

## Advances in Physics

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/tadp20>

### Analyzing and modeling real-world phenomena with complex networks: a survey of applications

Luciano da Fontoura Costa <sup>a b</sup>, Osvaldo N. Oliveira Jr. <sup>a</sup>, Gonzalo Travieso <sup>a</sup>, Francisco Aparecido Rodrigues <sup>c</sup>, Paulino Ribeiro Villas Boas <sup>a</sup>, Lucas Antiqueira <sup>a</sup>, Matheus Palhares Viana <sup>a</sup> & Luis Enrique Correa Rocha <sup>d</sup>

<sup>a</sup> Instituto de Física de São Carlos, Universidade de São Paulo, PO Box 369, 13560-970, São Carlos, São Paulo, Brazil

<sup>b</sup> National Institute of Science and Technology for Complex Systems, Brazil

<sup>c</sup> Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, PO Box 668, 13560-970, São Carlos, São Paulo, Brazil

<sup>d</sup> Department of Physics, Umeå University, 90187, Umeå, Sweden

Available online: 12 May 2011

**To cite this article:** Luciano da Fontoura Costa, Osvaldo N. Oliveira Jr., Gonzalo Travieso, Francisco Aparecido Rodrigues, Paulino Ribeiro Villas Boas, Lucas Antiqueira, Matheus Palhares Viana & Luis Enrique Correa Rocha (2011): Analyzing and modeling real-world phenomena with complex networks: a survey of applications, *Advances in Physics*, 60:3, 329-412

**To link to this article:** <http://dx.doi.org/10.1080/00018732.2011.572452>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any

instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## REVIEW ARTICLE

### Analyzing and modeling real-world phenomena with complex networks: a survey of applications

Luciano da Fontoura Costa<sup>a,b,\*</sup>, Osvaldo N. Oliveira Jr.<sup>a</sup>, Gonzalo Travieso<sup>a</sup>,  
Francisco Aparecido Rodrigues<sup>c</sup>, Paulino Ribeiro Villas Boas<sup>a</sup>, Lucas Antikueira<sup>a</sup>,  
Matheus Palhares Viana<sup>a</sup> and Luis Enrique Correa Rocha<sup>d</sup>

<sup>a</sup>Instituto de Física de São Carlos, Universidade de São Paulo, PO Box 369, 13560-970, São Carlos,  
São Paulo, Brazil; <sup>b</sup>National Institute of Science and Technology for Complex Systems, Brazil;  
<sup>c</sup>Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, PO Box 668, 13560-970,  
São Carlos, São Paulo Brazil; <sup>d</sup>Department of Physics, Umeå University, 90187 Umeå, Sweden

(Received 2 November 2009; final version received 9 March 2011)

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 being 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. A diversity of phenomena are surveyed, which may be classified into no less than 11 areas, providing a clear indication of the impact of the field of complex networks.

**PACS:** 89.75.Fb Structures and organization in complex systems; 02.10.Ox Combinatorics; graph theory; 89.75. He Networks and genealogical trees; 89.75.Da Systems obeying scaling laws; 89.75.Kd Patterns

**Keywords:** complex networks; complex system; network analysis; network theory; collective phenomena

Contents	PAGE
1. Introduction	330
2. Building networks	333
3. Biological networks	336
3.1. Biomolecular networks	336
3.1.1. Protein–protein interaction networks	337
3.1.2. Metabolic networks	341
3.1.3. Genetic networks	342
3.2. Medicine	344
3.3. Ecology	344
3.4. Neurosciences	345
4. Internet	348
5. World Wide Web	352
6. Social networks	355
6.1. Personal relations	356
6.1.1. Acquaintances	357
6.1.2. Sexual relations	358
6.1.3. Trust	359

\*Corresponding author. Email: ldfcosta@gmail.com

6.1.4. Movie actors	359
6.1.5. Sports	360
6.1.6. Scientific collaboration	360
6.2. Citations	361
6.3. Music	363
6.4. Terrorism networks	364
6.5. Communication	364
6.5.1. Electronic mail	365
6.5.2. Telephone	365
7. Epidemic spreading	367
8. Economy	369
8.1. Trade networks	369
8.2. Currency	370
8.3. Industry	370
8.4. Wealth	371
8.5. Tourism	371
8.6. Financial market	372
9. Computer science	373
9.1. Software architecture	374
9.2. Data sharing	375
9.3. Circuits	375
10. Linguistics	375
11. Engineering	379
11.1. Transportation	379
11.1.1. Airports	379
11.1.2. Roads and urban streets	380
11.1.3. Some remarks on transportation networks	381
11.2. Electric power transmission systems	382
12. Earth sciences	383
12.1. Earthquakes	383
12.2. Climate networks	383
13. Basic sciences	384
13.1. Physics	384
13.1.1. Energy landscapes	384
13.2. Chemistry	384
14. Collaboration network of the papers cited in this review	385
15. Conclusions and perspectives	385
Acknowledgements	390
Notes	390
References	390
Appendix	406

1. Introduction

The many achievements of physics over the last few centuries have been based on reductionist approaches, whereby the system of interest is reduced to a small, isolated portion of the world, with full control of the parameters involved (e.g. temperature, pressure, electric field). An interesting instance of reductionism, which is seldom realized, is the modeling of nonlinear phenomena with *linear* models by restricting the parameters and variables in terms of a linear approximation. In

establishing the structure of matter with the quantum theory in the first few decades of the twentieth century, for example, reductionism was key to reaching quantitative treatment of the properties of atoms, molecules and then sophisticated structures such as crystalline solids. Indeed, deciphering the structure of matter was decisive for many developments – not only in physics, but also in chemistry, materials science and more recently in biology [1]. Nevertheless, with reductionist approaches, only limited classes of real-world systems can be treated, for the complexity inherent in naturally occurring phenomena cannot be embedded in the theoretical analysis.

There is now a trend in science to extend the scientific method to become more integrationist and to deal explicitly with nonlinear phenomena. The impressive evolution of the field of complex networks fits perfectly within such a scientific framework [2–7]. Its origins can be traced back to Leonhard Euler’s solution of the Königsberg bridges problem, e.g. [8,9], 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 [10], among others, particularly for the so-called random networks, including the model now known by their name (ER) [6,11–13]. This type of network is characterized by the feature that in a network with  $N$  initially isolated nodes, new connections are established with uniform probability between any pair of nodes. Such networks are well described in terms of their average degree, implying that 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. [14–16], that graph theory started to be systematically applied to represent and model natural phenomena, more specifically social relations. These efforts were mainly 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 [6]. 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 power law organization of the Internet [17] and the identification of such a kind of connectivity in the World Wide Web (WWW) [18], giving rise to the *scale-free* paradigm [6,19]. Subsequent investigations showed that many natural and man-made networks also exhibited scale-free organization, including protein–protein interaction networks [20], domain interaction networks [21], metabolic networks [22], food webs [23], networks of collaborators [24], networks of airports [25] and roads [26].

The success of complex networks is therefore to a large extent a consequence of their natural suitability to represent virtually any discrete system. Moreover, the organization and evolution of such networks, as well as dynamical processes on them [27,28], involve nonlinear 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 comprehensive surveys [27,29–32]. For instance, the behavior of biological neuronal networks, one of the greatest remaining scientific challenges, is largely defined by connectivity, e.g. [33–35]. Because of its generality for representing connectivity in 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 [36,37]. This potential has been confirmed with applications of basic complex networks, encompassing areas such as ecology, genetics, epidemiology, physics, the Internet and WWW, computing, etc. [2,7]. In fact, applications of complex networks are redefining the scientific method through incorporation of dynamic and multidisciplinary features of statistical physics and computer science.

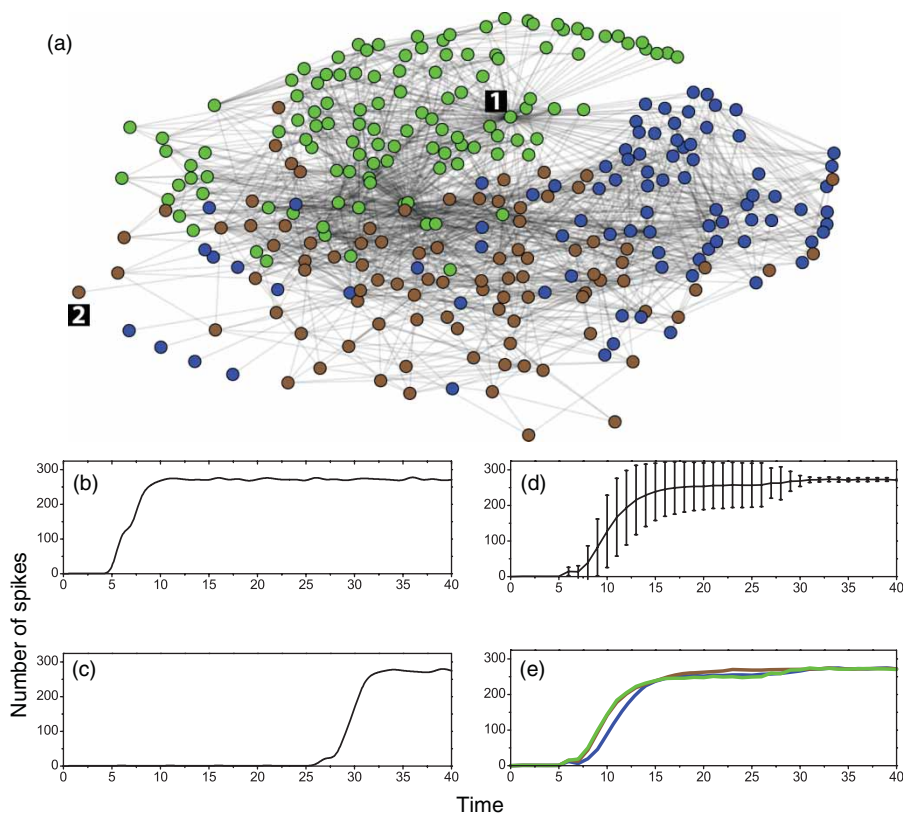


Figure 1. (a) Neuronal network of *C. elegans*, in which each node represents a neuron and two nodes are connected if there is a synaptic connection between the corresponding nodes (network obtained from [38]). The number of spikes in terms of time, considering four stimulation sources [39]: (b) the largest hub (node 1 in (a)); (c) a node with degree 1 (node 2 in (a)); (d) the average and standard deviations considering all nodes as source; and (e) averages taken for each of the communities identified by the corresponding colors.

The importance of studying network topology for establishing structure–function relationships may be illustrated with Figure 1 that deals with the *Caenorhabditis elegans* neuronal network [38]. Integrate-and-fire dynamics has been simulated on this network, leading to the onset of avalanches of activity as a consequence of continuous injection of stimuli from one or more nodes. A typical avalanche signature is shown in Figure 1(b), which was obtained by activating the network from the largest hub. In case the network is activated from a node with smaller degree, such as a node with degree 1 (Figure 1(c)), the onset of the avalanche is substantially delayed. The average and standard deviation of the avalanches obtained while stimulating the network from every possible node are shown in Figure 1(d). It is also possible to inject activity into the network by pumping spikes from a subset of nodes, as illustrated in Figure 1(e), which shows the average signatures with respect to each of the communities identified by the corresponding colors. It is clear from these results that the dynamics on a complex network depends strongly on the origin of the injected stimulation.

Despite the various reviews on complex networks, relatively little material has been provided on the complete modeling approach, from real-world data to dynamics simulation and interpretation. The current survey addresses precisely this issue, which has a central role in complex system research. More specifically, we aim at reviewing works that use complex networks to represent, model, simulate and validate the topological and dynamical aspects of real-world problems.

Advances in these fields have been catalyzed by an ever growing availability of data, which are often made available publically. Such data can be used to build models of topology and dynamics through sound application of complex networks concepts, and the resulting models can then be used to gather insights and predictions about the behavior of the systems under investigation. One distinguishing feature of the present survey is to allow for an integrated discussion of the components and challenges related to complex systems. A practical issue in this context relates to the accuracy and completeness of real-world data. Given that the same problem can be approached at varying scales (e.g. space or time), it is interesting to verify whether these scales have been fully explored in the current approaches. Another important question is associated with the possible existence of universal topological or dynamical features in complex systems, which could be revealed upon using the complex networks approach. Also important is to try to infer the state of the art for the main fields of application, in addition to the identification of the challenges posed in those areas. This survey is aimed at a comprehensive review of 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.

## 2. Building networks

Each category of real-world systems should be represented as complex networks by using a specific mapping procedure. In this section we identify and describe six main approaches which can be used for this purpose. Table 1 summarizes these types of network mapping as well as the main kind of data to which they are normally applied.

The first kind of network construction, henceforth called *communication mapping*, is obtained by monitoring communications between nodes of a network [40]. This method is generally applied to social data, where people or groups of people send and receive messages from one another. It

Table 1. Six types of methods for network construction.

Mapping method	Examples of networks
Communication	Email, phone, social, mail
Coexistence	Domains, collaborations, books, music, movies
Reference	Wikipedia, web, articles, forms, emails, software
Confluence	Cities, highways, undergrounds, circuits, power-grid
Correlation	Climate, financial market, neuroscience
Adjacency (temporal and spatial)	Earthquake, landscape, linguistic

Note: The second column shows examples of real systems, in which these methods can be applied.

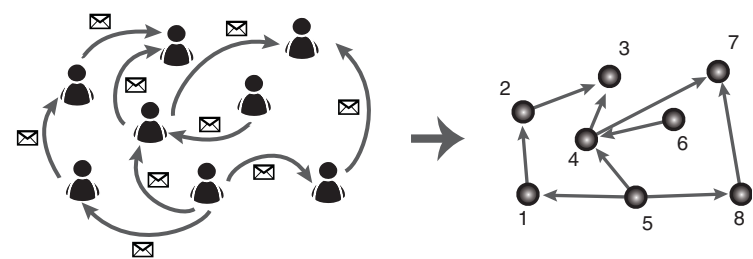


Figure 2. An example of communication mapping where vertices (people) send and receive messages. In a given period of time, the communication between vertices is monitored (left side), and the corresponding network is constructed according to the number of messages exchanged (right side).

