly capture a functional organization and thus need a complementary knowledge

of the underlying system. Although an engineered system might look similar to a given biological network, the second usually exhibits a high tolerance to the failure of single units through different homeostatic mechanisms that is seldom displayed by the former.

**Small world patterns.** Recent research on a number biological, social and technological graphs revealed that they share a common feature: the so called *small world* (SW) property (Watts & Strogatz, 1998; Newman, 2000). Small world graphs have a number of surprising features that make them specially relevant to understand how interactions among individuals, metabolites or species lead to the robustness and homeostasis observed in nature. The SW pattern can be detected from the analysis of the *path length*  $d$ , defined as the average minimum distance between any pair of nodes. For ER graphs, we have very short distances. Specifically, it can be shown that  $d_{ER} \approx \log(n)/\log(z)$ . Graphs where  $d \approx d_{ER}$  are said to be 'small-world'. A SW can be obtained from a regular lattice (where nodes are linked to  $z$  nearest neighbors) if a small fraction of nodes are rewired to randomly chosen nodes. Thus a small degree of disorder generates short paths (as in the random case) but retaining the regular pattern (Watts & Strogatz, 1998). Small world networks propagate information very efficiently.

The so called *clustering coefficient*  $C$  measures the probability that two neighbors of a given node are also neighbors of one another. For an Erdős-Renyi graph (ER),  $C_{ER} \approx z/n$  and is thus a very small quantity. Watts & Strogatz (1998) noticed that  $C \gg C_{random}$  when looking at real networks. The small-world phenomenon is often defined as  $d \approx d_{ER}$  and  $C \gg C_{ER}$ , but small distances can be obtained without the need of  $C \gg C_{ER}$  not the only mechanism (Dorogovtsev & Mendes, 2001).

**Degree distributions.** A different type of characterization of the statistical properties of a graph is given by the degree distribution  $P(k)$ . Although the ER graph displays a Poisson distribution, most complex networks are actually characterized by highly heterogeneous distributions: they 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  $\xi$ . Three main classes can be defined (Amaral *et al.*, 2000). (a) When  $\xi$  is very small,  $P(k) \sim \phi(k/\xi)$  and thus the link distribution is single-scaled. Typically, this would correspond to exponential or Gaussian distributions; (b) as  $\xi$  grows, a power law with a sharp cut-off is obtained; (c) for large  $\xi$ , scale-free (SF) nets are observed. The last two cases have been shown to be widespread and their topological properties have immediate consequences for network robustness and fragility (Albert *et al.*, 2000). The three previous scenarios are observed in: (a) power grid systems and neural networks (Amaral *et al.*, 2000), (b) protein interaction maps (Jeong *et al.*, 2001), metabolic pathways (Jeong *et al.*, 2000), ecological networks (Solé & Montoya 2001;

Dunne *et al.* 2002; Jordano *et al.* 2002) and electronic circuits (Ferrer i Cancho *et al.*, 2001) and (c) Internet topology (Jeong *et al.*, 2000; Caldarelli *et al.*, 2000), scientific collaborations (Newman, 2001) and lexical networks (Ferrer i Cancho & Solé, 2001a).

**Redundancy and degeneracy.** Complex biological networks (see below) exhibit an extraordinary homeostasis against random failure of a given unit. Partially based on the experience from technological systems, it was often assumed that such robustness was essentially due to redundancy: the failure of a given unit (such as a gene in a gene network) would be compensated by a copy of it. Nonetheless, the analysis of biological networks reveals that robustness is mainly associated with features that are fairly different from redundancy (Wagner, 2001; Edelman and Gally, 2001). Instead, it has been shown that, in many cases, totally different components can perform similar functions. This feature is known as *degeneracy* (Tononi *et al.*, 1999; Edelman and Gally, 2001): unlike redundancy, which involves structurally identical elements, degeneracy involves structurally different elements that yield the same or different functions depending on the context. Mounting evidence suggests that it is actually a ubiquitous property of biological nets.

Degeneracy is very common in natural systems (Edelman and Gally, 2001) but totally unknown within the context of technological evolution. In man-made systems, redundancy is the standard solution to the problem of random failure of single components: by introducing copies of sub-parts of the system, failure of one of them can be compensated by its copy. This was assumed to be the origin of resilience against mutation in genetic networks, although later evidence indicates that the source of homeostasis in cellular nets is a very different one (see below). To a large extent, degeneracy is intimately associated with tinkering in evolution: it reflects the re-use that is made of different parts of a system in order to achieve similar functions.

**Modularity.** Modularity pervades biological complexity (Hartwell *et al.*, 1999). Many cell functions are carried out by subsets of units that define functionally meaningful entities. An example are modules of genes involved in development (von Dassow *et al.*, 2000; Solé *et al.*, 2000; Solé *et al.*, 2002).

Modularity allows the adaptation of different functions with a small amount of interference with other functions and is likely to be a prerequisite for the adaptation of complex organisms, although it arises most likely as a byproduct of adaptability rather than being an adaptation itself (Wagner, 1995). Modularity can arise in two ways: by parcellation or by integration. Parcellation consists of the differential elimination of cross-interactions involving different parts of the system. Instead, if the network is originally formed by many independent, disconnected parts, it is conceivable that modularity arises by differential integration of those independent charac-

ters serving a common functional role.

### III. PROTEOME MAPS AND GENE NETWORKS

Let us start our exploration from molecular cell biology. Complex genomes involve many genes that are associated with at least one regulatory element and each regulatory element integrates the activity of at least two other genes. The nature of such regulation started to be understood from the analysis of small prokaryotic regulation subsystems (Lodish et al., 2000) and the current picture indicates that the webs that shape cellular behavior are very complex, sharing some common traits with neural networks and related computational systems (Bray, 1995).

Gene regulation takes place at different levels and involves the participation of proteins. The whole cellular network includes three levels of integration: (a) the genome (and the regulation pathways defined by interactions among genes), (b) the proteome, defined by the set of proteins and their interactions and (c) the metabolic network, also under the control of proteins that operate as enzymes. Unlike the relatively unchanging genome, the dynamic proteome changes through time in response to intra- and extracellular environmental signals. The proteome is thus particularly important: proteins unify structural and functional biology. They are both the products of gene activity and regulate reactions or pathways.

A key issue concerning the evolution of cellular nets is raised by their robustness against single-unit failure. The analysis of the effects of mutations in different organisms revealed an extraordinary level of homeostasis: in many cases the total suppression of a given gene in a given organism leads to a small phenotypic effect or even no effect at all (Ross, 1999; Wagner 2000). By following the analogy with engineered systems, it would be suggested that such robustness would come from the presence of a high degree of redundancy. Under mutation, additional copies of a given gene might compensate the failure of the other copy. However, the analysis of redundancy in genome data indicated that redundant genes are rapidly lost and that redundancy is not the leading mechanism responsible for mutational robustness (Wagner, 2000).

