arXiv:0711.3199v2 [physics.soc-ph] 16 May 2008

Analyzing and Modeling Real-World Phenomena with Complex Networks: A Survey of Applications

Luciano da Fontoura Costa[1] Osvaldo N. Oliveira Jr.[1] Gonzalo Travieso[1]
Francisco Aparecido Rodrigues[1] Paulino Ribeiro Villas Boas[1]
Lucas Antiqueira[1] Matheus Palhares Viana[1]
Luis Enrique Correa da Rocha[2]
[1]Instituto de Física de São Carlos, Universidade de São Paulo
São Paulo, São Carlos, SP, PO Box 369, 13560-970,
phone +55 16 3373 9858, FAX +55 16 3371 3616, Brazil
[2]Department of Physics, Umeå University, 90187 Umeå, Sweden
`luciano@ifsc.usp.br`

May 19, 2008

Abstract

The success of new scientific areas can be assessed by their potential in contributing to new theoretical approaches and in applications to real-world problems. Complex networks have fared extremely well in both of these aspects, with their sound theoretical basis developed over the years and with a variety of applications. In this survey, we analyze the applications of complex networks to real-world problems and data, with emphasis in representation, analysis and modeling, after an introduction to the main concepts and models. Whenever possible, the main results in the studies surveyed are discussed within the context of the corresponding area. Details are given about the size of the networks, in addition to the measurements and theoretical models used to study the associated real-world problems. A diversity of phenomena are surveyed, which may be classified into no less than 21 areas, which is clear indication of the impact of the field of complex networks.

Contents

1 Introduction

From its earliest origins, science has lain somewhere between theory and application, abstraction and real-world. Driving strongly on mathematics, which provides the basis for formalization and systematization of knowledge, science delves deeply into natural problems, seeking for explanation and prediction. This bridge between theory and application has achieved such a level of integration and complementation that it becomes virtually impossible to separate one from the other. Instead, a synergistic integration is achieved where practical problems provide insights for new theories while theories suggest new ways of looking at Nature. The *Scientific Method*, illustrated in Figure 1, is key for progresses in science. Starting from real world observations or theoretical reasoning, the scientific method involves cycling through the stages illustrated in the figure, with special attention to the *validation* stage, which is the most distinguishing feature of the scientific method, involving confrontation between the simulation results/predictions and further experiments.

The impressive evolution of the field of complex networks fits perfectly within such a scientific framework. Its origins can be traced back to Leonhard Euler's solution of the Königsberg bridges problem (e.g., [1]), after which the theory of graphs has been useful for theoretical physics, economy, sociology and biology. However, most of such studies focused on *static* graphs, i.e. graphs whose structure remained fixed. Important developments on dynamic networks were addressed by Erdős and Rényi [2], among others, particularly for the so-called random networks, including the model now known as *Erdős and Rényi* — ER [3, 4, 5]. This type of network is characterized by the feature that in a network with N initially isolated nodes, new connections are progressively established with uniform probability between any pair of nodes. Such networks are well described in terms of their average degree, implying they have a relatively simple structure. Despite the formalism and comprehensiveness of the theoretical results obtained by Erdős and collaborators, random networks ultimately proved not to be good models for natural structures and phenomena. Indeed, heterogeneous structuring, not the relative uniformity and simplicity of ER networks, is the rule in Nature. Therefore, it was mainly thanks to the efforts of sociologists along the last decades (e.g., [6, 7]) that graph theory started to be systematically applied to represent and model natural phenomena, more specifically social relations. These efforts were to a great extent related to the concept of the *small-world* phenomena in networks, which are characterized by small average shortest path lengths between pairs of nodes and relatively high clustering coefficients. Interestingly, the small world property turned out to be ubiquitous. The next decisive development in graph applications took place quite recently, including Faloutsos *et al*. characterization of the Internet power law organization [8] and the identification of such a kind of connectivity in the WWW [9], giving rise to the *scale-free* paradigm [10]. Subsequent investigations showed that many natural and human-made networks also exhibited scale-free organization, including protein-protein interaction networks [11], domain interaction networks [12], metabolic networks [13], food webs [14], networks of collaborators [15], networks of airports [16] and roads [17].

The success of complex networks is therefore to a large extent a consequence of their natural suitability to represent virtually any discrete system. Going back to the scientific method in Figure 1, we note that most scientific advances until the 20th century, including several of the major achievements in physics, were largely characterized by reductionism. In other words, attention was concentrated on some specific parts (e.g., a particle, a molecule or cell) or aspect of a problem. Therefore, the 'real world' block in Figure 1 was typically restricted to a very small and isolated portion of the world. Great effort has been focused in isolating such portion of Nature from the environment. It is not coincidence that many research laboratories in physics, chemistry and biology are organized in that way, i.e. have as the main objective the isolation of some small subsystems from the rest of the world. By focusing attention and resources on isolated portions of the problem, while fully controlling the parameters involved (e.g., temper-

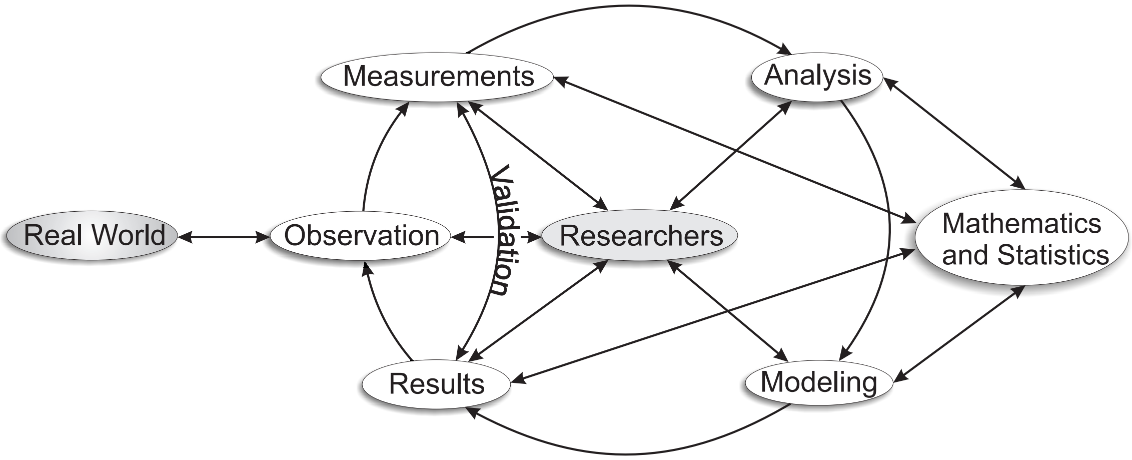

Figure 1: Schematic representation of the scientific method. Starting from real world observations, the scientific method involves cycling through the stages until a reasonable model is obtained.

ature, pressure, electric field, etc.), the reductionist approach was ultimately responsible for many scientific advances along the last two or three centuries. An interesting instance of the reductionist approach which is seldom realized is the fact that many researches have considered only *linear* models. Because many natural systems are non-linear, such problems were often treated by restricting the parameters and variables in terms of a linear approximation, which is itself a kind of reductionism (the linear approximation is usually valid only inside a small local region of the problem). One of the great trends in science, which has received growing attention, is the extension of the scientific method to become more integrationist and to deal explicitly with non-linear approaches. In this context, complex networks can pave the way for both such trends. First, they are naturally oriented to integrative, connectionist models of nature. Second, the organization and evolution of such networks, as well as dynamical processes on them [18, 19], involve non-linear models and effects.

The connectivity of networks is ultimately decisive in constraining and defining many aspects of systems dynamics. The key importance of this principle has been highlighted in many comprehensive surveys [18, 20, 21, 22, 23]. For instance, the behavior of biological neuronal networks, one of the greatest remaining scientific challenges, is largely defined by connectivity (e.g., [24, 25, 26]). Because of its virtually unlimited generality for representing connectivity in the most diverse real systems in an integrative way, complex networks are promising for integration and unification of several aspects of modern science, including the inter-relationships between structure and dynamics [27, 28]. Such a potential has been confirmed with a diversity of applications for complex networks, encompassing areas such as ecology, genetics, epidemiology, physics, the Internet and WWW, computing, etc. In fact, applications of complex networks are redefining the scientific method through incorporation of dynamic and multidisciplinary aspects of statistical physics and computer science.

This survey is aimed at reviewing in a comprehensive fashion the myriad of applications of complex networks, discussing how they have been applied to real data to obtain useful insights. In order to ensure a coherent, integrated presentation of the related works, the survey has been organized according to main areas and subareas. The survey starts with a section describing the basic concepts related to complex networks theory, such as measurements.

2 Basic Concepts

2.1 Network Representation

A *graph* or *undirected graph* G is an ordered pair $G = (\mathcal{N}, \mathcal{E})$, formed by a set $\mathcal{N} \equiv \{n_1, n_2, \ldots, n_N\}$ of *vertices* (or nodes, or points) and a set $\mathcal{E} \equiv \{e_1, e_2, \ldots, e_E\}$ of *edges* (or lines, or links) $e_k = \{n_i, n_j\}$ that connect the vertices [29, 30, 31]. In the Physics literature, a graph has also been called a network [23].[1] When the edges between pairs of vertices have direction, the graph is said to be a *directed graph*. In this case, G is an ordered pair $G = (\mathcal{N}, \mathcal{E}^{\rightarrow})$, where $\mathcal{N}$ is the set of vertices and $\mathcal{E}^{\rightarrow}$ is the set of ordered pairs of arcs (or arrows). If an edge $e_k = (n_i, n_j)$ is a directed edge extending from the node n_i to n_j, n_j is called the *head* and n_i is referred to as the *tail* of the edge. Also, n_j is a *direct successor* of n_i, and n_i is a *direct predecessor* of n_j. A walk (of length k) is a non-empty alternating sequence $n_0 e_0 \ldots e_{k-1} n_k$ of vertices and edges in G such that $e_i = \{n_i, n_{i+1}\}$ for all $i < k$. If $n_0 = n_k$, the walk is closed. A *path* between two nodes is a walk through the network nodes in which each node is visited only once. If a path leads from n_i to n_j, then n_j is said to be a *successor* of n_i, and n_i is a *predecessor* of n_j. A *cycle* is a closed walk that starts and ends at the same node, in which no edge is repeated. A graph $G^{\star} = (\mathcal{N}^{\star}, \mathcal{E}^{\star})$ is a *subgraph* of $G = (\mathcal{N}, \mathcal{E})$ if $\mathcal{N}^{\star} \subseteq \mathcal{N}$, $\mathcal{E}^{\star} \subseteq \mathcal{E}$ and the edges in $\mathcal{E}^{\star}$ connect nodes in $\mathcal{N}^{\star}$. If it is possible to find a path between any pair of nodes, the network is referred to as *connected*; otherwise it is called *disconnected*.

The intensity of connections can also be represented in the graph by associating weights to edges. Thus, the weighted graph $G^w = (\mathcal{N}, \mathcal{E}, \mathcal{W})$ is formed by incorporating, in addition to the set of $\mathcal{N}$ vertices and $\mathcal{E}$ edges, the set of $\mathcal{W} \equiv \{w_1, w_2, \ldots\}$ weights, i.e. real numbers attached to the edges. The weighted graph G^w can also be directed. In this case, instead of edges, the nodes are linked by arcs. Therefore, the most general graph is the direct, weighted graph $G^{w\rightarrow}$ [23].

The special category of *geographical networks* is characterized by having nodes with well-defined coordinates in an embedding space. In this case, the network $G = (\mathcal{N}, \mathcal{E}, \mathcal{D})$ incorporates additional information, given by the set $\mathcal{D} \equiv \{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_N\}$, where $\vec{p}_i$ is an $n-$dimensional vector which gives the position of the node i, generally in the $\mathbb{R}^n$ space.

Graphs can be represented by using *adjacency lists* or *adjacency matrices*. In the former case, the graph is stored in a list of edges (represented through head and tail). This data structure is frequently used to reduce the required storage space, allowing the use of sparse matrices. This list may also have a third element which represents the intensity of the connection. In the latter case, the graph is represented by an adjacency matrix A whose elements a_{ij} are equal to 1 whenever there is an edge connecting nodes i and j, and equal to 0 otherwise. When the graph is undirected, the adjacency matrix is symmetric. In order to represent weighted networks, a generalization of the adjacency matrix is required. In this case, weighted networks are represented by the so-called *weight matrix* W, where the matrix element w_{ij} represents the weight of the edge connecting the nodes i and j. The weight matrix can be transformed into an adjacency matrix through a thresholding operation, $\delta_T(W) = A$, associating each element of W whose value is larger than T with 1 (0 is otherwise associated) [23].

2.2 Network Measurements

In order to characterize and represent complex networks, many measurements have been developed [23]. The most traditional ones are the average node degree, the average clustering coefficient and the average shortest path length. The *degree* k_i of a node i is given by its

---

[1] Strictly speaking, in graph theory a *network* is a directed graph (digraph) with nonnegative capacities associated with each edge and distinguished source and sink vertices [31].

number of connections. Using the adjacency matrix,

$$k_i = \sum_{j=1}^{N} a_{ij}. \tag{1}$$

The *average node degree* is a global measurement of the connectivity of the network,

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i. \tag{2}$$

If the network is directed, it is possible to define, for each node i, its *in-degree*, $k_i^{in} = \sum_{j=1}^{N} a_{ji}$, and *out-degree*, $k_i^{out} = \sum_{j=1}^{N} a_{ij}$, as well as the corresponding averages considering the whole network. The total degree is given by $k = k^{in} + k^{out}$.

Another measurement related to connectivity is the degree distribution $P(k)$, which gives the probability that a node chosen uniformly at random has degree k. This has been found to follow a power law for many real world networks, as discussed further. For directed networks, there are two distributions, for ingoing links, $P(k^{in})$, and outgoing links, $P(k^{out})$. The *clustering coefficient* is related to the presence of triangles (cycles of order three) in the network [32]. The local clustering coefficient of a node i (with degree $k_i > 1$) is given by the ratio between the number of edges among the neighbors of i, denoted by e_i, and the maximum possible number of edges among these neighbors, given by $k_i(k_i - 1)/2$. Thus,

$$c_i = \frac{2e_i}{k_i(k_i - 1)} = \frac{\sum_{j=1}^{N} \sum_{m=1}^{N} a_{ij} a_{jm} a_{mi}}{k_i(k_i - 1)}. \tag{3}$$

The corresponding global measurement frequently used to characterize the graph is the *average clustering coefficient*, which is given as

$$\langle c \rangle = \frac{1}{N} \sum_{i=1}^{N} c_i. \tag{4}$$

The *length* of the path connecting the vertices i and j is given by the number of edges along that path. The *shortest path* (or *geodesic path*) between the vertices i and j is any of the paths connecting these two nodes whose length is minimal [33]. By considering the whole network, it is possible to represent the geodesic distances by a distance matrix D, in which the entry d_{ij} represents the length of the shortest paths between the nodes i and j. The *average shortest path length* is obtained from such a matrix,

$$\ell = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} d_{ij}, \tag{5}$$

where the sum considers $i \neq j$ and disregards pairs that are not in the same connected component.

All the measurements discussed above can be extended to weighted networks, in which case the *node strength* is defined with the weight matrix W [34],

$$s_i = \sum_{i=1}^{N} w_{ij}. \tag{6}$$

The average strength is defined considering the nodes in the whole network, i.e.

$$\langle s \rangle = \frac{1}{N} \sum_{i=1}^{N} s_i. \tag{7}$$

The *weighted clustering coefficient* of a vertex i can be defined as [34],

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j>k} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{ik} a_{jk}, \tag{8}$$

where the normalizing factor $s_i(k_i - 1)$ ensures that $0 \leq C_i^w \leq 1$. The average weighted clustering coefficient is given by,

$$\langle C^w \rangle = \frac{1}{N} \sum_i C_i^w. \tag{9}$$

The average shortest path length for weighted networks is determined similarly as in Equation (5), considering the weight of the edges. In this case, the *weighted shortest path length*, d_{ij}^w, is defined as the smallest sum of the edges lengths throughout all the possible paths from i to j [18].

Another important structural aspect of to complex networks characterization is related to the analysis of how vertices with different degrees are connected. The determination of the degree correlation can be achieved by considering the Pearson correlation coefficient of the degrees at both ends of the edges [35]:

$$r = \frac{\frac{1}{M} \sum_{j>i} k_i k_j a_{ij} - \left[\frac{1}{M} \sum_{j>i} \frac{1}{2}(k_i + k_j) a_{ij} \right]^2}{\frac{1}{M} \sum_{j>i} \frac{1}{2}(k_i^2 + k_j^2) a_{ij} - \left[\frac{1}{M} \sum_{j>i} \frac{1}{2}(k_i + k_j) a_{ij} \right]^2}, \tag{10}$$

where M is the total number of edges and ($0 \leq r \leq 1$). If $r > 0$ the network is assortative (vertices with similar degrees tend to be connected); if $r < 0$, the network is disassortative (highly connected vertices tend to connect to few connected); for $r = 0$ there is no correlation between vertex degrees, and the network is called non-assortative.

The measurements discussed above can be used for local analysis, in terms of node measurements, or global analysis, in terms of average measurements for the whole network. However, an intermediate analysis is possible by taking into account the modular structures in networks. Such structures, called *communities*, are common in many real networks, formed by sets of nodes densely connected among themselves while being sparsely connected with the remainder of the network [36, 37]. Communities play an important role in network structure, evolution and dynamics, defining modular topologies. Unfortunately, their identification is a NP-complete problem [38], so that many heuristic algorithms have been proposed for their identification [23, 36, 38].

Depending on the application, subgraphs can be fundamental to characterize network structures. For instance, modular structures in networks can be associated to different functions, such as scientific collaboration areas. In addition to communities, other types of subgraphs can be found in complex networks, such as motifs [39], cycles [40] and chains [41]. Motifs are subgraphs that appear more frequently in a real networks than could be statistically expected [39, 42] (see Figure 2, page 47). Other types of motifs, whose structure is not regularly defined, are said to be chain motifs (handles and tails) [41] or border trees [43].

A fundamental issue related to networks measurement is the sampling bias. If the data used to generate a particular network present a large quantity of noise or incompleteness, it becomes critical to consider measurements that are little sensitive to perturbations. Such measurements must reflect the differences in distinct networks structures [44]. This analysis is fundamental to complex network theory and constitutes a promising research field.

The consideration of multivariate statistical methods and pattern recognition plays a fundamental role in the understanding of network structure. For instance, the ability of a given model to reproduce real-world networks can be performed by canonical variable analysis with

Bayesian decision theory [23]. In addition, structures of networks can be properly characterized and classified by using these methods. For instance, the simplicity of networks can be determined by searching for nodes with similar measurements [45].

2.3 Network Models

In addition to the characterization of networks in terms of informative sets of measurements, it is important to construct models capable of reproducing the evolution and function of real systems or some of their main features. Among the many complex networks types, the models of Erdős and Rényi, Watts and Strogatz, and Barabási and Albert are among the most popular. The *random graph* of Erdős and Rényi (ER) uses what is possibly the simplest way to construct a complex graph: starting with a set of N disconnected vertices, edges are added for each pair of vertices with probability p [3, 29]. Consequently, the degree distribution follows a Poisson distribution for large N, with average degree $\langle k \rangle = p(N-1)$ and average clustering coefficient $\langle c \rangle = p$. Random graphs are simple but unsuitable to model real networks because the connectivity of these networks is characterized by heterogeneous connections and abundance of cycles of order three much larger than in random graphs [32].

The model developed by Watts and Strogatz, called *small-world networks*, overcomes the lack of cycles of order three in random graphs, but does not provide the non-uniform distribution of connectivity [32]. To construct a small-word network, one starts with a regular lattice of N vertices in which each vertex is connected to κ nearest neighbors. Next, each edge is randomly rewired with probability p. When $p = 0$ there is an ordered lattice with high number of cycles of order three but large average shortest path length, and when $p \to 1$ the network becomes a type of random graph.

In order to explain the uneven distribution of connectivity in several real networks, Barabási and Albert developed the so-called scale-free networks model, henceforth abbreviated as BA model, which is based on two rules: *growth* and *preferential attachment* [20]. The process starts with a set of m_0 vertices and, at each subsequent step, the network grows with addition of new vertices. For each new vertex, m new edges are inserted with some previous vertices. The vertices which receive the new edges are chosen following a linear preferential attachment rule, i.e. the probability of the new vertex i to connect with an existing vertex j is proportional to the degree of j, i.e.

$$\mathcal{P}(i \to j) = \frac{k_j}{\sum_u k_u}. \tag{11}$$

This evolution is related to the “the rich get richer” paradigm, i.e. the most connected vertices have greater probability to receive new vertices. Several networks are believed to be well-modeled by the Barabási and Albert approach, which means there is preferential attachment in these networks, as we shall discuss in the next sections.

Network models have been increasingly considered to investigate different types of dynamics. Indeed, the relation between the function and structure of networks is a fundamental issue that can lead to the understanding of many real-world phenomena, such as the association between biological networks and the products of such interaction or between society and disease spreading. In this way, many dynamical process have been studied by complex networks researchers, including synchronization [46, 47], spreading [48, 49], random walks dynamics [50, 51, 52], resilience [10, 53, 54], transportation [55] and avalanches [56]. A good review about dynamical process in complex networks is presented in [18].

3 Social Networks

Since ancient times, the way individuals establish relations among themselves has been crucial to guide the cultural and economical evolution of society. Hidden and clear relationships have always defined different social, diplomatic, commercial and even cultural networks. In various of those ancient networks, it was possible to qualitatively identify relevant structural properties [57], such as the importance of strategic individuals to intermediate or decide negotiations, or experience the power of ideological/religious thoughts within groups of people.

Though the quantitative study of social systems dates back to the seventeenth century [58], the systematic study of social relations using mathematical methods possibly began in the first decades of the last century with the study of children friendship in a school in 1926 [59], later followed by Mayo [6] with an investigation into interactions among workers in a factory. The motivation in understanding the "social networks" increased in the following years, especially with analysis of empirical data. We shall not review those early works, but they were important because methods were employed which are now adopted by the so-called complex network community. In addition, some of the problems now treated with complex networks had already been addressed. These included the phenomena described by Simon in his seminal paper on a model to generate highly skewed distribution functions [60], the study of citation networks by Price [61], which converged in a model of network growth able to generate power law degree distributions [62], Freeman's measurement of centrality (betweenness), which quantified the amount of geodesic paths passing through a node [63, 64], and the *small-world* effect, which emerged in the famous Milgram social experiment [65, 66], to name but a few. Further information and results from the sociological point of view can be found in specialized books [6, 7, 67, 68, 69] and papers in journals such as "Social Networks" [70].

Although sociometric research has contributed to the understanding of society, the collected data are still subjected to criticism because of the difficulty to define and associate intensities to some types of relation between two persons. Personal relations based on feelings, thoughts, trust or friendship deeply depend on the cultural environment, the sex and/or age of the actors involved and even on the current political and economical context. As an example, if the Milgram classical experiment [65, 66] had been done in Brazil, it would be needed to be redesigned since in this country most people are called by the first name and relationships are less formal. In other words, the concept of acquaintance seems to vary according to the country. In sexual networks, Liljeros *et al.* [71] suggested that man may overtell their sexual partners because of social expectations. In music, the level of similarity between two bands may be quite divergent if assigned by musical experts or enthusiastic fans [72].

Trying to overcome this bias, sociologists prepared extensive questionnaires and cross-compared the responses obtained to achieve reliable data. However, the interview process is time-consuming and expensive if one wants to get a significant sample size. On the other hand, researchers have also investigated systems, such as the collaboration and citation social networks, where the rules specifying the relation between the actors are quite clear, which guarantees some common ground for defining the network. Both types of approaches have benefited from the increase of World Wide Web popularity [73, 74]. The pleasure to be world visible, to contribute to global knowledge, to share thoughts, or only to make unusual friends or find partners have contributed to an increasing number of members in all types of virtual social environments. Tools like blogs, photoblogs, messengers, emails, social network services and even a complete social environment such as "Second Life" [75] are now widespread for all ages and genders [74, 76]. Although these virtual networks reflect only a piece of the world population and somehow specific types of relations, they usually provide a significant statistical sample and possibly unbiased features of the social relations they represent. In addition, there are extensive electronic databases about music, theater, sports, scientific papers and other fields which have contributed to construct reliable social networks faster and more accurately than ever [73, 76]. One needs nevertheless to be cautious when inferring behavioral and social conclusions from those specific networks, especially when dealing with dynamical variables.

In the following sections, we shall focus on describing the results on social networks by using mainly the methods adopted or developed by the complex network community in the last

10 years [18, 20, 22, 23], considering real-world data. In other words, we present the most important and common structural properties observed in several types of social networks such as degree distributions, community structure and the evolution of topological measurements, reporting their relation with social features when available.

3.1 Personal Relations

Personal relations are possibly the most important and oldest network type from the sociological point of view. Since people can establish contact with other individuals in uncountable ways, networks of this kind ultimately provide information about the structure of society. Personal relations can be divided into several classes ranging from friendship to professional life, including sexual [71] and trust networks [77, 78], or email [79] and blogs [80, 81]. In search of a universal behavior, the concept of personal relations was extended to other species such as wasps [82] and dolphins [83]. Since this topic includes the majority of social networks ever studied, we separate the various subjects into sub-sections for the sake of better organization.

3.1.1 Movie Actors

An important class of social acquaintances is related to professional actors participating in a movie. In terms of the artistic scene, such a network can provide a glimpse of the popularity of one actor along his career as well as individual fame. Despite the vast selection of movies in the Internet Movie Database (`http://us.imdb.com`), the film actors network is structurally small-world with high clustering coefficient. This means a well-defined tendency of actors to play with common partners [32]. Possibly because of the number of actors in a single movie and the number of movies an actor takes part along the career, the average degree is significantly large ($\langle k \rangle \sim 29$). The popularity of some actors and the short life time of the majority could explain the power-law regime ($\gamma \sim 2.3$) observed for large k in the distribution of actors partners [84]. Amaral and collaborators [85] pointed out that this distribution indeed is truncated by an exponential tail which could be an effect of aging, i.e., all actors naturally end their careers at some point. This conclusion is emphasized by the results of Zhang *et al.* [86] who argue that when considering multiple edges between actors, a stretched exponential distribution is fits to the data (the same functional form was used to obtain a good fit on a network of Chinese recipes [86]).

3.1.2 Acquaintances

Popular folklore is fascinated by the so-called “six degrees of separation” concept, resulting from the Milgram’s famous social experiment, which suggests that any two randomly chosen people are separated by six intermediate individuals on average. Since the experiment was done completely inside the USA, one of the remaining question was whether the six degrees also applied to the whole world. Taking advantage of the email system, Dodds *et al.* organized a similar worldwide experiment involving any interested person [87] and found, in accordance with Milgram’s results, that the diameter of social acquaintances varied between 4.05 and 7, whether only completed chains were considered or not (in the latter case, they estimated the value). The carefully organized experiment detected that successful chains depend on the type of relationship between senders and receivers, and not on the connectivity of the individuals. The most useful category of social tie was medium-strength friendships that originated in the workplace. Geography clearly dominated in the early stages of the chain, while occupation tended to dominate the final stages.

Although interesting, this kind of social experiment depends largely on peoples’ motivation to participate (being limited by lack of interest, time or incentive) [87]. To overcome this experimental bias, already-established ties of acquaintances have been usually investigated.

The nowadays very popular “blog” services were used by Bachnik *et al.*, who found power-law in out- and in-degree distributions (exponents ranging from nearly 2 to 3) and small-world property for three size-scale (number of members) providers [80]. The “blogs” databases are large, but are also considerably sparse, with most members isolated. On the other hand, Zakharov investigated the LiveJournal service and found a small-world network with a large giant component containing nearly $4 \cdot 10^6$ users [81]. Possibly the biggest social network ever investigated, this structure took 14 days and 2 computers to be crawled. He found a power-law ($\gamma = 3.45$) for $k > 100$ which corresponds to actively participating users, while not so active users with few connections led to a plateau in the distribution. The effect of opinion formation in a network like that would be particularly interesting to investigate since it has a high level of clustering (~ 0.3) and connection reciprocity (nearly 80% of the edges are bi-directional). However, this network has a limitation in the total number of friends per user (750) and only 150 of them are listed on the users’ info page, which clearly affects the flow of information. Upon using a new diffusion method, Zakharov identified some communities considerably fast [81].

Using phone calls records within a period of 18 months, involving approximately 20 per cent of a country’s population, Onnela *et al.* [88, 89] constructed a network with $4.6 \cdot 10^6$ nodes and $7 \cdot 10^6$ edges. The nodes represent users and a connection is established when two users called and received calls from each other. Moreover, the weights represented the duration of calls between two users. The degree and strength distributions could be approximated by power-laws (respectively, $\gamma = 8.4$ and $\gamma = 1.9$). A careful analysis of degree and weight correlations was carried out [88], in which strong ties appeared between members of circles of friends, while most connections between different communities were weak. As a consequence, the removal of strong ties has basically local effects (within communities). In contrast, when weak connections are removed, communication between different communities may be disrupted, causing collapse of the whole network. It is worth noting that information diffusion was tested in such weighted network using an equivalent SI (Susceptible Infected) epidemiologic model. They found that the majority of the nodes were first infected through ties of intermediate strength with a peak at 100 seconds [89], which means that most of the nodes were infected after 100 seconds.

A student affiliation network was built by Holme and collaborators, from officially registered classmates during a period of 9 years in Ajou University [90]. Considering the weight of the network as inversely proportional to the number of students attending the course (which is related to the social proximity between the students) and assuming that old courses contribute less than new ones to this social strength, they found that students become more peripheric with time in the network, whose core comprises mainly fresh students. Interestingly, fellow students become strongly tied over time owing to a decrease in the number of classmates. Using questionnaires data from middle and high school students in the USA, Gonzalez *et al.* analyzed the resulting friendship network [91]. They found that when considering only mutual connections (bi-directional edges), the network results in a set of various connected components. The friendship networks were identified as assortative with a degree distribution with a sharp cut-off. However, the c-networks built with connections between communities identified with the k-clique percolation method were disassortative, with scale-free degree distribution. From a comparison of node pairs with given ethnicities in relation to the random network, they found that the common behavior for each ethnic group is to nominate friends of the same ethnicity than from any of the other ethnicities. Interestingly, an asymmetry with respect to the composition was observed in case of inter-ethnic nominations. For instance, both blacks and whites show increasing homophily as they get into minorities; however, minor black groups get more integrated than the white ones when involved in other ethnic majorities [91].

3.1.3 Email

The establishment of networks incentived communications between individuals. Guimerà *et al.* obtained an email network for the University at Rovira i Virgili [79]. This intrinsically directional network was converted to an undirected version by assigning connections between two individuals whenever they sent and received messages from the same partner. As in the LiveJournal case, the network displayed small-world and high clustering (0.25) in the giant component, but with an exponential cumulative degree distribution ($P(k) \propto \exp(-k/k^*)$ with $k^* = 9.2$ for $k > 2$). On the other hand, the community structure was self-similar with heavily skewed community-size distribution $\gamma \sim 0.48$ in the range 2–100 and sharp subsequent decay [79, 92]. The self-similarity means that the organization is similar at different levels, i.e. individuals form teams, teams join to form departments, departments join to form colleges, and so on. Valverde *et al.* [82] constructed a network using data from the email traffic between members of Open Source Software Communities (OSS) and found that the weight distribution follows, on average, a power-law with two regimes ($\gamma \sim 1.5$ and $\gamma \sim 2.4$) in the case of small communities and with one regime $\gamma \sim 2.27$ for a large community [82]. The latter community also presented a power-law in the strength distribution ($\gamma \sim 1.73$).

3.1.4 Trust

From the hierarchical point of view, the network of trust can be viewed as a special sub-network of the acquaintances network, where the strongest connections in the latter could be related to connections of the former network. The lack of reliable data about trustful partnerships motivated the studies to concentrate on electronic trust ties. The PGP (Pretty Good Privacy) encryption algorithm lets one user certify another one by signing his public encryption key; in other words, the first user creates a directional connection to the second one if he trusts that the second is really what he says he is. By looking the emergent in- and out-degree, Guardiola *et al.* found a power-law degree distribution with exponents, respectively, 1.8 and 1.7 [77]. Interestingly, the network is composed of many strongly connected components in which the clustering coefficients are independent of the component size. The network was found extremely fragile against attack while the strongly connected structure remained essentially unaffected. Boguñá *et al.* considered only bidirectional signatures in the PGP web of trust [78] which resulted in a stronger sense of trust since both sides must sign each other's keys. The giant component presented a two regime power-law degree distribution with exponents 2.6 (for $k < 40$) and 4. The clustering is large and nearly constant as a function of the degree, but the network is assortative. Finally, using the Girvan and Newman community detection algorithm, a scale-free community distribution ($\gamma \sim 1.8$) was identified.