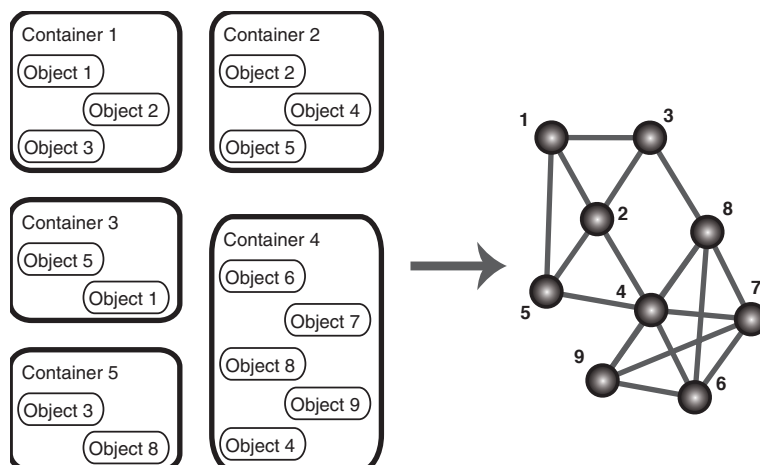


Figure 3. Example of coexistence mapping. Several containers are identified containing different objects. When two objects appear inside the same container, a connection is established between them. This procedure can be used to map such systems as collaboration networks, where the containers correspond to articles and the objects are the authors. Also, movies, musical groups and books can be represented as networks using this approach.

can also be used in other areas, such as computer science, where computers exchange information among themselves, as well as communication systems (e.g. letters, phone, etc.). The construction process is simple and needs only to monitor the vertices considered during a certain time period and connect them according to the number of exchanged messages, as illustrated in Figure 2. An arc is therefore created between vertices  $i$  and  $j$  whenever  $i$  sends at least one message to  $j$  during a given time period. In this kind of network, it is also possible to associate a weight to each arc  $i \rightarrow j$  which corresponds to the number of messages sent from  $i$  to  $j$ .

A large number of systems can be transformed into networks by using *coexistence mapping*. These systems are composed of containers which share a number of objects, as illustrated in Figure 3. Each object is mapped into a vertex, and two nodes are connected if they appear inside the same container. Real-world systems such as movies, musical groups and scientific papers define containers composed by actors, musicians and authors, respectively.

Another important method to build networks, henceforth called *reference mapping*, can be applied to data containing objects interrelated through a finite set of references (see Figure 4). One example is the network of citations among scientific papers: each work builds upon previously published works which are accordingly cited. Other examples are the network of pages connected by hyperlinks in the WWW and the email network where users have a contact list of destination email addresses. This method can also be applied to software components referencing each other and social forums where people are asked to identify their acquaintances.

In general, maps are represented as networks using the *confluence mapping*. The pathways represent the edges of the network, while the confluence of two or more pathways are represented by nodes, as illustrated in Figure 5. This method can be used to represent real-world systems as networks, including urban streets, highways, power grids, bone channels and ant galleries.

Complex networks theory is also an important tool for analysis of multidimensional time series. In these networks, nodes are time series and two nodes are connected according to a similarity metric, such as the Pearson correlation coefficient or mutual information, between their time series. This method will be referred to as *correlation mapping*. The estimated connectivity patterns may depend on choices of the metrics adopted for establishing the connections. Figure 6 shows how



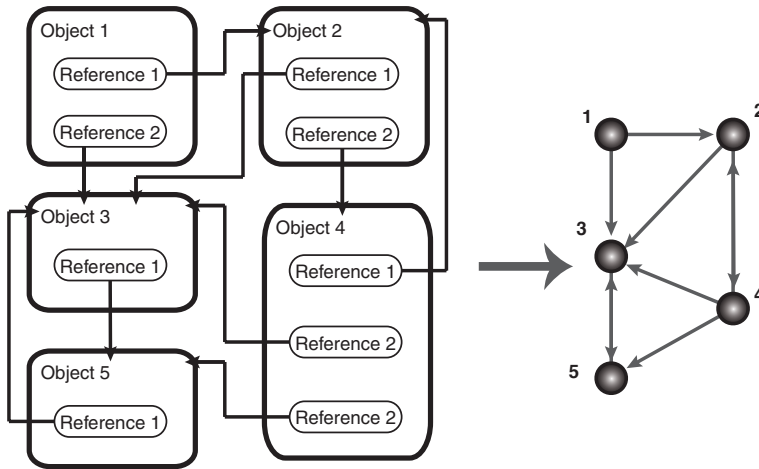


Figure 4. A reference network built from five objects referencing each other. An object is allowed to refer to any other object in the same environment (e.g. webpages in the cyberspace and papers among the set of scientific publications). In principle, the number of citations to other objects is unlimited, and usually connections are directed.

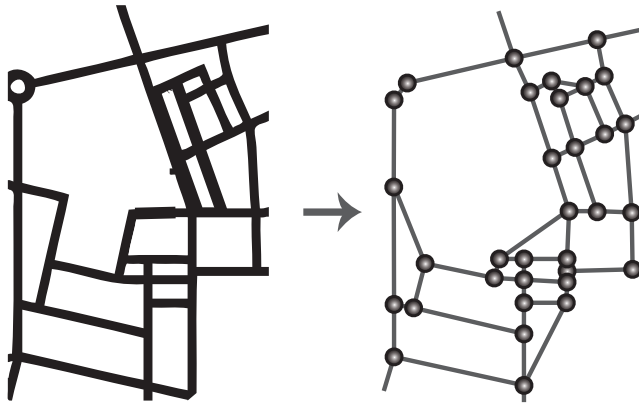


Figure 5. Example of confluence mapping. This method considers edges given by a kind of pathway, with intersections of these pathways corresponding to nodes. For example, the pathways can represent the streets of a city, the highways of a country, the rivers of a region, a digital integrated circuit and so on.

to construct a network from a set of time series (Figure 6(a)). Initially, it is necessary to define the time lag and the metrics to characterize the strength of connections. Afterwards, the calculation of the metrics between each pair of time series results in a weighted, fully connected network (see Figure 6(b)). Some connections can be removed by thresholding, as shown in Figure 6(c). This methodology for constructing networks from a set of time series can be used for climate data [41], financial time series [42], functional magnetic resonance imaging (fMRI) recordings [43] and other applications.

The *adjacency mapping* method takes into account the time or spatial adjacency between objects. When two or more objects are contiguous, edges are created linking the nodes, as shown in Figure 7 for linear adjacencies such as in texts (a) and spatial adjacencies such as in landscape and earthquakes (b).

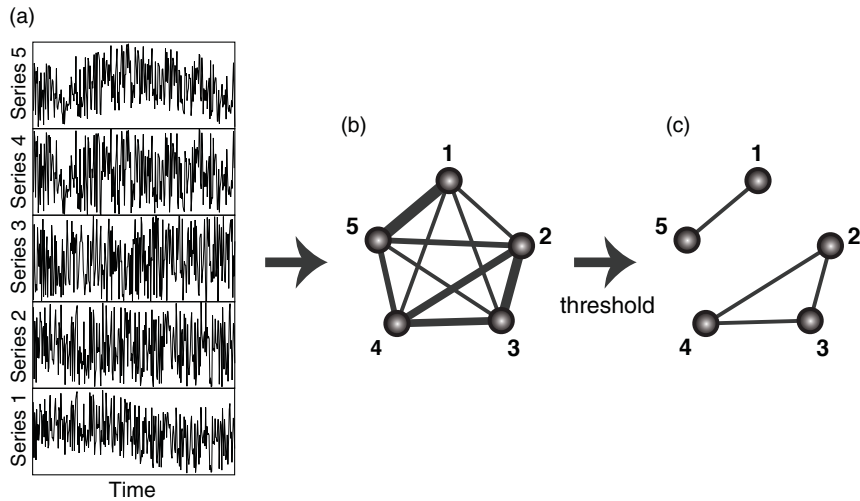


Figure 6. Example of correlation mapping. (a) A set of time series. (b) A weighted fully connected graph, in which each node represents a time series and the edge weights represent the correlation between the corresponding time series. (c) An unweighted network obtained after a thresholding procedure.

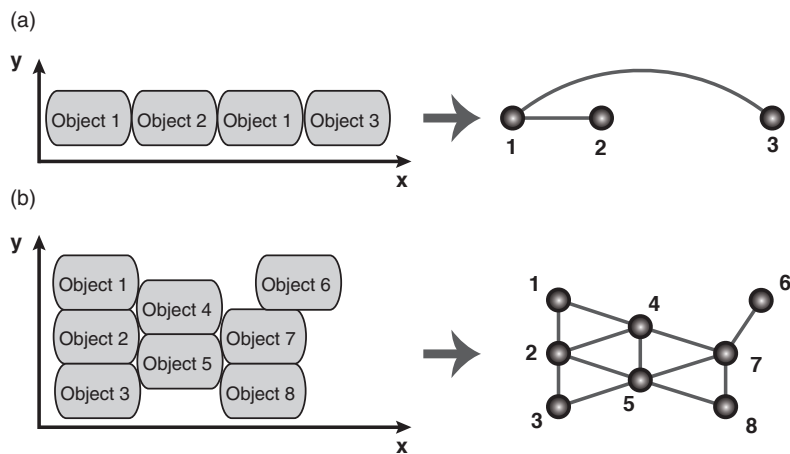


Figure 7. Illustration of the adjacency mapping for linear (a) and spatial adjacencies (b). In this mapping procedure, two or more objects are connected whenever they are contiguous.

3. Biological networks

3.1. 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 [44], 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* [45,46] one has to assemble the parts unfolded by genome projects. For instance, Vogelstein *et al.* [47] 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 [48]: (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 (TF) to the gene that regulates this transcription [8].

Modeling these biological systems may be performed with complex network theory, because of its universality and ability to mimic systems of many interacting parts [46,49]. Biological networks comprise metabolites, enzymes, proteins or genes as vertices, which are linked depending on their interaction. By using complex network theory, fundamental properties of biological networks have been discovered, including power law connectivity distributions, small-world properties, motif and community structures [1]. Furthermore, the crucial concepts to understand biological systems, namely emergence, robustness and modularity, are also ingredients of complex networks theory [45,49].

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

### 3.1.1. Protein–protein interaction networks

In modeling protein interactions, one considers a network comprising proteins (nodes) connected by physical binding (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 [50].

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 [20,51,52], which allows one to infer their importance for cellular maintenance.

Protein interaction networks have also shown to be small-world, with a small average shortest path length and large average clustering coefficient. They present skewed, heavy-tailed degree distributions [20,51,52], hierarchical organization [53], i.e. small groups of nodes organize in a hierarchical manner to make large groups, and modular structure [54–56] composed by densely connected groups of nodes. 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.* [20] showed a correlation between lethality and connectivity by analyzing the yeast protein–protein network, composed by 1870 proteins and 2240 interactions. 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 [20]. 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 [19]. 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 [56]. Even though 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 [57].

The importance of hubs was also addressed by Maslov and Sneppen [58] who showed that the functional modules inside the cell, responsible for biological functions (e.g. transport and regulation of metabolism), are organized around individual hubs. They analyzed the network of *Saccharomyces cerevisiae*, composed by 3278 proteins and 4549 physical interactions. 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 [58]. The functional importance of proteins can also be addressed in terms of network entropy [59] and betweenness centrality [60]. Joy *et al.* [60] 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.* [61] 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 [62]. Wuchty *et al.* analyzed the conservation of 678 proteins of yeast with an orthologous in five eukaryotic species, namely *Arabidopsis thaliana*, *Caenorhabditis elegans*, *Drosophila melanogaster*, *Mus musculus* and *Homo sapiens*. They observed that some motifs are conserved from simple organisms to more complex ones [61]. 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 [48,63], instead of single proteins, such structures are conserved throughout natural selection. In fact, as the cell is organized in modules, the structures should be conserved along evolution and not only the individual cellular components [63,64]. The conservation of network motifs can be used to predict protein interactions. Using machine-learning algorithms, Albert and Albert [65] showed that conserved properties of protein networks can be used to identify and validate protein interactions.

The Baker's yeast (*S. cerevisiae*) has about 6300 genes, which encode about the same number of proteins [66]. Therefore, the determination of interaction among such proteins requires checking 6300 times 6300 pairs, which is close to 40 million potential interactions. In 1989, Stanley Fields proposed a revolutionary technique to detect protein–protein interactions using the GAL4 transcriptional activator of yeast *S. cerevisiae* [67]. This “Yeast two-hybrid”(Y2H) method [68,69] 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 [70,71]. Further information about this method and its extensions can be found in [71,72].

Y2H allowed the global analysis of protein–protein interactions and the birth of *interactome* (one of the next steps after the genome) [73,74]. The success of this method lies on the identification of interactions without antibodies or the need to purify proteins [67]. 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 [75,76]. This limitation has motivated the development of other methods, including biochemical techniques such as affinity [50] and molecular size-based chromatography, affinity blotting, immuno-precipitation and cross-linking [73,77–79]. Also, computational methods are applied to identify protein–protein interactions [80], such as those based on the genome sequence [81]. Sprinzak *et al.* [82] proposed a method to assess the quality of protein interaction databases using the cellular localization and cellular-role properties to estimate the true positives in databases. Krogan *et al.* [50] made available a database of protein interactions based on tandem affinity purification which identified 2708 proteins and 7123 interactions

Table 2. Public databases of protein–protein interactions.

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

of *S. cerevisiae*. Yu *et al.* [83] developed an empirically controlled method to obtain a so-called “second-generation” high-quality, high-throughput Y2H data set, which covers about 20% of all yeast binary interactions. They noted that instead of correlating with essentiality, protein connectivity correlates with genetic pleiotropy. Moreover, their results contrast with computational analyses, which have suggested that high-throughput Y2H data sets contain more false-positives than literature-curated or high-throughput data sets. In Table 2, the main protein–protein interaction databases and their web addresses are presented.

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 [81] and phylogenic profiles [84]. 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 [85]. 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 [86]. Hishigaki *et al.* [85] 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. Vázquez *et al.* [87] extended this method by minimizing the number of physical interactions among different functional categories and considering the connections between unknown proteins, thus achieving 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 in terms of network

measurements [88]. Eisenberg and Levanon [89] proposed that the Barabási–Albert model is relevant for protein interaction modeling. They applied a cross-genome comparison and observed a correlation between proteins' age and their connectivity.

Even though these models reproduce some protein interaction properties, other biological processes must be considered in modeling. This may be performed with addition/elimination of interactions between proteins or gene duplication, which increases the number of proteins and interactions [90]. Vázquez *et al.* [91] employed a model in which each node in the network represents a protein that is expressed by a gene, and the network evolves according to two mechanisms: (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 model has two parameters  $p$  and  $q$ , representing the creation of a connection by node duplication and the loss of interactions, respectively. Solé *et al.* [92] proposed a similar model with the same concepts, which is given in detail in [93]. Solé and Fernández showed that the models by Vázquez *et al.* [91] and Solé *et al.* [92] reproduce the modular structure and the degree of correlation observed in protein interaction networks [94]. In the case of degree of correlations, hubs tend to connect with poorly connected proteins. Also, the modular scale-free structure in protein networks emerge naturally from the duplication-divergence rule [94] or optimization principles [95].