The degree distribution displayed by the protein interaction map is given by a power law (Jeong *et al.*, 2001; Wagner, 2001; Maslov and Sneppen, 2002) i.e.  $P(k) \sim k^{-\gamma}$  with  $\gamma \approx 2.5$ , with a sharp cut-off for large  $k$ . The link seems obvious here: the high degree of homeostasis against random failure would come from the highly heterogeneous distribution of interactions. This conjecture is actually supported by comparing the phenotypic effects of mutated genes with their degree: there is a clear positive correlation between degree and phenotypic effects of mutations (Jeong *et al.*, 2001). Consistently

with the SF scenario, mutations involving some key genes can have very important consequences. This is the case, in particular, of the p53 tumor suppressor gene (figure 3) which is known to play a critical role in genome stability and integrates many different signals related to cell-cycle or apoptosis (programmed cell death) (Vogelstein *et al.*, 2000). This and other tumor-suppressor genes prevent cell proliferation (thus keeping cell numbers under control) but can also promote apoptosis.

What can be concluded here? On the one hand, history is an obvious ingredient of the genome/proteome evolution: the growth of these nets takes place by gene duplication (Ohono, 1970) and simple models based on gene duplication plus re-wiring successfully reproduce the observed properties of proteome maps (Solé *et al.*, 2001; Vazquez et al, 2001; Pastor-Satorras *et al.*, 2002). In some sense, tinkering is widely used by starting from previous genes and interactions. The mutational robustness displayed by cellular nets might actually provide the best example of the success of degeneracy in evolution (Edelman and Gally, 2001). It is also clear that many genes are recruited at different stages through development and thus re-use of available building blocks is widespread.

The analysis of metabolic pathways reveals that some metabolites that are known to be older are highly connected (Wagner and Fell, 2001; Wagner, this issue) thus suggesting preferential attachment at least in early stages of evolution (see also Podani *et al.*, 2001). But models indicate that SF structure might spontaneously emerge provided that the rates of duplication and rewiring are appropriately tuned. In this sense, it might well happen to occur that, as a consequence of the duplication process, together with a sparse density of connections (associated to low cost constraints but also to dynamical constraints) the high robustness provided by the scaling structure might actually be a byproduct of network growth (Vazquez et al., 2001; Solé *et al.*, 2001) under optimization of communication. In that case, the reach of a sparse, but connected web of interactions would automatically provide an *emergent* source of robustness for free.

A final point to be mentioned is the fact that the need of an integrated, cellular order in the adult organism must coexist with some degree of flexibility associated to the developmental program. The building of a multicellular organism requires a control program that is exemplified by tumor-suppressor genes (such as p53). But this program has to integrate some weaknesses in order to allow for embryogenesis, growth and wound healing (Israel, 1996). Perhaps not surprisingly, the homologue of p53 in invertebrates (such as *Drosophila*) is expressed throughout development, particularly in early stages (Jin et al., 2000). It might be the case that one of the side effects of reaching a complex network with a high degree of homeostasis from a sparse, well-communicated graph of interactions, is that inevitably sooner or later the node

that fails is a key one (and cancer develops). Under these conditions, cellular functions are irreversibly altered and the cellular context (which imposes some control on cell states) is no longer a constraint.

#### IV. ECOLOGICAL NETWORKS

The effects of biodiversity loss are highly mediated by the architecture of ecological interactions. Ecosystems (defined broadly as the set of interacting species within a delimited habitat) are likely to be only rivaled in complexity by neural systems or the global economy (Brown 1994). Despite such complexity, early studies have found some interesting topological patterns that seem to be related with their stability and functionality (May 1973; see Pimm 1991 for a review). In this sense, recent characterization of complex topologies in several ecosystems may provide an amazing framework for predicting how perturbations might propagate throughout ecosystems.

It is clear that tinkering plays a leading role in conforming local biotas. Local assemblages comprise only some of the variety of species richness present on the Earth (between 10 and 100 million of species), conforming a myriad of ecosystems where species are connected via different interaction types (e.g. trophic, competitive or mutualistic). Thus, the ecosystems' tinkerer works only with available species from the regional species pool, and the sequence of species invasions partially determines the composition of resulting communities. For instance, some field experiments have shown that priority effects are above competition, that is, if one species arrives before another that is a better competitor, the former should persist in the community, and the later will not be established.

The food web is the basic description of ecosystems, where nodes are species (or sets of species with similar diets) and links are feeding relationships. Trophic structure influences the performance of several ecosystem functions (e.g., productivity, nutrient capture and cycling) and determines the effects of the propagation of disturbances through the entire system. In particular, the effect of one species on the density of another tends to diminish with their separation in the food web, measured by the shortest path connecting them. Typically, if species A and species B are more than 3 links away, a disturbance in the density of A does not influence B (Abrams *et al.* 1996). But most species within communities are separated by only 2 or 3 links from each other (Williams *et al.* 2001; Montoya & Solé 2002), so perturbations might propagate through the entire network. In most of these food webs the clustering is clearly higher than the expected one from randomness (Montoya & Solé 2002), providing evidence of SW patterns despite their relative small size (less than 200 nodes).

Food webs exhibit a hierarchy of discrete trophic levels in such a way that a species within a trophic level only feeds on species belonging to the immediately lower level (e.g. carnivores feed on herbivores that eat plants). If this hierarchy were strict, no clustering would be observed. Thus, the origins of clustering in this type of systems is the presence of omnivorous species, that is, species that feed on more than one trophic level, including intra-trophic level predation. Omnivory also contributes to shorten paths between species, and it has been shown both empirically and theoretically that omnivory tends to increase the stability of ecosystems by reducing population fluctuations (McCann 2000).

Ecological networks also display heterogeneous degree distributions, fitting in most cases a truncated power-law  $P(k) \sim k^{-\gamma} \phi(k/\xi)$ , where  $\phi(k/\xi)$  introduces a sharp cut-off. In particular, species-rich food webs (Solé & Montoya 2001; Camacho *et al.* 2002; Montoya & Solé 2001, 2002; Dunne *et al.* 2002) and most plant-animal mutualistic networks (Jordano *et al.* unpublished) display this type of distribution, with  $\gamma \sim 1.0$ . This exponent is different from those observed in cellular, social and technological networks ( $\gamma \sim 2.1 - 3.0$ ). Both features (sharp cut-offs and low  $\gamma$ ) suggest the existence of constraints in the assembly of ecological networks, as the limits to the addition of new links due to different phenological attributes of species (Jordano *et al.* unpublished), or due to dynamical constraints related with the persistence of preys (a prey with many connections is more affected by changes in the food web) or with the efficiency of predators (the more connected you were, the less efficient you will be; Montoya, Solé & Pimm *unpublished*). These constraints add to cost-related constraints (Amaral *et al.* 2000) in limiting the prevalence of preferential attachment as the mechanism underlying ecological organization. In contrast, these constraints do not seem to work on small-size food webs, where degree distributions are closer to a random wiring with typical Poisson distributions (Montoya & Solé 2001; Dunne *et al.*, 2002).