3.1.5 Sexual Relations

Even though sexual relations between individuals do not necessarily correspond to acquaintance relations, sexual partners can become social partners and *vice-versa*. Analyzing a random sample of individuals aged between 18 and 74 years old in Sweden, Liljeros *et al.* [71] found, in a study within a period of 12 months, a cumulative power-law distribution of partners with exponents 2.54 and 2.31, respectively, for women and men. The exponents are quite close, but the average number of partners is larger for men than for women, which is explained by the fact that men may inflate their number of partners because of social expectations. Interestingly, for the whole life-time another cumulative power-law distribution was found, but with smaller exponents, 2.1 (women) and 1.6 (men). The scale-free structure of such network was suggested to emerge from the increased skills in acquiring new partners as the number of previous partners also grows, in the same way of "the rich get richer" paradigm. The main result is that the core-group (i.e. groups of risk) concept may be arbitrary in such networks since there is no well-defined boundary between core groups and other individuals and therefore, safe-sex campaigns should be focused on highly connected individuals to prevent propagation of sexually transmitted diseases. However, by using statistical methods in the same and different data, Jones *et al.* [93] argued that the method used by Liljeros and collaborators to find the exponent of the number of partners distribution had statistical problems. As a consequence, the preferential attachment process may not be the only cause generating real sexual networks as proposed. They explained that targeting at-risk core groups has a proven efficiency in reducing disease incidence although degree-based interventions have already been proposed in the past and could be also adopted to lower the reproductive rate of sexually transmitted diseases. Finally, they suggested that other structural properties such as concurrency and local clustering have significant impact on epidemic processes since infinite-variance networks have different internal structures, affecting the spreading processes. In a review paper [94], Liljeros explores the complexity of the mechanisms contributing to spreading of sexually transmitted diseases, and concludes that one single solution for the problem is far from being found, even though the underlying structure of the sexual web contributes to this dynamics. Indeed, broad and targeted interventions have both been proven to be effective.

Finally, other networks were constructed using smaller datasets [95, 96] which described populations of different countries. The questionable scale-free behavior of such partners distribution has motivated interesting discussions on the internal structure [97] and dynamics of sexual interactions [96]. For instance, González and collaborators suggested a model to reproduce structural features of a homo- and a heterosexual network. They found that the clustering coefficient, the number of triangles and squares are relatively small, which can be better explained with their model than with the BA model [96]. Also worth analyzing is the network of romantic communication, though it is not necessarily related to sexual contacts. Holme and collaborators [98] observed that during the period considered in their study, both the number of members and the average degree grew with time but with decreasing growth rate. This is partly explained by the fact that old members log on for the first time during the sampling period. Reciprocity depends on the type of relationship established in the virtual community and is rather low. Furthermore, the number of triangles is smaller than the number of 4-circuits and the network apparently presents dissasortative mixing. The cumulative degree distributions was observed to be highly skewed, but not with pure power-law form [98].

3.1.6 Sports

The sports field is particularly interesting because of its dynamical nature, as teams constantly change players, there are dozens of collective sports involving interaction between people, championships range from local to world size scale, and so on. However, little attention has been given to such dynamical systems in terms of networks. As one of the tests of their community detection algorithm, Girvan and Newman investigated the community structure of the United States college football in the 2000 season [37]. The nodes represented teams connected by edges expressing regular-season games between those teams. Although teams are divided into conferences which imply more games between those members, interconference games are usual, which generates a complex structure involving all participants. With their algorithm, they were able to identify the conferences with high precision. The north-American college football league was also a motivation for a ranking system based on a directed complex network where the direction represented wins or losses of a specific team (the nodes) [99]. The method was based on the idea that if a team A can beat a team B and B can beat a team C, consequently, A most probably will beat C. The resulting ranking for a specific season was compared with the official method with good agreement.

A complex network study of Brazilian soccer players was carried out by Onody *et al.* [100]. Initially, a bi-partite network was built where one type of node represented the teams while the other represented the players. They found an exponential law $P(N) \sim 10^{-0.38N}$ in the probability that a player has worked in a given number of clubs N. To investigate the topological properties, they merged those nodes in such a way that two players became connected if they were in the same team at the same time. The final network exhibited an exponential degree distribution such that $P(k) \sim 10^{-0.011k}$. The time evolution of that network showed that although its size became more than five times larger in the period from 1975 to 2002, the network maintained the small-world characteristic. They suggested that the clustering coefficient had a small decrease because of the exodus of a considerable number of players in the last decades while the average degree became 20% higher, possibly because of the increase in the transfer rate and/or in the professional life-time. The network became more assortative with time, suggesting a growing segregationist pattern where teams with similar importance preferentially transfer players between themselves.

3.1.7 Comics

Differently from the actors network where the social ties are professional, the comics network are composed by characters whose connections can be constrained by other factors such as moral rules, as pointed out by Gleiser [101]. Therefore, though both networks are related in some level, they have fundamental differences which can be captured by structural measurements.

The social relation between Marvel characters was investigated in terms of a bi-partite network (characters and books) where two characters were connected through the appearance into the same comic book, with connections representing both “professional” and familiar ties. Although the Marvel universe tries to reproduce human relations, Alberich and collaborators concluded that the clustering coefficient and average degree showed a considerable smaller value than usual real social networks [102]. The resulting network has a small diameter. Perhaps as a result of personal fame, Captain America and Spider Man are examples of hubs which contribute to the power-law with cutoff observed in the distribution of the number of partners in this collection of stories [102]. Contrasts between this specific artificial network with real ones were suggested by the results of Gleiser [101]. The analysis of degree correlations indicated no correlation up to $k = 200$ and disassortativity after this value. The clustering coefficient as a function of the degree clearly showed the existence of a hierarchy of nodes. In order to extract the intensity of relations between the characters, a weighted network counting the number of times two characters appeared together was adopted, leading to a power-law distribution with ($\gamma \sim 2.26$). He also found that the hubs are basically heroes that connect different communities. The fact that villains appear only around hubs can be a result of the Comics Association rules, which state, for instance, that the good has always to succeed against the evil. These constraints imply that this artificial social network differs from a real social net.

A different network related to comics stories was built by considering the cross-talk or comic dialogue (Xiangsheng) folk art of China [103]. The network is small, but exhibits interesting features. It is shown that most of the comic dialogues involve two players and the number of dialogues an actor has played follows a power law ($\gamma \sim 2$). The projection onto a non-bipartite network showed that the collaboration is highly clustered and could contain a hierarchical, modular structure. Moreover, the network has some disassortativity and is small-world. These results suggested that the main actors tend to repeat dialogues with the same collaborators.

3.1.8 Non-human relations

While investigating networks of wasps in which the weights were given by the number of dominances (hierarchy) of one wasp over the other, Valverde and collaborators found networks with weight distribution similar to the small OSS communities [82]. Another non-human social experiment was carried out by Lusseau [83], where social acquaintance of dolphins was defined as preferred companionship, i.e. individuals that were seen together more often than expected by chance. The data took 7 years of observations and showed that the emerging network is small-world with high clustering coefficient. The data is relatively small, but it is argued that the tail of the number of acquaintances distribution follows a power-law ($\gamma \sim 3.45$). The dolphin network is resilient to random attacks, but targeted attacks increase the diameter of the network, although the random attacks are not enough to fragment the network into small connected components. This effect is possibly a consequence of the non-power-law interval for a small number of acquaintances. The hubs were identified to be mainly adult, old females [83].

3.2 Music

Professional music relations could be included in the section of *personal relations*, but we have singled music out because different types of networks have been studied. In terms of professional relations, Gleiser and Danon [104] proposed a network where two musicians were connected if they played in the same jazz band. This connection mechanism resulted in an assortative, small-world network with high clustering and degree distribution following $P(k) \sim (1 + 0.022k)^{-1.38}$. For the information collected dating back to the 1930's and 1940's, interesting social aspects were extracted from the network. Studying the community structure with the Girvan and Newman algorithm, they found a clear segregation pattern of black and white musicians in this collaboration network and that the cumulative distribution of community sizes presents a power-law with $\gamma = 0.48$ [104]. The network of collaboration between rappers was constructed by Smith [105] considering two rappers connected whenever they recorded together. The resulting network is also a small-world with high clustering and a partnership distribution following a power-law ($\gamma = 3.5$).

No correlations were found between topological measurements such as betweenness and node degree, or between an index of record sales and starting release year, i.e. high level of collaboration is not related to commercial popularity. Differently from the jazz music network, the community analysis used a weighted network and then a "clearing algorithm" to convert the weighted network to an unweighed version such that the most important edges were extracted [105]. Considering only the connections, the fast modularity community structure algorithm [106] was applied and could only identify small and peripheral rap groups. When the clique percolation method was applied [107], groups and geographical regions were identified correctly but the same did not apply to music labels. The level of communities (groups, music labels, regional/community affiliation) identified could be controlled by a parameter in the clearing algorithm and then applying the clique percolation method. Both jazz and hip-hop networks are relatively small, with considerable high average degree.

Another collaborative network was built by Park *et al.* [108] who used data extracted from the `allmusic.com` database. They constructed a network of similarity between artists using the same database and investigated the topological properties of both networks and the resulting networks obtained when intersecting them, i.e. the networks obtained from only common nodes in both networks. The method is interesting for comparison of the same dataset for different social ties. However, the overall properties did not change considerably, and the collaboration network before and after the intersection process presented power-law $\gamma \sim 3$, while the similarity network had exponential degree distribution with $P(k) \sim \exp(-0.12k)$ (before) and $P(k) \sim \exp(-0.15k)$ (after intersection). The small-world and high clustering features were maintained after the intersection process. Interestingly, only 464 common edges (about 4% of the total) were identified in the intersection of both networks which indicates that having worked together does not necessarily translate into being classified as musically similar. The similarity network was assortative while the collaboration was partly assortative. The cumulative fraction of betweenness displayed a power-law in all networks studied.

The similarity between artists was extensively investigated by Cano *et al.* [72] by using four online databases which differ in the way similarity is assigned (by user habits or musical knowledge and by musical experts). Those networks are larger than the musicians networks above, though preserving the small-world property with high clustering. They found that user rating networks resulted in power-law in-degree distributions $\gamma \sim 2.4$, while experts classification mechanisms resulted in networks with exponential decay. The out-degree distributions follow the same shape with cut-offs due to clear limitations in the web-pages usability constraint (the recommendations should fit in one web-page). The power-law behavior for the degree distribution was confirmed in an experiment where users sent playlists such that two artists were connected if they appeared together. An exponential decay emerged in the case of a user selecting the most similar artist to a given one in a list of 10 possibilities [72].

The Brazilian popular music network was constructed by connecting two song writers with a common singer [109]. The network presents small-world feature, high clustering and high average connectivity. The cumulative degree distribution is fitted by an exponentially truncated power-law $\gamma = 2.57$.

At the level of personal ties, a network of jazz bands was constructed where two bands were connected if they had a musician in commonding to another segregation between black and white musicians [104]. That network presented small-world, high clustering, with the degree distribution following an exponential function $P(k) \sim \exp(-(k/32.8)^{1.78})$. Using the Girvan and Newman community detection algorithm, they found a polarization in two big communities representing the major records locations, namely, Chicago and New York. The New York community split up into two other communities corresponding to another segregation between black and white musicians [104]. In these networks, an example of errors in the acquired data is the inclusion of the same musician appearing with different names, as pointed out in the jazz network [104]. Similarly, errors occur in the Brazilian musician network, which includes anonymous musician, that concentrate a large number of connections [109].

3.3 Collaborations

The topic of scientific collaboration can be understood as a sub-topic of personal relations, since it is related to professional (academic) ties between scientists. It deserves a special section for its relevance in terms of knowledge dissemination. From the several databases investigated, common features were identified, even though the sizes and some structural properties depend on the field of interest. Newman built networks from databases of sizes varying from 10^4 to 10^7 papers and compared the topological properties in the same time interval. The distribution of number of papers is well fitted by a power-law for two databases ($\gamma \sim 2.9$ for Medline and $\gamma \sim 3.4$ for NCSTRL) and by an exponentially truncated power-law for the Los Alamos Archive [15, 110]. The evolutionary computation (EC) co-authorship network also had a power-law ($\gamma \sim 2$) [111]. The distributions of number of authors displayed a power-law shape in all four databases investigated by Newman ($2 \lesssim \gamma \lesssim 6$) and in the EC data ($\gamma \sim 5.27$), with the largest collaboration project containing 1681 people. Purely theoretical papers appear to be the work of two scientists on average, according to results for mathematics and sociology. The latter were also found to contain a small number of collaborations although suggesting an increase in the collaboration rate in the last decades [112, 113]. On the other hand, experimental and interdisciplinary (e.g. neuroscience) subjects have higher average collaboration rate, reaching about 9 authors per paper in high-energy physics.

The histogram of the number of collaborators per author fits a power-law in many databases with exponents varying from $\gamma \sim 2.1$ to $\gamma \sim 2.58$ [111, 112, 113]. The only difference appears in some networks investigated by Newman, which exhibited two power law slopes ($\gamma \sim 2$ and $\gamma \sim 3$). Interestingly, Cotta and collaborators pointed out that the most prolific authors are not necessarily the most connected, but in fact they have diverse interests which motivate collaborations. They also found that the number of collaborators strongly correlates with the number of papers [111].

Only the network of sociologists has been found not to be small-world, possibly because of the large number of research areas with little collaboration between them. The conclusion is reinforced by the small clustering observed [113]. The Medline database presents small clustering, as a result of the tree-like structure ("principal investigator" with several post-docs) akin to biomedical research [15, 110]. The other databases are characterized by a high degree of clustering [111]. The clustering coefficient decays and the degree of an author increases with time in the case of neuroscience and mathematics [112]. The growth of the giant component of these two latter subjects is consistent with the increasing collaboration rate over the last years. Newman verified that for most authors, the paths between them and other scientists go through just one or two of their collaborators (funneling effect). The average distance to all authors decreased with increasing number of collaborations [15, 111, 114]. In terms of knowledge diffusion, the average distance is important because it measures the centrality of an author in terms of its access to information. On the other hand, betweenness is a measure of author's control over information flowing between other scientists [15, 114].

Girvan and Newman tested their algorithm to detect communities in a database of co-authorship of papers, books and technical reports of the Santa Fe Institute between 1999 and 2000 [37]. They found that scientists are grouped into two types of similarity, either of research topic or methodology. The same method was used by Arenas and collaborators to identify communities in two networks of co-authors [92]. In the first, they considered the panels contributions of Statistical Physics conferences in Spain over a period of 16 years and took the final co-authorship network. They found that collaboration is more common within the same institution and that the emergent community size distribution follows a power-law ($\gamma \sim 1$). The arXiv network has a two-slope power law in the same distribution with exponents $\gamma \sim 1$ (with strength $s < 60$) and $\gamma \sim 0.5$ ($60 < s < 1000$) [92]. The community in collaboration network was also addressed by considering hierarchical measurements in a scientific institutional collaboration network [115]. In this case, the authors identified different patterns of authorship emerging from different research areas.

Li *et al.* considered a different weight function while investigating the collaboration network between econophysics authors from 1992 to 2003/2004 [116]. The nodes represented scientists and the weights between two scientists were obtained by summing three functions, where each function represents one type of relation between them. The first counts the number of co-authorships, the second counts the number of citations and the last, the acknowledgements. Each function expresses a different contribution to the final weight value according to their relevance in the final report. They found that the degree and weight Zipf plots have an undefined shape and did not change when considering two snapshots (with one year of difference between them) of the collaboration database, though individual nodes changed considerably their importance in the network (measured by their betweenness) [116, 117]. The network has a small giant component (with about 25% of the total nodes), small clustering and large average shortest paths when compared to other collaboration networks. They found that the weight affects slightly the edge and node betweenness distributions. By using only the amount of collaboration between co-authors, they investigated the community structure in the giant component of such a network. They used the Girvan and Newman algorithm and hierarchical clustering. The first algorithm provided the best results, classifying scientists belonging to the same university, institute or interested in similar research topics. The second one classified modules, but the result was not consistent with real data. They found that the members of each community changed considerably depending on whether the weights are taken into account [118].

Barabási and collaborators suggested that for the evolution of the collaboration network, the small and finite time interval might affect the results leading to incorrect conclusions because of incomplete data. Such a trend was identified by observing that the average shortest path decreased with time (and size of the network) while it was expected to increase [112]. Although important, proceedings editorship cannot be seen in the same level of peer-reviewed papers. Cotta and collaborators argued that proceedings collaborations should be removed since they bias considerably the shortest path distribution, as they include authors from different thematic subjects and create long-distance edges. They also suggested to analyze separately the distinct types of scientific reports since different types of collaboration are reflected in such reports [111]. Some care should be taken when considering the original data because two authors might appear with the same name or one author might identify himself in different ways on different papers. To overcome this problem, Newman ran the measurements twice, first considering the network obtained by the author's surname and first initial only, and a second version considering the surname and all initials [110].

3.4 Religion

The importance of religion in the ancient and contemporary society is unquestionable; despite some fluctuations along the years, religion always played a central role in the history of humanity. Choi and Kim used a Greek and Roman mythology dictionary to analyse the relation between the mythological and fascinating characteres of that time [119]. Using a directed network, they associated outgoing links with a specific entry of a character in the dictionary and the incoming links were related to characters appearing in the corresponding explanation field of that specific character. It means that Heracles, for instance, has the most number of outgoing links which reflects that he gave origin to many great events by himself. On the other hand, Zeus concentrates a high number of incoming links since he appeared as a supporting character in different myth tales. The degree distributions resulted in power-laws with exponents between 2.5 and 3.0. The close relation between different characters is visible by looking the small-world feature and the high clustering of common neighbors. Furthermore, a hierachical structure was observed, which had two main fundamental origins: genealogical tree and the native class a myth character belongs, e.g., gods, titans, heroes. The distribution of local cyclic coefficients displayed two peaks that correspond to tree-like and triangular patterns [119].

In the medieval church, there was a big concern about heresy, which is fundamentally an opinion at variance with established religious beliefs professed by a baptized church member. Since heresies were against the official beliefs of the church, their propagation was punished and avoided. The inquisitors recognized that they should target the most connected people, i.e. the people responsible of propagating those ideas, which resulted on much more effective results [57]. A qualitative analogy of historical facts and disease spreading on networks led Ormerod and Roach to suggest that the spreading of medieval heresy resembles a disease diffusion in a scale-free network [57].

At an individual level, Amaral *et al.* analyzed a group of 43 Mormons and suggested that a Gaussian distribution could explain their acquaintance network [85]. This result agrees with another friendship network of high school students, also mentioned in the same paper [85].

3.5 Organizational Management

The growth of interdisciplinary research can be identified by data related to national or continental research funding proposals. Based on data from the National Natural Science Foundation of China (NSFC), Liu and collaborators constructed a network of related research areas [120]. Since each proposal to the foundation provides two fields to specify different areas, they took each research area as a node, with two nodes being connected if they appeared in the same proposal. The cumulative degree distribution followed the function $P(k) =$

$\exp(-0.13k)$ for the period between 1999 and 2004. The growth of interdisciplinarity could be identified by the increase in the average degree and clustering, and by the average distance decrease. Interestingly, the network tends to be disassortative with time.

At the level of organizations, Barber *et al.* [121] and Lozano *et al.* [122] constructed networks based on the European-Union Framework Programs (FP) on Research and Technological Development. In the first case, they got a bi-partite network from the first four FPs, with one node-type representing the research projects and the other type representing the organizations. The connections were assigned by considering organizations which collaborate in common projects. Network properties were extracted from the intersection graph. The distributions of project (number of organizations taking part in one project) and organization (number of projects in which an organization takes part) sizes exhibited power-laws, respectively, with $\gamma \sim 4$ and $\gamma \sim 2$ for larger values, which remain stable over the three last periods considered. This indicates that the organizations participating in a particular number of programs in each FP were not altered in spite of changes in the research activities. The network is highly clustered, small-world, and the number of triangles increases linearly with the degree of the node [121]. Using data from the 6th FP, Lozano *et al.* constructed a different network where two organizations were connected whenever they had collaborated at least in two projects. The degree distribution presents an undefined shape and the analysis focused on detecting communities by a new algorithm, which was able to classify some communities according to their type (industries or research centers), nationality or services providers [122].

The network of the United States House of Representatives from the 101st to 108th Congresses was investigated by Porter and collaborators [123]. Initially, they built a network with two types of nodes which corresponded to Representatives and committees, with edges connecting each Representative to the committees. Then, another network was created whose nodes represent the committees while the edges indicate common membership between committees. They found evidence for several levels of hierarchy within the network of committees and identified some close connections between committees without incorporating any knowledge of political events or positions. They identified correlations between committee assignments and political positions of the Representatives. Finally, they verified that the network structure across different Congresses has changed, especially after the shift in the majority party from Democrats to Republicans in the 104th House.

The networks of product development for different industries were characterized by Braha and Bar-Yam [124, 125, 126]. In these networks, two tasks are connected by arcs if the first task feeds information to the second task. The resulting network was highly clustered with small average distance, which reflects the optimization of the networks and unavoidable iteration nature of the design process. They found that in-degree distributions always presented a scale-free behaviour with cutoff, while out-degree distributions were scale-free with and without cutoff, which can be related to the fact that transmission of information is often less constrained than reception [124, 125, 126].

4 Communication

The study of the structure and function of social networks has always been constrained by practical difficulties of mapping the interactions of a large number of individuals. The construction of these networks is based on questionnaire data, which reaches only a few number of individuals and depends on the personal opinion about their ties. With the advent of the Internet and with the use of phone and mobiles, a large amount of data can be recorded for further analysis. These communication networks are useful not only for providing better data for the study of social networks but also for economical reasons. Here we discuss the email, call graph, and wireless networks.

4.1 Email

Email has become one of the most important means of communication nowadays and email exchanges provide plentiful data on personal communication in an electronic form which leads to an automatic way to construct social networks. Since email networks are one of the major means of computer virus spreading, the main reason to study them is computer security, i.e. finding ways to stop virus spreading.

There are two ways to construct email networks: (i) vertices are email addresses and there is a directed arc from vertex i to vertex j if i sent at least one email to j (this network is obtained by the log files of email servers) [127, 128, 129]; (ii) vertices are also email addresses but there is a directed arc from vertex i to vertex j if j is in the address book of vertex i (this network is obtained by email address book of several users of a specific institution) [129, 130]. Therefore former kind of network is naturally weighted and directed, where the arc weights are given by the number of emails sent. Ebel *et al*. [127], however, also studied these networks by considering them as with undirected arcs where email addresses are connected if at least an email was exchanged between them.

All collected email networks were obtained from university email servers [79, 127, 129, 130] except the email network from the HP Labs mail server [128]. These networks can be very large, with the largest being the email network of the Kiel University [127] with 59 812 vertices, and the smallest the email network of University at Rovira i Virgili in Tarragona, Spain [79], which contains 1 667 vertices. All studied email networks have the small-world and the scale-free properties [79, 127, 128, 129, 130]. Tyler *et al*. [128] and Gimerà *et al*. [79] also found that email networks are composed of communities and proposed methodologies to find them. Braha and Bar-Yam [131] showed that vertex degree and betweenness of such kind of networks change dramatically from day to day, suggesting a reinterpretation of “hubs” in dynamical networks.

A model of an evolving email network was proposed by Wang and Wilde [129] and it is based on addition and deletion of links between users, and on the user email checking time. Zou and Gong [132] proposed a model for simulating the email virus spreading and immunization, which is also based on the user email checking time and his/her probability of opening email attachments. They showed that viruses spread more quickly on scale-free networks than on small-world and random networks, and that the immunization defense is more effective on the first kind of network than on the other two.

There is a free software, called EmailNet, for email traffic mining which generates the corresponding email network for further analysis. This software also provides a interface for the visualization of the email networks and is available at `http://ipresearch.net/emailnet/` [133].

4.2 Telephone

Call graphs are constructed from telephone calls completed during a specified time period. The vertices are the telephone numbers, and a connection from vertex i to vertex j is present there was a call from i to j. The edge has naturally a direction, with i at the tail and j at the head. Such graphs can be very large. Aiello, Chung and Lu [134] found power laws for the in-degree and out-degree distributions, as well as for the distribution of components with a given size (excluding the giant component) in a telephone call graph. A similar study was done for mobile phones by Nanavati *et al*. [135]. They analyzed local calls in four different regions for two different periods (one week and one month); despite social differences in the regions and the different periods used, the results are similar. Average degree varied from 3.6 to 8.1, while the clustering coefficient was from 0.1 to 0.17. The distribution of in-degrees and out-degrees is close to a power-law, with exponents 2.9 and 1.7, respectively. There is a significant correlation between the in-degree and the out-degree of vertices, which means that people that receive many calls also generate many calls. With regard to nodes connected by a call, assortativity is present for the in-degree of the caller with in- and out-degree of the called person; for the out-degree of the caller, weak disassortativity with in- and out-degree of the called was detected. Expressed in another way, vertices that receive many calls tend to call vertices that also receive and generate many calls, and vertices that receive few calls tend to connect with vertices that receive and generate few calls; on the other hand, the number of calls generated by the caller is not a good predictor for the connectivity of the called person. A giant strongly connected component and a power law distribution of the remaining components were also detected. Onnela *et al*. [88, 89] considered the call graph of a mobile operator, but included only reciprocal calls in the graph, i.e. an undirected edge is present between nodes i and j only if i called j and j called i during the considered time interval; calls for numbers of other operators were not included. The largest connected component of this network has a degree distribution with a tail of the form $P(k) = a(k + k_0)^{-\gamma}$, but with significantly larger exponent ($\gamma = 8.4$) than the networks discussed above. The number of calls between two telephone numbers and the total duration of the calls were used as two different weighting features. The distribution of strengths shows that most users make few, brief calls, but some make a large number of calls, with some pairs of users chatting for hours. The network is assortative, confirming previous results for a social network. The authors also found that the overlap of the neighbors of two vertices is proportional to the weight of their connection, therefore corroborating experimentally the “weak ties hypothesis” of Granovetter [136].

5 Economy

Trade, currency, industrial production, wealth distribution and tourism are important target studies of economy. These systems can be considered as formed by discrete parts that interact in a defined way. For instance, trade/commerce is the voluntary exchange of goods, services, or both. In this way, a network can be generated by this activity associating nodes to countries and considering the in- and out-degree as representing, respectively, the imports and exports between these countries. In this section, we present how complex networks theory can be used to model economical relationship.

5.1 Trade networks

Economic relations are now magnified owing to the increasing number of commercial partners due to globalization. Trades at a personal level have substantially increased with e-commerce, but the gross market between countries is still a business of companies. Trying to unveil this fascinating economical complex system of trades between countries, Serrano and Boguñá studied the so-called world trade web [137]. A network was built by assigning a node to each country and considering the in- and out-degree as representing, respectively, the imports and exports between countries. For the non-weighted directed network, they found the in- and out-degree correlation to be very high ($r = 0.91$), with 0.61 of reciprocity. A power-law function emerged with exponent $\gamma \sim 2.6$ for $k > 20$ in the number of commercial partners, regardless of whether the in-, out- or undirected-degree distribution were considered. The intense trade activity between countries exhibits the small-world property, a high clustering coefficient, and large average degrees. A positive correlation between the number of trade channels and the country’s economical wealth was identified, with some exceptions for developing countries. Interestingly, the system seems to present a hierarchical architecture of highly interconnected countries belonging to influential areas, which in turn connect to other influential areas through hubs. This last property is suggested by the strong clustering coefficient dependence with the degree (power-law behavior with $\gamma \sim 0.7$) and the disassortative degree structure [137].

In another paper, Serrano and Boguñá [137] showed that the topology of the "World Trade Web" presented scale-free, small-world effect and modular structure, similar to the Internet and WWW networks.

5.2 Currency

In 2004 Li *et al*. [138] extended the studies of Serrano and Boguñá to “World Exchange Arrangements Web,” with a method to build a bipartite network where the nodes can represent countries or currencies. In the network, one country is linked to one currency if the currency circulates in the country. The authors showed that the network possesses a scale-free behavior with exponent $\gamma = 1$. A non-bipartite network was also analyzed where the nodes were the currencies only. Two currencies were linked if there was one or more countries where these currencies circulate. This network also presented scale-free behavior with coefficient 1.3, small-world effect, disassortative and modular structure. The power law behavior of the degree distribution had also been observed by Gorski *et al*. [139] in the complex network built from FOREX database (the largest financial market in the world). The robustness of currency complex networks was investigated recently by Naylor *et al*., who used hierarchical methods to show that the currency network topology minimizes the effects caused by economic crises, such as the Asian crisis from 1 August 1997 to 31 October 1998.

5.3 Industry

Andrade *et al*. [140] showed that the topology of the oil refineries network is scale-free. In order to build the network, the authors took devices and unitary processes (e.g., valves, pumps, tanks) as nodes, which are linked by pipes. They observed a small-world effect and hierarchical organization on two networks studied, and argued that the network topology may provide a useful tool to design, characterize and evaluate refinery plants.

5.4 Wealth

The effects of network topology on the evolution of a dynamic process have been widely studied. For example, Souma *et al*. [141] built a model to investigate the wealth distribution with networks with small-world effect. The authors considered a multiplicative stochastic process and studied the effect of deletion and rewiring of edges in the wealth distribution. Their results indicated two phases for the wealth distribution, depending on two parameters, the probability of rewiring edges and fraction of rewiring edges. The distribution was log-normal for the first phase and power law for the second phase. It had an intermediate state characterized by log-normal distribution with a power law tail, which is also observed in real world economy. The break of the wealth clusters also occurred if the edges were rewired. A similar study was performed by Matteo *et al*. [142] using an additive stochastic process. The authors showed that the shape of the wealth distribution was defined by the degree distribution of the network used. The results were consistent with real data of income distribution in Australia.

5.5 Tourism

Tourism comprises a multitude of activities that form one of the world’s fastest growing international economic industries. The sector is so strongly characterized by intense linkages through which informational exchanges occur, that it seems natural to apply network theoretical methods to its study [143, 144]. Nonetheless only quite recently quantitative techniques have been applied, mainly in analyzing tourism destinations, complex localized clusters of public and private companies and organizations [145]. The major characteristics of tourism destination networks have been measured both from a static and a dynamic point of view. The results showed that topological measurements can highlight their main structural features such as product clusters, structural divides and central organizations [146, 147]. The underlying social and economic system evolutionary history has been also assessed by using network dynamical growth models [148]. In general the topology of the networks is comparable to the one exhibited by similar systems: a marked scale-free structure. However, some differences are found, mainly due to the relatively poor connectivity and cauterization. These results are interpreted by considering the formation mechanisms and the peculiarities of the economic sector and of the real cases examined. Clustering and assortativity coefficients are also proposed as quantitative estimations of the degree of collaboration and cooperation among the destination stakeholders [146]. Moreover, the network has been used to perform a series of numerical simulations to investigate possible scenarios of information and knowledge diffusion among the components of a tourism destination. It has been shown that the highest improvement in efficiency for this process is obtained by optimizing the network in order to increase the degree of local clustering [149].

Tourism networks have also been addressed by analyzing the relation between their structure and dynamics. Costa and Baggio [150] studied the Elba (Italy) tourism destination network by complex networks topological measurements and considered dynamical process over the network, i.e. the inward/outward activations and accessibilities, according to the superedges framework [28]. They showed that the type and size of the companies has a strong influence over their respective activations and accessibilities. On the other hand, geographical position of companies tend to not influence in such dynamical aspects. By the characterization of the tourism networks structure, the authors concluded that Elba tourism network is fragmented and heterogeneous.

6 Financial market

A corporation, by definition, has a group of owners, called shareholders, that share its stocks. A stock representing part of the assets and profits of a company can be bought and sold in a stock exchange, and any person can in principle own part of a publicly traded company. Stock prices are constantly changing, following unstable market and political conditions all around the globe. Thus, the financial market is a highly complex, evolving system, difficult to grasp and predict, being sensitive to economical instabilities such as that of Black Monday (October 19, 1987), when the Dow Jones Industrial Average (DJIA) decreased in an alarming way (this index measures the performance of companies in the USA stock market). The organization of the financial market may be better understood by representing this system as a network. Given the high accuracy of financial data available, it is possible to construct networks that reliably reflect the real market, in contrast to other fields that lack high quality data, such as the Internet and social networks. Financial networks can be constructed, for example, from stock prices or stock ownerships. In the first case, each pair of nodes (stocks) is connected to each other by a weighted edge that encodes the distance between two stocks. This distance can be computed as a function of the correlation coefficient, taken between the time series of the prices of the two stocks. A hierarchical structure is frequently obtained from this complete network, which is usually a minimal spanning tree, also called in this context “asset tree”. A minimal spanning tree is a connected subnetwork with no cycles, which includes every node of the original network with minimal cost (i.e. minimal sum of edge weights). On the other hand, a network of stock ownership encodes the stockholders of a group of companies. It can be a directed network indicating who is the owner of a stock and who is the company that shares the stock, where the owner can be an individual or even another company. Examples of the characterization of these two types of networks are discussed in the following.