The rate of duplication and deleting of interactions was estimated by Wagner [90] who concluded that every 300 million of years, around half of all existing protein interactions are replaced by new interactions. Wagner [96] 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.* [97] proposed a model of protein interaction with 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 [97].

As domains can recombine to form multi-domain proteins, domain recombination may be the main mechanism to modify protein function and to increase the proteome complexity [98]. Protein interactions can be validated by domain–domain interactions [99] and proteins with similar activities are likely to contain similar domains [100]. 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 built from (i) protein complexes, (ii) Rosetta Stone sequences and by using (iii) protein interaction networks [99,101,102]. 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.* [99] suggested a measure to determine the potential of interaction between domains considering the protein interaction network.

Wuchty [101] 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. In a subsequent paper, Wuchty [102] 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.* [21] 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. The authors 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.

The dynamics of protein–protein interaction networks has also been analyzed, especially using random walk dynamics. For instance, Rodrigues and Costa investigated the self-avoiding random walks dynamics in terms of the accessibility measurement [103] and outward activation [104]. They showed that this type of dynamics is strongly related to protein lethality, where proteins exhibiting high outward activation and central proteins tend to be essential for organism survival. Random dynamics was also used by Manke *et al.* [105] to define network resilience. They used the dynamical entropy for a Markov process and verified that knockouts of proteins with large contribution to network entropy are preferentially lethal.

Complex networks theory has also been used to analyze 3D structures of proteins resulting from linear chains of amino acids, which is a classical example of biological self-organization. Bagler and Sinha [106] studied the role of topological parameters in the kinetics of protein folding in two length scales – protein contact networks (PCNs) and their corresponding long-range interaction networks (LINs), which are constructed by ignoring the short-range interactions. They found that while the assortativity coefficient provides a positive correlation with the rate of protein folding at both short- and long-contact scales, the clustering coefficients of the LINs exhibit a negative correlation.

### 3.1.2. Metabolic networks

Metabolism is the complete set of chemical reactions in living cells to allow for growth and reproduction, to maintain their structures and to respond to the environment. The essential energy and substrates for the function of a cell are obtained by breaking large molecules. 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, 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) have the links connecting all reactants that participate in a same reaction, (iii) links connect two reactions if they share a reactant. Major catalogs available online are the Kyoto Encyclopedia of Genes and Genomes<sup>1</sup> and the EcoCyc<sup>2</sup>.

Jeong *et al.* [107] investigated metabolic networks of 43 organisms from all three domains of life and found that metabolic organization is not random, but follows a power law degree distribution. 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 diameter increases logarithmically with the addition of new vertices. A possible explanation is that as the 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 [108]. 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. 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 [109]. 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 [107]. These substrates tend to be the hubs in metabolic networks.

Ravasz *et al.* [110] 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 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.

Guimerà and Amaral [22] proposed a methodology to find functional modules in complex networks and classify vertices according to their patterns in intra- and inter-module connections. They found that 80% of substrates are only connected to substrates within their modules. Also, substrates with different roles are affected by different evolutionary constraints and pressures. In contrast to the result by Jeong *et al.* [107], 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).

### 3.1.3. Genetic networks

Living cells are governed by gene expression programs involving regulated transcription of thousands of genes. Cell signaling and differentiation may be investigated by patterns of gene expression, which can be represented as complex networks [111]. Transcription is controlled at many levels and the gene regulation network fits into a network of networks that represent not only interactions among TFs, but also the factors that modulate these interactions biochemically [112]. 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 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 immunoprecipitation (ChIP), which was discussed in [113,114].

Transcriptional regulatory networks control 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 TF protein to a gene transcriptionally regulated by that TF, referred to as target gene (TG) [115]. Such complex systems can be analyzed as a multilayered system divided into four basic levels [116]: (i) the first level encompasses the collection of TFs, its TGs with DNA recognition site and regulatory interaction between them; (ii) the second level involves regulatory motifs, which are patterns of interconnections that appear more frequently in real networks than in randomized networks [62]; (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 with frequencies much higher than in random networks, they are expected to have special functions in information processing performed by the network [117]. The complexity of networks can be reduced by considering their motifs. Figure 8 presents the three main types of motifs found in regulatory networks in the yeast of *S. cerevisiae* and the bacteria *E. coli*. In transcriptional regulatory networks, the feed-forward motifs are defined by a TF that regulates a second one, in such a way that they jointly regulate a final TG. On the other hand, the simple input module and the multiple input regulate their targets by a single or multiple TFs, respectively. All these targets are controlled by the same sign (all positive or all negative)



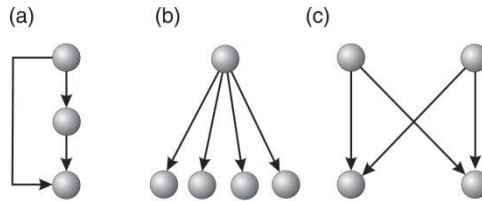


Figure 8. Three types of motif found in transcriptional regulatory networks of *S. cerevisiae* and *E. coli*: (a) feed-forward loop, (b) simple input module and (c) multiple input.

and have no additional transcriptional regulation [115]. Other important motifs are found in the transcriptional regulatory networks of *S. cerevisiae*, for which a detailed discussion is presented in [118].

Modules are organized by 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, separable modules [119]. Ma *et al.* [120] identified 39 modules in transcriptional regulatory networks of *E. coli*, which showed well-defined functions, such as regulatory gene for nitrite and nitrate reductases and production of membrane proteins. Also, Dobrin *et al.* [121] showed that in *E. coli* multiple input and feed-forward, loops motifs do not exist in isolation, but share TFs or TGs. A review about the modular topology in transcriptional regulatory networks is given in [116].

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 [119]. The exponential character of the incoming connectivity indicates that most TGs are regulated by a similar number of factors. On the other hand, the scale-free distribution of the outgoing connections points to a few TFs participating in regulation of a large number of TGs. In addition, such genes tend to be lethal if removed [122]. 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.* [123], who observed that the cell cycle of the regulatory network in yeast was extremely stable and robust. Klemm and Bornholdt [124] 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 [125].

The evolution of transcriptional regulatory networks occurs via three processes: (i) duplication of the TF, which results in both copies regulating the same gene, (ii) duplication of the TG with its regulatory region, where the TG will be regulated by the same TF, and (iii) duplication of both the factor and the target [116]. This mechanism of duplication and inheritance determines a large fraction of the interactions in regulatory networks. Furthermore, interactions are gained or lost after gene duplication, which may result in network motifs and specific connectivity distribution in the network. Teichman and Babu [126] showed that the evolution of regulatory networks is mainly defined by gene duplication and that 45% of regulatory interactions in *E. coli* and *S. cerevisiae* arose from duplication with inheritance of interactions. Balcan *et al.* [117] proposed an information-theoretic approach for modeling transcriptional regulatory networks based on statistics of binding sequences of given specificity in random promoter regions and a “sequence-matching” rule. They showed that the degree distributions, clustering coefficient, degree correlations, rich-club coefficient and the k-core structure obtained from the model agree with the measurements in the real network of *S. cerevisiae*.

Genetic networks have also been investigated according to their dynamics. Elowitz and Leibler [127] studied a synthetic oscillatory network of transcriptional regulators, for which an artificial network was constructed, namely “the repressilator”. This is a cyclic negative-feedback loop composed of three repressor genes and their corresponding promoters. They observed that the state of the network is transmitted to the progeny cells, despite a strong noise component. Gene regulation has also been modeled as Boolean networks, since the gene can be represented as having just two possible states: active or inactive. The first work on this modeling was proposed by Kauffman [128], who considered genes as the nodes of the network where each gene received inputs from a fixed number  $K$  of randomly selected genes. The system can either reach a fixed point, i.e. point attractor, or go successively and repeatedly through a set of configurations, forming a cycle attractor. Although Kauffman’s model agreed with experimental data then available, more recent results suggest that early empirical data have led to incorrect conclusions. More recent models are reviewed in the surveys by Bornholdt [129] and by Drossel [130].

Though the structure of biological networks has been extensively studied, the understanding of temporal activity of a cell is still incipient. Databases with time evolution of gene expression activity are being developed [131,132]. When data on cell network processing and intercommunication become available, new insights about the mechanisms of life will become possible.

### 3.2. Medicine

Complex networks have been useful in medicine, since networks representing biomolecular systems can be applied to understand disease principles and spreading [133]. Barabási [134] suggested the treatment of diseases by taking into account different levels of complex biological networks. For obesity, the analysis should consider biomolecular networks, because some genes or metabolic disfunctions may be of 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. Cancer metastasis has also been analyzed in terms of complex networks [135], with nodes representing the primary cancer site and the sites of subsequent metastases connected by links that measure the strength of co-occurrence. They investigated the evolution of the network links over time, which could be applied to make retrograde predictions of a primary cancer type given a sequence of metastases, as well as anterograde predictions of future sites of metastasis.

In addition to genesis of diseases, complex networks can be useful in epidemiology. Indeed, the spreading of infections caused by virus or bacteria has been studied with dynamical analysis in networks. Studies on spreading virus are discussed in Section 7.

Complex networks theory is potentially suitable to be applied in anatomy. For instance, the bone structure in mammals comprises a complex network of channels (Havers and Volkmann channels), which are required to nourish and maintain the bone marrow cells. The vertices represent intersection of two or more channels and these interconnecting channels are expressed as edges [136]. Costa *et al.* [137] characterized a network from a cortical bone structure in terms of complex networks measurements. The channel branching density obeyed a power law, implying the existence of distribution hubs, thus indicating a high level of resilience against random fails.

### 3.3. Ecology

The characterization, modeling and simulation of ecological systems are major challenges of natural sciences today [138]. In ecological communities, each species interacts in different ways, forming networks of interacting species. Important ecological relations are competition, parasitism, mutualism and predator–prey relationships. Food webs are composed by  $S$  trophic species

connected directly by  $L$  trophic links [139]. The links are represented by arrows from  $i$  to  $j$  if species  $i$  is eaten by species  $j$ . Such arrows 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 [140]. The analysis of such networks suggested that the ratio  $L/S$  was independent of  $S$  and therefore this relation was scale-invariant. However, subsequent works suggested that this belief is not correct [141]. Food webs are small-world networks, obeying the two degree of separation rule [142]. This property implies that the loss of biodiversity and species invasions may affect ecosystems strongly. 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 [139]. Ecological relationships were also explored by analyzing dynamical processes, such as metapopulation dynamics [143,144] and epidemic spreading [145]. In addition, food webs have been studied in terms of robustness [146] using concepts of scaling and universality. This particular dynamics is important to quantify the extinction of species owing to changes in the habitat and global warming [147]. Using the May–Wigner stability criterion [148], Sinha and Sinha [149] showed that increasing complexity in terms of size and connectivity results in greater instability, leading to extinction of a larger number of species. Therefore, previously unconnected ecosystems that start to interact by anthropogenic or natural means are highly vulnerable to large losses in species.

In addition to predator–prey relationships, species can interact according to mutualism – plants and animals establish beneficial interactions, where both individuals derive a benefit [150]. 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 [151]. The structure of networks representing mutualistic relationships is different from those obtained for antagonistic interactions (predator–prey, herbivore–plant). While antagonistic networks are highly compartmentalized [152], 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) [153] and cleaner and client species (cleaning networks) [154], with the cleaning networks being more asymmetric than the plant–animal networks [154]. The nested structure of mutualistic networks is related to phylogenetic relationships [151]. Indeed, phylogenetic-related species tend to have similar biological roles in mutualistic networks [151].

### 3.4. Neurosciences

Complex networks theory is a useful framework for the study of large-scale brain networks. Network models, topological measurements, and dynamical analyses (e.g. avalanches [155,156]) can be considered for brain studies. Representing brain connections with networks is useful to study brain diseases, which are related to network attacks and failures, and brain functions, where it is possible to associate a particular brain architecture with specific brain functions. Brain networks can be investigated at many scales, ranging from individual neurons to large brain areas [157–159]. In the former case, neurons are connected by synapses. In the latter, regions are linked by pathways or functions. Brain networks are generally directed and unweighted, though the description of neuronal systems may be enhanced by incorporating 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 [160]. A limitation of studies in brain connectivity is the proper choice and integration of the spatial and temporal resolutions, because a given data set may reflect individual neuron activities, neural assemblies' dynamics and other macroscopic brain activities [161,162]. 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.

At the microscopic scale, the connectivity of the brain is related to the geometry and spatial position of the individual neuronal cells, e.g. [158,163,164]. For instance, networks involving more intricate neurons are likely to imply enhanced connectivity, therefore affecting the corresponding dynamics, e.g. [34]. An interesting open question concerns how to predict connectivity and dynamics of neuronal networks from the geometric features of specific types of neurons [34,158], since the form of neurons is highly correlated to their functions [34]. One of the first studies using graph theory to represent brain connections was performed by Felleman and Van Essen [165] in 1991. They mapped the hierarchical organization of the cerebral cortex of primates, reporting 305 connections among 32 visual and visual-association areas. Watts and Strogatz [38] in 1998 analyzed neuroanatomical networks for the nervous systems of *C. elegans*. This network, formed by  $N = 282$  neurons and  $M = 3948$  synaptic connections, revealed a small-world 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 [166], 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 [167], cat [168] and monkey [165] are neither regularly nor randomly connected, i.e. they are small-world networks. Recently, the human cortex has also been mapped as a network by using diffusion spectrum imaging [169].

Costa and Sporns [170] characterized cortical networks with hierarchical measurements and identified principles for structural organization of networks. In addition, Costa *et al.* [171] 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 have also been analyzed dynamically. Costa and Sporns [172] 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 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 homogeneous opposite activation were observed in cases (ii) and (iii). The effects from uniform random walks on the dynamic interactions between cortical areas in the cat brain thalamocortical connections were investigated, from which such connections were found to be organized to guarantee strong correlation between the out-degree and occupancy rate [173].