A deep understanding of the architecture of ecological networks allow us to identify keystone species within ecosystems, that is, species whose removal trigger many secondary extinctions and the fragmentation of the network into disconnected subwebs in which species are more prone to extinction. Keystone species are typically the most-connected ones, and food webs with skewed degree distributions (that is, distributions with long tails) are predicted to exhibit an extreme fragility under their successive loss in contrast to the high robustness when species are removed randomly (Solé & Montoya 2001; Dunne *et al.* 2002). Also, functions performed by the ecosystems are more likely to be lost after the removal of such keystones, although in some cases some less-connected species perform key specific functions, so their removal may also have large effects on ecosystem functionality (e.g. the loss of plants that fix atmospheric

nitrogen). A key question not yet resolved is how this duality fragility-robustness is affected when dynamics is introduced.

Species play specific roles in ecosystems, both affecting population dynamics (e.g. a carnivore controlling the density of several herbivores thus guarantying diverse plant assemblages), and enhancing several ecosystem functions (e.g., biomass, energy use or nutrient retention). Can the roles of eliminated species be replaced by other species?, and if so, by what kind of mechanisms? Removal experiments and simulations have shown the presence of some degree of adaptability, depending on the degree of redundancy of the manipulated ecosystem. In particular, more redundancy implies the maintenance of some ecosystem functions (Rutledge *et al.* 1976; Walker 1992; Naem 1998; Fonseca & Ganade 2001) and reduce the risk of cascading extinctions after random extirpations (Borvall *et al.* 2000).

What is called *redundancy* in the ecological literature is what we have previously defined as degeneracy, because “replacement” species and removed species have always different traits. In fact, evolution tends to reduce redundancy levels (i.e., species with identical traits) while increasing degeneracy by promoting species that are complementary and overlapping in their resource requirements but different in their environmental tolerances (Ulanowicz 1986; Morgan-Ernest & Brown 2001). But often this mechanism of compensation is observed with a large delay or simply it does not happen, particularly when keystone species are lost (e.g., Brown & Heske 1990).

Thus, it is clear that a topological (static) approximation is a feasible first step in the understanding of community homeostasis. However, is such homeostasis a side-effect of an optimization of any function of the ecosystem? Ecosystems perform several functions that evolve along their spatio-temporal organization. Nevertheless, what functions, if any, are optimized, it is far from being clear. Some authors have suggested that ecosystems are likely to maximize their efficiency in transferring energy and materials. But these approximations are highly speculative and have several caveats. For instance, some mechanisms promoting community homeostasis imply a decrease in efficiency, as the presence of omnivorous species, which are typically inefficient in consuming their preys. However, simple optimization mechanisms observed in other networks are an interesting area for exploring the relationships between maximization of some ecosystem functions and homeostasis.

## V. LEXICAL NETWORKS

The emergence of human language is one of the major transitions in evolution (Maynard-Smith & Száthmáry,

1997). We humans possess a unique symbolic mind capable of language which is not shared by any other species (Deacon, 1997). Human language allows the construction of a virtually infinite range of combinations (i.e. sentences) from a limited set of basic units. The process of sentence generation is astonishingly rapid and robust and indicates that we are able to rapidly gather words to form sentences in a highly reliable fashion.

The study of lexical networks (networks in which nodes are words and links are formed between strongly correlated words; see Fig. 5) (Ferrer i Cancho & Solé, 2001a) and other linguistic networks (Sigman & Cechi, 2002; Steyvers & Tenenbaum, 2002) has shown that scaling is a strong regularity in human language. Links in lexical networks capture syntactic relationships between pairs of words. The degree distribution follows  $P(k) \sim k^{-\gamma}$  with  $\gamma \approx -3$ . Hubs in lexical networks are function words (e.g. prepositions, articles, determiners,...). Hubs and thus function words are crucial for the lexical network small-worldness. Function words constitute the most stable set of words of a language over time. Their high connectivity explains why they are less extinction prone in the same way omnivorous species are in ecological networks.

New function words are neither created from scratch nor borrowed from other languages. Function words result from grammaticalization processes (Traugot & Heine, 1991) in which non-function words become function words and function words become more grammatical (Hopper, 1991). Languages largely result from tinkering: Prepositions typically derive from terms of body part or verbs of motion while modals typically derive from terms of possession or desire (Traugot & Heine, 1991). Thereafter, hub formation in lexical network evolution, operates on a restricted set of candidate words. Every language creates new functions from such a restricted set so convergence is likely not only regarding the result of the grammaticalization process but also regarding its starting point.

The fact that the degree (i.e. number of links) of a word is positively correlated with its frequency (Ferrer i Cancho & Solé, 2001a) and that among the most frequent words there are the etymologically oldest ones (Prün, 1999; Zipf, 1949), may suggest preferential attachment is at play. Nonetheless, there are many reasons for thinking that minimization of word-word distance is involved. On the one hand, speaking is a complex task. The speaker must combine a big amount of words (in the order of many thousands (Miller & Gildea, 1987)) for forming sentences. Following links in a lexical network leads to syntactically well-formed sentences. If two words are to be linked during speech production, the smaller the distance between them, the smaller the amount of intermediate words required for performing the linkage. Average word-word distance in lexical networks is about  $d \approx 2.6$ , indicating that most of the words are reachable through less than two intermediate words (Ferrer i

Cancho & Solé, 2001).

From the other hand, it has been proposed that optimizing the small-worldness of a network under linking cost constraints may be the origin of the  $\gamma \approx -3$  scaling exponent (Ferrer i Cancho & Solé, 2001b). Small world behavior is a desirable property of a linguistic network and linking restrictions are likely to be at play. If linking cost is not taken into account, the optimal configuration would be a clique (every word connected to every word; Fig. 6 A) which would imply that all words are function words but none of them is a lexical word (which can be seen as having linkers but no words to link). Linking cost may not be the only restriction at play. A star network (every vertex connected to the same vertex; Fig. 6 B) provides the minimum distance possible at the minimum linking expense. In this case, there are words to link, but only one linker. Nonetheless, real languages have a rich repertoire of linkers in order to account for different types of relations (part-hole, action-receiver,...).

The finding of a  $\gamma \approx -3$  exponent helps us to understand why syntax is a robust trait of human communication. The degree of expression of such a feature is not correlated with intelligence and thereafter not surprisingly present even when intellectual skills are extremely poor, which is the case of the *idiots savants* (Deacon, 1997). According to previous work on error tolerance of scale-free networks (Albert *et al.*, 2000), a lexical network will be very robust against random removal of words but fragile against removal of the most connected vertices. Agrammatism, a kind of aphasia in which many of the most connected words seem to have disappeared, is characterized by a decrease in the ability to build syntactically complex sentences (Caplan, 1999). Unless the words that glue words for building complex sentences are removed (the “shortpathers”), a complex phrase (e.g. a circumlocution) can replace a missing word and the expressive power will be maintained.