Mantegna [151] studied the stocks traded in the New York Stock Exchange (NYSE) from

1989 to 1995, and obtained a taxonomy that organizes stocks according to economical activity. A network of stocks was computed, where each pair of nodes was connected by a weighted edge that encodes the distance between two stocks. This distance was calculated as a function of the correlation coefficient, taken between two daily time series of stock prices, as explained in Section 6. The hierarchical structure was identified in terms of the minimal spanning tree of the network of stocks. Bonanno *et al.* [152] employed the same methodology to study stock prices in the NYSE, recorded from 1987 to 1998. The authors compared the topology of the minimal spanning tree resulting from real data with an analogous one obtained from simulated data using market models. A power-law for the real data was observed, with exponent 2.6, a feature not captured well in simulations, since high and low degrees differ significantly. This means that the hierarchical structure obtained from real data is not captured by the stock price models.

Onnela *et al.* [153] obtained a similar exponent ($\gamma = 2.1$) to the one computed by Bonanno *et al.*, although during crash periods the exponent changes, such as in the Black Monday, where $\gamma = 1.8$. The authors used daily stock prices recorded from 1980 to 1999 in the NYSE. The asset tree underwent other particular changes in periods of crisis, such as the decreasing in its length. Onnela *et al.* [154] investigated another network obtained from correlations of stock prices. In this case, the network is not necessarily a tree, being constructed selecting only the $N - 1$ edges with lower weights of the complete network (note that a minimal spanning tree also has $N - 1$ edges). This network, generally called “asset graph”, does not clearly exhibit a power-law degree distribution. Moreover, results indicate that the asset graph is more robust and stable than the asset tree, since the former changes less than the latter in consecutive periods and also changes less under extreme conditions. However, the asset graph does not necessarily have a hierarchical organization, thus preventing a taxonomic analysis of the financial market. Furthermore, Onnela *et al.* [155] focused on the details of the construction of the asset graph, and measured how the clustering coefficient changed in this process. The authors were interested in finding what portion of the edges in the asset graph represent information, and not noise. They found a different behavior of the clustering coefficient in the empirical graph, in comparison to a random graph, and this observation was applied to estimate that only 10% of the edges in the asset graph convey real information.

Battiston *et al.* [156] considered a network of European firms where a directed edge (i, j) exists if firm i is a stockholder of firm j, and its weight is the amount of investment from i to j. Another version of this network was taken by joining firms whose headquarters are in the same region. The firm network shows a power-law distribution of investments, close to the distribution of firm activities, thus showing a correlation between investment and activity. On the other hand, the regional network has a log-normal distribution of activity, investments and degree, though again investment and activity are correlated. Garlaschelli *et al.* [157] studied shareholders of stocks traded in the NYSE, in the Italian stock market and in the NASDAQ (National Association of Securities Dealers). A network description of investors and assets was used, where companies were linked to investors (through a directed edge) that held some of their shares. Note that this is not a bipartite network since a company can be a shareholder of another company. The three datasets showed power-law distributions for both node in-degree (called portfolio diversification) and node in-strength (sum of the weights of ingoing links, called portfolio volume). Souma *et al.* [158] characterized a similar directed network, this time considering only Japanese companies along with shareholding links. Six datasets were used, ranging from 1985 to 2003, and for all of them a scale-free distribution of out-degrees was observed, with varying exponents along the years of the data.

7 Computer Science

Computer science is a relatively recent area, as the establishment of the first departments and award of the first degrees occurred in the 1960s [159, 160]. While its fundamental topics involve computer architecture, data storage and processing, and system control (programming and algorithm developing) [161], it has become essential for developments in a variety of areas. For computer science provides tools to help solving problems of physics, mathematics, chemistry, biology, medicine, economy, etc. For instance, the major problems in biology, e.g., protein folding, function prediction, phylogeny, and modeling of biological systems (see Section 13) can only be solved with computers and optimized algorithms.

7.1 Software Architecture

The importance of computers and software to our lives is unquestionable. Almost everything we do needs a computer and specific software. The cost of software development increases with its complexity, and may exceed a million dollars (e.g., computer games, compilers, and operating systems). The sub-area of software engineering is aimed at providing methodologies and tools for designing and building software efficiently, which can be achieved by decomposing a problem into many small, distinct but interlocking pieces, named software components [162, 163]. Deciding the size of the software components is difficult and must be planned carefully. Levels of granularity are defined for these components, e.g., subroutines, classes, source files, libraries, packages, etc. There are several ways to represent software as a complex network depending on the size of the software components. The topology of networks and the hierarchical relationships among the software components were studied by Valverde, Solé and Cancho in [162, 164], where the nodes were program classes and the edges were the inheritance and composition among the classes. Networks for various pieces of software were found to exhibit the same organization pattern, including eMule, Openvrml, GTK, VTK and the Linux Kernel. According to the authors, the networks studied display scale-free topology and small-world features. Furthermore, software architecture networks can be generated by a local optimization process, instead of preferential attachment or duplication-rewiring. Myers [165] found similar results for six software architecture networks VTK, CVS, abiWord, the Linux Kernel, MySQL and XMMS), where the distribution of incoming and outgoing links is scale-free.

The same applied to the Linux Kernel, Mozilla, XFree86 and the Gimp networks [163], in which the networks were built in a different way. These software packages were written in the C/C++ computer languages, which have two main types of files, “.c” (or “.cpp”) that are the source files and “.h” the headers files. The former contains the source code of the program and the header files contain the definition of variables and constants and description of data structures. Each source file includes (through the "#include" clause) a certain number of header files. The authors considered the header files as nodes of a network, with two nodes being connected if they were included in the same source file. The topology of these networks is scale-free and displays the small world effect even when the hubs are deleted.

Challet and Lombardoni studied software components networks [166] for Linux packages and found clear evidence of the asymmetry between incoming and outgoing links distribution. While the first presents a scale-free behavior, with exponent $\gamma = 2.0$, the latter does not. The authors also studied bug propagation in these networks with a similar model to *Susceptible, Infected and Resistant* of disease spreading. In the case of one faulty node with propagation only to the next neighbors, it is easy to fix the problem with a debug process. In contrast, fixing the problem is much harder when non-local fault propagation occurs, as in an illegal memory access. In other papers on software networks, Shannon’s information entropy was used to measure structural complexity [167], and scale-free distribution and small-world effects were found for networks built from Internet based repositories of open software [168]

7.2 Data Sharing

Iamnitchi *et al.* [169] considered a scientific file sharing network, where nodes are scientists and two scientists are connected if one of them is interested in the data of the other. Based on the small-world topology of scientific collaboration networks, where edges are created between researchers that are co-authors in a paper, the authors claimed that scientific file sharing networks may be also small-worlds, and thus they proposed a mechanism of data location that takes advantage of the presence of local clusters. Indeed, three data sharing networks (a physics data sharing community, WWW data sharing between Internet hosts and Kazaa traffic between users) have small-world properties [170, 171, 172]. Leibowitz *et al.* [172] studied Kazaa traffic in 2003, a period where Internet traffic was already dominated by peer-to-peer applications. Data of an Israeli Internet Service Provider (ISP) were employed, which showed that Kazaa traffic is much concentrated in the transport of a few popular files, suggesting that caching can be a solution to decrease redundant traffic. Moreover, some of these popular files lose their popularity in a few days, whereas other files remain constantly popular (at least in the time period considered by the authors).

7.3 Spam Filtering

A network of email contacts was used by Kong *et al.* [173] to create a spam filtering technique. It is a distributed collaborative system that relies on queries made between neighboring users (i.e. users that exchange emails) to decide whether a suspicious message is a spam. Thus, the system is based on a “trust” algorithm considering spams already identified by other users. Simulations showed a spam detection rate near 100% with almost no false positives.

7.4 Circuits

The small-world effect was observed by Ferrer i Cancho *et al.* [174] in networks of electronic circuits, where each node represented an electronic component (e.g., integrated circuits, resistors capacitors) and edges were wires in a broad sense. The authors showed that the degree distribution of this network follows a power law and the most common degree is $k \approx 4$, corresponding to four neighbors in 2D plane.

Another example of circuit network was used in 2004 by Barabási et. al [175]. They used a network formed by a circuit of a microprocessor Simple12 to show that the intensity of the electric current fluxes on the edges of the network varies in a wide range. Up to four orders of magnitude. The authors also showed that the standard deviation and the average value of the flux are related by the power law $\sigma =< f >^{0.5}$, and that this relation appear in other real transportation networks, with distinct exponents.

7.5 Wireless

In wireless networks, mobile stations (mobile nodes) are connected to each other through access points to the Internet. Each mobile node can connect at different access points at different times. Hsu and Helmy [176, 177] studied the patterns of connectivity of mobile nodes to access points, and constructed a network based on encounter between nodes. Two nodes are said to encounter each other if there is an overlap of their connections to the same access point. Each mobile node is a node in the constructed network, and a link is created between two nodes if there is an encounter between them in a specified interval of time. The authors found the network to be sparse, with nodes having encounter with just a few neighbors in average (with a node connected to about 1.88% to 5.94% of all the nodes, depending on the data set considered), following a bi-Pareto distribution. The graph has short average distances and high clustering coefficient. The high clustering is attributed to the fact that nodes have a “home” access point, and all nodes with the same access point have a high probability of being connected, giving rise to cliques. The average distance is small due to the presence of some nodes in these cliques that connect to nodes in other access points, generating a small world effect.

7.6 Image processing

Image processing is based in analysis, characterization and classification of digital images, which involves acquisition, pre-processing, extraction of features and characterization, and pattern recognition [178]. Some of these steps are particularly difficult to be implemented. For instance, segmentation techniques required to identify objects in images can involve complex algorithms, to the point that sometimes such an identification may becomes impossible.

Since images are composed by adjacent pixels, complex networks theory can be applied for their processing and analysis. The network can be constructed by associating pixels to nodes and connecting them according to some criteria, such as image neighborhood and similarity of the local properties of the pixels (e.g., gray-level values). Costa [179] suggested an image segmentation approach by considering complex networks theory. An image is mapped into a network and a community identification method is applied in order to obtain the objects. Since objects can have different gray levels, the pixels belonging to same objects tend to be densely connected whereas pixels in different objets are sparsely connected, defining respective network communities [180].

Other image processing methods can also be implemented by considering complex networks methodologies and concepts, such as color segmentation, texture analysis [181] and image classification. In the latter case, a image would be represented by a network and respective measurements would be extracted. Then, discriminant analysis can be applied to classify images according to the obtained network measurements. In addition to images, video can also be addressed in similar fashion, where each video frame can be characterized independently as an image, while interconnections can also be considered between successive frames.

8 Internet

The origins of the Internet can be traced back to 1969, when a military networking system called ARPAnet was created [182]. It has since evolved into a worldwide computer network no longer restricted to academy, through which many services are provided, with email and the World Wide Web (WWW) being the most popular. Thus, since the Internet is such a multipurpose integrated tool, it is important to understand how it globally works to keep it safe in case of failure or attack, or to improve its performance. A straightforward approach to study the Internet is to represent it as a graph, with hosts, routers and servers being the nodes and the physical links connecting them (optical fibers or copper cables, for example) being the edges.

Complex networks researchers have employed a graph-based approach to model the Internet structure and simulate its traffic. The complete mapping of the Internet is difficult to achieve, because it is not centrally administrated and changes constantly. Thus, researchers usually employ coarse-grained maps that contain only the links between Autonomous Systems (AS), which are subnetworks separately administrated, or use maps including only the connections at the router level. Nevertheless, such incomplete graphs have allowed important findings about the structure and dynamics of the Internet, as discussed below.

Faloutsos *et al.* were possibly the first to study power-laws in the Internet [8]. Three snapshots of the Internet at the AS level, collected between 1997 and 1998, and one instance at the router level, collected in 1995, were examined. Power-laws were found for the degree distribution, the degree rank, the number of pairs of nodes within h edges (called "hops") and the graph eigenvalues. Faloutsos *et al.* employed AS maps constructed from BGP (Border Gateway Protocol, an inter-domain router protocol) routing tables, an approach that was criticized by Chen *et al.* [183], which used instead BGP and IRR (Internet Routing Registry, a routing database). Siganos *et al.* [184] extended the study of [8] and found that those power-laws hold for more snapshots of the Internet, ranging from 1997 to 2002, as well as for the more complete data of [183]. BGP-based AS maps from a three year interval (1997-1999) were also verified to be scale-free [185, 186], and the same applied to a dataset considering the routes traced from only one host to any other address in the network [187].

Bianconi [188] assessed the number of cycles of length 3, 4 and 5 in 13 snapshots of the Internet at the AS level, collected between 1997 and 2001, and found a scaling behavior of these quantities over time. Moreover, cycles of size 5, called pentagons, are more frequent in these AS maps than in random networks with the same degree distribution. Therefore, pentagons could be regarded as characteristic motifs of the Internet.

The correlation profile, i.e. the comparison of the degree correlations of a network with its null (randomized) version, was used by Maslov *et al.* [189] to identify differences in networks with the same degree distribution, such as the Internet at AS level (in 2000) and molecular networks. A power-law degree correlation function, that gives the average degree of the neighbors of a node with degree k, was found in another map of the Internet, which is probably a consequence of a hierarchical structure [185].

The rich-club phenomenon was identified in an AS network, where the highest degree nodes are well connected with each other [190]. This property has also been addressed not only considering the degree, but across hierarchical degrees [191]. Such a tendency to have a strong core in the Internet was also reported in [192, 193], where the nucleus of the Internet, i.e. its innermost k-core (a k-core is the largest subgraph whose nodes have at least degree k), is a well connected cluster with approximately 100 nodes and diameter of 2. The authors used an extended data set, combining BGP-based AS data with the one obtained in the DIMES project [194], and divided the network into three parts: (i) the aforementioned nucleus, (ii) a fractal connected component that is not in the nucleus but encompasses the largest part of the Internet and (iii) dendrite-like components, which are connected only to the nucleus. The decomposition of the Internet was done using a method called k-shell decomposition. The k-shell is a set of nodes that contains all the nodes with degree equal to k, which are recursively removed from the network, where the decomposition is made by creating k-shells starting with $k = 1$ until $k = k_{max}$. The modularity of the Internet was also investigated using spectral analysis of AS level networks. Eriksen *et al.* [195] found modules that approximately match single countries, and Rosato *et al.* [196] observed that the Internet has large and highly clustered regions joined by a few links.

Peer-to-peer (P2P) overlay networks (virtual networks at application level that use the Internet infrastructure) were characterized recently. A power-law degree distribution was identified in a snapshot of the connected hosts of Gnutella taken in late 2000, although a more recent snapshot of early 2001 deviates from a pure power-law [197]. The P2P network eDonkey also displayed a power-law for in- and out-degree distributions [198]. In this case, a directed edge (i, j) exists if a host i makes a query related to a file provided by j. As a guide for the improvement of Gnutella, it was found that it does not use efficiently the underlying Internet: although 40% of its nodes are inside the top ten ASs, less than 5% of Gnutella's connections join nodes inside the same AS [197].

One common analysis of the Internet involves the study of its vulnerability to failure or attacks. The vulnerability of complex networks, including the Internet at AS level, was studied in [10] with regard to random node removal (node failure) or hub removal (network under attack). In this case, the diameter and size of connected clusters were monitored while nodes were removed. Results showed that the scale-free networks (BA model, Internet and WWW) are extremely efficient against random failures, differently from the ER model. However, the hub-based attack rapidly breaks scale-free networks into small isolated groups. In [54] attacks were performed in complex networks (including the Internet at the AS level) removing nodes with high degree or high betweenness centrality, both calculated only at the beginning of the attack (i.e. in the original network) or at each step of attack. Similar strategies were employed for edge removal. At each node or edge removal, the average inverse geodesic length and the size of the largest connected subgraph were obtained. It was shown that the recalculated betweenness is the most harmful to the Internet considering edge-based attacks. For node-based attacks, the different strategies are equally harmful to the Internet. The average inverse geodesic length was also used in another study [199], where it was called "efficiency". The drop in efficiency when a node is removed was calculated for every node of Internet backbone maps of USA and Europe. The authors identified the nodes with the highest efficiency reduction, which was a possible explanation for the problems the Internet users faced on September 11, 2001. Another interesting observation is that these important nodes do not necessarily have high degrees.

A scaling behavior of the form $\sigma \sim \langle f \rangle^{\gamma}$ between the mean flux per node $\langle f \rangle$ and the flux dispersion per node σ in the Internet router topology was reported by Menezes and Barabási [200], where the flux is the transmission rate of data packets. To explain the observed scaling law the authors used diffusion of random walkers and transfer of packets along shortest paths. The results pointed to an scaling exponent of the Internet ($\gamma \simeq 0.5$) reflecting an endogenous behavior. That is to say, Internet is rather more affected by internal decisions, like the origin and destination of a packet, than by external agents such as the variation of the number of Internet active hosts. In [201] a method is presented to separate internal from external dynamics. The processing capacity of nodes was also considered by Duch and Arenas [202] to study Internet traffic, using data similar to the one employed in [200]. The authors assumed a queue model with Poisson packet arrival and exponential processing time (per node). Varying the parameter of this model and the time window for measuring data, they showed that γ can vary significantly for the Internet.

Centers of activity were identified in the French-side of the Internet, in a scientific network called Renater [203]. In this case, the exchange of data between all routers was considered, even if they were not physically connected, and the Random Matrix Theory (RMT) was applied in a matrix of correlations between edge flows (measured in bytes per 5 minutes) to identify the most correlated connections. The Internet has a high heterogeneity of connection quality between hosts separated by the same distance [204]. This result was measured using the Round-Trip-Time (RTT) and the geographical distance between pairs of hosts over the Internet. As for the network growth, it was discovered that in 2002 the overall Internet performance, measured in RTT per unit distance, was improved in comparison with the year 2000 [204]. In addition, new links were preferably placed among old nodes than new nodes in the Internet, which indicates that redundancy plays an important role on its growth [185].

Rosato *et al.* [196] performed simulations of traffic flow in a Internet map, and noticed that congestion is reached before in the Internet than in a random network. The authors argued that the hubs can be responsible for this behavior, since these nodes participate in a large number of transmissions between other nodes. Using a model of self-regulated packet generation with different levels of routing depth, Valverde and Solé [205] simulated Internet's traffic and detected a critical path horizon that separates low-efficient from highly-efficient traffic. This transition happens when the size of the routing tables approaches the network diameter. Furthermore, the authors hypothesized that the Internet is working close to this critical path horizon.

When delivering data packets, routers in the Internet choose neighboring routers to obtain an estimation of the shortest path between origin and destination. However, this procedure does not take into account the load present in each router, which can lead to unexpected processing delays in routers or even congestion in extreme cases. Echenique *et al.* [206] showed in simulations with an AS map of 2001 that if the queue load of routers is taken into account, along with the cost of paths, the routing protocol of the Internet can be improved. Moreover, the authors noticed that an increased clustering coefficient can also lead to improved efficiency as congested nodes can be avoided when alternative routers in the neighborhood are available. Krioukov *et al.* [207] studied the scaling properties of Internet routing. A good scaling happens when the size of routing tables grows slowly (e.g., logarithmically). However, the routing in the Internet is far from ideal, since for optimal routing it is necessary to maintain a global view of the evolving network in each router. In this case, good scaling of table sizes is, in principle, impossible. Although the Internet needs substantial routing modifications, simple solutions may arise, as Boguñá *et al.* [208] analytically proved: using only local information nodes can work efficiently, especially in networks with power-law degree distribution and high clustering coefficient, such as the Internet. Tadić *et al.* [55] also argue that it is not possible to substantially improve the efficiency of Internet traffic by considering global information. More specifically, the authors stated that neighborhoods at a maximum distance of two are sufficient. Local information was also highlighted by Thadakamalla *et al.* [209] as important parameters for transport strategies in geographical scale-free networks such as the Internet.

An accurate network model should contain the same characteristics of its real counterpart. Each Internet model usually reflects a few of these features, which leads to a variety of complementary models [210]. Yook *et al.* [211], for example, introduced a model based on empirical observations of the router and AS levels of the Internet. Their model includes incremental growth, where the probability of connecting a new node i to an older node j is linearly proportional to the degree k_j and to the inverse of the distance d_{ij} between them. Moreover, positions of nodes form a scale-invariant fractal set, positively correlated with the population density. Park and Newman [212] analytically studied ensembles of networks with the restriction proposed in [189], i.e. that the maximum number of edges between a pair of nodes is one. This mechanism leads to the negative degree correlation (disassortative mixing) reported by Pastor-Satorras *et al.* [185], which relates the average degree of the neighbors of a node to a given degree k. Nevertheless, Park and Newman argue that this approach is not responsible for all the degree correlations verified in the Internet.

Rosato and Tiriticco simulated the growth of the Internet at AS level [213] using a mechanism of triad formation introduced in [214]. This mechanism uses preferential attachment to create the first edge that connects a new node i to an existing node j. Moreover, node i is sequentially connected to the neighbors of its first assigned neighbor j, thus building a "triangle", or to any other node following preferential attachment. This model leads to a scale-free network with adjustable clustering coefficient. The authors found that simulations of this model fits Internet data from 1998 to 2000, for quantities such as diameter, clustering coefficient, degree distribution and average geodesic length.

Zhou and Mondragón modeled the Internet at the AS level using an interactive growth mechanism and nonlinear preferential attachment [215]. Their approach, called positive-feedback preference (PFP) model, is also based on data features such as the initially slow growth of node degree and the establishment of links between old nodes. Simulations of this model showed that it can reproduce many characteristics of the Internet, including the degree distribution, the rich-club phenomenon and the disassortative mixing. A numerical analysis was carried out by Zhou [216] to identify which mechanisms of the PFP model are responsible for the reprodution of each Internet's topological feature.

Serrano *et al.* [217] considered in their model the number of users (hosts) in each AS and the capability of adaptation of ASs according to their size. The authors assumed exponential growth for the number of users, the number of ASs (the nodes) and the number of links between ASs (the edges of the network), all inspired by empirical observations of Internet data [34]. Another important feature of this model is that ASs can have connections strengthened by increasing their edge weights. Furthermore, the model can optionally include a distance constraint to avoid connections between distant ASs. The parameters used for the model led to a power-law degree distribution with exponent $\gamma = 2.2 \pm 0.1$, which is in agreement with previous works [8, 185]. The model explains the average shortest path length, the clustering coefficient and the k-core decomposition of the Internet.

Internet models have also been compared by considering multivariate statistical methods, as proposed by Rodrigues *et al*. [210]. The authors suggested the consideration of canonical variable analysis and bayesian decision theory to determine which model is most suitable to reproduce the Internet structure. Such approach allows one to consider a large set of measurements in order to obtain an accurate description of the network structure. Rodrigues *et al.* analyzed Internet models at the autonomous systems level and concluded that none of the considered models were able to reproduce the Internet structure. The proposed approach is suitable to determine which model is more suitable to any type of complex networks.

9 World Wide Web

The World Wide Web (WWW, or just Web) is perhaps where Internet users spend most of their time, despite the existence of other highly popular Internet applications such as electronic mail. The WWW relies on the HyperText Transfer Protocol (HTTP), which allows a Web browser running on a client computer to communicate with Web servers distributed over the Internet. Web servers usually contain hundreds (sometimes many more) files that can be accessed by anyone with a Web browser connected to the Internet. These files can be virtually any computer file, such as those containing text, image or video; nevertheless, the files that give the WWW a networked structure are the pages, which can contain pointers (called hyperlinks) to other pages available in the WWW. Thus, one can think of the WWW as a giant network of pages interconnected by hyperlinks. These files are the most important of the WWW because they are multimedia files that group rich text (i.e. with formatting options), images, videos and even computer programs (e.g., scripts and applets).

The WWW viewed as a network is an interesting subject for complex networks researchers and for those whose main concern is the Web itself [218]. It is huge, with billions of pages, despite its relatively recent birth (released in 1993 to public access), and has an almost uncontrolled growth mechanism, where individuals or organizations create their interconnected pages depending solely on their will. Grasping the WWW organization and how users jump from page to page while navigating (surfing) the Web is crucial to optimize search engines and facilitate information access. Mainly because of its uncontrolled nature, the WWW is difficult to map completely, even for sophisticated search engines [219]. Usually a WWW map is constructed by a computer program called “crawler” that navigates through pages storing their hyperlinks, i.e. identifying who points to whom. Unfortunately, many pages are not considered by the crawler for reasons such as authorization requirements to obtain a page, dynamic pages that require filling a form, crawler limitations, broken links or unreachable Web servers. Nevertheless, the Web viewed as a network has been studied by many researchers. Usually, Web maps are viewed as directed networks, where each edge points from one node to another, i.e. the page that contains the hyperlink to the referenced page. Sometimes the Web is not considered at the page level, but at the site level, where each site (roughly a group of pages under the same domain address) is considered as a node, and pairs of sites are connected whenever a hyperlink is found between some of its pages.

In a number of papers the WWW was reported as a scale-free network [9, 84, 220, 221]. With a network of 325,729 pages taken from the nd.edu domain, Albert, Jeong and Barabási [9, 84, 220] verified that the out- and in-degree distributions followed a power-law of the form $P(k) \sim k^{-\gamma}$, with exponents $\gamma_{out} = 2.45$ and $\gamma_{in} = 2.1$, respectively. Broder *et al.* [221]

and Donato *et al.* [222] used a considerably bigger Web map, with 200 million pages, and obtained $\gamma_{in} = 2.1$, in agreement with previous results. Pennock *et al.* [223] divided the distribution of pages by category, such as the set of company websites or the set of newspaper websites. Although the in-degree distributions of these subnetworks displayed the heavy tail characteristic or power-laws, the frequency of nodes with low degree k_{in} significantly deviates from a pure power-law. Nevertheless, other power-laws were found in the WWW for the distribution of the sizes of connected components [221] and for the distribution of pages in a site [224]. Moreover, self-similarity was verified in the WWW using the nd.edu map [225]. In this case, the network was divided into boxes of nodes of size l_b, where every pair of nodes in a box is separated by a minimum distance $l < l_b$, with the WWW displaying self-similarity at different scales. It was found that the number of boxes required to group the network and the size of the box also follows a power-law. Furthermore, the PageRank, which is a webpage ranking used by Google based on a random surfer model [226], follows a power-law with exponent 2.1 [222]. PageRank is not statistically correlated with the in-degree, which means that Webpages with high in-degree (i.e. referenced by many other pages) do not necessarily appear at the top of a Google search. The small-world effect in the WWW was observed by Adamic [227] at the site level (with approximately 250 thousand sites). Hence, this WWW map presents high average clustering coefficient and low average distances between nodes. Bianconi *et al.* [228] developed a theory to assess if a directed network has more or less loops than a random network. Using the nd.edu map, the authors observed that the WWW has more short loops than its randomized version, which includes loops of length 3 (i.e. the clustering coefficient). In another experiment, random versions of observed data that preserve the original degree distribution were used to estimate the average geodesic length $\langle d \rangle$ of the entire WWW [9, 220]. The size of the WWW was thought at that time (late 1990s) to be $N = 8 \times 10^8$, which gave $\langle d \rangle \simeq 19$, a relatively small value. The authors predicted a logarithmic growth of $\langle d \rangle$, indicating that the average distance between nodes in the Web grows slowly with N. However, Broder *et al.* [221] verified that $\langle d \rangle$ is very different in their Web map of 200 million pages, with $\langle d \rangle \simeq 500$. Moreover, they found that a path between randomly chosen pairs of nodes exists in only 24% of the times, thus confirming the absence of a small-world effect in their data. Even in the biggest strongest connected component (a central group of pages) they obtained $\langle d \rangle \simeq 28$.

Before Bianconi *et al.* [228] studied cycles in the Web, Caldarelli *et al.* [229] performed a similar study in some WWW maps. The latter introduced the grid coefficient, a measurement that quantifies the level of cycles of length n of a network. The authors also characterized cycles of length 4 (called quadrilaterals) in a number of networks, including the WWW map of the nd.edu domain. This Web map had far more quadrilaterals than its randomized counterpart, which indicates that some order of regularity, represented by grid-like structures, is present in this network (a similar behavior was found in the Yeast protein interaction network and in a scientific collaboration network). Moreover, the authors noticed that the grid coefficients follow a power-law distribution in these networks and in an AS map of the Internet. This observation leads to the claim that the WWW and the other networks have a hierarchical arrangement of cycles. Broder *et al.* [221] reported an intricate organization of a WWW map collected in 1999, which was divided into four groups of roughly similar size: (i) the largest strongest connected component (SCC), where every page can be reached starting from any other page; (ii) the IN group, which contains pages that can lead to SCC (even indirectly), but cannot be reached from SCC; (iii) the OUT group, with pages that do not lead to SCC, but can be reached from SCC; and (iv) the TENDRILS, with pages separated from SCC that can only reach IN and OUT. Donato *et al.* [222] investigated the size of these components in a WWW map collected in 2001, which yielded very different values than the ones reported by Broder *et al.* Other modular structures in the WWW were also reported in the literature, regarding groups of pages about similar topic. In [230] communities of related topics were found based on the maximum-flow/min-cut theorem and in [231] groups of connected pages with high curvature (a geometrical interpretation of the clustering coefficient) were associated with specific topics.

The analysis of WWW-based social networks can be performed by considering networks of personal pages, such as the blogs where hyperlinks are frequently created between pages of friends or acquaintances. Some connectivity properties between 200,000 pages of a Chinese blogspace revealed a scale-free degree distribution and the small-world effect [232]. Indeed, Milo *et al.* [233] showed that the WWW and social networks are included in the same superfamily, based on the significance profile of motifs (this profile is defined in Section 17.2). This observation may be useful to boost the understanding of the WWW by using models of social organization.

Another network related to the Web, though not built using pages as nodes, was characterized by Shen and Wu [234]. This network, called folksonomy, uses semantic tags created to describe the content of a database of hyperlinks (bookmarks). A folksonomy represents the connectivity between tags, where two tags are linked if they are associated with the same content. The employed data of 9,804 tagged bookmarks exhibit scale-free and small-world properties.

The dynamics in the WWW is usually associated with user surfing, i.e. the sequence of pages a user (or a group of them) visits when following hyperlinks in the WWW. Dezsö *et al.* [235], for example, divided the pages of a news portal into stable pages (the overall fixed structure of the portal) and news pages (that are created with high frequency). The rate of visits to stable pages is constant, while news pages receives a large number of visits after a few hours and decays as a power-law. Moreover, unlike Poisson processes, a power-law distribution also describes consecutive visits to the site performed by a single user. Huberman *et al.* [236] compared real user activity data with a model of user surfing that gives the probability distribution of the number of visits to a Web page. The model, which reproduced well real data, considers that there is a cost associated with a surfing activity, and that the user continues surfing while this cost remains below a threshold. Another surfer model, the aforementioned PageRank, is used by Google to sort its search results [226]. PageRank simulates the behavior of a user randomly surfing the Web, where the user goes from page to page following hyperlinks and sometimes performs random jumps (with probability d) to any other page in the Web. The frequency of visits of the random surfer per page is the PageRank score. Menezes and Barabási [200] recorded the visits to a group of sites during 30 days, and the mean flux per node $\langle f \rangle$ and the flux dispersion per node σ were estimated. Similarly to what happens with the Internet (see Section 8), a scaling behavior of the form $\sigma \sim \langle f \rangle^{\gamma}$ was found in the WWW, this time with exponent $\gamma \simeq 1$. Results indicate that the fluctuations in the WWW are dominated by external driving forces, such as the variation in the number of surfers, rather than by internal choices made by surfers.

Kleinberg *et al.* [237] and Kumar *et al.* [238] modeled WWW growth applying intuitive ideas of user behavior, where a user tends to create links to pages related to some topic of interest. While the network grows, some pages appear about a topic and users interested in this topic start creating links to those pages. This makes it easier for other users to find those pages, thus also establishing links to them. Ultimately, it allows the formation of groups of pages about similar topics. The model performs this procedure by using four stochastic processes, one for node creation and one for edge creation, along with two other processes for node/edge removal. Kumar *et al.* [239] created a model for generating an evolving WWW, where new nodes are continuously added, connecting to previous nodes with either uniform probability or with preference to older nodes. The latter case simulates a behavior where newly created nodes are likely to be unknown by most of page creators. Moreover, some nodes reproduce the neighborhood of a randomly picked node. All these models agreed with observed degree and bipartite core distributions.