The small-world topology has a direct influence on the dynamical complexity of the network. As discussed by Barahona and Pecora [174] and Hong and Choi [175], information propagates faster on many small-world networks of indirect, 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.* [176] showed that non-identical

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. On the other hand, Percha *et al.* [177] showed that small-world neural networks suffer a transition from local to global phase synchronization depending on the rewiring parameter for network construction ( $p \approx 0.3\text{--}0.4$ ) [177]. Indeed, the authors suggested that neural systems may form networks whose structures lie in the critical regime between local and global synchronization. In this way, the appearance of connections in injured regions may lead to the onset of epileptic seizures. Sinha *et al.* [178] showed that the different ratios of long-range connections in small-world networks can result in strikingly different patterns. Recent studies pointed to the functional role of neuronal-glia ratio in neuro-dynamical patterns, e.g. epilepsy [179]. Chatterjee and Sinha [180] analyzed the hierarchical structure of the *C. elegans* nervous system through the k-core decomposition and showed evidence that the assortativity increases as one goes to the innermost core of the network. The authors suggested that such assortative nature of the inner core can help in increasing communication efficiency while turning the network more robust. Therefore, the topological organization of neuroanatomical networks is directly related to network function. A good review on the dynamical modeling of brain functions is [181].

Other approaches to brain functional networks are based on the concept of functional or effective connectivity. The network is obtained by recordings of brain physiological functions instead of brain anatomy. Aertsen *et al.* [182] introduced graph theory in the study of brain functions. Three basic approaches have been used to obtain the connection structures: (i) electroencephalographic (EEG), (ii) magnetoencephalographic (MEG), and (iii) fMRI. The data from the two first methods are suitable for graph analysis because of the high spatial resolution [43].

De Vico Fallani *et al.* [183] investigated the cortical network dynamics during foot movements and showed that circular motor areas act as network hubs, presenting a large number of outgoing links. In another work, De Vico Fallani *et al.* [184] investigated the cortical network structure of spinal cord of injured patients (SCI), comparing with health subjects. They analyzed the structure of cortical connectivity obtained by the direct transfer function (DTF) applied to the cortical signals estimated from high resolution EEG recordings during the attempt to move a paralyzed limb by a group of five SCI patients. With DTF, the directional influence between each pair of channels could be determined in a multivariate data set, then allowing a directed network to be constructed [184]. The analysis of network structure showed that spinal cord injuries do not affect the global efficiency. Nevertheless, significant differences in the cortical functional connectivity were observed between SCI and normal patients, which can be related to the internal organization of the network and its tolerance to failures (these networks present hubs). In addition, SCI patients revealed a higher local efficiency than healthy subjects, which can be associated with more robustness of the cortical networks of these patients.

Eguíluz *et al.* [185] showed that the functional network obtained with fMRI for seven subjects follows a power law. Similar studies were performed by Salvador *et al.* [186] with five subjects, but using the frequency rather than the time domain. Achard *et al.* [187] concluded that the brain regions are so resilient to random failures as to target attacks (removal of the largest hubs), thus indicating that the brain networks may not be scale-free as suggested by Eguíluz *et al.* [185].

MEG and EEG techniques were also used to study the relationship between brain networks topology and brain pathologies. The matrices obtained by pairs of synchronization likelihood values were converted in unweighted networks by taking a threshold. Stam *et al.* [188] 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.* [189] compared the networks of spatial patterns of functional connectivity of the brain measured at rest by MEG obtained for 17 patients with brain tumors and of 15 healthy patients. They concluded that brain tumors 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 potentially associated with brain damage.

In summary, complex networks theory has proven to be particularly 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 also allow a better understand of brain diseases.

#### 4. Internet

The origins of the Internet can be traced back to 1966, when the military networking system ARPANET was created [190,191]. It has since evolved into a worldwide computer network no longer restricted to academy, through which many services are provided, with email and the WWW being the most popular. Thus, since the Internet is such a multi-purpose 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 indeed employed a graph-based approach to represent the Internet structure [17,191–194]. Nevertheless, the complete mapping of the Internet is difficult to achieve, since 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. These graphs have allowed important findings, as discussed in the following paragraphs.

One common analysis of the Internet involves the study of its vulnerability to failure or attacks [191]. The vulnerability of complex networks, including the Internet at the AS level, was analyzed by Albert *et al.* [19] with regard to a random node removal (node failure) or hub removal (network under attack). In this case, the diameter and size of the connected clusters were monitored while nodes were removed. The results showed that the scale-free networks (Barabási–Albert model, Internet and WWW) are extremely efficient against random failures, differently from the Erdős–Rényi model. However, the hub-based attack rapidly breaks scale-free networks into small isolated groups. This latter finding was challenged by Doyle *et al.* [195], who argued that this belief is not true at the router level, since they showed that the Internet is more accurately modeled by attack-tolerant models than by the scale-free model. In the work reported in [196], attacks were also performed in the Internet at the AS level by 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. Analogous 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 monitored. It was shown that attacks using the recalculated betweenness are 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 [197], where it was called efficiency. The drop in efficiency when a node is removed was calculated for every node of Internet backbone maps of the USA and Europe. The authors identified the nodes with the highest efficiency reduction, which could explain the problems Internet users faced on September 11, 2001. Another interesting observation made by the authors is that these important nodes do not necessarily have high degrees.

A scaling behavior of the form  $\sigma \sim \langle f \rangle^\xi$  between the mean flux per node  $\langle f \rangle$  and the flux dispersion per node  $\sigma$  in the Internet router topology was reported by Menezes and Barabási [198], 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 the shortest paths. The results pointed to a scaling exponent of the Internet ( $\xi \simeq 0.5$ ) reflecting an endogenous behavior. That is to say, the Internet is 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. Furthermore, in Ref. [199], a method was presented to separate internal from external dynamics. The processing capacity of nodes was considered by Duch and Arenas [200], using data similar to the one employed in Ref. [198]. The authors assumed a queue model with Poisson packet arrival and exponential distribution of the processing time per node. Varying the parameter of this model and the time window for measuring data, they showed that  $\xi$  can actually be larger than 0.5. Similarities between Internet traffic and heartbeat dynamics were also found [201], with congested traffic resembling diseased heart dynamics. Conversely, the healthy heart was associated with the unblocked Internet.

Rosato *et al.* [202] performed simulations of traffic flow in an Internet map and noticed that congestion is reached for lower traffics 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 [203] simulated Internet's traffic and detected a critical path horizon that separates a low-efficiency traffic from a high-efficiency 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. Nevertheless, Kim and Motter [204,205] found that there is a large amount of capacity available (more than 94%) in a router-level Internet map, mainly a consequence of large fluctuations in traffic. They employed the router map of the ABILENE backbone, MIT and Princeton University networks, as well as the corresponding average traffic measured in June 2006. Finally, Abe and Suzuki [206] found that Internet traffic follows Omori's law, originally associated with earthquakes. Omori's law states that the number of aftershocks following a main shock, during a time interval, follows a power law (in the case of the Internet, shocks were considered as sudden congestions).

When delivering data packets, routers in the Internet choose neighboring routers to obtain an estimation of the shortest path between the 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.* [207] showed in simulations with an AS map of 2001 that if the queue load of the 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. In a related research, Krioukov *et al.* [208] 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.* [209] analytically proved: nodes can work efficiently using only local information, especially in networks with power law degree distribution and high clustering coefficients, such as the Internet. Tadić *et al.* [210] 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.* [211] as important parameters for transport strategies in geographical scale-free networks such as the Internet.

As for the network growth, it was discovered that the overall Internet performance in 2002, measured in RTT (Round Trip Time) per unit distance, was improved in comparison with the year 2000 [212]. In addition, new links were preferably placed among old nodes than between new nodes in the Internet, which indicates that redundancy plays an important role on its growth, making the network more robust [193]. Zhang *et al.* [213] also analyzed the evolution of the AS-level Internet and made the following observations: (i) the number of nodes grows exponentially (also predicting that it will double every 5.32 years), (ii) the size of the maximum  $k$ -core (with largest  $k$ ) is stable since 2003, indicating that the nucleus of the Internet is not growing, and (iii) the maximum degree of the Internet is also relatively stable, showing that the mostly connected autonomous system is not receiving new connections. In another work, Bianconi *et al.* [214] investigated the scaling of loops of length  $h \leq 5$  and verified that the statistical distributions of loops of order 3, 4 and 5 remain stable during the Internet evolution. It seems that the Internet is a type of network for which the two-point correlation is sufficient to describe the whole local and global topology properties.

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 [191]. Yook *et al.* [215], 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, the positions of nodes form a scale-invariant fractal set, positively correlated with the population density. Park and Newman [216] analytically studied ensembles of networks with the restriction proposed in Ref. [217], i.e. that the maximum number of edges between a pair of nodes is 1. This mechanism leads to the negative degree of correlation (disassortative mixing) reported by Pastor-Satorras *et al.* [193], which inversely associates the average degree of the neighbors of a node with its degree  $k$ . Nevertheless, Park and Newman argue that this approach is not responsible for all the degree correlations verified in the Internet.

Rosato and Tiritico [218] simulated the growth of the Internet at the AS level using a mechanism of triad formation introduced in Ref. [219]. 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 “triangles”, 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 fit Internet data from 1998 to 2000, for quantities such as diameter, clustering coefficient, degree distribution and average geodesic length. In another model, Zhou and Mondragón [220] and Zhou [221,222] represented the Internet at the AS level using an interactive growth mechanism and nonlinear preferential attachment. Their approach, called positive-feedback preference model, is also based on data features such as the initial 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 as well as disassortative mixing.

Serrano *et al.* [223] 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 [224]. 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 led to a power law degree distribution, which is in agreement with previous works [17,193]. The model explains the average shortest path length, the clustering coefficient and the  $k$ -core decomposition of the Internet. Holme *et al.* [225] modeled the AS level composed of geographically constrained subnetworks built for the sake of economic profit. Their approach is able to match the degree distribution of real data. Internet models have also been compared with multivariate statistical methods, as proposed by Rodrigues *et al.* [226]. The authors employed canonical variable analysis and Bayesian decision theory to determine which model is most suitable to reproduce the Internet structure. This approach allows one to consider a large set of measurements and obtain an accurate description of the network structure. Rodrigues *et al.* analyzed Internet models at the AS level and concluded that none of the models considered was able to accurately reproduce the several intricacies of the Internet structure.

Another growth model for the AS-level Internet using preferential attachment (as in the Barabási–Albert model) was described in Ref. [227]. After a new node is created, edges are removed or created following several rules that adjust the local assortativity of nodes. The local assortativity quantifies the contribution a node gives to the overall network assortativity. Simulations were performed and the degree distribution, assortativity and rich-club phenomenon of the model fitted well the real data. Leskovec *et al.* [228] used a very different approach where the Kronecker matrix multiplication is employed. This model starts with an initial graph representing the self-similar structure to be reproduced in the growing network. Then, its adjacency matrix is (Kronecker) multiplied by itself until a large stochastic adjacency matrix is obtained, from which a sample is taken following the independent probabilities of each edge appearing in the final graph. The model was able to reproduce a series of empirically observed features of the Internet at the AS level, including degree distribution, diameter, eigenvalues of the adjacency matrix and clustering coefficient.

Sampling the Internet is an open and crucial problem, as underestimated maps may misguide empirical analyses. Viger *et al.* [229], for instance, proposed the estimation of Internet features through statistical inference on sampled maps. Some work has already been done to assess the implications of using incomplete maps. Using 2-month traceroute measurements, Magnien *et al.* [230] observed that Internet maps collected by this method do not reach a steady topology, with new IP addresses appearing at a high rate. The authors conjectured that dynamic IP addresses, routers with random addresses and slowly discovered paths might cause the instability of traceroute-based maps. The quality of Internet sampling through traceroute-like methods was also analyzed by Guillaume *et al.* [231], with results showing that a high number of distributed exploration units (i.e. sources and targets) contributes positively to a more complete discovery of nodes and links. Moreover, the estimation of the degree distribution and average distance has a good level of confidence even when the number of exploration units is small. Dall'Asta *et al.* [232] presented a model to evaluate shortest path-based sampling and found that the accuracy of the sampled network is intimately related to the betweenness centrality distribution, with heavy tails yielding more accurate samples.

Latora and Marchiori [233] 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 authors illustrated this concept in the communication network Infonet, which is formed by the US and European Internet backbones. Such data were collected from the Russ Haynal's ISP Page<sup>3</sup>, which provides a selection of web links to the major pieces of the Internet's infrastructure. After building the network, the authors deactivated all nodes, one by one, and showed that the most important nodes are New Jersey and New York. The results indicated that the destruction of these nodes can reduce network efficiency by more than 50%. It is interesting

to observe that, in this case, the most connected nodes were 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, reduce network efficiency by 57% and 53%, respectively.

The Internet is one of the first and most studied subjects in the field of complex networks, for which various topological and dynamical features have been already observed. These include large unused bandwidth and far from optimal routing. Many models were proposed for the Internet, which at most reproduce some of its experimentally observed features. Despite the incompleteness of Internet maps, the results from analyses made with different data compiled at distinct time periods and using various methodologies for collection, are consistent most of the time. Nonetheless, a more accurate methodology for data collection is still necessary. Researchers and professionals interested in improving Internet's performance and security may benefit from these relatively recent discoveries made in the complex networks field. Although new insights about Internet's structure and dynamics frequently appear, its distributed, centrally uncontrolled nature, and its constant evolution, still pose challenges for a complete understanding of its main governing rules.

## 5. 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 instant messengers. 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 of (sometimes many more) files that can be accessed by anyone with a Web browser connected to the Internet. These can be virtually any computer file, such as those containing text, image or video; the files, however, that give the WWW a networked structure are the pages, which may contain pointers (hyperlinks) to other pages available in the WWW. Thus, one can think of the WWW as a giant network of pages interconnected by hyperlinks [234]. These files are the most important for 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 [235]. It is huge, with billions of pages, despite its relatively recent birth (released in 1993 to public access) and it 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 access to information. Mainly because of its uncontrolled nature, the WWW is difficult to map completely [236]. Usually a WWW map is constructed by a computer program called "crawler" that navigates through pages storing their hyperlinks, i.e. identifying source and target pages. 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. 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 points 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.