## VI. TECHNOLOGY GRAPHS

Social and economic complexity are organized around three major networks: the transportation network, the power network and the communication network (Inose, 1972). These are also three components of biological complexity, involving the processing of information, energy and matter at very different scales. Artificial networks offer an invaluable reference when dealing with the rules that underlie the building process in complex systems. The Internet, in particular, powerfully exemplifies the importance of topology and homeostasis in SF nets (Albert *et al.*, 2000) and how it relates with biological complexity. The spread of computer viruses, for example, is closely related to the epidemic patterns displayed by natural viruses (Pastor-Satorras and Vespignani, 2000;

Lloyd and May, 2001). Not surprisingly, man-made artifacts have been used as a metaphor of biological complexity at different levels. The brain, for example, has been compared to mechanical engines, the telegraph and telephone webs or to different computer architectures, as technological changes followed each other through time. These metaphors are often flawed by the lack of a real mapping between both systems (there is little overlap between the telegraph network and cortical maps beyond the presence of a wiring diagram) but sometimes the relationships are surprisingly strong (Nelson and Bower, 1990).

Technological graphs result from predefined planning. As noted by Jacques Monod: “the difference between artificial and natural objects appears immediate and unambiguous to all of us. A rock, a mountain, a river, or a cloud -these are natural objects; a knife, a handkerchief, a car- are artificial objects, artifacts” (Monod, 1970). But technological design is not completely free of the constraints imposed by complexity (Kauffman, 1995). On the one hand, conflicting constraints effectively limit the reaching of optimal solutions: the process of design is itself a search process on a rugged landscape with many implicit variables.

The study of electronic circuits reveals that small world is present in technological artifacts (Ferrer i Cancho *et al.*, 2001). Modularity in these systems is not only a direct consequence of the need of different sub-parts performing different (but complementary) functions but also the result of needing to reuse existing circuits. The engineer avoids designing a new large circuit from scratch. The engineer is encouraged to work as a tinkerer when the size of the circuit crosses a certain threshold. Optimization is present at different levels of the circuit design (Ferrer i Cancho *et al.*, 2001). For instance, minimization of both average path length and physical distance are present and this can easily lead to SW structure (Mathias and Gopal, 2001). Interestingly, the larger ones clearly exhibit power laws in their degree distributions (Ferrer i Cancho *et al.*, 2001).

A very important class of networks derived from software architecture maps (Perry and Wolf, 1992), has been recently shown to display both SW and SF patterns as the non-planned result of a design optimization process (Valverde *et al.*, 2002). Software architecture is reflected by a diagram of the software components and their interrelations, which is produced during the design process (Pressman, 1992). Sometimes the diagram is not explicitly provided, but it is possible to reconstruct it from source code (reverse engineering). This map can be interpreted as a graph where nodes are software components and the links are relationships between software components.

A large effort has been dedicated to understand the nature of software and why to efficiently build and maintain large software systems is so difficult (and costly). For

years, the software community has been promoting the need of software measurement tools that help to quantify if a software project is being developed "well" and controlling the deviations from the established engineering plan (Zuse, 98). Early measures of software were centered in intra-module aspects like program length or number of lines of code (LOC). Recently, there is a growing interest in analyzing software structure or software architecture measurement (inter-module). The modern software engineer conceives software systems at a higher conceptual level and is more concerned about how the different components are assembled and interrelated to each other (Gamma *et al.*, 1994).

A closer look at different software architecture maps indicates that there is a noticeable difference in the architecture of small-scale (i.e: a Gauss-Seidel linear system solver, sorting an arbitrary sequence of numbers) and the architecture of large-scale software (i.e: an operating system or a modern videogame). In both cases, different conflicting constraints (economization of memory storage, efficiency of processes, ease of integration of changes and new features) must be satisfied simultaneously while developing software. For small-scale software it is possible to produce very optimized structures because the constraints are not too hard to satisfy. The architecture of small-scale software abounds in hierarchies and simple connectivity patterns (Fig. 7). The degree distribution is Poissonian.

Large-scale software shows clustering, low mean distance between software components (about six) and scale-free degree distribution, the exponent ranging from  $-2$  to  $-3$  (Valverde *et al.*, 2002). For large-scale software the constraints are much harder and history plays an important role. In order to reduce the complexity of software development, the system is partitioned into modules that group similar or related software components. Usually, the modules are developed separately. One of such modules roughly corresponds to a single connected component in the software architecture map, but other modules could have more than one component. Even a single connected component can span several modules. Briefly, the modules are a logical partition of the software architecture that may or may not coincide with the set of connected components.

There are few components with an huge number of links (hubs) in most software architectures. Surprisingly, these hubs are considered a bad practice by software en-

gineering principles (Brown *et al.*, 1998). The existence of hubs is an indication that only a sub-optimum solution can be reached. Sometimes the cost of introducing new components is higher than adapting an existing one to provide the required functionality. As the number of dependencies of the software component grows up, it progressively loses its intended original meaning, assuming more and more responsibilities from other system components. This class of components limits reuse and are expensive in terms of memory storage. Also, they tend to be very affected by changes made to other components of the system.

There is a certain degree of redundancy in large-scale software. Because different parts of the system are developed in parallel by a team of software engineers, there is a chance that the same subproblem will be solved twice. This is also known as "reinventing the wheel" and it is considered one of the main causes of productivity loss. The duplication of software functions is not recommended by software engineering because the increased effort required to fix errors and to extend the system functionality. In fact, it is encouraged to seek and locate duplicated portions of code and substitute them by a single software component. Moreover, a good practice of software engineering promotes reusing of large portions of code (or better, entire software components) not only within a software system but from project to project (Pressman, 1992). In any case, it seems that such traditional claims for reusing are very difficult to catch effectively and the way to achieve it still being pursued by modern software engineering.

## VII. DISCUSSION

The heterogeneous character of most complex biological networks reveals a surprising example of convergence. In evolutionary theory, convergence refers (within the ecological context) to the observation that organisms living in similar habitats resemble each other in outward appearance. These similar looking organisms may, however, have quite different evolutionary origins. Convergent evolution takes place at very different levels, from organisms to molecules, and here we propose the idea that a new type of convergent evolutionary dynamics might be at work underlying a very wide class of both natural and artificial systems.