Bornholdt and Ebel [240] used Simon's model created in 1955 to explain Zipf's law. This model accurately predicts the in-degree exponent found in the WWW [84, 221], which is $\gamma_{in} = 2.1$. The authors defined the following procedure, where a class $[k]$ contains the nodes with the same degree k and $f(k)$ is the number of nodes in the class $[k]$: (i) with probability α create a node and connect it to any other node, or (ii) with probability $(1-\alpha)$ connect any node to a node of class $[k]$ with probability $P_{[k]} = [kf(k)] \,/\, [\sum_i if(i)]$. Tadić [241] suggested two general rules to generate the WWW: (i) growth, where at each time a new node j is added to the network, and (ii) rearrangements, where at each time a number of outlinks is created from j to an old node and a number of links between old nodes are created or removed. This model agreed with empirical data regarding distributions of in- and out-degree and the distribution of the size of connected components. Pennock *et al.* [223] changed the BA model of preferential attachment including a uniform attachment probability. Thus, the probability of connecting a new node i to an old node j is combination of the usual preferential attachment with a uniform attachment rule. This model encompasses the rich-get-richer mechanism, along with the creation of hyperlinks that are more influenced by personal interests of the page owner than by the popularity of a page. Moreover, the model is especially able to capture degree distributions of specific subgraphs in the Web, such as those explained in Section 9 (e.g., set of company websites). Menczer [242] expanded this model and included a parameter of lexical similarity between Webpages, where the probability of connecting two nodes includes a power-law that depends on the lexical distance. Thus, this approach considers the content of the documents and their connectivity. Besides approximating the degree distribution of a 100-thousand-document snapshot of the Web, this model could also estimate the content similarity distribution of connected Webpages.

10 Citations

Although dynamically described by diffusion processes, information transfer can also be quantified from static structures such as citation networks. Any information contained in a paper or technical report contributes to the global knowledge of the reader. However, only the most relevant manuscripts tend to be remembered when one has limited time and space. In this respect, an information filtering process is already performed at this stage. By using data from citation databases, one can investigate the information flow through specific groups of people. The citation networks so far constructed derive from technical reports (especially scientific papers) as nodes and citations between those reports as directed edges. Those networks are constantly growing. As the in-degree represents the citations from other reports, this measurement carries information about the importance of the report, i.e. the number of times that report was found to be relevant to other works. However, once published, the report cannot aggregate new references and the out-degree is fixed over time. Because of this intrinsic network property, the out-degree distribution of scientific papers from the ISI database in the period from 1991 to 1999 is characterized by a maximum value with strong fluctuations at the left-hand side of the distribution. Most importantly, there is a universal behavior at the two rightmost classes, corresponding to journals of limited and unlimited number of pages per paper [243]. In contrast, patent-to-patent citations present a power-law with exponent $\gamma = 0.62$ [244]. Van Raan found a power law with exponent $\gamma \sim 3.7$ in the number of references per publication in 2001 [245]. This interesting unidirectional growth mechanism has curious properties: though the average number of citations increased over time [243, 246], the average number of citations for papers published in a given year has decreased slowly with time [246]. This suggests that the relevance of a paper might decrease over time. In order to further analyze such a behavior, Redner constructed two networks, one considering only the citation of papers published in 1981 and available at the ISI database, and another network covering Physical Review D papers published and cited in a period of nearly 22 years [246]. By means of a Zipf plot (cumulative distribution), he found a power-law citation distribution $P(k) \sim k^{-\gamma}$ with $\gamma \approx 3$ for large k. Other power-law distributions with exponent $\gamma = 2.7$ were verified in the SPIRES database by Bilke and Peterson [247], now with exponent $\gamma \sim 3.1$ for papers published and cited in 2001 [245]. Such distributions, together with other evidence, led Redner to suggest that minimally and highly-cited papers obey different statistics where most part of papers are forgotten a short time after publication. In fact, nearly 47% of the papers in the ISI database were not cited and 80% were cited only 10 times or less. Conversely, about 0.01% were cited over 1000 times [246]. On the other hand, Tsallis and Albuquerque [248] suggested that the entire distribution found by Redner might emerge from the same phenomenon since it could be well fitted by a single function derived from the non-extensive formalism. Power-law citation distributions were also found on an extensive US patents database granted between January 1963 and December 1999. Chen and Hicks [244] showed that the patent-to-patent citation distribution had exponent $\gamma = 2.89$, and $\gamma = 2.31$ when the data were restricted to a specific field related to both basic and applied research. Interestingly, while analyzing only citations of papers in US patents, they concluded that papers with explicit funding acknowledgements tended to be cited relatively more often ($\gamma = 1.97$) than papers without explicit acknowledgements ($\gamma = 2.16$), possibly because those articles in the former category are more likely to have impact on inventions described by patents. Using a random walk procedure, Bilke and Albuquerque [247] found that the spectral dimension of the ISI citation network is about 3.0, having a tree-like structure with the small-world property.

The level of similarity between two reports can be inferred not only by keywords, but also by the number of common references. Since the references are supposed to be directly related to the subject of the report, similar reports should cite similar previous works. In this type of network, this property is obtained by the neighborhood of a specific node, but it could also be obtained by constructing a different network as done by van Raan using databases from several citation indexes in 2001. He considered that two publications were connected if they had at least one reference in common [245]. Although the network was unweighted, the same data could now be used to quantify the level of similarity between two reports. The degree (named *bibliographically coupled publication cluster size*) distribution is also described by a power-law function and an exponential cut-off for values larger than 1000. Analogously to other works, a memory effect was identified by considering the age of the references, the distribution function changed from a power-law to an exponential as long as only the older references were considered to build the network. He suggested that reports with low degrees are related to very specific themes which are typically recent, and therefore cite recent articles. In contrast, old references are generally more general and connect more parent publications, giving rise to a more uniform distribution.

The databases reviewed in this section were considerably large and reliable in terms of an interaction rule between their components. However, some inaccuracies are present since no standardization is adopted and individual entries are made by different people at different times. Redner suggested that inaccuracies as incorrect page numbering for citations, citations of specific pages and input errors in transferring citation data supplied by authors, have minor effects in the citation distribution [246].

11 Transportation

Transportation networks are important for the development of a country and may be seen as indicator of economic growth. The tourism industry and transport of goods and people are particularly dependent on transportation networks, which include airports, railways, highways, subways, and public transport. Studying this kind of network can help one to understand the movement of people around the world and predict how diseases can spread, in addition to design optimal networks for the flow of people. Ultimately, it may give insights on how to improve the economy of a country.

11.1 Airports

In airport networks, cities with airports are considered vertices, and flights between them are the arcs. This type of network is naturally directed and weighted because of the direction and number of passengers in flights. The arcs are thus directed and their weights are given by the number of passengers in flights or the number of flights itself in a day or in a week. Guimerà *et al.* [16], however, argued that the connections considering the network of all airports in the world is almost symmetrical, with minor asymmetries arising from a small number of flights following a "circular" path. Therefore, there is no need to consider the arc directions.

He *et al.* [249] studied the Chinese airport network in 2002 without considering either the directions or the number of passengers in a flight. They concluded that this network is small-world without the scale-free property, as the vertex degree distribution is exponential. They explained the topology of the network with a model based on the population and gross domestic product of the cities. Li and Cai [250] also analyzed the Chinese airport network using a connectivity matrix of size $N \times N \times 7$, where N is the number of airports for the 7 days of the week. Besides the scale-free and small-world properties, they observed for the Chinese airport network that (i) the daily and weekly cumulative vertex degree distribution of undirected and directed Chinese airport networks obey two regime power-laws with different exponents (the so-called Pareto law [251]); (ii) the cumulative distribution of flight weights has power-law tails; (iii) the diameter of a subcluster, consisting of an airport and all of its neighbors, is inversely proportional to its density of connectivity; and (iv) the transportation efficiency of the subclusters of the airport network increases with the density of connectivity.

The worldwide airport network [16, 252] is a scale-free, small-world network. Nevertheless, the most connected cities are not necessarily the most central, in contrast to other scale-free networks. This is a consequence of the community structure of this network [16], on which a global role of cities was developed. This measurement indicated that the vertices connecting different communities are hubs in their own community. Guimerà *et al.* [16] pointed out that the community structure cannot be fully explained only by geographical constraints, but geopolitical considerations have to be taken into account. A model to account for these findings was developed by Guimerà and Amaral [252], which explains why the most connected cities are not the most central. In the model, not all cities of a country can establish connection with cities of another country.

Bagler [253] studied the airport network of India considering edge weights, and found that this network is hierarchical with the small-world property, and characterized by a truncated power-law degree distribution. The network has disassortative mixing, in contrast to the worldwide network of airports. This difference can be related to the local and global scales of airport networks.

In addition to analyzing the networks, it is important to design optimal networks for this kind of transport. Barnier and Brisset [254] proposed a new route network to optimize the operational cost for the French airports network, based on graph coloring for the air traffic flow management.

Problems associated with airport networks are related to difficulties in the air traffic flows, e.g., caused by heavy fogs or snowstorms. In order to assess the performance of the network under such circumstances, Chi and Cai [53] analyzed the errors caused by an attack to the structure of the US airport network. Analogously to other scale-free networks, the US airport network is tolerant to errors and random attacks, but extreme vulnerable to a target attack to hubs. While topological properties, including average vertex degree, clustering coefficient, diameter, and efficiency are almost unaffected by the removal of a few airports with few con-

Table 1: Papers that considers different airport networks.

Network	**Reference**
Chinese airport network	[249, 250]
French airport network	[254]
Indian airport network	[253]
Japan airport network	[255]
US airport network	[53]
Worldwide airport network	[16, 252]

nections, the same properties are drastically altered if a few hubs are removed.

11.2 Roads and urban streets

These networks deserve special attention in our daily life since they directly influence our travel times and transport costs. In the last years, we have witnessed an increasing number of vehicles on the roads which resulted in slower traffic and more frequent jams. Although possible in the past, building new roads or streets is not feasible today because of spacial constraints imposed by the local buildings. The best solution is, therefore, to make better planning for the traffic or even to manage it in real-time so that the wasted time on the roads is minimized.

In the context of complex network theory, many questions related to the representation of the metric distances of the roads arise. These include what kind of graph representation to use, which topological features to study, as well as the correlations between structural measurements and the dynamics of the traffic flows. The representation of road networks can be: *primal* or *dual* graph (e.g., [256, 257]). In the former case, while intersections are nodes, the roads are edges. This is a natural way of representing these networks since it captures the most important feature of geographical dimensions, i.e. the distance, and it was used in [17, 258]. A more detailed study of networks of this kind can be found in [258], where the authors analyze twenty samples of street patterns of several world cities. There is a limitation, however, with the primal representation, since it does not express the difficulty to walk on them [259], which can be obtained by the *dual* representation, where roads are nodes and intersections are edges. Although this representation does not give the geographical distance between two arbitrary points in the network, it expresses the information (i.e. the number of road changes) needed to travel between those points. Therefore, the less the number of road changes to reach a specific destination, the easier to find it.

Rosvall *et al.* [259] expressed the difficulty of navigation on these networks in terms of the “search information” [260]. The higher this measurement, the more difficult it is to find a destination. With this approach, the authors showed that new cities, e.g., Manhattan, are better planned than old cities, e.g., Stockholm, since the search information for the former is less than for the latter city. Another conclusion of [259] is that it is preferable to replace a big number of streets with a few long, provided it reduces the number of road changes necessary to connect any two points of the network. The accessibility of places in towns and cities was also investigated by considering self-avoiding random walk dynamics [261]. In this case, Travençolo and Costa showed that the dynamics of transportation through towns and cities is strongly affected by the topology of the connections and routes.

Following the second definition for road networks, Kalapala *et al.* [257] also found that this kind of network has, besides the small-world property [256], topological and geographic scale invariance. In other words, for sufficiently large geographical areas, the degree distribution is scale-free, with journeys having identical structures, regardless of their length. Indeed, the driver starts on small roads, moves to progressively larger, faster roads, until the fastest is reached, where he/she covers most of the trip before descending back to progressively small roads. The authors also proposed a simple fractal model to reproduce the above features. Geographical scale-free models can be found in the review [255].

In [17], Gastner and Newman compared geographical networks such as an airline network, the US interstate highway network, and the Internet at the autonomous system level. They showed that networks, which are fundamentally two-dimensional such as the highways in the primal graph representation, have a close relation between their geographical and topological features, as e.g., the geographical and geodesic distance. A model to reproduce their findings was also proposed.

The dynamics on road networks has been studied for years, with many models proposed [262, 263, 264, 265, 266], ranging from macroscopic models based on the kinetic gas theory or fluid dynamics to microscopic approaches with equations for each car in the network. In [267], the authors analyzed the German highway network using cellular automata as a model for the traffic flow, and showed that this model is suitable for real time traffic forecasting. With this model, it was also possible to find the bottlenecks of the highway network.

11.3 Other transportation networks

There are many other transportation networks, but their properties are quite similar to the airport and road networks mentioned above. Therefore, we shall not discuss networks such as the railways [268, 269, 270], subway networks [271], and public transportation [272, 273, 274].

12 Electric power transmission systems

The electric power transmission system is one of the most complex human-made networks. It comprises transmission lines and several substations, that include power plants; transmission substations, which connect high-voltage transmission lines; and load centers, which delivery the electricity to consumers (e.g., [275]). Colloquially known as power grid, the electric power transmission system has a complex structure including redundant paths to route power from any generator to any load center. The main reason for this redundancy is to guarantee that every load center can be supplied by any generator. In other words, if one generator fails, the load centers will receive the necessary power from the other generators. In the same way, if one transmission substation fails, the others have to be able to handle the additional load and keep the whole network working. Nonetheless, even with the redundancy of lines, some cascading failures and blackouts happen and several load centers stop receiving power. One of the most serious was the Northeast Blackout (on August 14th, 2003) which affected 50 million people in the U.S.A and Canada and resulted in a huge loss of money (about 30 billion US dollars) [276]. Obviously, this kind of network is crucial for the economy of a country and deserves special attention in engineering and science.

The first studies in the power transmission systems relied on creating simple dynamical models which simulate each component of the network to understand the blackout dynamics of the whole system [277, 278, 279, 280, 281]. The networks were simple structures such as rings, trees or mathematical grids, and the blackouts were considered instantaneous events caused by cascading failures of the transmission lines. When one transmission line failed, all the power flow was redistributed to the other lines, but new failures may happen due to overflow, leading to a cascading effect. These dynamical processes were simulated in [277, 278, 279, 280, 281], from which it was inferred that the size of the blackouts follows a power-law tail, which means that big blackouts are not so uncommon.

Although useful for predicting blackouts and finding the critical components of the power transmission network, the approach above was limited to simple network topologies which do not correspond to those of the real power transmission networks. In fact, these power grids

Table 2: Electric power grid network databases. Complete networks contains all substations regardless of voltage.

Network	**Size**	**Reference**
Western US power network	4,941	[32]
US power network	14,099	[275]
Complete US power network	314,123	[282]
Italian power network	341	[283]
Chinese power network	8,092	[284]

have the small-world property, high clustering coefficient (e.g., [276]), degree distribution in an exponential form [85, 275], and the "bow-tie" configuration [282]. For a power network of 314 123 nodes, Chassin *et al.* [282] considered any vertex regardless of voltage (note that the size of the power networks used in the other studies did not take into account nodes of small voltages, see Table 2 for more power network data), and showed that the network has a radial form. The generators are in the center, with transmission substations in the middle and the load centers at the border, in the so-called "bow-tie" configuration [221]. The degree distribution obeyed a power law with a cutoff term.

In the analysis of the topology of these networks, it was shown that the removal of highly connected nodes (without distinguishing their types) can lead to blackouts of certain regions of the networks [285, 286, 287]. Each substation was assumed to have a transmission capacity that depends on the number of shortest paths passing through it. Albert *et al.* [275], however, showed that only the removal of highly connected transmission substations can provoke blackouts, since the other transmission substations can fail because of the additional overload. This is not the case of generator failures, because even the removal of the highly connected generator is unable to cause blackouts [275]. This is a consequence of the redundancy of these networks where all generators can be routed to all load centers. When one generator fails, the others provide the additional power to supply the whole network.

A more realistic and complex model was developed by Anghel *et al.* [288]. It describes an electric transmission network under random perturbations (e.g., line or substation failures and overloads) and the operator's response to the contingency events (a system to repair the damaged lines or substations). In this model, each random event was stochastic and could be triggered at any time. The model was not only able to predict blackouts but also to find the optimal strategy for minimizing the impact of random component failures.

13 Biomolecular Networks

In June 2000, the Human Genome Project and Celera Genomics decoded the human genome, providing the so-called "book of life", after which the genetic code of many organisms has been discovered. The Genome project is only a starting point [289], as the post-genomic era should be concerned with modeling biological interactions instead of analyzing the genetic code by itself. As the behavior of living systems can seldom be reduced to their molecular components, in the post-genomic area of *system biology* [290] one has to assemble the parts unfolded by genome projects. For instance, Vogelstein, Lane and Levine [291] concluded that more significant results can be obtained by analyzing connections of the p53 gene (a tumor suppressor) than by studying this gene separately.

Processes in living organisms are basically divided and linked in three levels of complexity [292]: (i) *metabolic and signaling pathways*, which are determined by (ii) *the network of interacting proteins*, whose production is controlled by (iii) *the genetic Regulatory network*.

Understanding life processes therefore requires one to: (a) analyze how energy is obtained from cell biochemical reactions by interactions among metabolites, products and enzymes; (b) understand how proteins take part in various processes, as in the formation of protein complexes; (c) understand how information is transmitted from a transcription factor to the gene that regulates this transcription [1].

Modeling these biological systems may be performed by considering complex network theory, because of its universality and ability to mimic systems of many interacting parts [293]. Biological networks comprise metabolites, enzymes, proteins, or genes as vertices, which are linked depending on their interaction. Using complex network theory, fundamental properties of biological networks have been discovered, including power law connectivity distributions, small world properties and community structures [294]. Furthermore, the crucial concepts to understand biological systems, namely emergence, robustness and modularity, are also ingredients of complex networks theory [290, 293].

In the following, we shall discuss major developments in applying complex network theory to biological systems.

13.1 Protein-protein Interaction Networks

In modeling protein interactions, one considers a network comprising proteins (nodes) connected by physical interaction (undirected edges). This requires reliable databases, which have become available since the 1980s because of high throughput methods, and a set of measurements to characterize the network structure and dynamics. The connections are generally not weighted, but some databases provide indexes of reliability associated with each link as the confidence score [295].

The study of protein interactions using complex networks theory may be divided into four basic areas: (i) characterization of the network structure; (ii) prediction of protein functions; (iii) modeling of the interactome and (iv) modeling and characterization of protein-domain interactions.

The structure of protein interaction networks has been studied to determine the relative importance of vertices and their organization in modular structures or subgraphs. Some of these structures are conserved along the evolution of many organisms, which allows one to infer their importance for cellular maintenance. The structure of protein interactions was believed to be completely random. However, in 2001 Jeong *et al.* [11] discovered that these networks were far from uniform, upon analyzing the *S. cerevisae* protein interaction networks. In fact, a few proteins are able to physically attach to a huge number of other proteins, but most proteins play a very specific role, interacting with just one or two other proteins. The distribution of connections of Yeast follows a power-law with an exponential cutoff whose exponent is about 2.5 [11]. Other systems, such as the simple bacterium *H. pylori* [296] and the insect *D. melanogaster* [297], also present a power-law dependence for the connectivity in the protein-protein interaction. Therefore, it seems that the scale-free nature of protein interaction networks is a general feature of all organisms.

The protein interaction networks have also been shown to be small-world, with a small average shortest path length, and to present many loops of order three [298]. Their structure is formed by modules of highly connected proteins [107, 299, 300]. The hubs possibly play a critical function in the network maintenance, being strongly related to lethality. They are believed to be the oldest proteins in the network. Jeong *et al.* [11] showed a correlation between lethality and connectivity. While among the proteins with five or fewer links, 93% are non-lethal, among the proteins with more than 15 links, 62% are lethal. The highly uneven structure of protein interaction networks imparts robustness against random failures [11]. When random failures occur in scale-free networks, there is only a small probability that the removed node is a hub, and the network structure is not affected severely [10]. In contrast, when highly connected proteins are removed, the network breaks up in several disconnected components, which reinforces the relation between lethality and connectivity. This explains why scale-free networks are very abundant in nature, being associated with the evolution of cellular processes and genes [300]. In spite of the fact that the loss of hub proteins can lead to death of the organism, some network-based strategies to restore cellular functions caused by specific mutations have been developed [301].

The importance of hubs was also addressed by Maslov and Sneppen [302] who showed that the functional modules inside the cell are organized around individual hubs. In addition, the proteins are uncorrelated according to their connectivity which indicates that the neighborhood of highly connected proteins tends to be sparser than the neighborhood of less connected proteins [302]. Schnell and Fortunato [303] investigated the relation between structural features of hubs and their number of connections. They concluded that the disorder of a protein (or of its neighbors) is independent of its number of protein-protein interactions. The presence of hubs and the local cohesiveness of networks are also related, as suggested by Ravasz *et al.* [298]. The latter authors showed that the relations between the scale-free architecture and the high degree of clustering in protein networks are consequences of a hierarchical organization, suggesting that small groups of nodes organize in a hierarchical manner to make large groups.

The functional importance of proteins can also be addressed in terms of network entropy [304] and betweenness centrality [305]. Joy *et al.* [305] observed that proteins with high betweenness, but low connectivity, are abundant in the yeast interaction network, with a positive correlation between the fraction of lethal proteins and their betweenness centrality.

Some structural properties of protein interaction networks seem to be conserved during evolution. Wuchty *et al.* [306] observed that some types of subgraphs referred to as network motifs may be preserved during natural evolution. These motifs were found more frequently in real networks than in their random versions, generated by the rewiring method that maintains the network degree distribution [39]. Wuchty *et al.* analyzed the conservation of 678 proteins of Yeast with an ortologus in five eukaryote species, namely *Arabidopsis thalianam*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*. They observed that some motifs are conserved from simple organisms to more complex ones [306]. Each of the motifs conserved was suggested to perform specific roles, e.g. in forming protein complexes, where smaller parts are represented by fully connected n-node motifs. Because some biological functions emerge from modules of many connected proteins [292, 307], instead of single proteins, such structures are conserved throughout the natural selection. In fact as the cell is organized in modules, the structures should be conserved along the evolution, and not only the individual cellular components [307, 308]. The conservation of network motifs can also be used to predict protein interactions. Using machine learning algorithms, Albert and Albert [309] showed that conserved properties of protein networks can be used to identify and validate protein interactions.

The Baker's Yeast (*Saccharomyces cerevisiae*) has about 6,300 genes, which encode about the same number of proteins [310]. Therefore, the determination of interaction among such proteins requires checking about 6,300 times 6,300 pairs, which is close to forty million potential interactions. In 1989, Stanley Fields proposed a revolutionary technique to detect protein-protein interactions using the GAL4 transcriptional activator of yeast *Saccharomyces cerevisiae* [311]. This "yeast two-hybrid"(Y2H) method [312, 313] is based on the fact that a protein with DNA-binding domain may activate transcription when it simply binds to another protein containing an activation domain. Basically, Y2H detects interaction using two hybrid proteins referred to as the "bait" and the "prey". The "bait" contains a query protein X fused to a DNA binding domain and the "prey" is a fusion of a second protein Y to a transcriptional activation domain. If the proteins X and Y interact, then their DNA-binding domain and activation domain will combine to form a functional transcriptional activator, which will proceed to transcribe the reporter gene that is paired with its promoter [314, 315]. Further information about this method and its extensions may be found in [315] and [316].

The Y2H allowed the global analysis of protein-protein interactions and the birth of *interactome* (one of the next steps after the genome) [317, 318]. The success of this method lies on the identification of interactions without antibodies or the need to purify proteins [311]. However, the main drawback is that Y2H generates many false positives, i.e. interactions identified in the experiment but that never take place in the cell [319, 320]. This limitation has motivated the development of many other methods, including biochemical techniques such as affinity [295] and molecular size-based chromatography, affinity blotting, immuno-precipitation and cross-linking [317, 321, 322, 323]. Also, computational methods are applied to identify protein-protein interactions [324], as those based on the genome sequence [325]. Sprinzak *et al.* proposed a method to asses the quality of protein interaction databases using the cellular localization and cellular-role properties to provide an estimative of true positives in databases [326]. They suggested that the reliability of the high-throughput Y2H is about 50% and that the yeast interactome should be composed of $10\,000$–$16\,600$ interactions. Recently, Krogan *et al.* [295] made available a database of protein interactions based on tandem affinity purification which identified $2\,708$ proteins and $7\,123$ interactions of the *Saccharomyces cerevisiae*. This database offers greater coverage and accuracy than the previous high throughput studies related to Yeast protein-protein interaction. In Table 3 the main protein-protein interaction databases and their web addresses are presented.

Table 3: Public databases of protein-protein interactions.

Database	**URL**
DIP	`http://dip.doe-mbi.ucla.edu`
IntAct	`http://www.ebi.ac.uk/intact`
HPID	`http://wilab.inha.ac.kr/hpid`
MIPS	`http://mips.gsf.de/services/ppi`
Biogrid	`http://www.thebiogrid.org`
BIND	`http://bind.ca`
CYGD	`http://mips.gsf.de/proj/yeast/CYGD/interaction`
iHOP	`http://www.ihop-net.org/UniPub/iHOP`
JCB	`http://www.imb-jena.de/jcb/ppi`
MINT	`http://mint.bio.uniroma2.it/mint`
PathCalling	`http://curatools.curagen.com/pathcalling_portal`
String	`http://string.embl.de`
InterDom	`http://interdom.lit.org.sg`

Assigning functions to unknown proteins is one of the most important problems in the post-genomic era. This may be performed by genome analysis methods that exploit domain fusion [325] and phylogenic profiles [327]. In the first method one observes that pairs of interacting proteins have homologs in higher species fused in a single protein chain. The second, on the other hand, considers that proteins participating in structural processes or metabolic pathways are functionally linked, evolving in a correlated fashion. Other methods are based on interaction partners, because proteins with the same functions tend to share connections [328]. This property arises from the duplication and mutation mechanisms of protein evolution. When a protein is duplicated, the daughter protein has the same features as the mother protein. However, as the daughters suffer mutations, they tend to differ in structure, but preserve similar functions and connections. Another common approach for predicting protein function is the majority rule assignment, which takes into account the empirical finding that 70–80% of protein interaction partners share at least one function [329]. Hishigaki *et al.* [328] proposed a methodology based on analysis of n-neighborhood proteins, i.e. for a given protein i, the n-

neighborhood is composed by proteins distant n edges from i. With this methodology, one can predict with 72.2% accuracy the subcellular localization, with 63.6% accuracy the cellular role and with 52.7% accuracy the biochemical function of yeast proteins. However, such methods do not consider the links between unknown proteins. Vazquez *et al.* [330] extended this method by minimizing the number of physical interactions among different functional categories and considering the connections between unknown proteins. This methodology provides a more precise identification of protein function.

Despite the many approaches to determine protein function using the complex network of interactions, there are many limitations owing to the large number of false positives and negatives in the interactions. Thus, some connections between proteins may not occur in vivo and the function associated with such interactions may not be real. Another problem is related to multi-functional proteins. In this case, it may be hard to determine the function of the proteins when their neighbors have many functions.

Simple protein interaction models have been proposed using network measurements. Eisenberg and Levanon [331] proposed that the BA model is relevant for protein interaction modeling. They applied a cross-genome comparison and observed a correlation between proteins age and their connectivity. Then, they suggested that the protein evolution is governed by the preferential attachment rule. Pržulj *et al.* [332], on the other hand, suggested that the protein evolution is governed by a geometrical model and the scale-free degree distribution may be caused by the high percentage of false negatives. The geometric model is obtained by distributing N protein randomly in a metric space and connecting two proteins according to the distance between them. These results were later confirmed by Costa *et al.* [23].

Even though these models reproduce some protein interaction properties, other biological processes must be considered in modeling. This may be performed for two aspects related to addition and elimination of interactions between proteins, and gene duplication increasing the number of proteins and interactions [333]. Vázquez *et al.* [334] proposed a model in which each node in the network represents a protein that is expressed by a gene, and the network evolves following two steps: (i) *duplication*: a randomly selected node i is duplicated as i' which is connected to i according to a probability p and has all connections of i, (ii) *divergence*: a node j connected to i and i' loses the connection (i, j) or (i', j) according to a probability q. Therefore, this mode has two parameters p and q, representing the creation of a connection by node duplication and the loss of interactions, respectively. Solé *et al.* proposed a similar model with the same concepts [335], which is given in detail in [336]. Solé and Fernández showed that the models by Vázquez *et al.* [334] and Solé *et al.* [335] reproduce the modular structure and the degree correlation observed in protein interaction networks [337]. In this case, hubs tend to connect with poorly connected proteins. Also, the modular scale-free structure in protein networks emerge naturally from the duplication-divergence rule [337].

The rate of duplication and deleting of interactions was estimated by Wagner [333] who concluded that every 300 million of years, around half of all existing protein interactions are replaced by new interactions. Wagner [338] also estimated the rate of link dynamics and gene duplication using empirical data, showing that the rate of link dynamics is at least one order of magnitude higher than gene duplication. With this observation, Berg *et al.* [339] proposed a model of protein interaction that considers link attachment and link detachment together with gene duplication. In this case, the link dynamics is guided by a preferential attachment rule supported by empirical data. Such a model reproduces the scale-free degree distribution and the correlation between interaction proteins, as observed experimentally [339].

As domains can recombine to form multi-domain proteins, the domain recombination may be the main mechanism to modify protein function and increase the proteome complexity [340]. Protein interactions can be validated by domain-domain interactions [341] and proteins with similar activities are likely to contain similar domains [342]. Therefore, the characterization of domain-domain interactions has many applications and is crucial to understand the evolution of protein interactions.

Domain-domain interaction networks are constructed considering (i) protein complexes, (ii) Rosetta Stone sequences, and by using (iii) protein interaction networks [341, 343, 344]. With the first methodology, domain information is inferred from the intermolecular relationships in protein complexes. The second considers domain fusion in different organisms, i.e. domains that appear separately in one organism and together in another one are potentially interacting. Finally, the last approach considers the protein interaction to determine domain interaction. In this case, all domains belonging to two interacting proteins are also interacting. Ng *et al.* [341] suggested a measure to determine the potential of interaction between domains considering the protein interaction network.

Wuchty [343] studied domain interaction networks with methodologies (i) and (ii). In both cases, he showed that protein domain interactions follow a power law and exhibit the small world property with a high average clustering coefficient. In a subsequent paper, Wuchty [344] also investigated which factors force domains to accumulate links to other domains. As the network of domain interactions consists of hubs responsible for network integrity, it is natural to investigate the importance of the most connected domains. Costa *et al.* [12] analyzed the structure of domain interactions and the relation between protein connectivity and essentiality. In this case, the essentiality of a domain is hypothesized in two ways: (i) *domain lethality in a weak sense*: a domain is lethal if it appears in a lethal protein and (ii) *domain lethality in a strong sense*: a domain is lethal if it only appears in a single-domain lethal protein. Costa *et al.* showed that correlations between domains degree and lethality in both the weak and strong senses are significantly higher than the correlation obtained for proteins, which shows the importance of domains in defining protein interaction and protein lethality [12].

13.2 Metabolic Networks

Metabolism is the complete set of chemical reactions that occur in living cells, allowing cells to grow and reproduce, maintain their structures, and respond to their environments. The essential energy and substrates for the function of a cell are obtained by breaking large molecules into small ones. Hence, this process is the basis of life. Metabolic networks are related to the chemical reactions organized into metabolic pathways, in which one chemical compound is transformed into another by the action of a sequence of enzymes. Basically, the reaction networks can have three possible representations: (i) directed and weighted graphs, whose vertices can be of three types: metabolites, reactions and enzymes, while there are two types of edges representing mass flow (from reactants to reactions) and catalytic reactions (from enzymes to reactions), (ii) connecting all reactants that participate in a same reaction, (iii) connecting two reactions if they share a reactant. Major catalogues available on-line are the Kyoto Encyclopedia of Genes and Genomes (KEGG, `http://www.genome.jp/kegg`) and the EcoCyc (`http://ecocyc.org/`).

Jeong *et al.* [345] investigated metabolic networks of 43 organisms from all three domains of life, and found that metabolic organization is not random, but follows the power law degree distribution —the probability that a given substrate participates in k reactions follows a power law, $P(k) \sim k^{-\gamma}$, with $\gamma \simeq 2.2$ in all 43 organisms. Besides, metabolic networks are small-world ($\ell \approx 3.2$), where two metabolites can be connected by a small path. For instance, Fell and Wagner [346] showed that the center of *E. coli* metabolism map is glutamate and pyruvate, with a mean path length to other metabolites of 2.46 and 2.59, respectively. An important property discovered by Jeong *et al.* is that the diameter of metabolic networks is almost the same for all the 43 organisms. This property is quite different from that for other types of network, where the diameters increase logarithmically with the addition of new vertices. A possible explanation is that as complexity of organisms grows, individual substrates get more connections to maintain a relative constant network diameter. Since metabolic networks are scale-free, few hubs concentrate a high number of connections [346]. Such networks are thus tolerant to random failures, but vulnerable to attacks in which there is a sequential removal of nodes starting with the most connected one. For this procedure increases the network diameter and quickly breaks the network into disconnected components. This expectation has been confirmed experimentally in *Escherichia coli*, where mutagenesis studies performed *in-silico* and *in-vivo* showed that the metabolic network is highly tolerant upon removal of a considerable number of enzymes [347]. Another feature of hubs is their preservation along evolution. Only 4% of all substrates encountered in the 43 organisms are present in all species, which can be considered generic as they are utilized by many species [345]. Such substrates are the hubs in metabolic networks.