Dynamics on the Web 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.* [237], for example, divided the pages of a news portal into stable pages, i.e. the overall fixed structure of the portal,

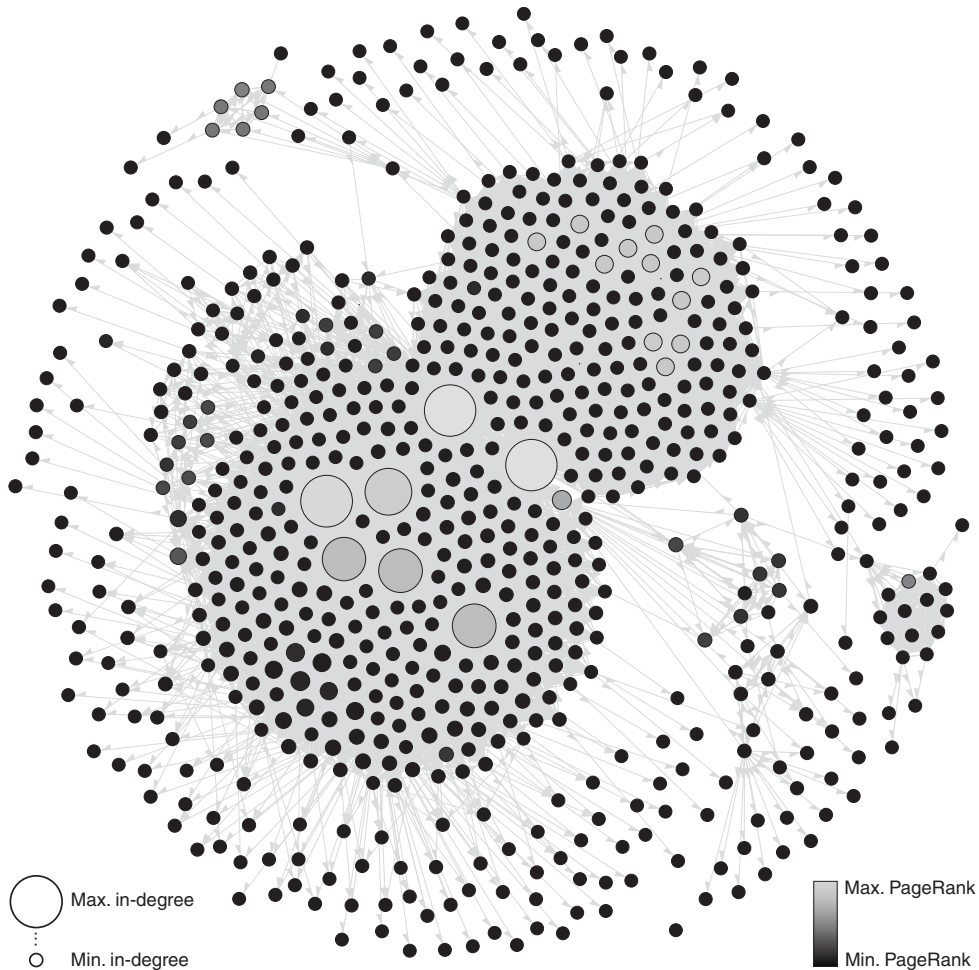


Figure 9. Network with  $N = 703$  nodes and  $E = 23,277$  edges encoding hyperlinks between webpages inside the [www.lincoln.ac.nz](http://www.lincoln.ac.nz) domain [240]. The PageRank was computed for each node and is indicated by different gray levels. Moreover, node sizes are proportional to the respective in-degrees. Notice that the in-degree is not correlated with the PageRank: nodes with similar in-degrees can have very different PageRank values. This means that a page with many hyperlinks pointing to it does not necessarily have high PageRank.

and news pages, which are pages frequently created. The rate of visits to stable pages is constant, while news pages receive 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.* [238] 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 explained real data well, 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, PageRank, is used by Google to sort its search results [239]. 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, which is then

used as a measure of page importance (see Figure 9 for an example). Menezes and Barabási [198] 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  $\sigma$  were estimated. Similarly to what happens with the Internet (see the corresponding section), a scaling behavior of the form  $\sigma \sim \langle f \rangle^\xi$  was found in the WWW, with exponent  $\xi \simeq 1$ . 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.* [241] and Kumar *et al.* [242] 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 other links to them. Ultimately, it allows for 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.* [243] 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 page creators. All these models agreed with observed degree and bipartite core distributions of real Web data.

Bornholdt and Ebel [244] employed the Simon model, created in 1955 to explain the Zipf's law. This model accurately predicts the in-degree exponent found in the WWW [245,246]. 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  $\alpha$  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 [247] 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 are created from  $j$  to an old node and a number of links between old nodes are created or removed. This model accounted for empirical data regarding distributions of in- and out-degree and the distribution of the size of connected components.

Pennock *et al.* [248] changed the Barabási–Albert model of preferential attachment by including a uniform attachment probability. Thus, the probability of connecting a new node  $i$  to an old node  $j$  is a 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 the set of company web sites. Menczer [249] 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 snapshot of the Web, this model can also estimate the content similarity distribution of connected webpages.

The issue of node deletion was addressed in [250], where real-world data spanning 12 months were analyzed and a preferential survival mechanism was discovered. The authors developed and analyzed a model comprising the well-known preferential attachment and the preferential surviving, with results showing that the power law exponent remains the same even when node exclusion occurs frequently. In another work [251], an interplay between experience and talent was found on the Web, where experience is the current number of ingoing hyperlinks a page has and talent is the probability of receiving more hyperlinks. Experiments showed that new talented

pages can “grow” (receive more connections) rapidly and surpass older pages. Barrat *et al.* [252] developed a directed network model of the Web where an in-strength preferential attachment is used for newly created nodes. The model is able to reproduce features of real-world data, such as scale-free degree and strength distributions and hierarchical organization. Another example models a directed network with continuous node creation and edge placement between older nodes which can reproduce in- and out-degree exponents of the WWW [253].

In conclusion, several models of WWW evolution have already been proposed, all of them reproducing some experimentally observed features [234]. One of the main challenges of Web researchers is to learn how to obtain a more complete, accurate set of WWW data, a problem faced by researchers of many other fields. Therefore, misleading observations might be made when analyzing Web maps, possibly a consequence of sampling limitations. For instance, Serrano *et al.* [254] compared four Web maps collected by different crawlers and found divergences when analyzing their out-degree distributions. There seems to be a larger agreement between reported studies concerning the power law in-degree distribution [18,245,246,255,256], despite some results showing the contrary [248]. Furthermore, although the small-world effect in the WWW has been observed by Adamic [257], Broder *et al.* [246] verified that a path between randomly chosen pairs of nodes exists in only 24% of the times, thus pointing to the absence of a small-world effect in their data. Therefore, Web data must be as complete as possible. Since more complete data means even bigger Web maps, some computational complexity issues may arise when carrying out experiments. Another important aspect of Web research is the efficiency of information retrieval, which is the reason why a better understanding of the structure and dynamics of WWW is crucial. Merging page interconnectivity and meta data (i.e. information about page content) is a promising approach because specific parts of the Web can be analyzed according to content, hence allowing more detailed insights that could help the design of search engines and crawlers, and also guide webmasters to better place their pages in the cyberspace.

## 6. Social networks

Since ancient times, the way individuals establish relations among themselves has been crucial to guide the cultural and economic 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 [258], 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 [259], 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 [260], later followed by Mayo with an investigation into interactions among workers in a factory [14]. The motivation for understanding “social networks” increased in the following years, especially with analysis of empirical data. We shall not review those early works, but they are important because the methods developed by social network researchers 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 by social researchers. These included the phenomena described by Simon in his seminal paper on a model to generate highly skewed distribution functions [261], the study of citation networks by Price [262], which converged in a model of network growth able to generate power law degree distributions [263], Freeman’s measurement of centrality (betweenness), which quantified the amount of geodesic paths passing through a node [264,265], and the *small-world* effect, which emerged in the famous Milgram social experiment [266,267], to name a few. Further

information and results from the sociological point of view can be found in specialized books [14–16,268–270]. For the state-of-the-art research, see the papers in the journals *Social Networks*<sup>4</sup> and *Journal of Social Structure*<sup>5</sup> for instance. Furthermore, two recent reviews discussed the historical perspective of social networks analysis [271] and the use of statistical physics to study social networks [272].

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 economic context. As an example, if the Milgram classical experiment [266,267] 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 addition, in sexual networks, Liljeros *et al.* [273] suggested that men may overtell their sexual partners because of social expectations. In music, the level of similarity between two bands may be divergent if assigned by musical experts or enthusiastic fans [274]. Complex networks theory has also been considered to model regional interaction patterns in archaeological contexts. This approach allows one to treat the interactions between sites in geographical space in terms of a network which minimizes an associated Hamiltonian [275,276].

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 WWW popularity [277,278]. The pleasure to be world visible, to contribute to global knowledge, to share thoughts, or only to make unusual friends or find partners has 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”<sup>6</sup> are now widespread for all ages and genders [278–280]. 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 [277,280]. 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 [27,29,31,32], considering real-world data. We present the most important and common structural properties 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.

### 6.1. 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 several 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 relations, including sexual [273] and trust networks [281], or email [282] and blogs [283,284]. In a search for a universal behavior, the concept of personal relations was extended to other species such as wasps [285] and dolphins [286]. Since this topic includes the majority of social networks ever studied, we separate the various subjects into subsections for the sake of better organization.

#### 6.1.1. *Acquaintances*

Popular folklore is fascinated by the so-called “six degrees of separation” concept, resulting from the Milgram’s famous social experiment, which suggested that any two randomly chosen people are separated by six intermediate individuals on average [266]. Since the experiment was done completely inside the USA, a question remained of whether the six degrees also applied to the whole world. Taking advantage of the email system, Dodds *et al.* [287] organized a similar worldwide experiment involving any interested individual. In accordance with Milgram’s results, they found that the diameter of social acquaintances varied between 4 and 7 (if 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 usual 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. Though it seems that we are most likely closely connected to other individuals, in an experiment with internet chain letters Nowell and Kleinberg [288] concluded that information not always flows through small-world principles. They observed the emergence of narrow but deep tree-like structures on samples of popular chain letters against the Iraq war and in favor of Public funding for the Radio in the USA.

The first social experiment relies largely on individual motivation to participate (being limited by lack of interest, time or incentive) [287]. On the other hand, the second experiment constrains the network to a specific dynamical behavior, i.e. chain letters. To overcome some of these experimental biases, more general already-established ties of acquaintances can be investigated by using virtual social connections. Email [289], “blog” services [283], LiveJournal service [284], massive multiplayer online game [290] are just a few examples of how social acquaintances can be inferred through virtual connections. The popularity of these online social networks has increased over this last decade (see ref. [291] for a historical perspective<sup>7</sup>). While most web sites simply disappear during the years, Cyworld is an example of successful online social environment, providing a network of more than 12 million nodes [292]. Another case of success is the Wealink service, whose popularity is characterized by a logistic growth in the number of nodes and edges [293]. The establishment of connections on this network is believed to be motivated by real-life friendships on the early stages since the assortativity values were equivalent to other collaboration networks. However, over time the network presented a transition to a disassortative regime, which is more characteristic of online social networks [293,294]. The origin of (dis)assortativity on different online social networks is still unknown. Communities from different countries or focused on distinct audiences may present similar characteristics [293,294]. Han and Kim [295] used two networks extracted from the Cyworld service to analyze the user’s concurrent behavior on two activities, namely writing bulletins and purchasing. They found that users who are highly active writers usually do not buy much, i.e. considering that individuals spend the same amount of time online, they clearly have different interests. An important consequence is the effect on the neighbors, since users topologically close to active writers are identified to also write more and buy little [295]. The influential or affinity connection mechanism was observed in an online community oriented to political preferences and also resulted in a broad degree distribution. This distribution

is suggested to depend on the time of participation of the user and how active he or she is, while active [296].

### 6.1.2. *Sexual relations*

Even though sexual relations between individuals do not necessarily correspond to acquaintance relations, sexual partners can become social partners and *vice versa*. Data collection in this subject is severely restricted mainly because of personal privacy. Large-scale information about the internal structure of sexual networks is still unavailable and needs more research. Nevertheless, several works have been done on this field. Bearman *et al.* [297] used questionnaires to collect information about dating relations between teenager students in the USA, and found that the network has one giant component mostly composed of a large ring with several smaller branches. An important consequence of this social structure is that the removal of one node could easily break down the giant component into smaller ones and reduce the spreading of diseases. Other networks were constructed using different data sets [298–300] which described populations of different countries and specific regions. González *et al.* [298] proposed a model to capture structural features of homosexual and heterosexual networks. The model is based on a group of mobile agents that collide and generate links between them. The authors found that the clustering coefficient, the number of triangles and squares are relatively small, which can be better explained by their model than by the scale-free model. Analyzing different periods of time, Blasio *et al.* [299] found evidence of non-random, sublinear preferential attachment for the growth in the number of partners for heterosexual Norwegians. It is worth noting that sexual behavior is slightly different for men and women. For instance, men tend to have more sexual partners than women on average [273,300] (except when compared to female sex workers [301]). In a specific region [300], apparently the number of sex contacts decreases with the number of partners for women while being kept constant for men. At the same time that individuals with many partners can potentially spread a disease to more people, the chance to transmit to a specific partner is lower because of less intense sexual activity.

Much effort has been devoted to accurately determine the functional form of the number of partners distribution. Analyzing a random sample of individuals aged between 18 and 74 years old in Sweden, Liljeros *et al.* [273] found a cumulative power law distribution of number of partners during a period of 12 months. 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. The scale-free structure of this 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) may be arbitrary in such networks since there is no well-defined boundary between core groups and other individuals. Therefore, safe-sex campaigns should be focused on highly connected individuals to prevent propagation of sexually transmitted diseases. These results were challenged by Jones *et al.* [302,303]. They argued that the fitting method used by Liljeros and collaborators has statistical problems. By employing other statistical methods in the same and different data, they concluded that the preferential attachment process may not be the only cause generating real sexual networks as proposed. They stressed 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 also be 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 [304], Liljeros explores the complexity of the mechanisms contributing to the spreading of sexually transmitted infections



and concludes that one single solution for the problem is far from being found. In real life, broad and targeted interventions have both been proven to be effective (see more about epidemics in Section 7).

Fichtenberg *et al.* [305] analyzed three important positions of people inside the network, i.e. if someone is in a (i) monogamous pair, (ii) connected to a larger network through a partner or (iii) in the center of a large network. These network positions were previously found to be correlated with chlamydia and gonorrhea infections. The authors found that the sexual network position is strongly associated with poverty, suggesting that the latter shapes network structure and therefore influences the rate of sexually transmitted infections. By using data from a forum-like web community related to commercial sex, Rocha *et al.* [301] constructed a time-ordered network of claimed sexual partners with 16,730 nodes. Among other findings (such as high clustering inside cities and sparse connectivity between cities), the authors reported that the preferential network growth is sublinear and the choice of new partners depends on a feedback from previous users, since positive reports attract new partners.