Property	Proteomics	Ecology	Language	Technology
----------	------------	---------	----------	------------



Tinkering	Gene duplication and recruitment	Local assemblages from regional species pools and priority effects	Creation of words from already established ones	Reutilization of modules and components
Hubs	Cellular signaling genes (e.g. p53)	Omnivorous and most abundant species	Function words	Most used components
What can be optimized?	Communication speed and linking cost	Unclear	Communication speed with restrictions	Minimize development effort within constraints
Failures	Small phenotypic effect of random mutations	Loss of only a few species-specific functions	Maintenance of expression and communication	Loss of functionality
Attacks	Large alterations of cell-cycle and apoptosis (e.g. cancer)	Many coextinctions and loss of several ecosystems functions	Agrammatism (i.e. great difficulties for building complex sentences)	Avalanches of changes and large development costs
Redundancy & Degeneracy	Redundant genes rapidly lost	R minimized and D restricted to non-keystone species	Great D	Certain degree of R but no D

TABLE I. Table I: A summary of the basic features that relate and distinguish different types of complex networks, both natural and artificial. Here different characteristic features of complex nets, as well as their behavior under different sources of perturbation, are considered.

Since very different systems seem to choose the same basic formula for their interaction maps, we can easily recognize a general trend that can be identified as a process of convergent evolution (see Table I). One particularly important point is the fact that similar network topologies (particularly the scale-free ones) emerge in biological and human-made systems. Although the first take advantage of the high homeostasis provided by scaling, 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.

The apparently universal character of these scaling laws in such disparate complex networks goes beyond homeostasis. We conjecture that the leading force here is an optimization process where reliable communication at low cost shapes network architecture in first place. This seems to be the case in all the previous systems analyzed. Once a small, critical average connectivity is reached, the graph experiences a percolation providing a spontaneous order linked to global communication. This occurs at a low cost, since the transition is sudden and effectively connects all parts of the system with a small number of links per unit.

There are two possible strategies for decreasing vertex-vertex distance at the percolation point: (a) increasing the average connectivity and (b) hub formation. (a) is a trivial strategy whose outcome under ideal circumstances is a clique (Fig. 7 A). (b) has the advantage of not im-

plying the addition of new connections. Link rearrangements suffice. In contrast, (b) is a more complex task than (a). The extreme outcome of (b) is a star network (Fig. 7 B). Stars that are formed by subsets of the entire set of vertices (i.e. hubs) can be easily identified in real scale-free networks (see Figs. 4,5,7). Scale-free distributions suggest the use of strategy (b) in real networks. If reaching a low density of links is more important than having a small average path length, skewed distributions, including scale-free nets, are easily obtained (Ferrer i Cancho and Solé, 2001) through random search. Nonetheless, other mechanisms, such as preferential attachment (Barabási & Albert, 1999) have been proposed for scaling. Unfortunately, preferential attachment can not straightforwardly explain the high clustering coefficient of real networks.  $C$  is a measure of cliquishness (more precisely, the abundance of cliques of 3 vertices). Thereafter, a high clustering coefficient can be seen as one side of the optimization process. Natural and artificial networks show that the higher the size of the network, the higher the significance of the clustering coefficient and also the power appearance of the degree distribution (Montoya & Solé, 2001; Ferrer i Cancho *et al.*, 2001). The coexistence of strategies (a) and (b) can be understood as the explosion of conflicting constraints once the networks size exceeds a certain threshold value. The fact that even engineers become tinkers in large systems illustrates how complicated is the achievement of optimal structures once they reach some complexity level. Clustering is likely to be unavoidable for small-worldness in

networks in which hub formation becomes a dramatically complicated task.

Modularity is an obvious source of clustering. We have suggested that homeostasis is a consequence of a more general principle. Actually, it can be a side-effect of optimization and not a direct consequence of functional parcellation in large networks.

The scale-free distributions observed in both natural and artificial graphs suggests that the homeostasis won by the second might well be a result of exploiting the SF topology resulting from optimization (Valverde *et al.*, 2002). This is something not (yet) exploited in current engineered systems, probably due to lack of degeneracy. Since degeneracy is a common feature of biological nets, it might have been exploited (or co-evolving) within heterogeneous architectures. We conjecture that there is a largely universal principle that pervades the evolution of scale-free nets (optimal communication) and that the observed topological features of bionets reflect this feature together with constraints arising from other causes, such as the need of modular organization. It is interesting to see that man made designs also evolve towards webs that strongly resemble their biological counterparts: as shown by the previous examples, often the paths toward optimization seem to cross the land of tinkering.

## Acknowledgments

The authors thank William Parcher and the members of the Complex Systems Lab for useful comments. This work has been supported by a grant PB97-0693, the Santa Fe Institute (R.V.S.) and grants from the Generalitat de Catalunya (FI/2000-0393, R.F.C.) and the Comunidad de Madrid (FPI-4506/2000, J.M.M.).

## VIII. REFERENCES

1. Abrams, P., Menge, B.A., Mittelbach, G.C., Spiller, D. & Yodzis, P. (1996) The role of indirect effects in food webs. *In: Food webs: integration of patterns and dynamics* (Polis, G.A. and Wine-miller, K.O., eds.) pp. 371-395, Chapman & Hall, New York.
2. Alberch, P. (1982) Developmental constraints in evolutionary processes in *Evolution and development*, ed. J.T. Bonner, (Berlin, Springer-Verlag) 313-332.
3. Albert, R. and Barabási, A-L. (2001) Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47- .
4. Albert, R., Jeong, H. and Barabási, A-L. (2000) Error and attack tolerance of complex networks. *Nature* **406**, 378-382.
5. Amaral, L. A. N., Scala, A., Barthélemy, M. and Stanley, H. E. (2000) Classes of behavior of small-world networks. *Proc. Nat. Acad. Sc. USA* **97**, 11149-11152.
6. Barabási, L-A. and Albert, R. (1999) Emergence of scaling in random networks. *Science* **286**, 509-512.
7. Bollobás, B. (1985) *Random Graphs*. Academic Press, London.
8. Borvall, C., Ebenman, B. & Jonson, T. (2000) Biodiversity lessens the risk of cascading extinctions. *Eco. Lett.*, **3**, 131-136.
9. Bray, D. (1995) Protein molecules as computational elements in living cells. *Nature* **376**, 307-312.
10. Brown, W. H. et al. (1998) *Anti-Patterns: Refactoring Software, Architectures, and Projects in Crisis*, John Wiley & Sons, New York.
11. Brown, J.H. & Heske, E.J. (1990) Control of a desert-grassland transition by a keystone rodent guild. *Science* **250**, 1705-1707.
12. Brown, J.H. (1994) Complex ecological systems. *In: Complexity: metaphors, models and reality* (Cowan, G.A., Pines, D. & Meltzer, D. eds.), pp. 419-443. Addison-Wesley, Reading.
13. ANTIP98 Brown, W. H. *et al.*, *Refactoring Software, Architectures, and Projects in Crisis*, John Wiley & Sons, New York, 1998.
14. Camacho, J., Guimerá, R. & Amaral, L.A.N. (2002) Robust patterns in food web structure. *Phys. Rev. Lett.* **88**, 228102.
15. Caplan, D. (1994) *Language. Structure, processing and disorders*. MIT Press.
16. Cohen, J.E. (1978) *Food webs and niche space*. Monographs in Population Biology 11, Princeton Univ. Press.
17. Deacon, T. W. (1997) *The symbolic species: the co-evolution of language and the brain*. New York: W. W. Norton & Co.
18. Dorogovtsev, S. N. and Mendes, J. F. F. (2001) *Evolution of random networks*. arXiv:cond-mat/0106144
19. Edelman, G. M. and Gally, J. A. (2001) Degeneracy and complexity in biology systems. *Proc. Natl. Acad. Sci. USA* **98**, 13763-13768.
20. Fell, D. and Wagner, A. (2000) The small world of metabolism. *Nature Biotech.* **18**, 1121.