The structures of a metabolic network are organized in a modular, hierarchical fashion [348]. The modules are discrete entities composed by several metabolic substrates densely connected by biochemical reactions. Ravasz *et al.* [349] showed that the average clustering coefficient of the metabolic networks of 43 organisms is independent of network size, being higher than for scale-free networks generated by the Barabási-Albert model of the same size. In addition, the average clustering coefficient follows a scaling law with the number of links, $c(k) \sim k^{-1}$, which indicates a hierarchical organization. Hence, metabolic networks are characterized by a scale-free degree distribution, average clustering coefficient independent of network size and hierarchical and modular organization. Ravasz *et al.* suggested a model of metabolic organization that reproduces these properties. It is built starting with a small cluster of four densely connected nodes, followed by generation of three copies of this structure with the three external vertices being connected to the central vertices of the old clusters. This process is repeated until a network with the same size as the real network is obtained. The networks thus generated display the properties of metabolic networks discussed above.

Guimerà and Amaral [13] proposed a methodology to find functional modules in complex networks and classify vertices according to their patterns in intra- and inter-modules connections. They found that 80% of substrates are only connected to substrates within their respective modules. Also, substrates with different roles are affected by different evolutionary constraints and pressures. In contrast to the result by Jeong *et al.* [345], Guimerà and Amaral showed that metabolites participating in a few reactions but connecting different modules are more conserved during evolution than the most connected substrates (hubs).

13.3 Genetic Networks

Living cells are governed by gene expression programs involving regulated transcription of thousands of genes. Cell signaling and differentiation can be investigated by pattern of gene expression, which can be represented by complex networks [350]. Transcription is controlled at many different levels and the gene regulation network fits into a network of networks that represent not only interactions among transcription factors but also the factors that modulate these interactions biochemically [351]. High-throughput methods and computational approaches have allowed important discoveries in genetic networks. In order to analyze the topology and dynamics of genetic networks, approaches were developed to evaluate the identification and expression level of interacting genes, how interactions change with time and the phenotypic impact when key genes are disrupted. The techniques aimed at elucidating transcriptional regulatory networks are mainly based on genome-wide expression profiling and the combination of chromatin immunopreciptation (ChIP), which was revised in [352, 353].

Transcriptional regulatory networks control the gene expression in cells and are composed of regulators and targets (nodes) and regulatory interactions (edges). Edges are directed from a gene that encodes a transcription factor protein (TF) to a gene transcriptionally regulated by that transcriptional factor, referred to as target gene (TG) [42]. Such complex systems can be analyzed as a multilayered system divided into four basic levels [354]: (i) the first level

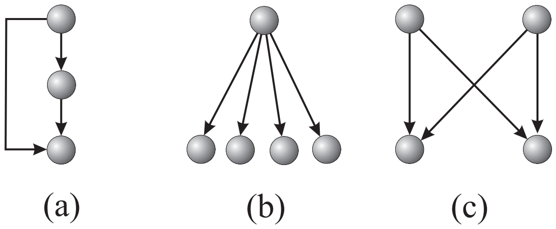

Figure 2: Three type of motifs found in transcriptional regulatory networks of *Saccharomyces cerevisiae* and *Escherichia coli*: (a) feed-forward loop, (b) simple input module and (c) multiple input.

encompasses the collection of transcriptional factors, its target genes with DNA recognition site and regulatory interaction between them; (ii) the second level is studied in terms of regulatory motifs, which are patterns of interconnections that appear more frequently in real networks than in randomized networks [39]; (iii) the third level considers the modular organization, where modules are group of nodes that regulate distinct cellular processes; and (iv) the last level consists of the regulatory network composed by the whole set of modules.

Because network motifs appear at frequencies much higher than in random networks, they are expected to have special functions in information processing performed by the network. The complexity of networks can be reduced by considering their motifs. Figure 2 presents the three main types of motifs found in regulatory networks in the yeast of *Saccharomyces cerevisiae* and the bacteria *Escherichia coli*. In transcriptional regulatory networks, the feed-forward motifs are defined by a transcription factor that regulates a second one such that both, jointly, regulate a final target gene. On the other hand, the simple input module and the multiple input regulate their targets by a simple or multiple transcriptions factors, respectively. All these targets are controlled by the same sign (all positive or all negative) and have no additional transcriptional regulation [42]. Other important motifs are found in the transcriptional regulatory networks of *Saccharomyces cerevisiae*, for which a detailed discussion is presented in [355].

Modules are organized by the interconnections of network motifs. In general, the modules refer to a group of physically or functionally linked genes. Distinct cellular processes are regulated by discrete and separable modules [356]. Ma *et al.* [357] identified 39 modules in transcriptional regulatory networks of *E. coli*, which showed well-defined functions. Also, Dobrin *et al.* [358] showed that in *E. coli* multiple input and feed-forward loops motifs do not exist in isolation, but share transcription factors or target genes. A review about the modular topology in transcriptional regulatory networks is presented in [354].

The structure of transcriptional regulatory networks exhibits two main properties: (i) the fraction of genes with a given incoming connectivity decreases exponentially, and (ii) the outgoing connectivity distribution follows a power law [356]. The exponential character of the incoming connectivity indicates that most target genes are regulated by a similar number of factors. On the other hand, the scale-free distribution of the outgoing connections points to a few transcription factors participating in regulation of a large number of target genes. In addition, such genes tend to be lethal if removed [359]. The structure of transcriptional regulatory networks changes with time or environmental conditions, because not all nodes are active at a given time. Dynamical processes can be analyzed in genetic networks, as in Li *et al.* [360] who observed that the cell-cycle of the regulatory network in yeast was extremely stable and robust. Klemn and Bornholdt [361] investigated the reliability of gene regulatory networks and found a distinction between reliable and unreliable dynamical attractors, showing that biological signaling networks are shaped by selective advantage of the ability of robust signaling processing. Other dynamical processes investigated in genetic networks are discussed in [362].

The evolution of transcriptional regulatory networks occurs via three processes: (i) duplication of the transcription factor, which results in both copies regulating the same gene, (ii) duplication of the target gene with its regulatory region, where the target gene will be regulated by the same transcription factor, and (iii) duplication of both the factor and the target [354]. Such a mechanism of duplication and inheritance determines a large fraction of the interactions in regulatory networks. Furthermore, a large number of interactions are gained or lost after gene duplication. This process can result in network motifs and specific connectivity distribution in the network. Teichman and Babu showed that the evolution of regulatory networks is mainly defined by gene duplication [363] and that 45% of regulatory interactions in *E. coli* and *S. cerevisiae* arose by duplication with inheritance of interactions.

Though the structure of biological networks has been extensively studied, the understanding of temporal activity of a cell is still in its beginning. Databases with time evolution of gene expression activity are being developed [364, 365]. When data on cell network processing and intercommunication become available, new insights about the mechanisms of life will be possible.

14 Medicine

Complex networks have been useful in medicine, as networks representing biomolecular systems can be applied to understand disease principles and spreading. In the former case, Barabási [366] suggested the treatment of diseases by taking into account different levels of complex biological networks. The basic level considers a complex network connecting cellular components (e.g. metabolic, protein-protein interaction and genetic networks), the middle level represents the networks of diseases (two diseases are connected if they have a common genetic or functional origin), and the top level represents the society, composed by human interactions, as friendship, geographical localization, and so on. In case of obesity, the analysis should consider biomolecular networks, because some genes or metabolic disfunctions can have fundamental importance. At the network disease level, obesity is related to other diseases, such as diabetes, asthma and insulin resistance. Finally, at the top level, the social interactions contribute to spreading of eating habits and lack of physical activities.

In addition to disease origin, complex networks can be useful in epidemiology. Indeed, the spreading of infections caused by virus or bacteria has been studied by dynamical analysis in networks. Works about spreading virus are discussed in Section 23.

Complex networks theory is potentially suitable to be applied in anatomy. For instance, the bone structure in mammals are composed by a complex networks of channels, which are required to nourish the bone marrow cells. In this case, vertices represent congruences of two or more channels and these interconnecting channels are expressed as edges [367]. Costa *et al*. [368] characterized a network constructed from a cortical bone structure in terms of complex networks measurements.

Bone structure in mammals involves a complex network of channels (Havers and Volkmann channels) required to nourish the bone marrow cells. This work describes how three-dimensional reconstructions of such systems can be obtained and represented in terms of complex networks. Three important findings are reported: (i) the fact that the channel branching density resembles a power law implies the existence of distribution hubs; (ii) the conditional node degree density indicates a clear tendency of connection between nodes with degrees 2 and 4; and (iii) the application of the recently introduced concept of hierarchical clustering coefficient allows the identification of typical scales of channel redistribution. A series of important biological insights is drawn and discussed

15 Ecology

The characterization, modeling and simulation of ecological systems are major challenges of natural sciences today. In ecological communities, each species interact in different ways, forming networks of interacting species. Important ecological relations are competition, parasitism, mutualism and predatory-prey relationships. Food webs are composed by S trophic species connected directly by L trophic links [369]. The links are represented by arrows from i to j if species i is eaten by species j. Such rows indicate the flow of resources. Trophic species is a collective designation for all species having a common set of predator and prey. Thus, species such as sponges and clams may belong to the same trophic species. The use of terms such as "detritus" and "dead organic material" indicates how hard it is to decide on what can be included or omitted in the food web chain. Before the 1990s, food webs were very small, with sizes varying from 5 to 48 [370]. The analysis of such networks suggested that the ratio L/S was independent of S and therefore this relation was scale-invariant. However, further works showed that this belief is not correct [371]. Food webs are small-world networks, obeying the two degree of separation rule [372]. This property implies that the loss of biodiversity and species invasions may affect ecosystems in a fundamental way. Also, Garlaschelli *et al.* showed that differently from biomolecular networks, food webs are not scale-free and do not present high average clustering coefficient [373, 374]. The modeling of food webs can be divided into three basic levels: static models, dynamical models, and species assembly and evolutionary models, which are discussed in the review by Drossel and McKane [369]. Ecological relationships were also explored by considering dynamical process, such as metapopulation dynamics [375, 376] and spreading diseases [377].

In addition to predatory-prey relationships, species can interact according to mutualism — plants and animals establish beneficial interactions, where both individuals derive a fitness benefit [378]. A classical example of mutualism involves unicellular algae and corals, where the coral provides the algae with shelter and inorganic nutrients, while the pigmented algae provide photosynthesis. Mutualism is fundamental in ecology and evolutionary biology, driving the evolution of much of the biological diversity [379]. The structure of networks representing mutualistic relationship are different from those obtained for antagonistic interactions (predator-prey, herbivore-plant). While antagonistic networks are highly compartmentalized [380], mutualistic networks are often nested, characterized by species with many interactions forming a kernel, and species with few interactions, which interact just with the highly connected species. This kind of structure was observed in mutualistic relationship between animals and plants (e.g. flower and bees) [381] and cleaner and client species (cleaning networks) [382], as the cleaning networks are more asymmetric than plant-animal networks [382]. The nested structure of mutualist networks is related to phylogenetic relationships [379]. Indeed, phylogenetic related species tend to have similar biological roles in mutualistic networks [379].

16 Neuroscience

Complex networks theory is a useful framework for the study of large scale brain networks. The network models, topological measurements, and the dynamical analysis (e.g. avalanches [56, 383]) can be considered for brain studies. The representation of brain connections by networks is useful to study brain diseases, which are related to network attacks and failures, and brain functions, where it is possible to associated a particular brain architecture with specific brain functions.

Brain networks can be investigated at many different scales, ranging from individual neurons to large brain areas [384, 385]. In the former case, neurons are connected by synapses. In the latter case, regions are linked by pathways or functions. Brain networks are generally directed and unweighted, though the description of neuronal systems may be enhanced by considering weighted networks.

An important question in neuroscience regards the relationship between brain function and the structure of neural connectivity. The brain is composed of a complex interconnection system whose organization is aimed at optimizing resource allocation and minimizing constraints [386]. A limitation of studies in brain connectivity is the proper choice and integration of the spatial and temporal resolutions, because a given dataset may reflect individual neuron activities, neural assemblies dynamics and other macroscopic brain activities [387, 388]. The main types of connectivity are: (i) *anatomical connectivity*, related to the connections between two brain sites; (ii) *functional connectivity*, which is defined as the temporal correlation between spatially distant neurophysiological events; and (iii) *effective connectivity*, which is a more abstract approach defined as the influence that a neural system may have over another in a direct or indirect way.

Anatomical connectivity data can be obtained at different scales, from neurons to brain regions. This analysis can be considered static when the time of analysis varies from seconds to minutes, but can be dynamic for large time scales, such as hours and days, because of learning or development activities. At a more microscopic spatial scale, the connectivity of the brain is related to the geometry and spatial position of the individual neuronal cells (e.g. [385, 389, 390]). For instance, networks involving more intricate neurons are likely to imply enhanced connectivity, therefore affecting the corresponding dynamics (e.g. [25]). An interesting open question is to predict the connectivity and dynamics of neuronal networks from the geometric features of specific types of neurons [25, 385]. In this case, the form of neurons is highly associated to their functions [25]. The first study with graph theory to represent brain connections was performed by Felleman and Van Essen [391] in 1991. They mapped the hierarchical organization of the cerebral cortex, reporting 305 connections among 32 visual and visual-association areas, in order to study the distributed hierarchical processing in the cerebral cortex of primates. Watts and Strogatz in 1998 analyzed the first neuroanatomical network with complex networks theory, i.e. the nervous systems of *C. elegans* [32]. This network, formed by $N = 282$ neurons and $M = 3{,}948$ synaptic connections, revealed a small-world characteristic structure, but not a scale-free distribution of links. The topology of such network is suitably represented by the Watts-Strogatz small-world model. Similar results were obtained by Hilgetag [392], who studied compilations of corticocortical connections data from macaque – the whole cortex, the visual and the somatosensory – and cat. Other studies suggested that the large scale organizations of the brain cortex of rat [393], cat [394] and monkey [391] are neither regularly nor randomly connected. Most cortical networks have a multi-cluster structure, presenting the small-world property (e.g. in *C. elegans* $\langle c \rangle = 0.28$ and $\ell = 2.65$, in macaque visual cortex $\langle c \rangle = 0.59$ and $\ell = 1.69$, in macaque cortex $\langle c \rangle = 0.49$ and $\ell = 2.18$, and in cat cortex $\langle c \rangle = 0.60$ and $\ell = 1.79$). Therefore, the neuronal and cortical connectivity obtained by these works are small-world networks. This particular topology might be chosen by selection pressure to minimize wiring costs. The small-world structure allows the network to have a modular organization and connections between such modules by adding several long-distance connections [395, 396]. Costa and Sporns [397] characterized cortical networks by considering hierarchical measurements and identified principles for structural organization of networks. In addition, Costa et al. [398] proposed a computational reconstruction approach to the problem of network organization in cortical networks and showed that the organization of cortical networks is not entirely determined by spatial constraints. Cortical areas has also been analyzed dynamically. Costa and Sporns [399] applied Metropolis dynamics on four configurations of the cat thalamocortical systems, i.e. (i) only cortical regions and connections; (ii) the entire thalamocortical system; (iii) cortical regions and connections with the thalamic connections rewired so as to maintain the statistics of node degree and node degree correlations; and (iv) cortical regions and connections with attenuated weights of the connections between cortical and thalamic nodes. They identified particular substructures determined by correlations between the activity of adjacent regions when only cortical regions and connections were taken into account. In addition, two large groups of nearly homogenous opposite activation was observed in cases (ii) and (iii). The same authors also investigated uniform random walks effects on dynamic interactions between cortical areas in the cat brain thalamocortical connections and shown that such connections are organized in in order to guarantee strong correlation between the out-degree and occupancy rate [400].

The small-world topology has a direct influence in the dynamical complexity of the network. As discussed by Barahona and Pecora [46] and Hong and Choi [47], information propagates faster on many small-world networks of undirect uniformly coupled identical oscillators. Thus, the topology of neuroanatomical networks provides a better propagation of activities than regular or equivalent random networks. Also, Lago-Fernández *et al.* [401] showed that nonidentical Hodgkin-Huxley neurons coupled by excitatory synapses present coherent oscillations in regular networks, fast system response in random networks, and both advantages in small-world networks. Therefore, the topological organization of neuroanatomical networks is directly related to network function. A good review about dynamical modeling of brain functions is [402].

Other approaches of brain functional networks are based on the concept of functional or effective connectivity. The network is constructed by recordings of brain physiological functions instead of brain anatomy. Aertsen *et al.* [403] introduced the application of graph theory to study brain functions. Three basic approaches are mainly considered in order to obtain the connection structures: (i) electroencephalographic (EEG), (ii) magnetoencephalographic (MEG), and (iii) functional magnetic ressonance imaging (fMRI). The data obtained by the two first methods are very suitable for graph analysis because of the high spatial resolution. Stam [404] presented the first work that considered MEG, in 2004. By performing MEG recordings of five healthy human subjects, the graph was formed by $N = 126$ vertices and $1{,}890$ edges. For frequency bands smaller than 8 Hz and larger than 30 Hz, the synchronization patterns displayed the features of small-world networks. Bassett *et al.* [405] detected small-world properties in brain functional networks obtained from 22 subjects using wavelet decomposition of MEG time series.

Dodel *et al.* [406] applied graph theoretical analysis as a new approach to identify functional clusters of activated brain areas during a task. Eguíluz *et al.* [407] showed that the functional network obtained by fMRI, considering seven subjects, follows a power law. The unweighted networks are obtained by correlation matrices, which are obtained by functional correlations between brains sites (called “voxels”). Two voxels are connected if their temporal correlation exceeds a pre-defined threshold r_c. The main drawback of this approach is related to the choice of the right r_c. For $r_c = 0.6$, 0.7 and 0.8, scale-free topologies were seen in the obtained networks, irrespective of the type of task considered. High clustering coefficient (~ 0.15), relative small-world ($\ell = 11.4$ for $N = 31,503$, $\ell = 12.9$ for $N = 17,174$, and $\ell = 6.0$ for $N = 4,891$), well defined coefficient of the power law (~ 2.2), and hierarchical organization ($cc(k) \sim k^{-\alpha}$) were observed. Therefore, the obtained functional networks were scale-free and small-world. Similar studies were performed by Salvador *et al.* [408] with five subjects, but using the frequency rather than time domain. In another work, Achard *et al.* [409] considered the resilience of brain functional networks and concluded that the brain regions are so resilient to random failures as to target attacks (removal of the largest hubs). In this way, the authors suggested that the brain networks are not scale-free as suggested by Eguíluz *et al.* [407].

MEG and EGG techniques were considered to study the relation between brain networks topology and brain pathologies. The matrices obtained by pairs of synchronization likelihood values were converted in unweighted networks by considering a threshold. Stam [404] chose the threshold to keep the network with average degree equal to 15. Stam *et al.* [410] compared a group of 15 Alzheimer patients to a healthy control group of 13 patients and showed that the networks obtained for the first group have less pronounced small-world features than for the healthy patients. Bartolomei *et al*. [411] compared the networks of spatial patterns of functional connectivity of the brain measured at rest by MEG obtained for 17 patients with brain tumours to 15 healthy patients. They concluded that brain tumours alter the functional connectivity and the network topology of the brain. While pathological networks are closer to random networks, healthy networks are closer to small-world networks. Therefore, randomization in network structure can be associated with brain damage.

In summary, the complex network theory is useful for neuroscience, opening up new opportunities and creating new challenges. For instance, one may try to understand how brain topology changes during animal growth or evolution. Also, the implications of the genetic and environmental factors for brain formation can be addressed by network representation. Studies related to brain structure and pathologies may help understand brain diseases.

17 Linguistics

The creation of language is one of the greatest accomplishments of the human species. Understanding the evolution and organization of a specific language is useful because it sheds light into cognitive processes, as the way people think strongly affects the organization of a language. Conversely, the language influences how humans think. It is also important to compare the properties of various languages, to study their evolution. Moreover, the use of linguistic data is crucial for automated systems such as Web search engines and machine translators.

Various linguistic structures can be treated as networks, including texts and thesauri. The Natural Language Processing (NLP) [412] community has traditionally used such network representations in developing applications for automatic language understanding and generation. For instance, Graph Theory concepts have been used to sentiment analysis [413] and Spectral Graph Theory concepts have been used for word sense disambiguation and text summarization [414]. More recently, linguistic networks were included in the context of complex networks research. Novel techniques frequently based on statistical physics are now used in language studies [415, 416], providing new approaches for NLP applications and linguistic theories. Indeed, there has been an increasing effort to join statistical physics and language studies, as demonstrated by the realization of a workshop inside STATPHYS 23, the largest conference on statistical physics in the world [417].

A linguistic network can be formed by, for example, a group of interconnected words or syllables. The many ways these elements can be linked in the network lead to a division of two main groups of linguistic networks: semantic and superficial. The former group comprises networks such as the ones constructed from dictionaries or lexicons, which usually contain information about semantic relationships between words, such as synonyms or antonyms. The latter group mainly uses the inner structure of words or the position of words in texts to build networks. For example, in a word-adjacency network, words are connected if they appear as neighbors in a text. These linguistic networks are frequently studied according to some well-known types of complex networks, namely small-world and scale-free networks. Consequently, the tools most often used to study linguistic networks are the measurements referred to as degree, clustering coefficient and length of the shortest paths. Examples of these investigations are discussed in the next subsections.

17.1 Semantic Networks

Semantic networks usually encode relationships between a subset of words of a language, as in a thesaurus where each entry (frequently a word) is followed by a list of words that express concepts similar to the entry. An English thesaurus used to build semantic networks is the

Roget's Thesaurus, originally published in 1852. The Wordnet database is another resource to construct semantic networks. This lexicon groups words into sets of synonyms and stores other semantic relationships between words, such as antonymy, hypernymy and hyponymy. A semantic network can also be obtained from word-association experiments, where someone sequentially and freely provides words that he/she thinks are semantically related.

Semantic networks are usually characterized through the verification of small-world and scale-free properties. As already mentioned (Section 2), a network with the small-world effect has a high clustering coefficient and a low average shortest path, while a scale-free network is characterized by a power-law degree distribution. English thesauri, for example, are known as small-world and scale-free networks [418, 419, 420]. In particular, a directed network was built in [420], and the k^{in} and k^{out} distributions were described with the Burr XII distribution, with two power-law regions. It was even suggested that associative memory is originated in a process of efficiency maximization in the retrieval of information (also related to the fact that the neural network is a small-world network) [419]. Cycles of tourist walks were used to give complementary information to clustering coefficient and shortest path (a tourist walker goes to the nearest site that has not been visited in the last s steps) [421]. Furthermore, the Wordnet network is scale-free and small-world [418, 422]. In this network, polysemic links (used for the different meanings of a word) are important for the small-world effect, and hubs typically represent abstract concepts. Word-association networks are known to be scale-free and small-world [418, 423, 424]. When the flow of sequentially induced associations is considered, other properties such as context biasing and edges asymmetry are present [423]. Evoked words from two economically distinct populations were used to build word-association networks, and it was showed that the scale-free structures obtained are independent of the population's economic level [424].

The process of lexical development for an individual was modeled in [418], using a preferential attachment strategy that produces both scale-free and small-world structures. The network obtained with this model encodes associations, which can be thought of as semantic, between words or concepts. Three main rules guide this model: (i) a process of differentiation, in which a new word/concept is defined as the variation of the meaning of an existing word/concept through the acquisition of part of the pattern of connectivity of an existing node; (ii) the probability to differentiate a node is proportional to its degree; and (iii) each node has a "utility" value which gives the probability that it will be connected to new nodes. As shown in [418], this model produces networks similar to those obtained from real-world data of word-associations.

17.2 Superficial Networks

The networks considered in this subsection do not employ semantic information; instead, they use syntactic structures or sequence of words in texts. One of these networks is based on a list of consonants in specific natural languages. The set of syllables of a language may be used to build another type of superficial network, where the nodes are the syllables and the edges indicate that two syllables co-occur in the same word. Another network of interest encodes on its structure the proximity of words in texts. In such a network, called a word-adjacency or co-occurrence network, words that appear near each other in a text are connected, as word proximity frequently indicates some kind of syntactic relationship [416]. Furthermore, syntactic relationships can be encoded in the network explicitly if a grammar formalism is employed, such as dependency grammars.

As was the case of the semantic networks, the concepts of degree distribution, clustering coefficient and length of the shortest paths are frequently used to characterize superficial networks. Indeed, a bipartite consonant network displayed a two regime power-law distribution for the occurrence of consonants over languages [425]. This network presents two sets of nodes, one for the 317 languages and one for the 541 consonants, where a consonant node is connected to a language node if the former is present in the latter. A Portuguese syllable network, built from the entire work of the eminent Brazilian writer Machado de Assis, also presents a power-law degree distribution, along with the small-world effect [426]. Although some superficial networks were studied regarding community structure (see [427] which uses a non-bipartite consonant network), the most investigated network properties are indeed the small-world effect and the scale-free degree distribution. The word-adjacency networks obtained from the British National Corpus (BNC) were shown to be small-world and scale-free [416], where the authors even associate language disorders to hubs disconnection. Word-adjacency networks built from technical and literary texts in English and Portuguese were shown to be scale-free and small-world [428]. Moreover, a George Orwell's novel (1984) was used to build a word-adjacency network [429], which exhibited a composite power-law behavior when the node degree was correlated to the average neighbor's degree and to the average clustering coefficient. Word-adjacency networks from different languages (English, French, Spanish and Japanese) were found to belong to a superfamily of networks based on the significance profile of subgraphs, called motifs [233]. The significance profile compares the number of each motif in a real network with its number in a random network with the same size and degree distribution. Syntactic networks were obtained for Czech, German and Romanian texts using a dependency grammar formalism, where binary relations are defined between lexical nodes [430]. The authors showed that these networks present the small-world effect, a scale-free degree distribution and disassortative mixing. Finally, a syntactic dependence network was built for the Romanian language [431], and it was shown that spectral methods to detect community structure [432] can be useful to group words of the same class.

Word-adjacency networks have been applied to natural language processing tasks. For instance, an automatic assessment of text quality was implemented with network measurements, in which correlation was established with text quality scores assigned by humans to essays written by high-school students in Portuguese [433]. A more specific task of quality assessment was studied in [434, 435], where network measurements were associated to the evaluation of the quality of Portuguese summaries. A similar approach was adopted in [436] to evaluate the quality of automatic translations, where the original and translated texts, or more specifically, their word adjacency networks, were compared and the degree of structural change from source to translation was measured and correlated with translation quality. Another NLP task studied with word-adjacency networks was the authorship characterization, where texts in English created by famous writers were related to network measurements [437].

A growth model for word-adjacency networks was proposed in [428], which adds to a network full sentences instead of single words. Every word of a new sentence is represented by a node that is connected to the other nodes of the sentence (i.e. the sentence is represented by a complete subgraph - a clique). The ratio between new and old words in a sentence decreases over time according to a power-law, while the old words are the hubs. This model is in good agreement with real data, and presents both scale-free and small-world properties. A preferential attachment approach for the construction of consonant networks was reported in [425]. The degree distribution fits well with the network obtained from a real consonant inventory.

18 Physics

Physics is a promising area in which many phenomena can be modeled as a network. For instance, in 2001, Bianconi and Barabási [438] discovered that a Bose-Einstein condensate was related to complex networks. The authors mapped the network in a Bose gas, where the nodes corresponded to energy levels and the links were particles. The condensation state was reached when a single node had a macroscopic fraction of the links. Burioni *et al*. [439] reported a complementary work with topological inhomogeneities in the network being considered.

18.1 Energy landscapes

In 2005 Doye and Massen [440] used networks to study the configurational space of the Lennard-Jones potential. To build the network, called reaction graph, the authors consider each minima of the potential function as a nodes. The surrounding of each node defines its basin of attractions. The nodes are connected when there is a direct state transition between their respective minima, not considering auto-loops and multiples edges. The results indicates that the network topology is scale-free and presents small-world effect. Moreover, the authors found an interesting negative correlation between the degree and the potential energy of the minima, indicating that the low energies minima acts as hubs, the signatures of scale-free networks. Further, the authors found that the scale-free nature of this network is not due to preferential attachment, but it is a direct consequence of the potential landscape topography, because the basins of attraction related with low-energy minimum are larger in general and more likely to have surrounding transition states. The authors also suggested that the small-world effects can be useful to explain the Levinthal paradox (which says that the time of a protein to arrive at its the native conformation is larger than the age of the universe) because it reduces dramatically the number of steps needed to find the native conformation. For example, in a typically problem with 10^{21} minima, algorithms based on global optimization converges, in average, after only 150 minima. It is important to observe that reaction graphs were already used in previous papers, such as [441], to investigate the isomers of water dimer, and [442] to study interconnection of the protein conformal space. Gfeller *et al*. [443] also studied energy landscapes through networks (weigthted in this case). Their findings include some analytical results for low-dimensional models and they discuss how the network approach is useful to study the isomerization problem. Recently, the landscape os spin-glasses was also mapped in networks by Seyed-allaei *et al*. [444]. The authors found that the degree distribution of the resulting network follow a Weibull distribution, instead of the power-law found previously for the Lennard-Jones potential [440]. The authors concluded that the topology of the landscape networks are not universal because they depend on the physical properties of the model system.

18.2 Astrophysics

Magnetic effects were also explored through complex network concepts. In [445] the authors suggested that the solar corona can be considered as a critical self-organized system. Hughes and Paczuski [446, 447] studied the spatial structure of a network generated from this critical system after the transient phase. To build this network, an approach similar to that of earthquakes studies was used, with the solar surface being a discrete mesh with each mesh cell being a node. Two nodes were connected when a magnetic arc (solar flare) emerged between the cells. The network obtained has scale-free topology with a power law distribution of degree and strength.

18.3 Ising Model

With the complex networks field becoming more popular, systems that were only studied in regular networks started to be described in networks with arbitrary topologies. This is the case of the Ising, Potts and XY models. Gitterman [448] showed that a phase transition occurs when there is a minimum number of long range connections per node. Barrat and Weigt [449] found that at high temperatures the system behaves as in the conventional 1D case. At low, non-zero temperatures, a ferromagnetic phase transition was found in the small-world Ising model. Therefore, the phase transition in the small-world Ising model is of a mean-field nature. In 2001, Hong *et al*. [450] used Monte Carlo simulation to confirm the results of Gitterman [448], Barrat and Weigt [449], and found critical exponents $\alpha = 0$, $\beta = 1/2$, $\gamma = 1$ and $\nu = 2$ (α is the critical exponent for specific heat, β is the critical exponent for magnetization. γ and ν are the critical exponent for the susceptibility)

The two and three dimensional cases of the small-world Ising model were studied by Herrero [451]. He noted that, in 2D and 3D, the small-world geometry changes the universality class of the phase transition. In 1D, the critical temperature is given by $T_c \propto |\log p|^{-1}$ where p is the probability of rewiring an edge. Herrero found that in 2D and 3D, the dependence between T_c and p is given by the power law $T_c - T_c^0 \propto p^{1/\nu d}$, where T_c^0 is the critical temperature in the regular lattice and d is the spatial dimension. The bidimensional Ising model also was studied more recently by Cai and Li [452].

The Ising spin system under a scale-free topology was investigated by Aleksiejuk *et al.* in 2002 [453] and Bianconi [454]. A ferromagnetic-paramagnetic phase transition was observed, with the critical temperature depending on the network size as $T_c \propto |\log N|$. The analytical solution for this model in a network with arbitrary degree distribution was found in [455, 456]. In 2004, a spin system under a directed version of the Barabási-Albert network was investigated by Sumour and Shabat [457], who used Monte Carlo simulation to show the existence of spontaneous magnetization of the system.

The p-Potts model was investigated by Dorogovtsev et. al in 2004 [458] for networks with degree distribution $P(k) = k^{-\gamma}$. The authors showed that when the second moment of this distribution diverges, the phase transition is continuous and of infinite order instead of the first-order phase transition.

18.4 Nanotechnology

In [459], a study was presented of nano self-assembled devices that showed random topology and small-world effects. These circuits have better performance in the synchronization process, small latency and density of classification task, in comparison to purely local circuits. The authors also discussed the marketing and technological viability of building self-assembled circuits with these features.

19 Chemistry

Recently, complex networks concepts have been applied in chemistry. Jiang *et al.* [460] showed how an ammonia plant can be mapped in a complex network, which displayed small-world properties. Moreover, the ammonia plant network is weakly self-similar and possesses a modular structure. Each detected community represents a modular section in chemical plants. Another interesting result is that the ammonia complex network exhibits excellent allometric scaling, guaranteeing a better fluid flow.