### 6.1.3. Trust

From the hierarchical point of view, the network of trust can be viewed as a special sub-network of the acquaintances network. The lack of reliable data about trustful partnerships motivated the studies to concentrate on electronic trust ties. For instance, the Pretty Good Privacy (PGP) encryption algorithm allows one user to certify another by signing his/her public encryption key; in other words, the first user creates a directional connection to the second one if he/she trusts that the second user is really whom he/she says. The evolutionary Prisoner's dilemma has been investigated in a substrate of this network and also in an email network (network of people who exchanged electronic messages). It was found that the connectivity between communities and their internal structure affect the evolution of cooperation, by raising or lowering it [306]. Online auction websites heavily depend on the user feedback to rate the trustworthiness of its traders. Nevertheless, only user feedbacks are not sufficient, since fraudulent attempts to increase rates have been witnessed. It has been shown that these problems can be diminished by examining the network built from online auction transactions using, for instance, information about  $k$ -cores [307]. Online trust networks were also studied by Yuan *et al.* [308]. The authors verified that five trust networks have the small-world effect and used this property in a trust-aware recommender system. Their model uses the average path length of the network as the maximum trust propagation distance for user recommendation.

### 6.1.4. Movie actors

Another important class of social acquaintances is related to professional ties between movie actors, where actors are connected if they have been cast in the same movie. In the artistic scenario, such a network can provide a glimpse on the career popularity of an actor as well as individual fame. Despite the vast selection of movies in the Internet movie database<sup>8</sup>, the film actors network presents the small-world effect [38]. This means that there is a well-defined tendency of actors to play with common partners (high clustering coefficient). 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. The popularity of some actors and the short life-time of the majority could explain the power law regime observed for large  $k$  in the distribution of actors partners [245]. Amaral *et al.* [309] pointed out that this distribution 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.* [310] who argue that when considering multiple

edges between actors, a stretched exponential distribution fits reasonably the data. The temporal evolution of the weighted actor network along 117 years was taken into consideration by Eom *et al.* [311], where they measured the growth rates of (i) the vertex degree and strength, (ii) the edge weights and (iii) the probability of connecting previously unconnected actors. The authors devised a nonlinear model to reproduce the real data that incorporate growth, addition of new internal edges and reinforcement of old edges through increasing weight.

#### 6.1.5. *Sports*

Systems related to sports are particularly interesting because of the intrinsic dynamical nature and the data available for diverse activities and cultural backgrounds. Teams constantly change players, there are dozens of collective sports involving interaction between people, championships range from local to world wide scale and so on. Nevertheless, relatively little attention has been given to such dynamical systems in terms of network theory.

The North American college football league provided motivation for a study involving a ranking system based on a directed complex network where the direction represented wins or losses of a specific team (the nodes) [312]. The method was based on the idea that if team A can beat team B and B can beat team C, then A most probably will beat C. The resulting ranking for a season was compared with the official results with good agreement.

#### 6.1.6. *Scientific collaboration*

The topic of scientific collaboration is linked to professional (academic) ties between scientists and has special relevance with regard to knowledge dissemination. Several areas have been analyzed, including the scientific collaboration in sociology [313], mathematics [314], medicine [24,315], neuroscience [314], statistical physics [316], econophysics [317], high-energy theory [24], astrophysics [24] and arXiv [316], as well as complex networks theory [318].

The relation between co-authorship and number of publications was addressed and its importance for science evolution was investigated [319]. Purely theoretical papers appear to be the work of two scientists on average, according to results for the fields of mathematics and sociology. The latter were also found to contain a small number of collaborations although with an apparent increase in the collaboration rate in the last decades [313,314]. On the other hand, experimental and interdisciplinary subjects have higher average collaboration rates, reaching about nine authors per paper in high-energy physics. An intermediate collaboration rate, about 4, was identified in topic-specific fields (e.g. “networks” and “granular media”) [320]. Cotta *et al.* [321] pointed out that the most prolific authors are not necessarily the most connected, but they have diverse interests which motivate collaborations. The number of collaborators is observed to strongly correlate with the number of papers [321].

Barabási *et al.* [314] observed that the growth of the giant component of the scientific collaboration networks in neuroscience and mathematics are consistent with the increasing collaboration rate over the last years. Palla *et al.* [322] showed that small groups of collaboration or friendship circles and institutions have different evolution patterns. While the communities around small groups with strong relationships are stable, the large ones present continuous changes where almost all their members are replaced by new ones after some time. One example of such changes happens in schools where most students could change after some years.

Newman observed 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 [24,321,323]. 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 or specific knowledge. On the other hand, betweenness is a measure of author's control over information flowing between other scientists [24,323].

One main problem in the analysis of networks is related to the finite size of the data sets available. Barabási *et al.* suggested that for the evolution of the collaboration network, the small and finite time interval available 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 [314]. Another type of scientific reports (e.g. proceedings editorship), though important, cannot be seen in the same level of peer-reviewed papers. Cotta *et al.* argued that proceedings collaborations should be removed from such analysis since they bias considerably the shortest path distribution, as they include authors from different thematic subjects and create long-distant edges. They also suggested to analyze separately the distinct types of scientific reports since different types of collaboration are reflected in such reports [321], which does not agree entirely with the weighting functions suggested above by Li *et al.* [317,324]. Some care should also be taken when analyzing 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 [315] and obtained qualitatively the same results.

## 6.2. Citations

Although dynamically related to 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. Indeed, there is a tendency to remember recent papers in detriment to older ones [325,326]. Therefore, 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 [327].

Citation networks usually derive from a set of technical texts (more commonly, scientific papers) where a node represents a single text and directed edges are created whenever a text contains a reference to another text. 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 an apparent universal behavior at the two rightmost classes, corresponding to journals of limited and unlimited number of pages per paper [328]. In contrast, patent-to-patent citations present a power law [329].

The unidirectional growth mechanism has important consequences and properties: though the average number of citations increased over time [328,330], the average number of citations for papers published in a given year has decreased slowly with time [330]. This suggests that the relevance of a paper might decrease over time. Analysing data from arXiv and Physical Review Letters, Hajra and Sen [331] concluded that most papers were cited within a period of 10 years after publication. It also indicates that trendy research topics were popular during the same period and afterwards became less important [331]. It was claimed that some citation networks follow a preferential attachment mechanism constrained by a decreasing exponential dependence on the age of each paper [326]. Redner [330] 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. By means of a Zipf plot (cumulative distribution), he found a power law citation distribution. Other power law distributions were verified in the SPIRES database by Bilke and Peterson [332]. Such distributions led Redner to suggest that minimally and highly cited papers obey different statistics where most papers are forgotten a short time after publication, as also stated by Wang *et al.* [326]. 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 [330]. On the other hand, Tsallis and Albuquerque [333] 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. While analyzing only citations of papers in US patents, Chen and Hicks [329] concluded that papers with explicit funding acknowledgements tended to be cited relatively more often than papers without explicit acknowledgements, 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 [332] found that the spectral dimension of the ISI citation network is about 3.0, with a tree-like structure and small-world feature. In general, all subjects possess broad citation distributions but different exponents, which makes it harder to compare across disciplines. Radicchi *et al.* [334] identified a universal pattern for the citation distribution on different disciplines (and different years) obtained after a rescaling by the average number of citations (degree).

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. This property is obtained by the neighborhood of a specific node, but it could also be obtained from a network such as that of van Raan who used databases from several citation indexes in 2001. He considered that two publications were connected if they had at least one reference in common [335]. 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 was also described by a power law with exponential cut-off functional form for values larger than 1,000. Analogous to other works, a memory effect was identified in the age of the references, and the distribution function changed from a power law to an exponential if the network was built only with older references. 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 self-citation issue was addressed by Hellsten *et al.* [336,337] by adopting the optimal percolation method [338,339]. They proposed a methodology over the self-citation evolution patterns of a specific author and compared this pattern to the co-authors and keywords in the articles. The method was used to detect emerging research fields and to trace mobility of scientists through different fields and critical moments in the academic career. For one author, changing co-authorship drives the changing research interests and move to new research topics [336] while for another author, this inter-field movement is an effect of maintenance of the same collaborators [337].

The databases reviewed in this section are considerably large and reliable in terms of an interaction rule between their components. However, some inaccuracies are present since little standardization is adopted and individual entries are made by different people at different times. Redner [330] 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 on the citation distribution.

### 6.3. Music

Professional music relations could be included in the section of *personal relations*, but we have singled out the music subject because different types of networks inspired on music have been studied. In terms of professional relations, Gleiser and Danon [340] proposed a network where two musicians were connected if they played in the same jazz band. For the information collected dating back to the 1930s and 1940s, 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 [340]. The network of collaboration between rappers was constructed by Smith [341] 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.

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. Different from the jazz music network, the community analysis used a weighted network and then a “clearing algorithm” to convert the weighted network to a non-weighted version such that the most important edges were extracted [341]. Considering only the connections, the fast modularity community structure algorithm [342] was applied and only small and peripheral rap groups could be identified. When the clique percolation method was applied [54], 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 before 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.* [343] who used data extracted from the *allmusic.com* database. They constructed a network of similarity between artists in this database and investigated the topological properties of both networks, in addition to the resulting networks obtained when intersecting them, i.e. the networks obtained from only common nodes in both networks. The method is useful for comparison of the same data set for different social ties. However, the overall features did not change considerably, and the collaboration network before and after the intersection process presented a power law, while the similarity network had exponential degree distribution. The small-world and high clustering features were maintained after the intersection process. 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 one was partly assortative. The cumulative fraction of betweenness displayed a power law in all networks studied.

The similarity between artists was also extensively investigated by Cano *et al.* [274] 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 explained earlier, although still showing the small-world effect. They found that user rating networks resulted in power law in-degree distributions, while experts classification mechanisms resulted in networks with exponential decay. The out-degree distributions follow the same functional form with cut-offs due to clear limitations in the webpages usability constraint (the recommendations should fit in one webpage). The power law behavior for the degree distribution was confirmed in an experiment where users sent play-lists and 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 [274].

Lambiotte and Ausloos [338,339] analyzed a website dedicated to sharing musical habits using a percolation-based method to identify social groups and music genres according to personal habits. Non-trivial connections between the listeners and music groups were identified, with some empirical subdivisions obeying standard genre classification, while others did not. By analysing the original bipartite network, a power law was observed for the distribution in the number of listeners of a specific music group, while an exponential was fitted of the case of the number of music groups per user [338]. They further improved the methodology mapping music groups into genres using online listeners records. Analyzing different tags given by listeners to classify music groups, they observed that similar groups tend to be listened by the same people [339].

#### 6.4. *Terrorism networks*

The use of graph theory to visualize the organization of criminal or secrete groups has some decades of tradition [344–346], but it became more popular after the terrorist attacks in 2001. In this year, Krebs [347,348] built a network where the nodes were the terrorists related directly or indirectly with the hijackers of September 11 attacks, while the links were the knowledge interplay among the hijackers. The data were collected from public released information in major newspapers<sup>9</sup>. According to the author, using network theory to prevent criminal activity is difficult, but it is an important tool for prosecution purposes. Latora and Marchiori [233] also studied the network from September 11 attacks and found that the most critical node had the largest number of direct connections with other terrorists. However, the second most critical node had degree of only half the maximum. This shows that even nodes with low connectivity can play a crucial role for the network efficiency. Latora and Marchiori noted that the importance of a particular node was given in terms of the change in network efficiency when this node is removed. Thus, the deletion of an important node leads to a large decrease in network efficiency. The efficiency criterion used was the average shortest path of the network. The importance of nodes in terrorism networks was also investigated by Muhammad and Hussain [349]. An interesting finding was reported by Maeno and Ohsawa [350,351], where they proposed a method to solve the node discovery problem in complex networks. Although some previous knowledge is required, this method can be applied in terrorism networks to identify an unobserved agent behind the perpetrators of terrorist attacks. Upon using public databases<sup>10</sup> and publicly released information, Will *et al.* [352] built a terrorism network related to David Coleman Headley, who is accused of planned attacks in Denmark and India. This network has a crucial difference from the network of the September 11 attacks, namely that the hijackers are very socialized and have many connections with people outside the network, while in the case of the September 11 attacks the terrorists rarely contacted the world outside the network. The authors showed that, due to this difference, the importance of the nodes is strongly related to classical centrality features, such as degree, betweenness and closeness.

#### 6.5. *Communication*

The study of the structure and function of social networks is usually constrained by practical difficulties of mapping the interactions of a large number of individuals. The construction of these networks is traditionally based on questionnaire data, which reaches only a few number of individuals and depends on the personal opinion about their ties (see the beginning of Section 6). With the advent of the Internet and the use of fixed and mobile phones, a large amount of data could be recorded for further analysis. These communication networks are useful not only to provide high-quality data of social systems, but also for economic reasons. In the following sections, we report on works addressing the electronic mail, call graph and wireless networks.

### 6.5.1. Electronic mail

Electronic mail (email) has become one of the most important modern means of communication, being largely used in business, social, technical and scientific relationships. Email exchanges therefore provide plenty of data on personal communication in an electronic form, amenable to build social networks automatically. Email networks are also one of the major means of computer virus spreading. Another important related aspect concerns email topic classification [353].

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  $i$  sends at least one email to  $j$  (this network is obtained by the log files of email servers) [354–356]; (ii) vertices are also email addresses but there is a directed arc from vertex  $i$  to vertex  $j$  whenever  $j$  is in the address book of vertex  $i$  (this network is obtained from the email book of several users of a specific institution) [289,356]. Ebel *et al.* [354], however, studied these networks by considering undirected arcs where email addresses are connected if at least an email was exchanged between the users.

Gimerà *et al.* [282] and Arenas *et al.* [316] found that email networks are composed of communities and suggested that they are self-similar, which 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). It is worth noting that not only the communities themselves are important, but also the central nodes participating on parallel communities [357]. Braha and Bar-Yam [358] 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 De Wilde [356] and is based on addition and deletion of links between users, and on the user email checking time. Zou and Gong [359] proposed a model for simulating the email spreading of virus 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 in the first kind of network than in the others.

Much interest has been recently devoted also to study the related subject of interval between two consecutive events. In case of sending or receiving emails, a cascading non-homogeneous Poisson process was proposed to explain the power law distribution of inter-event time on email communication [360,361]. It is suggested that circadian and weekly cycles together with cascading activity are the ingredients needed to reproduce the heavy tails of inter-event time identified in email communication networks.