21. Ferrer i Cancho, R., Janssen, C. and Solé, R. V. (2001) The topology of technology graphs: small world pattern in electronic circuits. *Phys. Rev. E* **64**, 32767.
22. Ferrer i Cancho, R. and Solé, R. V. (2001a) The small world of human language. *Procs. Roy. Soc. London B* **268**, 2261-2266.
23. Ferrer i Cancho, R. and Solé, R.V. (2001b). Optimization in complex networks. *cond-mat/0111222*.
24. Fonseca, C.R. & Ganade, G. (2001) Species functional redundancy, random extinctions and the stability of ecosystems. *J. An. Ecol.*, 89, 118-125.
25. Gamma, E., Helm, R., Johnson, R. and Vlissides, J. (1994) *Design Patterns*, Reading, MA, Addison-Wesley.
26. Goodwin, B.C. (1994) *How the Leopard Changed Its Spots: the Evolution of Complexity*. Charles Scribner's Sons, New York.
27. Gould, S. J. (1989) *Wonderful Life*. Penguin. London.
28. Gould, S. J. (2002) *The Structure of Evolutionary Theory*. Belknap Press, Cambridge MA.
29. Hartwell, L. H., Hopfield, J. J., Leibler, S. and Murray, A. W. (1999) From molecular to modular cell biology. *Nature* **42** supp, c47-252.
30. Hertz, J., Krogh, A. and Palmer, R. G. (1991) *Introduction to the theory of neural computation*, Addison-Wesley, Reading.
31. Hjelmfelt, A., Weinberger, E. D. and Ross, J. (1991) Chemical implementation of neural networks and Turing machines. *Proc. Natl. Acad. Sci. USA* **88**, 10983-10987.
32. Hopper, P.J. (1991) On some principles of grammaticization. in *Approaches to grammaticalization* (Traugot, E. C. & Heine, Bernd, eds.). Volume I. John Benjamins Publishing Company, Amsterdam/Philadelphia.
33. Inose, H. (1972) Communication networks. *Sci. Am.* n 3, September, pp. 117-128.
34. Israel, L. (1996). Tumor progression: random mutations or and integrated survival response to cellular stress conserved from unicellular organisms? *J. Theor. Biol.* **178**, 375-380.
35. Jacob, F. (1976) Evolution as tinkering. *Science* **196**, 1161-1166.
36. Jeong, H., Mason, S., Barabási, A. L. and Oltvai, Z. N. (2001) Lethality and centrality in protein networks *Nature* **411**, 41 (2001).
37. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. and Barabasi, A.-L. (2000) The large-scale organization of metabolic networks *Nature* **407**, 651-654
38. Jin, A., Martinek, S, Joo, W. S. et al. (2000) Identification and characterization of a p53 homologue in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA*. **97**, 7301-7306.
39. Kauffman, S. A. (1962) Metabolic stability and epigenesis in randomly connected nets. *J. Theor. Biol.* **22**, 437-467.
40. Kauffman, S. A. (1993) *Origins of Order*, Oxford, New York.
41. Kauffman, S. A. (1995) *At Home in the Universe*. Oxford U. Press, new York.
42. Lodish, H., Berk, A., Zipursky, S. L. and Matsudaira, P. (2000) *Molecular Cell Biology* (4th edition) New York, W. H. Freeman.
43. Maslov, S. and Sneppen, K. (2002) Specificity and stability in topology of protein networks. *Science* **296**, 910-913.
44. Maynard-Smith, J. & Szäthmáry, E., 1997 *The major transitions in evolution*. Oxford University Press.
45. McCann, K.S. (2000) The diversity-stability debate. *Nature*, 405, 228-233.
46. Miller and Gildea (1987) How children learn words. *Scient. Am.* **257**, 94-99.
47. Monod, J. (1970) *Le hasard et la nécessité*, Editions du Seuil, Paris.
48. Montoya, J. M. and Solé, R. V. (2002) Small world patters in food webs. *J. Theor. Biol.* **214**, 405-412 .
49. Montoya, J.M. & Solé, R.V. (2001) Topological properties of food webs: from real data to assembly models. Santa Fe Institute Working Paper 01-11-069.
50. Naeem, S. (1998) Species redundancy and ecosystem reliability. *Conservation Biology*, 12, 39-45.
51. Nelson, M. E. and Bower, J. M. (1990) Brain maps and parallel computers. *Trends Neurosci.* **13**, 403-408.
52. Newman, M. E. J. (2000) Models of the small world. *J. Stat. Phys.* **101**, 819-841.