In 2001 Amaral and Barthèlemy [461] showed that the small-world effect emerges on the phase space of polymer conformations. For this, each conformation was represented by a single node in a complex network and two nodes were connected if the Monte Carlo distance between them was unitary. Another work involving polymer analysis through complex networks was developed by Kabakçioglu and Stella [462]. Chemical reactions were studied as a complex network topology. For example, Lazaros and Argyrakis [463] investigated the classic model $A + A \rightarrow 0$ and $A + B \rightarrow 0$ under scale-free topology. Their results indicate that the speed of the reaction is rather faster than in lattice models. The authors also observed that the diffusion behavior on very sparse scale-free networks are the same as for regular lattices.

Another network built from chemical reactions can be found in [464], in which the UMIST Database for Astrochemistry was used to generate networks. The nodes were reactants and products of each relevant reaction found in the interstellar medium and in planetary atmospheres (including Earth, Mars, Titan, Venus). Among the results, two basic topologies were revealed and associated with the presence or absence of life. According to the authors, the Earth chemical network is the only one to present a scale-free topology, although all the others also displayed small-world properties. The community structure was investigated in the Earth chemical reaction network, with two main communities being identified. They are related to reactants and products of the reactions with OH and Cl, and one of the products is H_2O. The authors suggest that the community structure in this network results from action of the most reactive free radicals of Earth atmosphere, namely Cl and OH.

20 Mathematics

Complex networks have been used to represent sets of natural and prime numbers connected by some specific rules. Corso [465] used a fundamental theorem of number theory, which states that each natural number has a unique decomposition in prime factors, to build a network. Two natural numbers (nodes) were connected if they shared at least one prime factor. The networks were nonsparse and small-world ($d \approx 2$, $N \approx 10^5$). The degree distribution was invariant with network size (number of natural numbers considered to build the network) and presented plateaux related to families of prime numbers. Chandra and Dasgupta [466] used a similar method to connect two prime numbers according to a free parameter (α). Given an even number n, this number can be written as a sum of two prime numbers p and q: $n = p + q$. The probability of establishing a connection between p and q is $|p - q|^\alpha$. The authors observed a phase transition in the network topology by varying the free parameter. For $\alpha > \alpha_0$, the network exhibits a scale-free topology and high clustering coefficient, characterizing a small world effect. Outside this regime, the power law of the degree distribution was not observed, but the hubs still existed. The size of the network studied is 5×10^3 nodes.

21 Earthquakes

The study of earthquake occurrences and its spatial distribution may be done through complex networks. Abe and Suzuki [467, 468] proposed a method in which the earth surface was divided into cells that were the nodes of the network. Two successive earthquakes (i.e without any other between they) establish a new link in the network between the cells where these shocks occurred. Upon analyzing seismic data from Japan and California, the authors showed that both earthquake networks have a scale-free topology with different scale coefficients. This difference may characterize the distinct tectonic plates analyzed. The hubs were found to be related to the place where a mainshock occurred (i.e earthquakes with large magnitude). In another paper [469], the earthquake network was shown to have a very small average path length and high clustering coefficients, in comparison to the random network version. These results suggest the presence of the small-world effect and a hierarchical organization, as indicated by the power-law behavior of the clustering coefficient with respect to the node degree [470]. The seismology network is therefore different from the Barabási-Albert networks, which also has a scale-free degree distribution but no hierarchical structure. Structured networks contain features that are linked to the physical properties of the earthquake dynamics. The mixing property was investigated using the concept of nearest-neighbor average connectivity, in which Abe *et al*. showed that highly connected nodes are linked to each other, for both seismology networks, namely from California and Japan. Abe and Suzuki [471] discovered that in the growth of a seismology network after the main shock the clustering coefficient remains constant during the time Δt, which depends on how the network was built, topically of the order of hours. After Δt, the clustering coefficient has a steep drop and then decays slowly according to a power law before it becomes stationary again.

The seismology networks were also studied by Baiesi and Paczuski [472], who used correlation between events to classify them systematically as foreshocks, main shocks and aftershocks. The network was built with each earthquake (node) receiving one incoming link from its most correlated predecessor, while outgoing links referred to aftershocks. It displayed the scale-free behavior with exponent $\gamma = 2$. The metrics used consider the magnitude and the spatial and temporal distances between events. Two spatially distant earthquakes can be related if the temporal distance is short, which is consistent with the self-organized nature of earthquake dynamics.

22 Security and Surveillance

Latora and Marchiori [473] studied the consequence and prevention of terrorist attacks in a given network, and suggested a method to detect critical nodes (i.e. the most important nodes for efficient network functioning). The network efficiency is related to the shortest path between all nodes of the network. The importance of a particular node is given in terms of the change in network efficiency when this node is removed. The deletion of an important node leads to a large decrease in network efficiency. The authors illustrated this concept in the communication network Infonet, which is the US and European Internet backbones. After deactivating all nodes, one by one, the most important nodes detected were New Jersey and New York. The results indicated that the destruction of these nodes can reduce network efficiency by more than 50%. Another important result indicated that the most connected nodes are not necessarily the most important. For example, deactivation of Chicago, with degree 15, only reduces network efficiency by 28%, while New Jersey and New York, both with degree 9, reduces network efficiency by 57% and 53%, respectively.

In the same paper, the authors gave a second example in which they built a network where the nodes were the terrorists related directly or indirectly with hijackers of September 2001 attacks. The links are knowledge interplay among the hijackers. The data were obtained from information taken from major newspapers. Mohamed Atta, on the flight AA-11, was the most critical node, who also has the largest number of direct connections with other terrorists. The second most critical node was Salem Alhazmi although its degree is half of the maximum. This shows that even nodes with low connectivity can play a crucial role for the network efficiency.

23 Spreading of Viruses and Diseases

Epidemics of computer viruses have been studied with the aid of graphs and random graphs for at least three decades. In 1991, Kephart and White [474] extended epidemiological models to investigate the spreading of computer viruses using a directed random graph. They showed that proliferation can be controlled if the infection rate does not exceed a critical epidemic threshold. Kleczkowski and Grenfell [475] applied a cellular automata model to the spreading of diseases in small-world networks, where the dynamics of the clusters was investigated as a function of the network order parameter (the fraction of the long range links). In the model, N_a agents were placed in the nodes of a two-dimensional square lattice. The nodes were classified as infected, infectious and immune. At each time step ($\Delta t = 1\mathrm{week}$) each agent interacts with its z nearest neighbors. Moreover, the authors included a mixing pattern in the model, i.e. they allowed two agents to swap their positions at any time, in order to reflect the rules of the social structure. The results showed that upon increasing the mixing the disease spreads faster. The problem of disease spreading in small-world networks was also explored in [476], where the authors found the exact values for the epidemic thresholds as a function of the infection and transmission probabilities. Also studying this system, Kuperman and Abramson [477] found an interesting oscillatory behavior of the size of the infected subpopulation. The authors showed that the number of infected agents changed from an irregular, low-amplitude state, to a spontaneous, high amplitude state, when the order parameter changed from 0 to 1. Recently, Small *et al*. [478] showed that only with the introduction of a small-world topology in the epidemiological model the random fluctuation of the real data could be explained.

In 2001 the first papers appeared which took into account the scale-free topology of the networks. Pastor-Satorras and Vespignani [479] showed that, in contrast to the pure small-world lattices, the uncorrelated scale-free networks with exponent $0 \leq \gamma \leq 1$ do not have a critical threshold, thus indicating that in these networks diseases spread regardless of the infection rate of the agents. For the scale exponent in the range $1 < \gamma \leq 2$ the critical threshold appeared but no critical fluctuations were observed. Only for $\gamma > 2$ has the traditional behavior been observed. For correlated scale-free networks with $P(k'|k) \propto k$ there is a non zero critical threshold for the spreading dynamics [48, 49]. This result was modified in subsequent papers [480, 481], where the authors used analytical arguments to show that in absence of higher order correlations, the epidemic threshold is null for scale-free networks with $2 < \gamma \leq 3$. The analytical solution of the SIR model (susceptible-infectious-recovered), including bipartite nodes and non-uniform transmission rate, was found by Newman [482]. The effects imposed by the finite size of the networks to epidemic threshold was studied by Pastor-Satorras and Vespiagnani in [483].

The lack of epidemic threshold in social networks with scale-free topology is worrying because for any transmission rate the disease can propagate to all nodes of the network [479, 484, 485, 486]. This is the case of the sexual contact network of Sweden, as shown by Liljeros *et al*. [485]. The solution to eradicate virus spreading in these classes of network was proposed in [487, 488, 489], which is based on immunizing the hubs of the network (*target immunization*). Through computer simulation they showed that this approach is able to eradicate the disease, being more effective than a random immunization. Though efficient, this method requires a priori information about the whole network structure, such as the connectivity of all nodes. Instead of this global approach, Cohen *et al*. [490] suggested a local method referred to as *acquaintance immunization*, where a subset of nodes was randomly selected and, depending on the neighborhood of each node, immunization was performed. In [491], Gómez-Gardeñes *et al*. studied various immunization strategies for Internet maps at the Autonomous Systems (AS) and Routers networks. They suggested a new immunization method that is neither local nor global, in which each vertex looks at its neighborhood (at maximum distance d) and immunizes the highest connected neighbor. The results confirmed this to be the most efficient method.

Spreading on growing networks also was investigated. Hayashi *et al*. [492] studied the oscillatory pattern for the number of infected computers on scale-free networks as new users joined. The authors simulated the spreading dynamics on networks with exponent of degree distribution extracted from real data of received and sent emails. Random immunization was not able to eradicate the virus. The number of infected computers oscillated with a period depending on the fraction of vertices that receive immunization. It also occurs in the scale-free networks, but in this case, it is possible to find a set of parameters to stop the oscillatory pattern and eliminate the virus. For example, an immunization of 20% of the hubs in a growing scale-free network prevents virus infection.

Real data of *Mycoplasma pneumoniae* infection was used in [493] to build a mathematical model and to investigate the spread and control of this disease in closed communities. Other approaches, including percolation theory, have also been used to study disease spreading with complex networks [494, 495].

Recently, Li and Wang [496] studied the SIR dynamics on small-world networks where they took into account a delay-time recovery. The results indicate that the action aiming at the recovery of the network should be executed as soon as possible to avoid the spread of the infection. The spreading on geography networks also was considered. In 2007 Guo *et al*. [497] proposed a novel model for geographic networks where the recovery was based on the action

Table 4: Topological properties of the collaboration network.

Measurement	**Value**
Number of nodes	911
Number of edges	4479
Average degree	9.83
Maximum degree	70
Average clustering coefficient	0.78
Assortative mixing coefficient	0.92
Average shortest path length	35.64

radius (local region immunization). The results indicated the existence of a critical value for this radius above which the epidemic spreading vanished. Recently, a particular type of scale-free network with self-similar structure was considered by Zhang *et al*. [498]. The authors showed that the network had high clustering coefficient, "large-world" and dissortitative behavior. Using a renormalization approach, they solved the SIR model under this topology and found the existence of an epidemic threshold. The main conclusion of the authors is that the degree distribution does not suffice to characterize the epidemic dynamics.

24 Collaboration Network of the Papers Cited in this Review

In order to have a more complete perspective of the works addressing applications of complex networks, we constructed a network of collaborations taking into account the references listed in the present survey. More specifically, each author in each reference entry was mapped into a node, while co-authorship was considered to implement the connections (i.e. an edge was used to interconnect each pair of co-authors). A program was especially implemented in order to filter, organize the Bibtex database. The network contains 911 nodes and 4479 edges is illustrated in Figure 3. The main areas covered in the present survey are identified by colors. Several features can be seen in this figure. First, the overall network is highly fragmented, except for two larger connected components. The largest connected components (with 189 nodes), contains articles from every area. The second largest connected component (with 168 nodes) is mostly composed of biological articles. A third connected component with intermediate size (27 nodes) can be identified in the upper middle part of the figure. All other communities involve less than 19 nodes and are highly homogeneous with respect to the covered areas. This high fragmentation indicates that the applications of complex networks are still performed by distinct communities of researchers, in a little integrated manner. As a consequence, the concepts and methods used by each of the communities are not likely known or applied by the other communities. It is felt that important and interesting advances could be achieved through a more integrated research and more intense collaboration between the several communities in Figure 3.

Table 4 shows the main measurements of the topology of the collaboration network. The way in which the network was constructed naturally implies in a high clustering coefficient. The assortative mixing coefficient (r) is particularly high, implying that highly connected researchers tend to be co-authors.

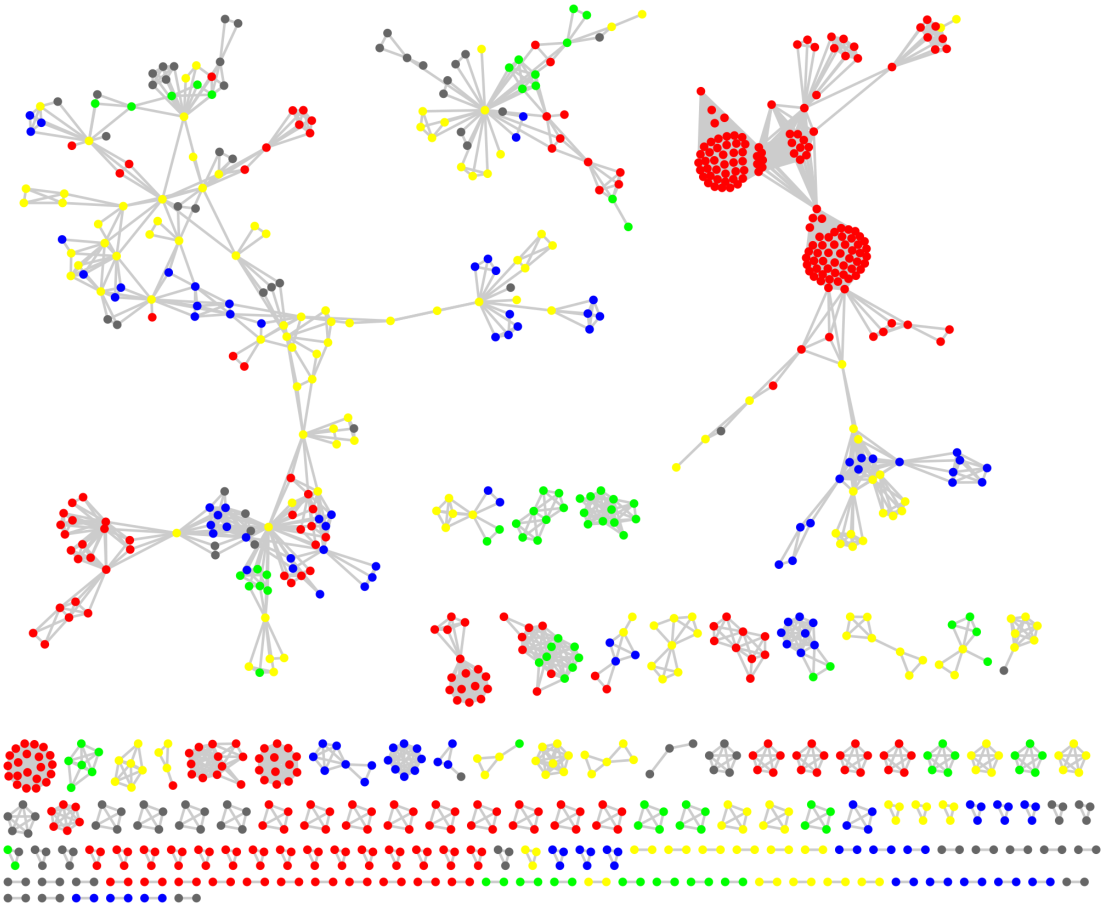

Figure 3: Collaboration network considering the papers cited in this review. The colors correspond to the main areas covered in this survey: biology (red), information networks (green), social networks (blue), technological networks (yellow) and other areas (gray). Graph visualization obtained by using Cytoscape [499].

25 Conclusions and perspectives

Frequently, the success of new areas of physics are judged not only from the theoretical contributions they imply, but also from their potential for applications to real-world data and problems. Despite its relatively young age, the area of complex network research has already established itself, especially through its close relationship with formal theoretical fields such as statistical mechanics and graph theory, as a general and powerful theoretical framework for representing and modeling complex systems. It has been capable of taking into account not only the connectivity structure of those systems, but also intricate respective dynamics (see, for instance, the surveys by Newman [22] and Boccaletti *et al.* [18]).

Judging by the large number of areas and articles reviewed in the present survey, complex networks have performed equally well (if not better) with respect to their application potential. Indeed, the generality and flexibility of complex networks extends to virtually every real-world problems, from neuroscience to earthquakes, encompassing at least the 21 areas considered in the present work. In addition, most of the studied reviewed report detailed and comprehensive investigations where not only complex systems are represented as networks, but their topological properties are also quantified with respect to a number of measurements. Also, frequently real-world networks are compared with some of the existing theoretical models such as scale-free and small-world. While the application potential of complex networks is clearly substantiated, it is important to conclude the present survey by conducting a global analysis of the reported applications.

Table 5 lists the 21 application areas considered here and the corresponding *number of reviewed papers*, the *size of the networks*, the *number of applied measurements*, as well as the *number of theoretical models* adopted for comparative purposes. Despite the bias implied by the choice of articles to be reviewed, this table still provides a representative snapshot of the state of the art in complex networks applications.

The first remarkable aspect in this table is the fact that protein applications is the area where complex networks have been more intensively applied (50 applications), while the applications in Internet (one of the important initial motivations for complex networks research) is limited to 39 articles. Other areas that received particular interest from complex network research include World Wide Web, transportation, transcription, neuroscience, linguistics and physics, all with at least 20 reviewed articles. The areas with the smallest number of applications (less than 10 revised articles) include actors, citations, acquaintances, trust, trust, sexual contacts, sports, music, financial market, metabolic, chemistry, mathematics, and earthquake. It is an interesting, difficult issue to identify which of these areas will remain little investigated, while others may become the focus of increasing attention.

Another important feature in Table 5 concerns the size of the networks. Though complex networks usually involve very large number of nodes (e.g. Internet and WWW), valuable results can be also obtained by applying complex networks concepts and methods to small networks. This agrees with the fact that 22 areas involve networks smaller or equal to 1000 nodes. Also interesting is the considerably large range of the sizes of networks used in each area. At the same time, several areas have considered very large networks, with 8 areas including networks in excess of 100000 nodes.

Interestingly, the measurements to characterize quantitatively the connectivity of the networks vary considerably from one area to the other, reaching as many as 20 different measurements in the case of organizational management. Most applications, however, involve only 4 or 5 distinct measurements such as node degree and clustering coefficient. Surprising trends can also be identified in Table 5 regarding the number of theoretical models considered or proposed as part of the investigation of the real-world structures. Reinforcing the major motivation and importance the Internet has had in complex networks, a total of 12 models appeared in the application articles. This seems to corroborate the fact that a definitive model of the Internet has not been reached yet. Several papers with applications have also considered 8 or more models, including linguistics, WWW, transportation and computer science.

The results contained in the revised works, as well as the several measurements and models adopted, provide unquestionable evidence about the importance and dynamics of the complex network field. Because of their intrinsic potential for representing, characterizing and modeling real-world complex systems, complex networks are poised to play a key role in an ever-increasing number of areas. The developments which may characterize the future application of complex networks include the use of additional measurements, required for a more comprehensive characterization of investigated structures, as well as additional theoretical models that may be created.

Acknowledgments

Luciano da F. Costa is grateful to FAPESP (proc. 05/00587-5), CNPq (proc. 301303/06-1) for financial support. Francisco A. Rodrigues acknowledges FAPESP financial support (proc. 07/50633-9). Paulino R. Villas Boas is grateful to CNPq sponsorship (141390/2004-2). Luis E. Correa Rocha is grateful to CNPq and to the Swedish Research Council (grant 621-2002-4135) for financial support. Lucas Antiqueira acknowledges FAPESP sponsorship (proc. 06/61743-

Table 5: Number of papers by research area and network sizes, total number of measurements and number of models covered by the works considered in our revision.

Research field	**Papers**	**Network size**	**Measurements**	**Models**
Actors	4	225 000	3	1
Citation	7	1 000–7 10^6	18	0
Acquaintance	7	1 500–375 000	12	0
Trust	2	60 000	5	0
Sexual contacts	7	1 500–3 000	2	1
Sports	3	100–13 500	4	0
Music	9	200–51 000	12	0
Collaboration	10	300–1.5 10^6	18	0
Communication	17	15 000–5 10^7	5	5
Economy	14	180	3	0
Financial market	7	200–30 000	10	1
Computer science	18	20–50 000	9	8
World Wide Web	30	1 000–271 10^6	7	9
Internet	39	30–1 10^6	12	12
Transportation	26	100–7 000	8	8
Transcription	20	500-2 000	5	2
Power grid	15	500–15 000	4	2
Protein interaction	50	2 000-20 000	5	3
Metabolic	9	500-1 000	6	0
Ecology	14	50–300	5	2
Neuroscience	31	100-1 000	5	0
Linguistics	28	100–500 000	13	8
Physics	26	210 000	2	2
Chemistry	3	400–3 200	5	2
Mathematics	2	100–5 000	5	1
Earthquake	6	1 200–60 000	6	2
Spreading disease	15	1 000–1 10^8	2	5

7). Matheus P. Viana is grateful to FAPESP sponsorship (proc. 07/50882-9). We are also grateful to Sumiyoshi Abe, M. Madan Babu, Rodolfo Baggio, Yaneer Bar-Yam, Regino Criado, Janos Kertesz, Adilson Motter, Sérgio Furtado dos Reis, Francesco Sorrentino, Dietrich Stauffer, Francisco Tamarit, Yanghua Xiao, and for comments and suggestions.

References

[1] A.-L. Barabási. *Linked: how everything is connected to everything else and what it means for business, science, and everyday life*. Plume, 2003.

[2] P. Erdős and A. Rényi. On the evolution of random graphs. *Bulletin of the International Statistical Institute*, 38(4):343–347, 1960.

[3] P. Erdős and A. Rényi. On random graphs, 1959.

[4] B. Bollobás. *Random graphs*. Cambridge studies in advanced mathematics ; 73. Cambridge University Press, Cambridge ; New York, 2001.

[5] B. Bollobás. *Graph theory : An introductory course*. Graduate texts in mathematics ; 63. Springer Verlag, New York, 1990.

[6] J. P. Scott. *Social Network Analysis: A Handbook*. Sage Publications, London, 2nd edition edition, 2000.

[7] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, Cambridge, 1994.

[8] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. *Computer Communication Review*, 29(4):251–262, 1999.

[9] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the World Wide Web. *Nature*, 401:130–131, 1999.

[10] R. Albert, H. Jeong, and A.-L. Barabási. Error and attack tolerance of complex networks. *Nature*, 406:378–382, 2000.

[11] H. Jeong, S. P. Mason, A.-L. Barabási, and Z. N. Oltvai. Lethality and centrality in protein networks. *Nature*, 411(6833):41–42, 2001.

[12] L. da F. Costa, F. A. Rodrigues, and G. Travieso. Protein domain connectivity and essentiality. *Applied Physics Letters*, 89(17), 2006.

[13] R. Guimerà and L. A. Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005.

[14] A. Krause, K. A. Frank, D. M. Mason, R. E. Ulanowicz, and W. M. Taylor. Compartments exposed in food-web structure. *Nature*, 426:282–285, 2003.

[15] M. E. J. Newman. The structure of scientific collaboration networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 98(2):404–409, 2001.

[16] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *The Proceedings of the National Academy of Sciences of the United States of America*, 102:7794–7799, 2005.

[17] M. T. Gastner and M. E. J. Newman. The spatial structure of networks. *European Physical Journal B*, 49:247, 2006.

[18] S. Boccaletti, V. Latora, Y. Moreno, M. Chaves, and D.-U. Hwang. Complex networks: structure and dynamics. *Physics Reports*, 424:175–308, 2006.

[19] L. da F. Costa, F. A. Rodrigues, and G. Travieso. Analyzing trails in complex networks. *Physical Review E*, 76(4):46106, 2007.

[20] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:48–98, 2002.

[21] S. N. Dorogovtsev and J. F. F. Mendes. Evolution of networks. *Advances in Physics*, 51:1079–1187, 2002.

[22] M. E. J. Newman. Structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[23] L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007.

[24] S. Strogatz. *Sync: The Emerging Science of Spontaneous Order*. Hyperion, 2003.

[25] L. da F. Costa and M. S. Barbosa. An analytical approach to connectivity in regular neuronal networks. *European Physical Journal B*, 42:573–580, 2004.

[26] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag. Organization, development and function of complex brain networks. *Trends in Cognitive Sciences*, 8(9):418–25, 2004.

[27] L. da F. Costa, O. Sporns, L. Antiqueira, M. G. V. Nunes, and O. N. Oliveira Jr. Correlations between structure and random walk dynamics in directed complex networks. *Applied Physics Letters*, 91:054107, 2007.

[28] L. da F. Costa. Superedges: Connecting structure and dynamics in complex networks. Preprint arXiv:0801.4068, 2008.

[29] B. Bollobás. *Random graphs*. Academic Press, Inc., 1985.

[30] R. Diestel. *Graph Theory*. Springer, 2000.

[31] D. B. West. *Introduction to Graph Theory*. Prentice Hall, 2001.

[32] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[33] D. J. Watts. *Small Worlds: The Dynamics of Networks between Order and Randomness*. Princeton University Press, 2003.

[34] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 101(11):3747–52, 2004.

[35] M. E. J. Newman. Assortative mixing in networks. *Physical Review Letters*, 89(20):208701, 2002.

[36] M. E. J. Newman. Detecting community structure in networks. *European Physical Journal B*, 38:321–330, 2004.

[37] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 99(12):7821–7826, 2002.

[38] L. Danon, J. Duch, A. Arenas, and A. A. Díaz-Guilera. Comparing community structure identification. *Journal of Statistical Mechanics: Theory and Experiment*, page P09008, 2005.

[39] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. Network motifs: Simple building blocks of complex networks. *Science*, 298(5594):824, 2002.

[40] J. Bagrow, E. Bollt, and L. da F. Costa. Network structure revealed by short cycles. Preprint arXiv:cond-mat/0612502, 2006.

[41] P. R. Villas Boas, F. A. Rodrigues, G. Travieso, and L. da F. Costa. Chain motifs: The tails and handles of complex networks. *Physical Review E*, 77(2):26106, 2008.

[42] S. S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nature Genetics*, 31(1):64–68, 2002.

[43] P. R. Villas Boas, F. A. Rodrigues, G. Travieso, and L. da F. Costa. Border trees of complex networks. Preprint arXiv:0706.3403, 2007.

[44] P. R. Villas Boas, F. A. Rodrigues, G. Travieso, and L. da F. Costa. Sensitivity of complex networks measurements. Preprint arXiv:0804.1104, 2008.

[45] L. da F. Costa and F. A. Rodrigues. Seeking for simplicity in complex networks. Preprint arXiv:physics/0702102v2, 2008.

[46] M. Barahona and L. M. Pecora. Synchronization in small-world systems. *Physical Review Letters*, 89(5):54101, 2002.

[47] H. Hong, M. Y. Choi, and B. J. Kim. Synchronization on small-world networks. *Physical Review E*, 65(2):26139, 2002.

[48] D. Volchenkov, L. Volchenkova, and Ph. Blanchard. Epidemic spreading in a variety of scale free networks. *Physical Review E*, 66(4):046137–046146, 2002.

[49] M. Boguñá and R. Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66(4):047104–047107, 2002.

[50] L. da F. Costa and G. Travieso. Exploring complex networks through random walks. *Physical Review E*, 75(1):16102, 2007.

[51] L. da F. Costa. Synchronization of non-linear random walk dynamics in complex network. Preprint arXiv:0801.2520, 2008.

[52] L. da F. Costa. Random and longest paths: Unnoticed motifs of complex networks. Preprint arXiv:0712.0415, 2007.

[53] L. P. Chi and X. Cai. Structural changes caused by error and attack tolerance in US airport network. *International Journal of Modern Physics B*, 18:17–19, 2004.

[54] P. Holme, B. J. Kim, C. N. Yoon, and S. K. Han. Attack vulnerability of complex networks. *Physical Review E*, 65(5):056109, 2002.

[55] B. Tadić, G. J. Rodgers, and S. Thurner. Transport on complex networks: Flow, jamming and optimization. *International Journal of Bifurcation and Chaos*, 17(7):2363–2385, 2007.

[56] L. da F. Costa. Avalanches of activation and spikes in neuronal complex networks. Preprint arXiv:0802.0421, 2008.

[57] P. Ormerod and A. P. Roach. The medieval inquisition: scale-free networks and the suppression of heresy. *Physica A*, 339:645–652, 2004.

[58] P. Ball. The physics of society. *The Proceedings of the National Academy of Sciences of the United States of America*, 415:371, 2002.

[59] B. Wellman. The school child's choice of companions. *Journal of Educational Research*, 14:126–132, 1926.

[60] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[61] D. J. de S. Price. Networks of scientific papers. *Science*, 149:510–515, 1965.

[62] D. J. de S. Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 27:292–306, 1976.

[63] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40:35–41, 1977.

[64] L. C. Freeman. Centrality in social networks: Conceptual clarification. *Social Networks*, 1:215–239, 1979.

[65] S. Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[66] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32:425–443, 1969.

[67] D. J. Watts. *Six Degrees: The Science of a Connected Age*. W. W. Norton and Company, New York, 2003.

[68] A. Degenne and M. Forsé. *Introducing Social Networks*. Sage Publications Inc, London, 1999.

[69] L. C. Freeman. *The Development of Social Network Analysis: A Study in the Sociology of Science*. Empirical Press, Vancouver, Canada, 2004.

[70] *http://www.elsevier.com/locate/inca/505596*. 2008.

[71] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aaberg. The web of human sexual contacts. *Nature*, 411(6840):907–908, 2001.

[72] P. Cano, O. Celma, M. Koppenberger, and J. M. Buldú. Topology of music recommendation networks. *Chaos*, 16:013107, 2006.

[73] J. D. Watts. A twenty-first century science. *Nature*, 445:489, 2007.

[74] B. Wellman. Computer networks as social networks. *Science*, 293:2031–2034, 2001.

[75] 2008. http://www.secondlife.com.

[76] M. C. González and A.-L. Barabási. From data to models. *Nature*, 3:224–225, 2007.

[77] X. Guardiola, A. Guimerà, R. Arenas, A. Díaz-Guilera, D. Streib, and L. A. N. Amaral. Macro- and micro-structure of trust networks. Preprint arXiv:cond-mat/0206240, 2002.

[78] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas. Models of social networks based on social distance attachment. *Physical Review E*, 70:056122, 2004.

[79] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas. Self-similar community structure in a network of human interactions. *Physical Review E*, 68(6):65103, 2003.

[80] W. Bachnik, S. Szymczyk, P. Leszczynski, R. Podsiadlo, E. Rymszewicz, L. Kurylo, D. Makowiec, and B. Bykowska. Quantitive and sociological analysis of blog networks. *Acta Physica Polonica B*, 36(10):3179–3191, 2005.

[81] P. Zakharov. Thermodynamic approach for community discovering within the complex networks: LiveJournal study. Preprint arXiv:physics/0602063, 2006.

[82] S. Valverde, G. Theraulaz, J. Gautrais, V. Fourcassie, and R. V. Solé. Emergent behaviour in agent networks: self-organization in wasp and open source communities. *IEEE Intelligent Systems*, 2006.

[83] D. Lusseau. The emergent properties of a dolphin social network. *Proceedings - Royal Society of London. Biological sciences*, 270:186–188, 2003.

[84] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999.

[85] L. A. N. Amaral, A. Scala, M. Barthélemy, and H. E. Stanley. Classes of small-world networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 97(21):11149–11152, 2000.

[86] P. P. Zhang, K. Chen, Y. He, T. Zhou, B. B. Su, Y. Jin, H. Chang, Y. P. Zhou, L. C. Sun, B. H. Wang, and D. R. He. Model and empirical study on some collaboration networks. *Physica A*, 360(2):599–616, 2006.

[87] P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *Science*, 301(8):827–829, 2003.

[88] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész. Analysis of a large-scale weighted network of one-to-one human communication. *New Journal of Physics*, 9:179, 2007.

[89] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strengths in mobile communication networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 104(18):7332–7336, 2007.

[90] P. Holme, S. M. Park, B. J. Kim, and C. R. Edling. Korean university life in a network perspective: Dynamics of a large affiliation network. *Physica A*, 373:821–830, 2007.

[91] M. C. González, H. J. Herrmann, J. Kertész, and T. Vicsek. Community structure and ethnic preferences in school friendship networks. *Physica A*, 379(1):307–316, 2007.

[92] A. Arenas, L. Danon, A. Díaz-Guilera, P. M. Gleiser, and R. Guimerà. Community analysis in social networks. *European Physical Journal B*, 38:373–380, 2004.

[93] J. H. Jones and M. S. Handcock. An assessment of preferential attachment as a mechanism for human sexual network formation. *Proceedings of the Royal Society of London B*, page 1027, 2003.