Another issue related to email networks is to provide tools for filtering spam. Kong *et al.* [362] employed a network of email contacts to create a spam filtering technique, which 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. 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.

### 6.5.2. Telephone

Telephone call graphs are obtained 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 established if, during the specified period, there was a call from  $i$  to  $j$ ; the arc has naturally a direction, with  $i$  at the tail and  $j$  at the head. Such graphs can be very large; for example, a one-day call graph used by Abello *et al.* [363] has more than 50 million vertices and 170 million edges. Nanavati *et al.* [364] did a similar study for mobile phones and found a significant correlation between the in- and out-degree of vertices, meaning that people that receive many calls also generate many

calls. With respect to nodes connected by a call, assortativity was present for the in-degree of the caller with in- and out-degree of the receiver; for the out-degree of the caller, weak disassortativity with in- and out-degree of the receiver was detected. That is to say, vertices that receive many calls tend to call vertices that also receive and generate many calls, and vertices that receive few calls tend to be connected 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 receiver.

Using records of mobile phone calls within a period of 18 months, involving approximately 20% of a country's population, Onnela *et al.* [365,366] constructed a network with  $4.6 \times 10^6$  nodes and  $7 \times 10^6$  edges. The nodes represent users, and a connection is established whenever a reciprocal call is established during the studied period, i.e. an undirected edge is present between nodes  $i$  and  $j$  only if  $i$  called  $j$ , and  $j$  called  $i$  during the time interval analyzed. The weights represent the duration of calls between two users. The number of calls between two telephone numbers and the total duration of the calls were used as two weighting criteria. The distribution of strengths showed that most users make few, brief calls, but some make a large number of calls, with some pairs of users chatting for hours [365]. The authors also observed that the overlap of the neighbors of two vertices was proportional to the weight of their connection, therefore corroborating experimentally the "weak ties hypothesis" of Granovetter [367]. The removal of the strong ties had basically local effects, within communities. In contrast, when weak connections were removed, communication between different communities could be disrupted, causing collapse of the whole network. Interestingly, information diffusion was tested in such weighted network using an equivalent SI (susceptible infected) epidemiological model. They found that the majority of the nodes were first infected through ties of intermediate strength with a peak at 100 s [366], which means that most of the nodes were infected after 100 s.

Lambiotte *et al.* [368] analyzed the call graph of a Belgian mobile phone company with 2.5 million customers and 810 million calls or text messages over a period of 6 months. In order to eliminate accidental calls, a link between customers  $i$  and  $j$  was included only if there were at least six reciprocal calls between them during the time period analyzed, which is a stronger statement of tie. This network had power law degree distribution with the probability of two customers being connected proportional to  $d^{-2}$ , where  $d$  is the geographic distance between them. Blondel *et al.* [369] concluded that the same network also exhibited hierarchical communities.

To understand telephone traffic, a model which represents the customer network behavior in real world has to be taken into account. Unlike classic traffic analysis where a fully connected customer network is considered, Xia *et al.* [370] employed a model for the customer network based on the scale-free property and showed that the structure of the customer network is more likely to cause call blocking than the limited capacity of the telephone network.

The data available for mobile phone companies can also be used to analyze the pattern behavior of the customers [371] and to infer the correspondent friendship network [372]. González *et al.* [371] used the trajectory of 100,000 mobile phone users over a period of 6 months to understand their mobility pattern behavior and found that, despite the previous prediction of random trajectories [373], the human trajectories have some regularities, including small traveling length and a high probability of returning to few locations. These patterns are important for disease prevention, urban planning and social modeling. Eagle *et al.* [372] employed data collected by mobile phones to obtain insights into the social dynamics of individuals and compared data observed (physical proximity and calling patterns) with self-report survey data. They showed that only with data observed it is possible to infer 95% of friendships and to predict individual job satisfaction.

Even though the network of mobile phones is huge with million vertices well-connected, Wang *et al.* [374] demonstrated that viruses do not represent a real threat. For their spreading rate is



slow when Bluetooth connections are considered but alerted to the phase transition which limits the number of infected users to a small fraction in the multimedia messaging service. If the phase transition point is reached, viruses will be a real threat to mobile networks.

In conclusion, communication networks, besides constituting an important means to represent social networks, also deserve special attention for economical reasons, for they include everybody. These networks are generally huge, with millions of vertices. The main results obtained for these networks are: (i) a definition of a hub has to be reinterpreted in email networks, because the vertex degree and betweenness change dramatically from day to day; (ii) in the case of call graphs, the overlap of two neighbors is proportional to the weight of their connection; (iii) the probability of two vertices being connected in call graphs is inversely proportional to the square of their geographic distance; (iv) the mobility behavior of customers in mobile call graphs is not random; instead, it exhibits some regularities.

## 7. Epidemic spreading

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 [375] 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 [376] 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 ( $\Delta t = 1$  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 epidemic spreading in small-world networks was also explored in [377], 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 [378] 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. Small *et al.* [379] showed that only the introduction of a small-world topology in the epidemiological model could explain the random fluctuation of the real data. More recently, an epidemiological model was studied with the shortcut networks [380], which are an alternative to the small-world network in which the random edges are added to nodes at fixed distance  $m$ . Li and Wang [381] studied the susceptible-infectious-recovered (SIR) dynamics on small-world networks with a delay-time recovery. As expected, the results indicated that actions to recover the network should be executed as soon as possible to avoid spreading of the infection.

Pastor-Satorras and Vespignani [382] 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 [383, 384]. This result was modified in subsequent papers [385, 386], where the authors used analytical arguments to show that in the 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, including bipartite nodes and non-uniform transmission rate, was obtained by Newman [387]. The effects imposed by the finite size of the networks to epidemic threshold was studied by Pastor-Satorras and Vespignani in [388].

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 [382,389–391]. This is the case of the sexual contact network of Sweden, as shown by Liljeros *et al.* [390]. The solution to eradicate spreading of virus in these classes of network was proposed in [392–394], 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.* [395] 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. Gómez-Gardeñes *et al.* [396] studied various immunization strategies for Internet maps at the AS and router levels. 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. Immunization on geographical networks was investigated in [397], where it was found that the recovery depends on the action radius (local region immunization). The results pointed to a critical radius above which the epidemic spreading vanished.

Spreading on growing networks was investigated by Hayashi *et al.* [398] who 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 where the exponent of the degree distribution was 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. The spreading over modular networks was also explored. He *et al.* [399] showed how the infection propagates in the networks according to the inter- and extra-community infection probabilities. The extra-community infection probability was found to play a crucial role in the spreading.

Real data are frequently used to validate the spreading simulations on complex networks [400,401]. For example, data from the *Mycoplasma pneumoniae* infection were used in [400] to build a mathematical model and to investigate the spread and control of this disease in closed communities. Other approaches, including percolation theory, have been used to study epidemic spreading with complex networks [402,403]. Small *et al.* [379] compared the data from the severe acute respiratory syndrome (SARS) cases in Hong Kong in 2002 with various models. The Hong Kong SARS data were also used in [404] where a 4-state model was used to simulate the disease transmission under a small-world topology. The model included as agents those who were infected but not yet infectious, and as the main result the authors found that outbreaks could be prevented if the patients with symptoms were isolated as soon as possible. Another model applied to SARS spreading of virus in Hong Kong can be found in [405]. In 2007, Small *et al.* [406] studied the distribution of avian influenza virus among wild and domestic birds and obtained a network with scale-free topology with no epidemic threshold. Therefore, local methods could not be used to eradicate the disease, thus pointing to attacks to the hubs as a possible strategy.

Recent studies in the field of epidemic spreading have changed their focus from the analysis of the SIR and SIS models to metapopulation approaches, where each node of the network represents

a set of individuals. This set can be related to spatial regions, such as cities, states or countries. Moreover, individuals are labeled according to the usual classes (S, I or R) and can move through the network links to adjacent nodes. In these approaches, it is important to know details about the environment, transportation infrastructures, movement patterns, and traffic networks [407,408]. An interesting work in this direction was reported by Colizza *et al.* [409], in which they used a metapopulation model to investigate the worldwide spread of the H5N1 avian influenza virus. The existence of commercial flights around the globe was found as the main factor responsible for the emergence of a global outbreak. In [410], the same approach was used to explain the global spreading of the SARS, which started in Hong Kong in 2003. They considered the data of the worldwide commercial airline traffic from the International Air Transport Association (IATA), associated with urban area census information<sup>11</sup>. Their results show that the metapopulation approach is powerful to predict global outbreaks with high precision. The effects of short-range connections on the global disease spreading have also been explored. Instead of considering only airlines connections, an additional network was introduced to represent local interplay between nodes, such as roads and highways [411]. This network was built from a Voronoi tessellation of the airports nodes. Although the flow in small-scale connection is an order of magnitude larger than the airline flows, the global behavior of the outbreaks was affected only slightly. However, far from the airports the changes on the spreading behavior can be drastic. This approach was used in [412,413] to simulate the spreading of the influenza virus H1N1 and H3N2, respectively. The results have high precision when compared with the data reported by health agencies, which demonstrate the predicting power of the method. In [413], a computational solution was reported for the disease spreading via the metapopulation approach, referred to as GLEaM (Global Epidemic and Mobility<sup>12</sup>). This software implements the concepts involved in the metapopulation simulations, where different diseases and immunization strategies can be considered.

## 8. Economy

Trade, currency, industrial production, wealth distribution and tourism are important subjects for studies in economy, whose features may be represented by systems containing discrete parts that interact in a defined way [414]. For instance, trade/commerce is the voluntary exchange of goods, services or both [415]. In the following subsections, we shall discuss cases where networks have been used to describe phenomena in economy.

### 8.1. Trade networks

Economic relations are currently growing owing to the ever 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 economic complex system of trades between countries, Serrano and Boguñá [416] studied the so-called world trade web. A directed network was built by assigning a node to each country, with two nodes being connected if there are relations of importation or exportation between countries.<sup>13</sup> 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 distributions 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 economic wealth was identified, with some exceptions for developing countries. 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. Moreover, Li *et al.* [417] studied the price dynamics over different network structures. The results indicate that **networks with the small-world effect are more sensitive to variations of prices**. The evolution of the world trade web was analyzed in [418], where globalization was shown to have an effect of concentrating trade along the years, although new trade centers tend to appear. The evolution of the world trade web in a more recent stage was studied in [419]. Statistical properties of the nodes tended to be stable along the years.

Ausloos and Lambiotte [420] studied the **correlation between the G7 country** gross domestic product (GDP) evolution of the 23 most developed countries. The authors built a network from a **correlation matrix** and found subgraphs with high cluster coefficient as indicative of globalization effects. Miśkiewicz and Ausloos [421] studied the **globalization process**, while Gligor and Ausloos [422] investigated the structure of **macroeconomics of the European Union** (EU) countries using network concepts, with the correlation matrix of the GDP per capita annual rates of growth between 1990 and 2005 being used to establish the weight of the connections among the 25 countries (nodes) in the EU. They introduced the overlapping index,  $O_{ij}$ , to find a hierarchy of countries in the network analyzed. **In non-weighted networks**, this quantity is proportional to the number of neighbors **shared by  $i$  and  $j$** . The extension of the overlapping index for a weighted network was used to identify clusters and the hierarchy of EU countries.

Industries and firms of the electricity sector have been studied with the complex network approach, in a model to find a stable network structure and predict possible bilateral transactions [423]. The model was applied to real networks, such as the 4-bus test system, the IEEE 30-bus and the IEEE 118-bus test systems.

## 8.2. Currency

In 2004, Li *et al.* [424] 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 has also been observed by Górski *et al.* [425] in a complex network built from the **FOREX database (the largest financial market in the world)**. The **robustness of currency complex** networks was investigated by Naylor *et al.* [426], who used **hierarchical methods** to show that the **currency network topology minimizes the effects caused by economic crises**, such as the Asian crisis from August 1997 to October 1998.

## 8.3. Industry

Andrade *et al.* [427] 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 and argued that the **network topology may provide a useful tool** to **design, characterize** and **evaluate refinery plants**. Ferrary and Granovetter [428] studied industries in the Silicon Valley, with the edges being established from the financial relations **between the industries**. The authors investigated how individual firms interacted with one another and **how they supported the robustness of the system**.

While applying complex networks concepts in chemistry, Jiang *et al.* [429] showed how an ammonia plant can be mapped in a complex network with small-world properties. The ammonia plant network is weakly self-similar and possesses a modular structure, with each community representing a modular section in chemical plants. Another interesting result is that the ammonia complex network exhibits excellent allometric scaling, guaranteeing a better fluid flow.

#### 8.4. Wealth

The effects of network topology on the evolution of a dynamic process have been widely studied. For example, Souma *et al.* [430] built a model to investigate the wealth distribution with networks displaying the 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, *viz.* the probability of rewiring edges and the average degree. 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 the 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.* [431] 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.

Game theory was also used to model the wealth distribution in networks with different structures. Hu *et al.* [432] showed that the wealth distribution in scale-free networks follows a power law, in agreement with real observations, while the distribution has an exponential decay for random ER networks. According to the authors, the social networks can be organized in groups for high and low income. The results indicate that the high-income group has a heterogeneous scale-free structure, while the low-income group has a random ER topology.

#### 8.5. Tourism

Tourism comprises a multitude of activities that form one of the world's fastest growing industries. The sector is so strongly characterized by intense linkages through which informational exchanges occur that it seems natural to apply network methods to its study [433,434]. Nonetheless, only recently have quantitative techniques been applied, mainly in analyzing tourism destinations, complex localized clusters of public and private companies and organizations [435]. For instance, Scott *et al.* [436] analyzed tourism in Australia and showed that complex network theory can help define the weaknesses in destination structures, which can be improved by policy and management approaches. The underlying social and economic system has also been assessed by using network dynamical growth models [437]. Moreover, networks were used to perform numerical simulations so as 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 to increase the degree of local clustering [438].

Tourism networks have also been addressed by analyzing the relation between their structure and dynamics. Costa and Baggio [439] studied the Elba (Italy) tourism destination network using topological measurements. They considered dynamical processes over the network, *i.e.* the inward/outward activations and accessibilities, according to the superedges framework [37]. The type and size of companies were found to affect the activations and accessibilities. On the other hand, the geographical position of companies does not tend to affect dynamical features. With the characterization of the tourism networks structure, the authors concluded that the Elba tourism network is fragmented and heterogeneous.