53. Newman, M. E. J. (2002) Random Graphs as Models of Networks. Santa Fe Institute Working Paper 02-02-005.
54. Ohono, S. (1970) *Evolution by gene duplication*. Springer, Berlin.
55. Pastor-Satorras, R. and Vespignani, A. (2001) Epidemic spreading in scale-free networks. *Phys. Rev. Lett.*, **86**, 3200-3203.
56. Pastor-Satorras, R., E. Smith and Solé, R. V. (2002) Evolving protein interaction networks through gene duplication. Santa Fe Institute Working Paper 02-02-008.
57. Perry, D. E. and Wolf, A. L. (1992) Foundations for the Study of Software Architecture. ACM SIGSOFT Software Engineering Notes, 17:4.
58. Pimm, S.L. (1982) *Food webs*. Chicago University Press, Chicago.
59. Pimm, S.L. (1991) *The balance of nature?*. Chicago University Press, Chicago.
60. Podani, J., Oltvai, Z. N., Jeong, H., Tombor, B., Barabási, A.-L. and Szathmáry, E. (2001) Comparable system-level organization of Archaea and Eukaryotes, *Nature Genetics* **29**, 54-55.
61. Pressman, R. S. (1992) *Software Engineering: A Practitioner's Approach*, McGraw-Hill.
62. Prüin, C. (1999). *G. K. Zipf's conception of language as an early prototype of synergetic linguistics*. *Journal of Quantitative Linguistics*, **6**, 78-84.
63. Rutledge, R., Basore, B. & Mulholland, R. (1976) Ecological stability: an information theory viewpoint. *J. theo. Biol.* **57**, 355-371.
64. Sigman & Cechi, 2002 *Global organization of the WordNet lexicon*. *Proc. Natl. Acad. Sci. USA.* **99**, 1742-1747.
65. Solé, R. V., Salazar, I. and Newman, S. A. (2000) Gene network dynamics and the evolution of development. *Trends Ecol. Evol.* **15**, 479-480
66. Solé, R. V. & Montoya, J. M. (2001) Complexity and fragility in ecological networks. *Proc. Roy. Soc. Lond. ser. B*, **268**, 2039-2045.
67. Solé, R. V., Salazar, I. and Garcia-Fernandez, J. (2002) Common Pattern Formation, Modularity and Phase Transitions in a Gene Network Model of Morphogenesis. *Physica A* **305**, 640-647
68. Solé, R. V., Pastor-Satorras, R. Smith, E. and Kepler, T. (2002) A model of large-scale proteome evolution. *Adv. Complex Syst.* **5**, 43-54.
69. Sompolinsky, H., Crisanti, A. and Sommers, H. J. (1988) Chaos in random neural networks. *Phys. Rev. Lett.* **61**, 259-262.
70. Steyvers, M. Tenenbaum, J. B. (2002). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. Submitted to *Cognitive Science*. <http://www-psych.stanford.edu/jbt/>
71. Strogatz, S. (2001) Exploring complex networks. *Nature* **410**, 268-276.
72. Traugot, E. C. & Heine, Bernd (1991). *Approaches to grammaticalization*. Volume I. John Benjamins Publishing Company, Amsterdam/Philadelphia.
73. Ulanowicz, R.E. (1986) *Growth and development: ecosystems phenomenology*. Springer, New York.
74. Valverde, S., Ferrer i Cancho, R. and Solé, R. V. (2002) Scale-Free Networks from Optimal Design, Santa Fe Institute Working Paper 02-04-019.
75. Vázquez, A., Flammini, A., Maritan, A. & Vespignani, A. (2001) Modelling of protein interaction networks. *cond-mat/0108043*.
76. Venter, J. C. *et al.* (2001) The sequence of the human genome. *Science* **291**, 1305 (2001)
77. Vogelstein, B., Lane, D. and Levine, A. J. (2000) Surfing the p53 network. *Nature* **408**, 307-310.
78. von Dassow, G., Meir, E., Munro, E. M. and Odell, G. M. (2000) *Nature* **406**, 188-194.
79. Wagner, G. (1995) Adaptation and the modular design of organisms. In: *Advances in Artificial Life*, F. Moran, J. J. Merelo and P. Chacon (eds.) Springer-Verlag, Berlin.
80. Wagner, A. (2001) The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *J. Molec. Evol.* **18**, 1283-1292.
81. Walker, B.H. (1992) Biodiversity and ecological redundancy. *Conservation Biology*, **6**, 18-23.
82. Watts, D. J. (1999) *Small Worlds*. Princeton U. Press
83. Watts, D. J. and Strogatz, S. H. (1998) Collective Dynamics in 'small-world' networks, *Nature (Lond.)* **393**, 440-442
84. Williams, R. J., Martinez, N. D., Berlow, E. L., Dunne, J. A. and Barabási, A.-L. (2001) Two degrees of separation in complex food webs. Santa Fe Institute Working Paper, 01-07-036.

85. Zipf, G. K. (1949). *Human behavior and the principle of least effort. An introduction to human ecology*. New York: Hafner reprint, 1972. [1st. edition: Cambridge MA: Addison-Wesley, 1949].
86. Zuse, H. (1998) A Framework for Software Measurement, Walter de Gruyter, Berlin, New York.

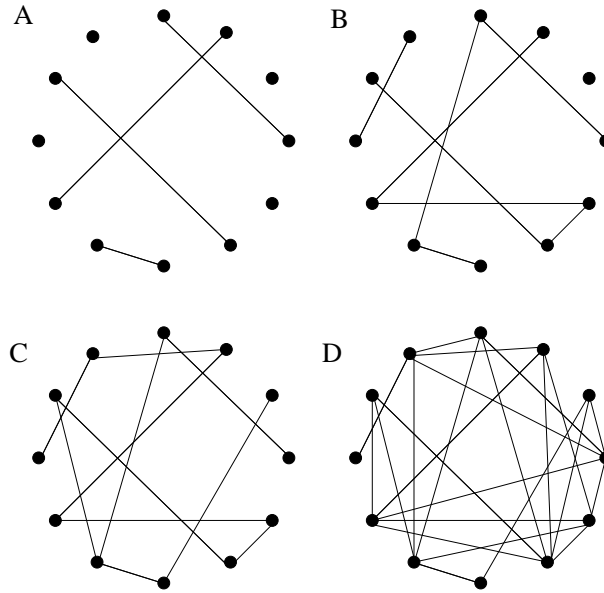


FIG. 1. Phase transition in random graphs: In (a-d) four different states of the wiring process are shown for a small graph with  $n = 12$  nodes. Here (a)  $z = 0.22$ , (b)  $z = 0.6$ , (c)  $z = z_c = 1$  and (d)  $z = 2$ . A phase transition occurs at  $z_c = 1$ , where the fraction  $\gamma$  of nodes included in the giant component rapidly increases. This picture corresponds to a very small system. Appropriate characterizations of the phase transition require large network sizes.

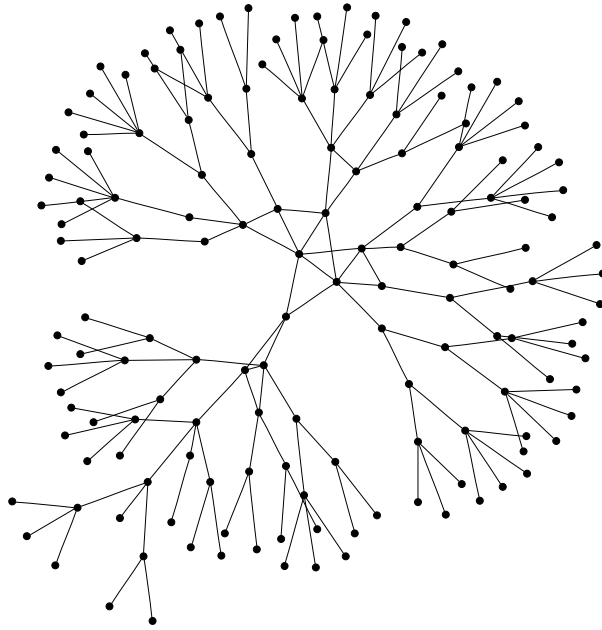


FIG. 2. Standard communication networks involve a large amount of hierarchy. In the case of the telephone network, the terminals are telephone sets and a node is a switching center for routing telephone calls. Here the graph of a local telephone net is shown. The tips of the tree represent telephones (based from Inose, 1972).