[94] F. Liljeros. Sexual networks in contemporary western societies. *Physica A*, 338:238–245, 2004.

[95] P. Ormerod and A. P. Roach. Network of sexual contacts and sexually transmitted HIV infection in Burkina Faso. *Journal of Medical Virology*, 78:724–729, 2006.

[96] M. C. González, P. G. Lind, and H. J. Herrmann. Model of mobile agents for sexual interactions networks. *European Physical Journal B*, 49(3):371–376, 2006.

[97] M. S. Handcock and J. H. Jones. Likelihood-based inference for stochastic models of sexual network formation. *Theoretical Population Biology*, 65:413–422, 2004.

[98] P. Holme, C. R. Edling, and F. Liljeros. Structure and time evolution of an Internet dating community. *Social Networks*, 26(2):155–174, 2004.

[99] J. Park and M. E. J. Newman. A network-based ranking system for US college football. *Journal of Statistical Mechanics: Theory and Experiment*, page 10014, 2005.

[100] R. N. Onody and P. A. de Castro. Complex network study of Brazilian soccer players. *Physical Review E*, 70(3):37103, 2004.

[101] P. M. Gleiser. How to become a superhero. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(09):P09020, 2007.

[102] R. Alberich, J. Miro-Julia, and F. Rosselló. Marvel universe looks almost like a real social network. Preprint arXiv:cond-mat/0202174, 2002.

[103] X. Xu, L. Yang, and F. Liu. Structure of the cross-talk collaboration network of China. *Physica A*, 376:738–746, 2007.

[104] P. M. Gleiser and L. Danon. Community structure in jazz. *Advances in Complex Systems*, 6, 2003.

[105] R. D. Smith. The network of collaboration among rappers and its community structure. *Journal of Statistical Mechanics: Theory and Experiment*, 2006.

[106] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Physical Review E*, 70:066111, 2004.

[107] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.

[108] J. Park, O. Celma, M. Koppenberger, P. Cano, and J. M. Buldú. The social network of contemporary popular musicians. *International Journal of Bifurcation and Chaos*, 17(7):2281–2288, 2007.

[109] D. de L. Silva, M. M. Soares, M. V. C. Henriques, M. T. S. Alves, S. G. de Aguiar, T. P. de Carvalho, G. Corso, and L. S. Lucena. The complex network of the Brazilian popular music. *Physica A*, 332:559–565, 2004.

[110] M. E. J. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical Review E*, 64(1):16131, 2001.

[111] C. Cotta and J. J. Merelo. The complex network of evolutionary computation authors: an initial study. Preprint arXiv:physics/0507196, 2005.

[112] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, 311(3-4):590–614, 2002.

[113] J. Moody. The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, 69:213–238, 2004.

[114] M. E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical Review E*, 64(1):16132, 2001.

[115] F. N. Silva, M. A. Rodrigues, and L. da F. Costa. Concentric characterization and classification of complex network nodes: Theory and application to institutional collaboration. Preprint arXiv:0710.1857, 2007.

[116] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, and J. Wu. Weighted networks of scientific communication: the measurement and topological role of weight. *Physica A*, 350:643–656, 2005.

[117] Y. Fan, M. Li, J. Chen, L. Gao, Z. Di, and J. Wu. Network of econophysicists: a weighted network to investigate the development of econophysics. *International Journal of Modern Physics B*, 18(17-19):2505–2512, 2004.

[118] P. Zhang, M. Li, J. Wu, Z. Di, and Y. Fan. The community structure of econophysicist collaboration networks. Preprint arXiv:physics/0505064, 2005.

[119] Y.-M. Choi and H.-J. Kim. A directed network of Greek and Roman mythology. *Physica A*, 382:665–671, 2007.

[120] J. Liu, Y. Dang, and Z. Wang. Complex network properties of Chinese natural science basic research. *Physica A*, 366:578–586, 2006.

[121] M. J. Barber, A. Krueger, T. Krueger, and T. Roediger-Schluga. Network of European Union-funded collaborative research and development projects. *Physical Review E*, 73:036132, 2006.

[122] S. Lozano, J. Duch, and A. Arenas. Community detection in a large social dataset of European projects. Maryland, April 2006. Sixth SIAM - International Conference on Data Mining.

[123] M. A. Porter, P. J. Friend, A. J. Mucha, and M. E. J. Newman. Community structure in the United States House of Representatives. *Chaos*, 16:041106, 2006.

[124] D. Braha and Y. Bar-Yam. Topology of large-scale engineering problem-solving networks. *Physical Review E*, 69(1):016113, 2004.

[125] D. Braha and Y. Bar-Yam. *Smart Business Networks*, chapter Information flow structure in large-scale product development organizational networks. Springer Verlag, 2005.

[126] D. Braha and Y. Bar-Yam. The statistical mechanics of complex product development: empirical and analytical results. *Management Science*, 53(7):1127–1145, 2007.

[127] H. Ebel, L. I. Mielsch, and S. Bornholdt. Scale-free topology of e-mail networks. *Physical Review E*, 66(3):35103, 2002.

[128] J. R. Tyler, D. M. Wilkinson, and B. A. Huberman. Email as spectroscopy: Automated discovery of community structure within organizations. communities and technologies, 2003.

[129] J. Wang and P. De Wilde. Properties of evolving e-mail networks. *Physical Review E*, 70(6):66121, 2004.

[130] M. E. J. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3):35101, 2002.

[131] D. Braha and Y. Bar-Yam. From centrality to temporary fame: Dynamic centrality in complex networks: Research articles. *Complexity*, 12(2):59–63, 2006.

[132] C. C. Zou and D. W. Gong. Email worm modeling and defense. *Computer Communications and Networks, 2004. ICCCN 2004. Proceedings. 13th International Conference on*, pages 409–414, 2004.

[133] M. van Alstyne and J. Zhang. Emailnet: A system for automatically mining social networks from organizational email communication, 2003.

[134] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 171–180, New York, NY, USA, 2000. ACM.

[135] A. A. Nanavati, S. Gurumurthy, G. Das, D. Chakraborty, K. Dasgupta, S. Mukherjea, and A. Joshi. On the structural properties of massive telecom call graphs: Findings and implications. In *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 435–444, New York, NY, USA, 2006. ACM.

[136] M. S. Granovetter. The strength of weak ties. *American Journal of Sociology*, 78(6):1360, 1973.

[137] M. A. Serrano and M. Boguñá. Topology of the world trade web. *Physical Review E*, 68(1):015101–015104, 2003.

[138] X. Li, Y. Y. Jin, and G. Chen. On the topology of the world exchange arrangements web. *Physica A*, 343:573–582, 2004.

[139] A. Z. Górski, S. Drożdż, J. Kwapień, and P. Oświęcimka. Complexity characteristics of currency networks. *Acta Physica Polonica B*, 37(11):2897–2906, 2006.

[140] J. S. Andrade Jr., D. M. Bezerra, J. Ribeiro Filho, and A. A. Moreira. The complex topology of chemical plants. *Physica A*, 360(2):637–643, 2006.

[141] W. Souma, Y. Fujiwara, and H. Aoyama. Small-world effects in wealth distribution. Preprint arXiv:cond-mat/0108482, 2001.

[142] T. Di Matteo, T. Aste, and S. T. Hyde. Exchanges in complex networks: income and wealth distributions. Preprint arXiv:cond-mat/0310544, 2003.

[143] R. Baggio. Complex systems, information technologies and tourism: a network point of view. *Information Technology and Tourism*, 8(1):15–29, 2006.

[144] R. Baggio. Symptoms of complexity in a tourism system, 2008.

[145] N. Scott, C. Cooper, and R. Baggio. Use of network analysis in tourism research. In *Proceedings of the Advances in Tourism Marketing Conference (ATMC)*, Valencia, Spain, September 2007.

[146] R. Baggio. The web graph of a tourism system. *Physica A*, 379(2):727–734, 2007.

[147] N. Scott, C. Cooper, and R. Baggio. Destination networks - theory and practice in four Australian cases. Annals of Tourism Research, (in press), 2008.

[148] R. Baggio and M. A. Corigliano. Tourism destinations: a network analysis of the Web space. In *Proceedings of the Advances in Tourism Marketing Conference (ATMC)*, Valencia, Spain, September 2007.

[149] N. Scott, C. Cooper, and R. Baggio. *Network Analysis and Tourism: From Theory to Practice*. Channel View, Clevedon, UK, 2008.

[150] L. da F. Costa and R. Baggio. The web of connections between tourism companies in Elba: Structure and dynamics. Preprint arXiv:0803.2510, 2008.

[151] R. N. Mantegna. Hierarchical structure in financial markets. *European Physical Journal B*, 11:193–197, 1999.

[152] G. Bonanno, G. Caldarelli, F. Lillo, and R. N. Mantegna. Topology of correlation-based minimal spanning trees in real and model markets. *Physical Review E*, 68(4):046130, 2003.

[153] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto. Dynamics of market correlations: Taxonomy and portfolio analysis. *Physical Review E*, 68(5):056110, 2003.

[154] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto. Asset trees and asset graphs in financial markets. *Physica Scripta*, T106:48–54, 2003.

[155] J.-P. Onnela, K. Kaski, and J. Kertész. Clustering and information in correlation based financial networks. *European Physical Journal B*, 38(2):353–362, 2004.

[156] S. Battiston, J. F. Rodrigues, and H. Zeytinoglu. The network of inter-regional direct investment stocks across Europe. *Advances in Complex Systems*, 10(1):29–51, 2007.

[157] D. Garlaschelli, S. Battiston, M. Castri, V. D. P. Servedio, and G. Caldarelli. The scale-free topology of market investments. *Physica A*, 350(2-4):491–499, 2005.

[158] W. Souma, Y. Fujiwara, and H. Aoyama. Change of ownership networks in Japan. Preprint arXiv:physics/0504045, 2005.

[159] A. V. Aho and J. D. Ullman. *Foundations of computer science*. Computer Science Press, Inc. New York, NY, USA, 1992.

[160] P. J. Denning. Computer science: The discipline. *Encyclopedia of Computer Science*, 2000.

[161] P. Bishop. *Computing Science*. Nelson, 1991.

[162] S. Valverde, R. Ferrer i Chanco, and R. V. Solé. Scale-free networks from optimal design. *Europhysics Letters*, 60(4):512–517, 2002.

[163] A. P. S. de Moura, Y.-C. Lai, and A. E. Motter. Signatures of small-world and scale-free properties in large computer programs. *Physical Review E*, 68(1):017102–017105, 2003.

[164] S. Valverde and R. V. Solé. Hierarchical small worlds in software architecture. Preprint arXiv:cond-mat/0307278, 2003.

[165] C. R. Myers. Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs. *Physical Review E*, 68(4):046116–046130, 2003.

[166] D. Challet and A. Lombardoni. Bug propagation and debugging in asymmetric software structures. *Physical Review E*, 70(4):046109–046113, 2004.

[167] Y. Ma, K. He, and D. Du. A qualitative method for measuring the structural complexity of software systems based on complex networks. In *Software Engineering Conference, 2005. APSEC '05. 12th Asia-Pacific*, pages 257–263, 2005.

[168] N. LaBelle and E. Wallingford. Inter-package dependency networks in open-source software. Preprint arXiv:cs/0411096, 2006.

[169] A. Iamnitchi, M. Ripeanu, and I. Foster. Locating data in (small-world?) peer-to-peer scientific collaborations. *Lecture Notes in Computer Science*, 2429:232–241, 2002.

[170] A. Iamnitchi, M. Ripeanu, and I. Foster. Data-sharing relationships in the Web. In *Proceedings of the Twelfth International World Wide Web Conference (WWW12)*, 2003.

[171] A. Iamnitchi, M. Ripeanu, and I. Foster. Small-world file-sharing communities. *INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies*, 2:952–963, 2004.

[172] N. Leibowitz, M. Ripeanu, and A. Wierzbicki. Deconstructing the Kazaa network. In *Proceedings of the The Third IEEE Workshop on Internet Applications*, pages 112–120, 2003.

[173] J. S. Kong, P. O. Boykin, B. A. Rezaei, N. Sarshar, and V. P. Roychowdhury. Let your cyberalter ego share information and manage spam. Preprint arXiv:physics/0504026, 2005.

[174] R. Ferrer i Cancho, C. Janssen, and R. V. Solé. Topology of technology graphs: Small world patterns in electronic circuits. *Physical Review E*, 64(4):046119–046123, 2001.

[175] A.-L. Barabási, M. A. de Menezes, S. Balensiefer, and J. Brockman. Hot spot and universality in networks dynamics. *European Physical Journal B*, 38(2):169–175, 2004.

[176] W. Hsu and A. Helmy. IMPACT: Investigation of Mobile-user Patterns Across university Campuses using WLAN Trace analysis. Preprint arXiv:cs/0508009v1, 2005.

[177] W. Hsu and A. Helmy. On nodal encounter patterns in wireless LAN traces. In *4th International Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks*, pages 1–10. IEEE, 2006.

[178] L. da F. Costa and R. M. Cesar Jr. *Shape Analysis and Classification: Theory and Practice*. CRC Press, 2001.

[179] L. da F. Costa. Complex networks, simple vision. Preprint arXiv:cond-mat/0403346, 2004.

[180] F. A. Rodrigues, G. Travieso, and L. da F. Costa. Fast community identification by hierarchical growth. *International Journal of Modern Physics C*, 18(06):937–947, 2007.

[181] T. Chalumeau, L. da F. Costa, O. Laligant, and F. Meriaudeau. Optimized texture classification by using hierarchical complex networks. In *Proceedings of Spie Conference on Machine Vision Applications in Industrial Inspection XIV*, volume 6070, San Jose, California, 2006. IS&T/SPIE 18th Annual Symposium on Electronic Imaging.

[182] C. J. P. Moschovitis, H. Poole, T. Schuyler, and T. M. Senft. *History of the Internet: A Chronology, 1843 to the Present*. ABC-Clio Inc, Santa Barbara, CA, 1999.

[183] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger. The origin of power laws in Internet topologies revisited. In *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies,*. IEEE Computer Society, 2002.

[184] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos. Power laws and the AS-level Internet topology. *IEEE/ACM Transactions on Networking (TON)*, 11(4):514–524, 2003.

[185] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the Internet. *Physical Review Letters*, 87(25):258701, 2001.

[186] A. Vázquez, R. Pastor-Satorras, and A. Vespignani. Large-scale topological and dynamical properties of the Internet. *Physical Review E*, 65:066130, 2002.

[187] G. Caldarelli, R. Marchetti, and L. Pietronero. The fractal properties of Internet. *Europhysics Letters*, 52(4):386–391, 2000.

[188] G. Bianconi. Number of cycles in off-equilibrium scale-free networks and in the Internet at the Autonomous System level. *European Physical Journal B*, 38(2):223–230, 2004.

[189] S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: Correlation profile of the Internet. *Physica A*, 333:529–540, 2004.

[190] S. Zhou and R. J. Mondragón. The rich-club phenomenon in the Internet topology. *Communications Letters, IEEE*, 8(3):180–182, 2004.

[191] J. J. McAuley, L. da F. Costa, and T. S. Caetano. Rich-club phenomenon across complex network hierarchies. *Applied Physics Letters*, 91:084103, 2007.

[192] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir. MEDUSA - new model of Internet topology using k-shell decomposition. In *Proceedings of the International Conference on Network Science*, 2006.

[193] S. Carmi, S. Havlin, S. Kirkpatrick, Y. Shavitt, and E. Shir. A model of Internet topology using k-shell decomposition. *The Proceedings of the National Academy of Sciences of the United States of America*, 104(27):11150–11154, 2007.

[194] DIMES - Distributed Internet MEasurements and Simulations. http://www.netdimes.org.

[195] K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen. Modularity and extreme edges of the Internet. *Physical Review Letters*, 90(14):148701, 2003.

[196] V. Rosato, L. Issacharoff, S. Meloni, D. Caligiore, and F. Tiriticco. Is the topology of the Internet network really fit to sustain its function? *Physica A*, 387(7):1689–1704, 2008.

[197] M. Ripeanu, I. Foster, and A. Iamnitchi. Mapping the Gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. *IEEE Internet Computing*, 6:50–57, 2002.

[198] J.-L. Guillaume and S. Le-Blond. Statistical properties of exchanges in P2P systems. In *Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications*, volume 3, pages 1423–1428, 2004.

[199] V. Latora and M. Marchiori. Efficient behavior of small-world networks. *Physical Review Letters*, 87:198701, 2001.

[200] M. A. Menezes and A.-L. Barabási. Fluctuations in network dynamics. *Physical Review Letters*, 92:028701, 2004.

[201] M. A. Menezes and A.-L. Barabási. Separating internal and external dynamics of complex systems. *Physical Review Letters*, 93:068701, 2004.

[202] J. Duch and A. Arenas. Scaling of fluctuations in traffic on complex networks. *Physical Review Letters*, 96(21):218702, 2006.

[203] M. Barthélemy, B. Gondran, and E. Guichard. Large scale cross-correlations in Internet traffic. *Physical Review E*, 66(5):056110, 2002.

[204] R. Percacci and A. Vespignani. Scale-free behavior of the Internet global performance. *European Physical Journal B*, 32(4):411–414, 2003.

[205] S. Valverde and R. V. Solé. Internet's critical path horizon. *European Physical Journal B*, 38(2):245–252, 2004.

[206] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno. Improved routing strategies for Internet traffic delivery. *Physical Review E*, 70:056105, 2004.

[207] D. Krioukov, k. claffy, K. Fall, and A. Brady. On compact routing for the Internet. *ACM SIGCOMM Computer Communication Review (CCR)*, 37(3):41–52, 2007.

[208] M. Boguñá, D. Krioukov, and k. claffy. Navigability of complex networks. Preprint arXiv:0709.0303, 2007.

[209] H. P. Thadakamalla, R. Albert, and S. R. T. Kumara. Search in spatial scale-free networks. *New Journal of Physics*, 9(6):190, 2007.

[210] F. A. Rodrigues, P. R. V. Boas, G. Travieso, and L. da F. Costa. Seeking the best Internet model. Preprint arXiv:0706.3225, 2007.

[211] S.-H. Yook, H. Jeong, and A.-L. Barabási. Modeling the Internet's large-scale topology. *The Proceedings of the National Academy of Sciences of the United States of America*, 99(21):13382–13386, 2002.

[212] J. Park and M. E. J. Newman. Origin of degree correlations in the Internet and other networks. *Physical Review E*, 68:026112, 2003.

[213] V. Rosato and F. Tiriticco. Growth mechanisms of the AS-level Internet network. *Europhysics Letters*, 66(4):471–477, 2004.

[214] P. Holme and B. J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65:026107, 2002.

[215] S. Zhou and R. J. Mondragón. Accurately modeling the Internet topology. *Physical Review E*, 70(6):066108, 2004.

[216] S. Zhou. Understanding the evolution dynamics of Internet topology. *Physical Review E*, 74(1):016124, 2006.

[217] M. A. Serrano, M. Boguñá, and A. Díaz-Guilera. Modeling the Internet. *European Physical Journal B*, 50(1-2):249–254, 2006.

[218] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D. J. Weitzner. Creating a science of the Web. *Science*, 313(5788):769–771, 2006.

[219] S. Lawrence and C. L. Giles. Accessibility of information on the Web. *Nature*, 400:107–109, 1999.

[220] A.-L. Barabási, R. Albert, and H. Jeong. Scale-free characteristics of random networks: the topology of the World Wide Web. *Physica A*, 281:69–77, 2000.

[221] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the Web. *Computer Networks*, 33:309–320, 2000.

[222] D. Donato, L. Laura, S. Leonardi, and S. Millozzi. Large scale properties of the Webgraph. *European Physical Journal B*, 38(2):239–243, 2004.

[223] D. M. Pennock, G. W. Flake, S. Lawrence, E. J. Glover, and C. L. Giles. Winners don't take all: Characterizing the competition for links on the Web. *The Proceedings of the National Academy of Sciences of the United States of America*, 99(8):5207–5211, 2002.

[224] B. A. Huberman and L. A. Adamic. Growth dynamics of the World-Wide Web. *Nature*, 401:131, 1999.

[225] C. Song, S. Havlin, and H. A. Makse. Self-similarity of complex networks. *Nature*, 433:392–395, 2005.

[226] S. Brin and L. Page. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30:107–117, 1998.

[227] L. A. Adamic. The small world web. In *Proceedings of the 3rd European Conf. Research and Advanced Technology for Digital Libraries, ECDL*, pages 443–452, 1999.

[228] G. Bianconi, N. Gulbahce, and A. E. Motter. Local structure of directed networks. Preprint arXiv:0707.4084v1, 2007.

[229] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani. Structure of cycles and local ordering in complex networks. *European Physical Journal B*, 38(2):183–186, 2004.

[230] G. W. Flake, S. Lawrence, C. L. Giles, and F. Coetzee. Self-organization and identification of Web communities. *IEEE Computer*, 35(3):66–71, 2002.

[231] J.-P. Eckmann and E. Moses. Curvature of co-links uncovers hidden thematic layers in the World Wide Web. *The Proceedings of the National Academy of Sciences of the United States of America*, 99(9):5825–5829, 2002.

[232] F. Fu, L. Liu, K. Yang, and L. Wang. The structure of self-organized blogosphere. Preprint arXiv:math.ST/0607361, 2006.

[233] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon. Superfamilies of evolved and designed networks. *Science*, 303:1538–1542, 2004.

[234] K. Shen and L. Wu. Folksonomy as a complex network. Preprint arXiv:cs.IR/0509072, 2005.

[235] Z. Dezsö, E. Almaas, A. Lukács, B. Rácz, I. Szakadát, and A.-L. Barabási. Dynamics of information access on the Web. *Physical Review E*, 73(6):066132, 2006.

[236] B. A. Huberman, P. L. T. Pirolli, J. E. Pitkow, and R. M. Lukose. Strong regularities in World Wide Web surfing. *Science*, 280:95–97, 1998.

[237] J. M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. S. Tomkins. The Web as a graph: Measurements, models and methods. *Lecture Notes in Computer Science*, 1627:1–17, 1999.

[238] R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins. Extracting large-scale knowledge bases from the Web. In *Proceedings of the 25th VLDB Conference*, pages 639–650, Edinburgh, 1999.

[239] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the Web graph. In *Proceedings of the 42st Annual IEEE Symposium on the Foundations of Computer Science*, volume 57, 2000.

[240] S. Bornholdt and H. Ebel. World Wide Web scaling exponent from Simon's 1955 model. *Physical Review E*, 64:035104, 2001.

[241] B. Tadić. Dynamics of directed graphs: The World-Wide Web. *Physica A*, 293:273–284, 2001.

[242] F. Menczer. Correlated topologies in citation networks and the Web. *European Physical Journal B*, 38(2):211–221, 2004.

[243] A. Vazquez. Statistics of citation networks. Preprint arXiv:cond-mat/0105031, 2001.

[244] C. Chen and D. Hicks. Tracing knowledge diffusion. *Scientometrics*, 59(2):199–211, 2004.

[245] A. F. J. van Raan. Reference-based publication networks with episodic memories. *Scientometrics*, 63(3):549–566, 2005.

[246] S. Redner. How popular is your paper? an empirical study of the citation distribution. *European Physical Journal B*, 4:131–134, 1998.

[247] S. Bilke and C. Peterson. Topological properties of citation and metabolic networks. *Physical Review E*, 64:036106, 2001.

[248] C. Tsallis and M. P. de Albuquerque. Are citations os scientific papers a case of nonextensivity? *European Physical Journal B*, 13:777–780, 2000.

[249] Y. He, X. Zhu, and D. R. He. Statistics and developing model of Chinese skyway network. *International Journal of Modern Physics B*, 18(17-19):2595–2598, 2004.

[250] W. Li and X. Cai. Statistical analysis of airport network of China. *Physical Review E*, 69(4):46106, 2004.

[251] V. Pareto. Cours d'économie politique. *Pichou, Lausanne*, 23, 1897.

[252] R. Guimerà and L. A. N. Amaral. Modeling the world-wide airport network. *European Physical Journal B*, 38:381–385, 2004.

[253] G. Bagler. Analysis of the airport network of India as a complex weighted network. Preprint arXiv:cond-mat/0409773, 2004.

[254] N. Barnier and P. Brisset. Graph coloring for air traffic flow management. *Annals of Operations Research*, 130:163–178, 2004.

[255] Y. Hayashi. A review of recent studies of geographical scale-free networks. *IPSJ Digital Courier*, 2(0):155–164, 2006.

[256] S. Porta, P. Crucitti, and V. Latora. The network analysis of urban streets: A dual approach. *Physica A*, 369(2):853–866, 2006.

[257] V. Kalapala, V. Sanwalani, A. Clauset, and C. Moore. Scale invariance in road networks. *Physical Review E*, 73:026130, 2006.

[258] A. Cardillo, S. Scellato, V. Latora, and S. Porta. Structural properties of planar graphs of urban street patterns. *Physical Review E*, 73(6):066107, 2006.

[259] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen. Networks and cities: An information perspective. *Physical Review Letters*, 94(2):28701, 2005.

[260] K. Sneppen, A. Trusina, and M. Rosvall. Hide-and-seek on complex networks. *Europhysics Letters*, 69(5):853–859, 2005.

[261] B.A.N. Travençolo and L. da F. Costa. Outward accessibility in urban street networks: Characterization and improvements. Preprint arXiv:0802.3665, 2008.

[262] K. Nagel and M. Schreckenberg. A cellular automaton model for freeway traffic. *Journal de Physique I France*, 2(2):2221–2229, 1992.

[263] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg. Towards a realistic microscopic description of highway traffic. *Journal of Physics A: Mathematical and General*, 33(48):L477–L485, 2000.

[264] D. Helbing. Traffic and related self-driven many-particle systems. *Reviews of Modern Physics*, 73(4):1067–1141, 2001.

[265] A. Schadschneider. Traffic flow: a statistical physics point of view. *Physica A*, 313(1-2):153–187, 2002.

[266] K. Nagel, J. Esser, and M. Rickert. Large-scale traffic simulations for transportation planning. *Annual Reviews of Computational Physics VII*, 7:151–202, 2000.

[267] A. Schadschneider, W. Knospe, L. Santen, and M. Schreckenberg. Optimization of highway networks and traffic forecasting. *Physica A*, 346:165–173, 2005.

[268] P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna. Small-world properties of the Indian railway network. *Physical Review E*, 67(3):36106, 2003.

[269] K. A. Seaton and L. M. Hackett. Stations, trains and small-world networks. *Physica A*, 339(3-4):635–644, 2004.

[270] K. P. Li, Z. Y. Gao, and B. H. Mao. A weighted network model for railway traffic. *International Journal of Modern Physics C*, 17(9):1339–1347, 2006.

[271] V. Latora and M. Marchiori. Is the Boston subway a small-world network? *Physica A*, 314:109–113, 2002.

[272] J. Sienkiewicz and J. A. Holyst. Public transport systems in Poland: From Bialystok to Zielona Góra by bus and tram using universal statistics of complex networks. *Acta Physica Polonica B*, 36(5):1771, 2005.

[273] J. Sienkiewicz and J. A. Holyst. Statistical analysis of 22 public transport networks in Poland. *Physical Review E*, 72(4):46127, 2005.

[274] C. von Ferber, Y. Holovatch, and V. Palchykov. Scaling in public transport networks. Preprint arXiv:cond-mat/0501296, 2005.

[275] R. Albert, I. Albert, and G. L. Nakarado. Structural vulnerability of the North American power grid. *Physical Review E*, 69(2):25103, 2004.

[276] W. J. Bai, T. Zhou, Z. Q. Fu, Y. H. Chen, X. Wu, and B. H. Wang. Electric power grids and blackouts in perspective of complex networks. *Communications, Circuits and Systems Proceedings, 2006 International Conference on*, 4, 2006.

[277] I. Dobson, B. A. Carreras, V. E. Lynch, and D. E. Newman. An initial model fo complex dynamics in electric power system blackouts. *System Sciences, 2001. Proceedings of the 34th Annual Hawaii International Conference on*, pages 710–718, 2001.

[278] B. Carreras, V. Lynch, M. Sachtjen, I. Dobson, and D. Newman. Modeling blackout dynamics in power transmission networks with simple structure. In *HICSS '01: Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 2*, page 2018, Washington, DC, USA, 2001. IEEE Computer Society.

[279] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman. Critical points and transitions in an electric power transmission model for cascading failure blackouts. *Chaos*, 12(4):985–994, 2002.

[280] B. A. Carreras, V. E. Lynch, I. Dobson, and D. E. Newman. Complex dynamics of blackouts in power transmission systems. *Chaos*, 14(3):643–652, 2004.

[281] B. A. Carreras, D. E. Newman, I. Dobson, and A. B. Poole. Evidence for self-organized criticality in a time series of electric power system blackouts. *Circuits and Systems I: Regular Papers, IEEE Transactions on [see also Circuits and Systems I: Fundamental Theory and Applications, IEEE Transactions on]*, 51(9):1733–1740, 2004.

[282] D. P. Chassin and C. Posse. Evaluating North American electric grid reliability using the Barabási-Albert network model. *Physica A*, 355(2-4):667–677, 2005.

[283] P. Crucitti, V. Latora, and M. Marchiori. A topological analysis of the Italian electric power grid. *Physica A*, 338(1-2):92–97, 2004.

[284] K. Sun. Complex networks theory: A new method of research in power grid. *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*, pages 1–6, 2005.

[285] A. E. Motter and Y.-C. Lai. Cascade-based attacks on complex networks. *Physical Review E*, 66:065102, 2002.

[286] Paolo Crucitti, Vito Latora, and Massimo Marchiori. Model for cascading failures in complex networks. *Physical Review E*, 69(4):045104, 2004.

[287] R. Kinney, P. Crucitti, R. Albert, and V. Latora. Modeling cascading failures in the North American power grid. *European Physical Journal B*, 46:101–107, 2005.

[288] M. Anghel, K. A. Werley, and A. E. Motter. Stochastic model for power grid dynamics. In *HICSS '07: Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, page 113, Washington, DC, USA, 2007. IEEE Computer Society.

[289] F. S. Collins, E. D. Green, A. E. Guttmacher, and M. S. Guyer. A vision for the future of genomics research. *Nature*, 422, 2003.

[290] H. Kitano. Systems biology: A brief overview. *Science*, 295, 2002.

[291] B. Vogelstein, D. Lane, and A. J. Levine. Surfing the p53 network. *Nature*, 408:307–310, 2000.

[292] Z. N. Oltvai and A.-L. Barabási. Life's complexity pyramid. *Science*, 298(5594):763, 2002.

[293] L. A. N. Amaral and J. M. Ottino. Complex networks. *European Physical Journal B*, 38:147–162, 2004.

[294] A.-L. Barabási and Z. N. Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.

[295] N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrín-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. S. Onge, S. Ghanny, M. H. Y. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, and J. F. Greenblatt. Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature*, 440:637–643, 2006.

[296] J. C. Rain, L. Selig, H. De Reuse, V. Battaglia, C. Reverdy, S. Simon, G. Lenzen, F. Petel, J. Wojcik, V. Schaechter, Y. Chemama, A. Labigne, and P. Legrain. The protein-protein interaction map of Helicobacter pylori. *Nature*, 409(6817):211–215, 2001.

[297] L. Giot, J. S. Bader, C. Brouwer, A. Chaudhuri, B. Kuang, Y. Li, Y. L. Hao, C. E. Ooi, B. Godwin, E. Vitols, G. Vijayadamodar, P. Pochart, H. Machineni, M. Welsh, Y. Kong, B. Zerhusen, R. Malcolm, Z. Varrone, A. Collis, M. Minto, S. Burgess, L. McDaniel, E. Stimpson, F. Spriggs, J. Williams, K. Neurath, N. Ioime, M. Agee, E. Voss, K. Furtak, R. Renzulli, N. Aanensen, S. Carrolla, E. Bickelhaupt, Y. Lazovatsky, A. DaSilva, J. Zhong, C. A. Stanyon, R. L. Finley Jr., K. P. White, M. Braverman, T. Jarvie, S. Gold, M. Leach, J. Knight, R. A. Shimkets, M. P. McKenna, J. Chant, and J. M. Rothberg. A protein interaction map of Drosophila melanogaster. *Science*, 302(5651):1727–1736, 2003.

[298] E. Ravasz and A.-L. Barabási. Hierarchical organization in complex networks. *Physical Review E*, 67(2):26112, 2003.

[299] A. W. Rives and T. Galitski. Modular organization of cellular networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 100(3):1128–1133, 2003.

[300] R. Albert. Scale-free networks in cell biology. *Journal of Cell Science*, 118(Pt 21):4947–57, 2005.

[301] A. E. Motter, N. Gulbahce, E. Almaas, and A.-L. Barabási. Predicting synthetic rescues in metabolic networks. *Molecular Systems Biology*, 4(168), 2008.

[302] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296(5569):910–913, 2002.