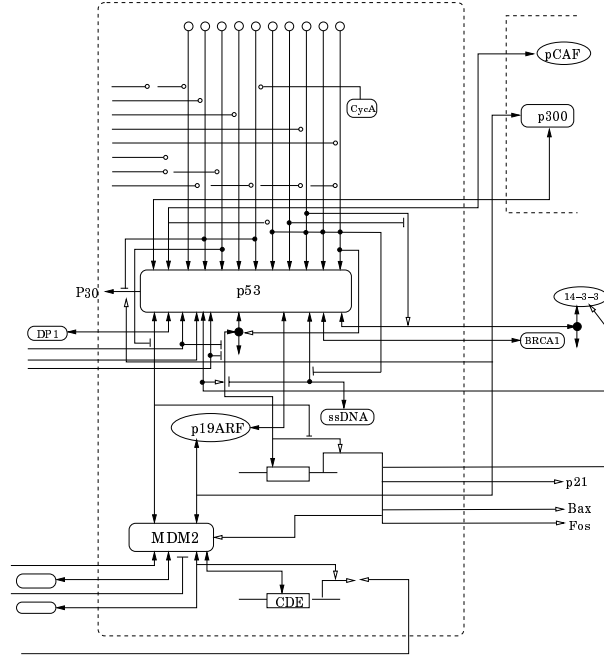


FIG. 3. The p53 network (redrawn from Kohn, 1999). Here a subgraph embedded in a much larger network (the interaction map of the mammalian cell cycle control and DNA repair systems) is shown. The activation of this network (due to different types of stresses) leads to the stimulation of enzymatic activities that stimulate p53 and its negative regulator, MDM2.



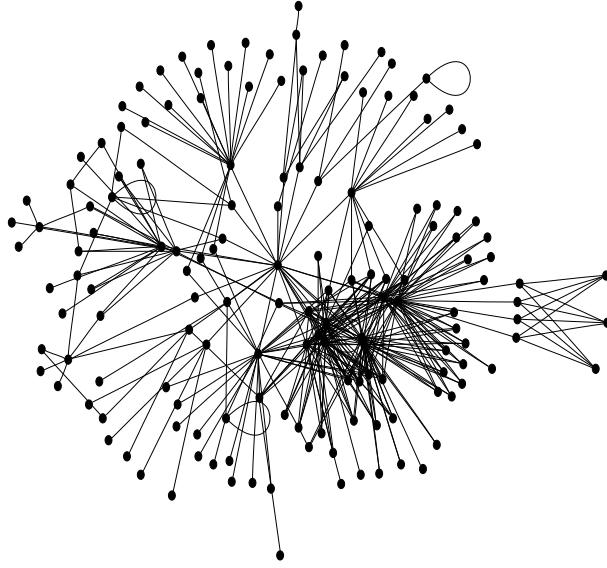


FIG. 4. The graph of species interactions of Silwood park web. Here each node represents one species. The central node in this representation is the Scotch broom *Cytisus scoparius*.

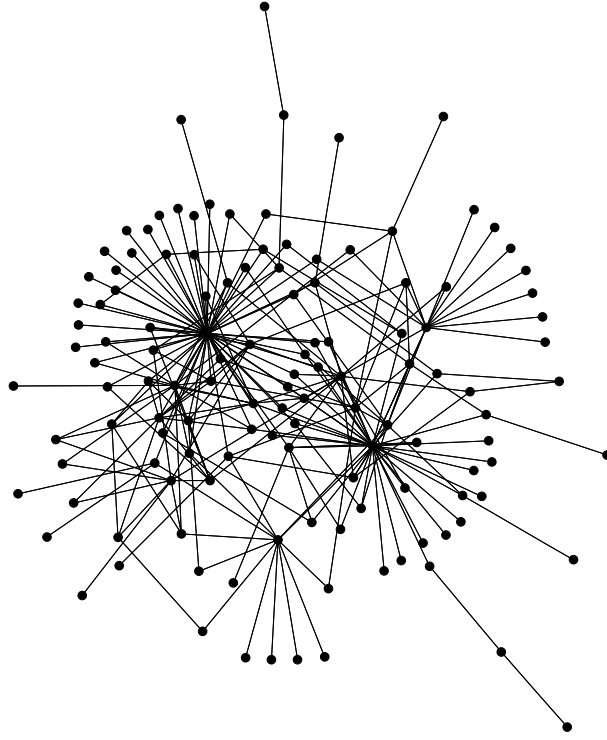


FIG. 5. The topology of Moby Dick. Two words appearing in Melville's book are linked if their mutual information is greater than a given threshold.

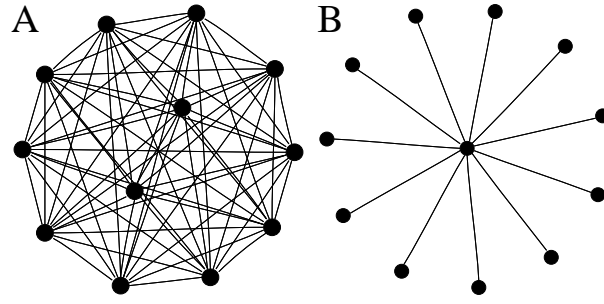


FIG. 6. Opposite ways of achieving small worldness. A The maximum linking expense (a clique) B The minimum linking expense (a star graph).

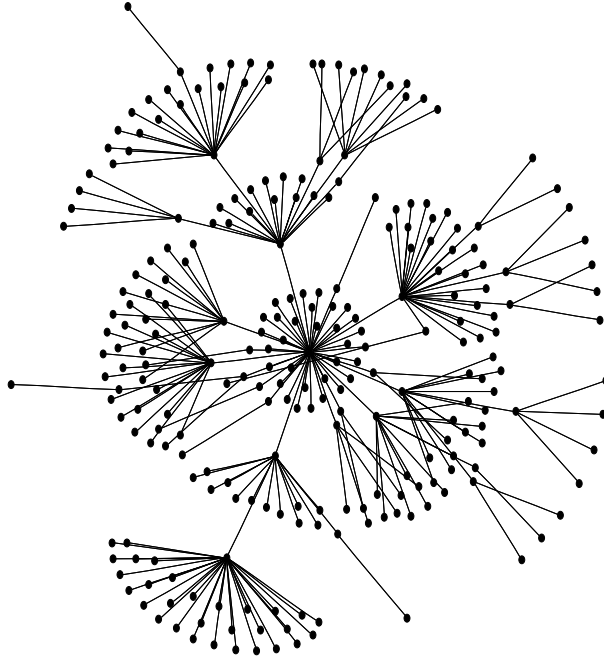


FIG. 7. An example of a medium-sized component of a software graph (extracted from the Java Development Kit 1.2 library). Here each node is a java class and edges indicate relationships among classes.