[303] S. Schnell, S. Fortunato, and S. Roy. Is the intrinsic disorder of proteins the cause of the scale-free architecture of protein-protein interaction networks? Preprint arXiv:q-bio/0606029, 2006.

[304] T. Manke, L. Demetrius, and M. Vingron. Lethality and entropy of protein interaction networks. *Genome Informatics*, 16(1):159–163, 2005.

[305] M. P. Joy, A. Brock, D. E. Ingber, and S. Huang. High-betweenness proteins in the yeast protein interaction network. *Journal of Biomedicine and Biotechnology*, 2005(2):96–103, 2005.

[306] S. Wuchty, Z. N. Oltvai, and A.-L. Barabási. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nature Genetics*, 35(2):176–179, 2003.

[307] L. H. Hartwell, J. J. Hopfield, S. Leibler, and A. W. Murray. From molecular to modular cell biology. *Nature*, 402(6761):C47–C52, 1999.

[308] J. F. Poyatos and L. D. Hurst. How biologically relevant are interaction-based modules in protein networks? *Genome Biology*, 5:R93, 2004.

[309] I. Albert and R. Albert. Conserved network motifs allow protein-protein interaction prediction. *Bioinformatics*, 20(18):3346–3352, 2004.

[310] N. Williams. Yeast genome sequence ferments new research. *Science*, 272(5261):481–500, 1996.

[311] S. Fields and O. Song. A novel genetic system to detect protein-protein interactions. *Nature*, 340(6230):245–246, 1989.

[312] P. L. Bartel and S. Fields. *The Yeast Two-Hybrid System*. Oxford University Press US, 1997.

[313] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *The Proceedings of the National Academy of Sciences of the United States of America*, 98(8):4569–4574, 2001.

[314] P. Legrain and L. Selig. Genome-wide protein interaction maps using two-hybrid systems. *FEBS Letters*, 480(1):32–6, 2000.

[315] S. Mukherjee, S. Bal, and P. Saha. Protein interaction maps using yeast two-hybrid assay. *Current Science*, 81(5):458, 2001.

[316] A. M. Lesk. *Introduction to Protein Architecture: The Structural Biology of Proteins*. Oxford University Press, 2001.

[317] D. Eisenberg, E. M. Marcotte, I. Xenarios, and T. O. Yeates. Protein function in the post-genomic era. *Nature*, 405:823–826, 2000.

[318] I. Xenarios and D. Eisenberg. Protein interaction databases. *Current Opinion in Biotechnology*, 12(4):334–339, 2001.

[319] R. Mrowka, A. Patzak, and H. Herzel. Is there a bias in proteome research? *Genome Research*, 11:1971–1973, 2001.

[320] R. Saito, H. Suzuki, Y. Hayashizaki, and O. Journals. Interaction generality, a measurement to assess the reliability of a protein-protein interaction. *Nucleic Acids Research*, 30(5):1163–1168, 2002.

[321] P. Bartel, C. T. Chien, R. Sternglanz, and S. Fields. Elimination of false positives that arise in using the two-hybrid system. *Biotechniques*, 14(6):920–924, 1993.

[322] S. Fields and R. Sternglanz. The two-hybrid system: an assay for protein-protein interactions. *Trends in Genetics*, 10(8):286–92, 1994.

[323] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, et al. A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature*, 403(6770):623–7, 2000.

[324] C. von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 417(6887):399–403, 2002.

[325] E. M. Marcotte, M. Pellegrini, H. L. Ng, D. W. Rice, T. O. Yeates, and D. Eisenberg. Detecting protein function and protein-protein interactions from genome sequences. *Science*, 285(5428):751–753, 1999.

[326] E. Sprinzak, S. Sattath, and H. Margalit. How reliable are experimental protein-protein interaction data. *Journal of Molecular Biology*, 327(5):919–23, 2003.

[327] M. Pellegrini, E. M. Marcotte, M. J. Thompson, D. Eisenberg, T. O. Yeates, G. M. Suel, S. W. Lockless, M. A. Wall, and R. Ranganathan. Assigning protein functions by comparative genome analysis: protein phylogeneticprofiles. *The Proceedings of the National Academy of Sciences of the United States of America*, 96(8):4285–8, 1999.

[328] H. Hishigaki, K. Nakai, T. Ono, A. Tanigami, and T. Takagi. Assessment of prediction accuracy of protein function from protein-protein interaction data. *Yeast*, 18(6):523–531, 2001.

[329] B. Schwikowski, P. Uetz, and S. Fields. A network of protein-protein interactions in yeast. *Nature Biotechnology*, 18:1257–1261, 2000.

[330] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani. Global protein function prediction from protein-protein interaction networks. *Nature Biotechnology*, 21(6):697–700, 2003.

[331] E. Eisenberg and E. Y. Levanon. Preferential attachment in the protein network evolution. *Physical Review Letters*, 91(13):138701, 2003.

[332] N. Pržulj, D. G. Corneil, and I. Jurisica. Modeling interactome: scale-free or geometric? *Bioinformatics*, 20(18):3508–3515, 2004.

[333] A. Wagner. The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes. *Molecular Biology and Evolution*, 18:1283–1292, 2001.

[334] A. Vazquez, A. Flammini, A. Maritan, and A. Vespignani. Modeling of protein interaction networks. *Complexus*, 1(1):38–44, 2003.

[335] R. V. Solé, R. Pastor-Satorras, E. Smith, and T. B. Kepler. A model of large-scale proteome evolution. *Advances in Complex Systems*, 5(1):43–54, 2002.

[336] R. Pastor-Satorras, E. Smith, and R. V. Solé. Evolving protein interaction networks through gene duplication. *Journal of Theoretical Biology*, 222:199–210, 2002.

[337] R. V. Solé and P. Fernandez. Modularity “for free” in genome architecture? Preprint arXiv:q-bio/0312032, 2003.

[338] A. Wagner. How the global structure of protein interaction networks evolves. *Proceedings: Biological Sciences*, 270(1514):457–466, 2003.

[339] J. Berg, M. Lässig, and A. Wagner. Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evolutionary Biology*, 4(51), 2004.

[340] C. Vogel, S. A. Teichmann, and J. Pereira-Leal. The relationship between domain duplication and recombination. *Journal of Molecular Biology*, 346(1):355–65, 2005.

[341] S. K. Ng, Z. Zhang, and S. H. Tan. Integrative approach for computationally inferring protein domain interactions. *Bioinformatics*, 19(8):923–929, 2003.

[342] D. Betel, R. Isserlin, and C. W. V. Hogue. Analysis of domain correlations in yeast protein complexes. *Bioinformatics*, 20 Suppl 1:I55–I62, 2004.

[343] S. Wuchty. Scale-free behavior in protein domain networks. *Molecular Biology and Evolution*, 18:1694–1702, 2001.

[344] S. Wuchty. Interaction and domain networks of yeast. *Proteomics*, 2(12):1715–1723, 2002.

[345] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[346] A. Wagner. The small world inside large metabolic networks. *Proceedings: Biological Sciences*, 268(1478):1803–1810, 2001.

[347] J. S. Edwards and B. O. Palsson. The Escherichia coli mg1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *The Proceedings of the National Academy of Sciences of the United States of America*, 97(10):5528–5533, 2000.

[348] S. Schuster, D. A. Fell, and T. Dandekar. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, 18:326–332, 2000.

[349] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási. Hierarchical organization of modularity in metabolic networks. *Science*, 297(5586):1551–1555, 2002.

[350] L. Diambra and L. da F. Costa. Complex networks approach to gene expression driven phenotype imaging. *Bioinformatics*, 21(20):3846–3851, 2005.

[351] U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, 2006.

[352] A. Blais and B. D. Dynlacht. Constructing transcriptional regulatory networks. *Genes & Development*, 19(13):1499–1511, 2005.

[353] E. D. Siggia. Computational methods for transcriptional regulation. *Current Opinion in Genetics & Development*, 15:214–221, 2005.

[354] M. M. Babu, N. M. Luscombe, L. Aravind, M. Gerstein, and S. A. Teichmann. Structure and evolution of transcriptional regulatory networks. *Current Opinion in Structural Biology*, 14(283):15193307, 2004.

[355] T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, et al. Transcriptional regulatory networks in Saccharomyces cerevisiae. *Science*, 298(5594):799–804, 2002.

[356] N. Guelzim, S. Bottani, P. Bourgine, and F. Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics*, 31:60–63, 2002.

[357] H. W. Ma, J. Buer, and A. P. Zeng. Hierarchical structure and modules in the Escherichia coli transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics*, 5:199, 2004.

[358] R. Dobrin, Q. K. Beg, A.-L. Barabási, and Z. N. Oltvai. Aggregation of topological motifs in the Escherichia coli transcriptional regulatory network. *BMC Bioinformatics*, 5(1):10, 2004.

[359] H. Yu, D. Greenbaum, H. Xin Lu, X. Zhu, and M. Gerstein. Genomic analysis of essentiality within protein networks. *Trends Genet*, 20(6):227–231, 2004.

[360] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed. *The Proceedings of the National Academy of Sciences of the United States of America*, 101(14):4781–4786, 2004.

[361] K. Klemm and S. Bornholdt. Topology of biological networks and reliability of information processing. *The Proceedings of the National Academy of Sciences of the United States of America*, 102(51):18414–18419, 2005.

[362] S. Bornholdt. Less is more in modeling large genetic networks. *Science*, 310(5747):449–451, 2005.

[363] S. A. Teichmann and M. M. Babu. Gene regulatory network growth by duplication. *Nature Genetics*, 36(5):492–496, 2004.

[364] N. S. Holter, M. Mitra, A. Maritan, M. Cieplak, J. R. Banavar, and N. V. Fedoroff. Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *The Proceedings of the National Academy of Sciences of the United States of America*, 97(15):8409–8414, 2000.

[365] O. Alter, P. O. Brown, and D. Botstein. Singular value decomposition for genome-wide expression data processing and modeling. *The Proceedings of the National Academy of Sciences of the United States of America*, 97(18):10101–10106, 2000.

[366] A.-L. Barabási. Network medicine - from obesity to the “diseasome”. *New England Journal of Medicine*, 357(4):404, 2007.

[367] L. da F. Costa, M. P. Viana, and M. E. Beletti. Complex channel networks of bone structure. *Applied Physics Letters*, 88:033903, 2006.

[368] L. da F. Costa, M. P. Viana, and M. E. Beletti. Hierarchy, fractality, small-world and resilience of Haversian bone structure: A complex network study. Preprint arXiv:q-bio.TO/0506019, 2005.

[369] B. Drossel and A. McKane. *Handbook of graphs and networks*, chapter Modelling food webs. Wiley-VCH, Berlin, 2003.

[370] J. E. Cohen, F. Briand, and C. M. Newman. *Community food webs*. Springer-Verlag New York, 1990.

[371] P. A. Murtaugh and J. P. Kollath. Variation of trophic fractions and connectance in food webs. *Ecology*, 78(5):1382–1387, 1997.

[372] R. J. Williams, E. L. Berlow, J. A. Dunne, A.-L. Barabási, and N. D. Martinez. Two degrees of separation in complex food webs. *The Proceedings of the National Academy of Sciences of the United States of America*, 99(20):12913–12916, 2002.

[373] D. Garlaschelli, G. Caldarelli, and L. Pietronero. Universal scaling relations in food webs. *Nature*, 423(6936):165–8, 2003.

[374] D. Garlaschelli. Universality in food webs. *European Physical Journal B*, 38(2):277–285, 2004.

[375] I. Hanski. Metapopulation dynamics. *Nature*, 396(6706):41–49, 1998.

[376] V. Vuorinen, M. Peltomäki, M. Rost, and M. J. Alava. Networks in metapopulation dynamics. *European Physical Journal B*, 38(2):261–268, 2004.

[377] P. R. Guimarães Jr, M. A. de Menezes, R. W. Baird, D. Lusseau, P. Guimarães, and S. F. Reis. Vulnerability of a killer whale social network to disease outbreaks. *Physical Review E*, 76(4):42901, 2007.

[378] D. H. Boucher. *The Biology of Mutualism: Ecology and Evolution*. Oxford University Press, 1988.

[379] E. L. Rezende, J. E. Lavabre, P. R. Guimaraes, P. Jordano, and J. Bascompte. Non-random coextinctions in phylogenetically structured mutualistic networks. *Nature*, 448(7156):925–928, 2007.

[380] P. R. Guimarães Jr, V. Rico-Gray, S. F. Reis, and J. N. Thompson. Asymmetries in specialization in ant-plant mutualistic networks. *Proceedings of the Royal Society B: Biological Sciences*, 273(1597):2041–2047, 2006.

[381] J. Bascompte, P. Jordano, C. J. Melián, and J. M. Olesen. The nested assembly of plant-animal mutualistic networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 100(16):9383–9387, 2003.

[382] P. R. Guimarães Jr, C. Sazima, S. F. Reis, and I. Sazima. The nested structure of marine cleaning symbiosis: is it like flowers and bees? *Biology Letters*, 3(1):51–54, 2007.

[383] J. M. Beggs and D. Plenz. Neuronal avalanches in neocortical circuits. *Journal of Neuroscience*, 23(35):11167, 2003.

[384] O. Sporns. Network analysis, complexity, and brain function. *Complexity*, 8(1):56–60, 2002.

[385] L. da F. Costa. Morphological complex networks: Can individual morphology determine the general connectivity and dynamics of networks? Preprint arXiv:q-bio.MN/0503041, 2005.

[386] C. Koch and G. Laurent. Complexity and the nervous system. *Science*, 284(5411):96–8, 1999.

[387] M. Hämäläinen, R. Hari, R. J. Ilmoniemi, J. Knuutila, and O. V. Lounasmaa. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Reviews of Modern Physics*, 65(2):413–497, 1993.

[388] N. K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–157, 2001.

[389] L. da F. Costa and L. Diambra. Topographical maps as complex networks. *Physical Review E*, 71(2):21901, 2005.

[390] L. da F. Costa, M. S. Barbosa, V. Coupez, and D. Stauffer. Morphological Hopfield networks. *Brain and Mind*, 4(1):91–105, 2003.

[391] D. J. Felleman and D. C. Van Essen. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47, 1991.

[392] C. C. Hilgetag. Anatomical connectivity defines the organization of clusters of cortical areas in the macaque and the cat. *Philosophical Transactions: Biological Sciences*, 355(1393):91–110, 2000.

[393] G. A. P. C. Burns. Analysis of the connectional organization of neural systems associated with the hippocampus in rats. *Philosophical Transactions: Biological Sciences*, 355(1393):55–70, 2000.

[394] J. W. Scannell, C. Blakemore, and M. P. Young. Analysis of connectivity in the cat cerebral cortex. *Journal of Neuroscience*, 15(2):1463–1483, 1995.

[395] J. Karbowski. Optimal wiring principle and plateaus in the degree of separation for cortical neurons. *Physical Review Letters*, 86(16):3674–3677, 2001.

[396] C. C. Hilgetag. Uniformity, specificity and variability of corticocortical connectivity. *Philosophical Transactions: Biological Sciences*, 355(1393):7–20, 2000.

[397] L. da F. Costa and O. Sporns. Hierarchical features of large-scale cortical connectivity. *European Physical Journal B*, 48(4):567–573, 2005.

[398] L. da F. Costa, M. Kaiser, and C. C. Hilgetag. Predicting the connectivity of primate cortical networks from topological and spatial node properties. *BMC Systems Biology*, 1:16, 2007.

[399] L. da F. Costa and O. Sporns. Diversity of cortical states at non-equilibrium simulated by the ferromagnetic Ising model under Metropolis dynamics. *International Journal of Bifurcation and Chaos in Applied Sciences and Engineering*, 17(7):2387–2398, 2007.

[400] L. da F. Costa and O. Sporns. Correlating thalamocortical connectivity and activity. *Applied Physics Letters*, 89:013903, 2006.

[401] L. F. Lago-Fernández, R. Huerta, F. Corbacho, and J. A. Sigüenza. Fast response and temporal coherent oscillations in small-world networks. *Physical Review Letters*, 84(12):2758–2761, 2000.

[402] M. I. Rabinovich, P. Varona, A. I. Selverston, and H. D. I. Abarbanel. Dynamical principles in neuroscience. *Reviews of Modern Physics*, 78(4):1213–1265, 2006.

[403] A. M. Aertsen, G. L. Gerstein, M. K. Habib, and G. Palm. Dynamics of neuronal firing correlation: modulation of “effective connectivity”. *Journal of Neurophysiology*, 61(5):900–917, 1989.

[404] C. J. Stam. Functional connectivity patterns of human magnetoencephalographic recordings: a Śsmall-worldŠ network? *Neuroscience Letters*, 355(1-2):25–28, 2004.

[405] D. S. Bassett, A. Meyer-Lindenberg, S. Achard, T. Duke, and E. Bullmore. From the cover: Adaptive reconfiguration of fractal small-world human brain functional networks. *The Proceedings of the National Academy of Sciences of the United States of America*, 103(51):19518, 2006.

[406] S. Dodel, J. M. Hermann, and T. Geisel. Functional connectivity by cross-correlation clustering. *Neurocomputing*, 44(46):1065–1070, 2002.

[407] V. M. Eguíluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian. Scale-free brain functional networks. *Physical Review Letters*, 94(1):18102, 2005.

[408] R. Salvador, J. Suckling, C. Schwarzbauer, and E. Bullmore. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):937–946, 2005.

[409] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore. A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience*, 26(1):63, 2006.

[410] C. J. Stam, B. F. Jones, G. Nolte, M. Breakspear, and P. Scheltens. Small-world networks and functional connectivity in Alzheimer's disease. *Cerebral Cortex*, 17(1):92, 2007.

[411] F. Bartolomei, I. Bosma, M. Klein, J. C. Baayen, J. C. Reijneveld, T. J. Postma, J. J. Heimans, B. W. van Dijk, J. C. de Munck, A. de Jongh, K. S. Cover, and C. J. Stam. Disturbed functional connectivity in brain tumour patients: Evaluation by graph analysis of synchronization matrices. *Clinical Neurophysiology*, 117(9):2039–2049, 2006.

[412] D. Jurafsky and J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall, New Jersey, 2000.

[413] B. Pang and L. Lee. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the ACL 2004*, pages 271–278, 2004.

[414] R. Mihalcea. Random walks on text structures. In *Proceedings of CICLing-2006*, volume 3878 of *LNCS*, pages 249–262, February 2006.

[415] S. N. Dorogovtsev and J. F. F. Mendes. Language as an evolving word web. *Proceedings of the Royal Society of London B*, 268:2603, 2001.

[416] R. Ferrer i Cancho and R. V. Solé. The small world of human language. *Proceedings of the Royal Society of London B*, 268:2261, 2001.

[417] V. Loreto and L. Steels. Social dynamics: Emergence of language. *Nature Physics*, 3:758–760, 2007.

[418] M. Steyvers and J. B. Tenenbaum. The large-scale structure of semantic networks: statistical analyses and a model for semantic growth. *Cognitive Science*, 29:41–78, 2005.

[419] A. E. Motter, A. P. S. Moura, Y.-C. Lai, and P. Dasgupta. Topology of the conceptual network of language. *Physical Review E*, 65:065102, 2002.

[420] A. J. Holanda, I. T. Pisa, O. Kinouchi, A. S. Martinez, and E. E. S. Ruiz. Thesaurus as a complex network. *Physica A*, 344:530–536, 2004.

[421] O. Kinouchi, A. S. Martinez, G. F. Lima, G. M. Lourenço, and S. Risau-Gusman. Deterministic walks in random networks: An application to thesaurus graphs. *Physica A*, 315:665–676, 2002.

[422] M. Sigman and G. A. Cecchi. Global organization of the WordNet lexicon. *The Proceedings of the National Academy of Sciences of the United States of America*, 99(3):1742–1747, 2002.

[423] L. da F. Costa. What's in a name? *International Journal of Modern Physics C*, 15:371–379, 2004.

[424] A. A. A. Ferreira, G. Corso, G. Piuvezam, and M. S. C. F. Alves. A scale-free network of evoked words. *Brazilian Journal of Physics*, 36(3A):755–758, 2006.

[425] M. Choudhury, A. Mukherjee, A. Basu, and N. Ganguly. Analysis and synthesis of the distribution of consonants over languages: a complex network approach. In *Proceedings of the COLING/ACL on Main conference poster sessions*, pages 128–135, 2006.

[426] M. M. Soares, G. Corso, and L. S. Lucena. The network of syllables in Portuguese. *Physica A*, 355:678–684, 2005.

[427] A. Mukherjee, M. Choudhury, A. Basu, and N. Ganguly. Modeling the co-occurrence principles of the consonant inventories: A complex network approach. Preprint arXiv:physics/0606132, 2006.

[428] S. M. G. Caldeira, T. C. Petit Lobão, R. F. S. Andrade, A. Neme, and J. G. V. Miranda. The network of concepts in written texts. *European Physical Journal B*, 49:523–529, 2006.

[429] A. P. Masucci and G. J. Rodgers. Network properties of written human language. *Physical Review E*, 74:026102, 2006.

[430] R. Ferrer i Cancho, R. V. Solé, and R. Köhler. Patterns in syntactic dependency networks. *Physical Review E*, 69:051915, 2004.

[431] R. Ferrer i Cancho, A. Capocci, and G. Caldarelli. Spectral methods cluster words of the same class in a syntactic dependency network. *International Journal of Bifurcation and Chaos*, 17(7):2453–2463, 2007.

[432] A. Capocci, V. D. P. Servedio, G. Caldarelli, and F. Colaiori. Detecting communities in large networks. *Physica A*, 352:669–676, 2005.

[433] L. Antiqueira, M. G. V. Nunes, O. N. Oliveira Jr., and L. da F. Costa. Strong correlations between text quality and complex networks features. *Physica A*, 373:811–820, 2007.

[434] T. A. S. Pardo, L. Antiqueira, M. G. V. Nunes, O. N. Oliveira Jr., and L. da F. Costa. Modeling and evaluating summaries using complex networks. In *Proceedings of the 7th Workshop on Computational Processing of Written and Spoken Portuguese (PROPOR)*, volume 3960 of *LNAI*, pages 1–10. Springer-Verlag, May 2006.

[435] T. A. S. Pardo, L. Antiqueira, M. G. V. Nunes, O. N. Oliveira Jr., and L. da F. Costa. Using complex networks for language processing: The case of summary evaluation. In *Proceedings of the International Conference on Communications, Circuits and Systems (ICCCAS'06) - Special Session on Complex Networks*, 2006.

[436] D. R. Amancio, L. Antiqueira, T. A. S. Pardo, L. da F. Costa, O. N. Oliveira Jr., and M. G. V. Nunes. Complex networks analysis of manual and machine translations. *International Journal of Modern Physics C*, 19(4):1–16, 2008.

[437] L. Antiqueira, T. A. S. Pardo, M. G. V. Nunes, O. N. Oliveira Jr., and L. da F. Costa. Some issues on complex networks for author characterization. In *Proceeedings of the Workshop in Information and Human Language Technology (TIL'06)*, 2006.

[438] G. Bianconi and A.-L. Barabási. Bose-Einstein condensation in complex networks. *Physical Review Letters*, 86(24):5632–5635, 2001.

[439] R. Burioni, D. Cassi, M. Rasetti, P. Sodano, and A. Vezzani. Bose-Einstein condensation on inhomogeneous complex networks. *Journal of Physics B*, 34:4697–4710, 2001.

[440] J. P. K. Doye and Massen C. P. Characterizing the network topology of the energy landscapes of atomic clusters. *The Journal of Chemical Physics*, 122(8):084105, 2005.

[441] T. Taketsugu and D. J. Wales. Theoretical study of rearrangements in water dimer and trimer. *Molecular Physics*, 100(17):2793–2806, 2002.

[442] F. Rao and A. Caflisch. The protein folding network. *Journal of Molecular Biology*, 342(1):299–306, 2004.

[443] D. Gfeller, P. De Los Rios, A. Caflisch, and F. Rao. Complex network analysis of free-energy landscapes. *The Proceedings of the National Academy of Sciences of the United States of America*, 104(6):1817–1822, 2007.

[444] H. Seyed-allaei, H. Seyed-allaei, and M. R. Ejtehadi. The energy landscape network of spin-glasses. Preprint arXiv:0710.5403v1, 2007.

[445] E. T. Lu and R. J. Hamilton. Avalanches and the distribution of solar flares. *Astrophysics Journal Letters*, 380(2):L89–L92, 1991.

[446] D. Hughes and M. Paczuski. Scale-free magnetic networks: comparing observational data with a self-organizing model of the coronal field. Preprint arXiv:astro-ph/0309230v1, 2003.

[447] D. Hughes and M. Paczuski. A heavenly example of scale free networks and self-organized criticality. *Physica A*, 342(1-2):158–163, 2004.

[448] M. Guitterman. Small-world phenomena in physics: the Ising model. *Journal of Physics A*, 33(47):8373–8381, 2000.

[449] A. Barrat and M. Weigt. On the properties of small-world network models. *European Physical Journal B*, 13(3):547–560, 2000.

[450] H. Hong, B. J. Kim, and M. Y. Choi. Comment on “Ising model on a small world network”. *Physical Review E*, 66(1):018101–018102, 2002.

[451] C. P. Herrero. Ising model in small-world networks. *Physical Review E*, 65(6):066110–066115, 2002.

[452] T.-Y. Cai and Z.-Y. Li. Ising model on a small world network. *International Journal of Modern Physics B*, 18(17-19):2575–2578, 2004.

[453] A. Aleksiejuk, J. A. Hołyst, and D. Stauffer. Ferromagnetic phase transition in Barabási-Albert networks. *Physica A*, 310(1-2):260–266, 2002.

[454] G. Bianconi. Mean field solution of the Ising model on a Barabási-Albert network. *Physics Letters A*, 303(2):166–168, 2002.

[455] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Ising model on networks with an arbitrary distribution of connections. *Physical Review E*, 66(1):016104–016108, 2002.

[456] M. Leone, A. Vázquez, A. Vespignani, and R. Zecchina. Ferromagnetic ordering in graphs with arbitrary degree distribution. *European Physical Journal B*, 28(2):191–197, 2002.

[457] M. A. Sumour and M. M. Shabat. Monte Carlo simulation of Ising model on directed Barabási-Albert network. Preprint arXiv:0411055, 2004.

[458] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Potts model on complex networks. *European Physical Journal B*, 38(2):177–182, 2004.

[459] C. Teuscher. On irregular interconnect fabrics for self-assembled nanoscale electronics. Preprint arXiv:cond-mat/060658, 2006.

[460] Z.-Q. Jiang, W.-X. Zhou, B. Xu, and W.-K. Yuan. A complex ammonia plant network in chemical engineering. *AIChE Journal*, 53(2):423–428, 2007.

[461] L. A. N. Amaral and M. Barthélemy. Small-world networks and the conformation space of a short lattice polymer chain. *Europhysics Letters*, 55(4):594–600, 2001.

[462] A. Kabakçioglu and A. L. Stella. Scale-free network hidden in a collapsing polymer. *Physical Review E*, 72(5):055102–055105, 2005.

[463] L. K. Gallos and P. Argyrakis. Absence of kinetic effects in reaction-diffusion processes in scale-free networks. *Physical Review Letters*, 92(13):138301, 2004.

[464] R. V. Solé and A. Munteanu. The large scale organization of chemical reaction networks in astrophysics. *Europhysics Letters*, 68(2):170–176, 2004.

[465] G. Corso. Families and clustering in a natural numbers network. *Physical Review E*, 69(3):036106–036110, 2004.

[466] A. K. Chandra and S. Dasgupta. A small world network of prime numbers. *Physica A*, 357(3-4):436–446, 2005.

[467] S. Abe and N. Suzuki. Scale-free network of earthquakes. *Europhysics Letters*, 65(4):581–586, 2004.

[468] S. Abe and N. Suzuki. Complex-network description of seismicity. *Nonlinear Processes in Geophysics*, 13(2):145–150, 2006.

[469] S. Abe and N. Suzuki. Small-world structure of earthquake network. *Physica A*, 337(1-2):357–362, 2004.

[470] S. Abe and N. Suzuki. Complex earthquake networks: Hierarchical organization and assortative mixing. *Physical Review E*, 74(2):026113–026117, 2006.

[471] S. Abe and N. Suzuki. Dynamical evolution of clustering in complex network of earthquakes. *European Physical Journal B*, 59(1):93–97, 2007.

[472] M. Baiesi and M. Paczuski. Scale-free networks of earthquakes and aftershocks. *Physical Review E*, 69(6):066106–066113, 2004.

[473] V. Latora and M. Marchiori. How the science of complex networks can help developing strategies against terrorism. *Chaos, Solitons and Fractals*, 20(1):69–75, 2004.

[474] J. O. Kephart and S. R. White. Directed-graph epidemiological models of computer viruses. *IEEE Computer Society Symposium on Research in Security and Privacy*, pages 343–359, 1991.

[475] A. Kleczkowski and B. T. Grenfell. Mean-field-type equations for spread of epidemics: The ‘small world’ model. *Physica A*, 274(1-2):355–360, 1999.

[476] C. Moore and M. E. J. Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):05678–05682, 2000.

[477] M. Kuperman and G. Abramson. Small world effect in an epidemiological model. *Physical Review Letters*, 86(13):2909–2912, 2001.

[478] M. Small, P. Shi, and C. K. Tse. Plausible models for propagation of the SARS virus. *IEICE Trans Fundam Electron Commun Comput Sci*, E87-A(9):2379–2386, 2004.

[479] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics and endemic states in complex networks. *Physcial Review E*, 63(6):066117–066124, 2001.

[480] M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Absence of epidemic threshold in scale-free networks with degree correlations. *Physical Review Letters*, 90(2):028701–028703, 2003.

[481] Y. Moreno, R. Pastor-Satorras, and A. Vespignani. Epidemic outbreaks in complex heterogeneous networks. *European Physical Journal B*, 26(4):521–529, 2002.

[482] M. E. J. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):016128–016139, 2002.

[483] R. Pastor-Satorras and A. Vespignani. Epidemic dynamics infinite size scale-free networks. *Physical Review E*, 65(3):035108–035111, 2002.

[484] X. F. Wang and G. Chen. Complex networks: small-world, scale-free and beyond. *Circuits and Systems Magazine, IEEE*, 3(1):6–20, 2003.

[485] F. Liljeros, C. R. Edling, H. E. Stanley, Y. Aberg, and L. A. N. Amaral. Sexual contacts and epidemic thresholds. *Nature*, 423(6940):605–606, 2003.

[486] T. Zhou, Z.-Q. Fu, and B.-H. Wang. Epidemic dynamics on complex networks. *Progress in Natural Science*, 16(5):452–457, 2006.

[487] Z. Dezső and A.-L. Barabási. Halting viruses in scale-free networks? *Physical Review E*, 65(5):055103–055106, 2002.

[488] R. Pastor-Satorras and A. Vespignani. Immunization of complex networks. *Physical Review E*, 65(3):036104–036111, 2002.

[489] R. Pastor-Satorras and A. Vespignani. Epidemics and immunization in scale-free networks. In S. Bornholdt and H. G. Schuster, editors, *Handbook of Graphs and Networks*. Wiley-VCH, Berlin, 2003.

[490] R. Cohen, S. Havlin, and D. ben Avraham. Efficient immunization strategies for computer networks and populations. *Physical Review Letters*, 91(24):247901–247904, 2003.

[491] J. Gómez-Gardeñes, P. Echenique, and Y. Moreno. Immunization of real complex communication networks. *European Physical Journal B*, 49(2):259–264, 2006.

[492] Y. Hayashi, M. Minoura, and J. Matsukubo. Oscillatory epidemic prevalence in growing scale-free networks. *Physical Review E*, 69(1):016112–016119, 2004.

[493] L. A. Meyers, M. E. J. Newman, M. Martin, and S. Schrag. Applying network theory to epidemics: Control measures for outbreaks of Mycoplasma pneumoniae. *Emerging Infectious Diseases*, 9(2):204–210, 2001.

[494] L. M. Sander, C. P. Warren, I. Sokolov, C. Simon, and J. Koopman. Percolation on heterogeneous networks as a model for epidemics. *Mathematical Biosciences*, 180(1):293–305, 2002.

[495] L. M. Sander, C. P. Warrena, and I. M. Sokolovb. Epidemics, disorder, and percolation. *Physica A*, 325(1-2):1–8, 2003.

[496] X. Li and X. A. Wang. On the stability of epidemic spreading in small-world networks: how prompt the recovery should be? *International Journal of Systems Science*, 38(5):401–411, 2007.

[497] W.-P. Guo, X. Li, and X.-F. Wang. Epidemics and immunization on euclidean distance preferred small-world networks. *Physica A*, 380:684–690, 2007.

[498] Z. Zhang, S. Zhou, T. Zou, and J. Guan. Finite epidemic thresholds in fractal scale-free 'large-world' networks. Preprint arXiv:08030361, 2008.

[499] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13:2498–2504, 2003.