### 8.6. Financial market

A corporation, by definition, has a group of owners, called shareholders, who share their 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 economic instabilities such as that of Black Monday (October 19, 1987, when the Dow Jones Industrial Average (DJIA) decreased sharply). The organization of the financial market may be well understood by representing it as a network. Given the high accuracy of financial data available, it is possible to build networks that reliably reflect the real market, in contrast to other fields that generally lack high-quality data, such as the Internet. Financial networks can be constructed, for example, from stock prices or stock ownerships. In the first case, each pair of stocks is connected to each other by a weighted edge that encodes the distance between them. This distance can be computed as a function of the correlation coefficient, taken between the time series of stock prices [42]. From this complete network (i.e. having all possible edges), a hierarchical structure is obtained, which is usually a minimum spanning tree, also called in this context “asset tree”. A minimum spanning tree is a connected subnetwork with no cycles, which includes every node of the original network with the lowest cost (minimum sum of the edge weights). Examples of analyses of these and similar types of networks are discussed in the following paragraphs.

Mantegna [42] was possibly one of the first to study financial markets in the context of networks. The author collected the daily time series of stock prices traded in the New York Stock Exchange (NYSE) from 1989 to 1995 – only stocks used to calculate the DJIA and the Standard and Poor’s 500 (S&P 500) indexes were employed. 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  $\rho_{ij}$ , taken between two daily time series of stock prices:  $d(i, j) = \sqrt{2(1 - \rho_{ij})}$ . Besides laying the foundation for a series of later financial market studies, Mantegna was able to detect a taxonomy that organizes stocks according to their economic activity in terms of the minimum spanning tree. Bonanno *et al.* [440] employed the same methodology to study stock prices recorded from 1987 to 1998 in the NYSE. The authors compared the topology of the minimum spanning tree resulting from real data with the ones obtained from simulated data using market models, such as the one-factor model. By observing the networks obtained from the artificial markets, the authors pointed out some deficiencies of the models, which failed to deliver the hierarchical structure observed in real markets. Also during crash periods, the exponent of the power law degree distribution changed, as on the Black Monday [441]. The asset tree underwent other particular changes in periods of crisis, such as the decreasing in its length.

Onnela *et al.* [442] investigated another network obtained from correlations of stock prices. In this case, the network is not necessarily a tree, being obtained by selecting only the  $N - 1$  edges with lower weights of the complete network (note that a minimum spanning tree also has  $N - 1$  edges). This network is generally called “asset graph”. Results indicated 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 under extreme conditions. However, the asset graph does not necessarily have a hierarchical organization, thus preventing a taxonomical analysis of the financial market. The construction of the asset graph was analyzed in [443], where the temporal evolution of the clustering coefficient was measured. The authors were interested in finding what portion of the edges in the asset graph represents 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 conveyed real information. The Korean, Japanese, Canadian, Italian and British stock markets were analyzed by Eom *et al.* [444] using minimum spanning trees. The authors discovered that common factors defined in the field of finance (i.e. factors that usually affect stocks, such as industrial, corporate and macroeconomic factors) can be used to explain highly connected nodes in the spanning trees.

Thresholding of stock correlations was used to construct networks from financial data. Pan and Sinha [445] investigated the stock price fluctuations in the National Stock Exchange (NSE) of India, and found that: (i) **stocks in emerging markets are more correlated than in developed ones** and (ii) **the Indian market evolved into clusters corresponding to business sectors**. The clustering property has also been investigated in the NYSE [446–449]. Bolgorian and Raei [450] considered thresholding in correlations between daily stock prices to build a network. They used the trading volume of the stocks listed **in the Tehran Stock Exchange between 2000 and 2009, and separated correlations for individual and institutional investors**. The authors defined a homogeneity index to assess the ratio between these two types of investment in terms of their number of links in each month. The dynamics of investment trends over time reveals that, in general, individual investors are more influenced by the stock fluctuations than institutional ones. A pronounced feature is that institutional investors are more conservative in booming periods, which is possibly related to their long-term investment approach. Another dynamics analysis, this time of link removal, was investigated by Garas *et al.* [451], where the authors employed a fully connected network with weights corresponding to the correlation strengths between NYSE stock prices. By sequentially removing either the weakest or the strongest links they found, in addition to clearly defined communities, an important structural role played by the weakest links to maintain the network connected.

**Financial markets** have a **distinctive feature** when considering the broad range of fields covered by this review: **available financial data are accurate**, being collected along several years. This means that researchers in this field can perform **highly reliable analysis of the evolution of financial markets**, thus helping to understand its underlying mechanisms and also allowing the creation of **new market models** (some network-based models have been proposed [452–454], although without direct validation with real-world data). The main approach to **create stock networks** is the correlation between time series of stock prices. This approach helps identifying the most correlated stocks, although a special care is needed to avoid **spurious correlations**. Nevertheless, research on financial markets employing a networked representation is still incipient, if we compare with more traditional subjects in complex networks, such as biological networks. In the context of networks research, **a great challenge for financial studies** is to **create models of market evolution**, thus helping the **prediction of important market** changes that could affect the **economy of a country or even of many countries**.

## 9. 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 [455,456]. While its more fundamental topics involve computer architecture, data storage and processing, and system control (programming and algorithm developing) [457], it has become essential for developments in a variety of areas. It also provides tools to help solving problems of physics, mathematics, chemistry, biology, medicine, economy, among others. For instance, major problems in biology, e.g. protein folding, function prediction, phylogeny and modeling of biological systems (see Section 3.1) typically can only be solved with the help of computers and optimized algorithms.



### 9.1. Software architecture

The importance of computers and software to our lives is undeniable. The cost of software development increases with its complexity and may exceed a million dollars (e.g. computer games, compilers and operating systems). The subarea 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 [458,459]. 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 and the hierarchical relationships among the software components have been studied by Valverde *et al.* [458,460], where the nodes are program classes and the edges are the inheritance and composition among the classes. Networks for various pieces of software were found to exhibit the same organization pattern, including eMule, Openvrm, GTK, VTK and the Linux Kernel. Similar results were reported by Zhang *et al.* [461] for a different set of software packages. All these networks display scale-free topology and small-world properties, which have also been observed for networks built with inter-package dependency in open software [462]. Open source software was studied by Myers [463], from which it was inferred a small-world, scale-free distribution for the network of six pieces of software, viz. VTK, CVS, abiWord, the Linux Kernel, MySQL and XMMS. Another model was proposed based on the local optimization process, which reproduced the features observed in the real networks. Other methodologies to create software networks can be found in [459].

Challet and Lombardoni [464] studied software networks of Linux packages and found clear evidence of the asymmetry between the distribution of incoming and outgoing links, in which only the first presents a scale-free behavior with exponent  $\gamma = 2.0$ . Bug propagation in these networks was analyzed with a model similar the SIR epidemic spreading model [27]. When bugs propagate only to the immediate neighbors of a faulty node, the problem can be easily fixed by a debug process. On the other hand, fix the problem of non-local fault propagation, as in an illegal memory access, can be very hard. In contrast to other studies mentioned here, the degree distribution for the network representing the Gentoo Linux software is a stretched exponential, which can not be accounted for by traditional models [465]. The features of the network could be explained by a model using a nonlinear preferential attachment with a term-dependent vertex age.

The most important issues of software design are the software's complexity and stability, which can be explored through the complex network theory. The complexity of a software network can be determined by Shannon's information entropy of properties related to vertex degree [466], which measures the homogeneity of the connectivity of the network. Therefore, simple software networks are those whose connectivities are homogeneous. The stability, on the other hand, can be obtained by  $I_{cc}$  – instability class-to-class [467] – a metric to measure the instability between classes. The results of  $I_{cc}$  for the evolution of the packages Sun Java2, Hipergate, JEdit and ToscanaJ show that Sun Java2 uses good programming practices along each new release, due to the decreasing values for  $I_{cc}$  in contrast to the Hipergate package since its  $I_{cc}$  values remain almost the same.

A new approach has been adopted by Cai and Yin [468] in applying network concepts to the field of software design and evaluation. The software execution was represented as an evolving complex network, where the nodes comprised the methods executed by a particular program. The sequence in which these methods were executed defined the directed links between them. A new class of graphs was also defined. The software mirror graphs incorporate new attributes with temporal nature, such as the parameter  $\beta_{i,j}$  measuring the number of steps from an execution of method  $i$  to an execution of method  $j$ . This parameter can be represented by a matrix



that evolves every time a new method is executed. The authors evaluated the evolution of the software mirrors graphs obtained from three software packages, viz. Intra3D, MyIE and Diagram Designer. The networks have about  $4 \times 10^5$  nodes. For the three cases, the networks obtained at the end of a complete execution exhibited scale-free and small-world properties. Interestingly, for the software mirror graph in intermediate stages, the degree distribution was not a power law, but an exponential that converged to the power law when the software reached the end of execution.

### 9.2. Data sharing

In data sharing networks, users or computers sharing files in the Internet are considered vertices, and a direct link exists from vertex  $i$  to vertex  $j$  if  $j$  receives a certain number of files from  $i$  [469–471]. Physics data sharing community [472], WWW data sharing between Internet hosts [470], the Kazaa traffic between users [471], and the Gnutella [473] and eDonkey [474] peer-to-peer (P2P) overlay networks are just a few examples of this kind of network. In the case of scientific file sharing network, the definition for the arcs above is slightly different, where vertices are scientists and two scientists are connected if one of them is interested in the data of the other [472]. A mechanism of data location in this kind of network has been proposed and takes advantage of the presence of local clusters (due to the small-world property) [472].

Leibowitz *et al.* [471] studying Kazaa traffic in 2003 – a period dominated by peer-to-peer applications – employed data of an Israeli Internet Service Provider (ISP) to show 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). The Gnutella P2P overlay network does not use efficiently the underlying Internet since 40% of Gnutella vertices are inside the top 10 ASs and less than 5% of Gnutella's connections join vertices inside the same AS [473].

### 9.3. Circuits

The topology of electronic circuits was also explored using the complex networks approach, normally with an electronic component (e.g. integrated circuits, resistors, capacitors) being represented by a node with edges being the wires in a broad sense. Several results [475,476] indicated that these networks have scale-free degree distribution and small-world behavior. In [477], 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.

## 10. Linguistics

The development of language is one of the greatest accomplishments of humanity. Understanding the evolution and organization of language is useful because it sheds light into cognitive processes, as the way people think strongly affects the organization of a language. Conversely, language influences how humans think. It is also important to compare the properties of various languages to study their co-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 [478]. The Natural Language Processing (NLP) [479] community has

traditionally used network representations to develop techniques for automatic language understanding and generation. For instance, graph theory concepts have been applied to sentiment analysis [480] and tools from spectral graph theory have been used in word sense disambiguation and text summarization [481]. 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 [482–486], providing new approaches for NLP applications and linguistic theories. Indeed, there has been an increasing effort to join statistical physics and language research, as demonstrated by the organization of a related satellite workshop inside STATPHYS 23, the largest conference for statistical physics [487]. A linguistic network can be formed, for example, by a group of interconnected words or syllables. The many ways these elements can be linked in a 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. In the following paragraphs, we review works dealing with both groups of linguistic 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 similar concepts (not necessarily synonyms) [488] – note that networks containing only synonyms are also commonly employed [489–491]. An example of English thesaurus used to build semantic networks is the Roget's Thesaurus, originally published in 1852. Using an English thesaurus, Motter *et al.* [490] built a network of 30,000 nodes, from which a series of topological measures were computed. The authors went one step further and discussed their results from the standpoint of cognitive science: since human memory is associative, the small-world feature indicates a process of efficiency maximization in the information retrieval. The Wordnet database is another common resource for semantic networks. This lexicon groups words into sets of synonyms and stores other semantic relationships between words, such as antonymy, hyperymy and hyponymy [492]. It was found that polysemic links (for words with different meanings) are extremely important for the English WordNet organization, conforming it as a small-world, and also that the hubs are the most polysemic words. These observations indicate the strong influence polysemy has on language organization, allowing abstraction and generalization.

In another work, the US presidential debate transcripts from 1960 to 2004 originated a set of semantic networks that were analyzed through eigenvector centrality [493]. In that case, semantic networks were obtained from further processing of word co-occurrence data. The main outcome of this work is that the authors were able to correlate winning candidates and well-structured speech in debates. Finally, semantic networks can also be obtained from word-association experiments, where someone sequentially and freely provided words that he/she thinks are semantically related. Costa *et al.* [494] found a hierarchical organization of word categories in a word association network with loops connecting non-adjacent levels. Possibly, these loops are the shortcuts that make distances shorter in associative memory (see also discussion of Ref. [490] above).

The process of lexical development for an individual was modeled by Steyvers and Tenenbaum [489], 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 [489], this model produces networks similar to those obtained from real-world data of word associations.

The networks we now discuss do not employ semantic information; instead, they use only shallow properties of texts or at most syntactic relationships. In word-adjacency or co-occurrence networks [495–500], words that appear near each other in a text are connected in the network, as word proximity frequently indicates syntactic relationship [483]. Ferrer i Cancho and Solé [483] analyzed word-adjacency networks obtained from the British National Corpus (BNC), and related the observed small-world feature to the need of navigation acceleration (analyses showed that these shortcuts are the particles). The authors also conjectured that the agrammatism, a language disorder, is a consequence of hub removal (also known as directed attack on scale-free networks).

Roxas and Tapang [501] employed network statistics of word-adjacency networks to classify texts between prose and poem. Common measures were applied, such as betweenness centrality and diameter. Using linear discriminant analysis, they found that the clustering coefficient, average path length and average degree are the best measures for prose/poem disambiguation. The authors also investigated how to identify the shift between poem and prose within a single written work. Fixed-text windows were employed for this purpose, from which network parameters were computed separately for each window following reading text flow. An interesting observation is that the clustering coefficient suddenly decreases in the transition from prose to poem (conversely, this measurement is increased from poem to prose).

A growth model for word-adjacency networks was reported in [495], 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 the power law, with the old words being the hubs. This model is in good agreement with real data and presents both scale-free and small-world properties.

Another type of superficial networks is the syntactic one. Usually, a grammar formalism is employed to build these networks, such as dependency grammars. Authors frequently perform structural analyses of these networks, such as the assessment of small-world and scale-free features [502–505]. On the other hand, Corominas-Murtra *et al.* [506] employed syntactic connections between words to study the language development in children. For each time interval of child utterances they built a single network and found that children with less than 2 years old have a syntactic organization resembling a tree, while after that age the syntactic structure changes dramatically to a larger, denser network, indicating a remarkable change in grammar structure. These more developed structures also show small-world and scale-free features, along with disassortative behavior. The authors developed a stochastic model to reproduce these empirically observed features. The model chooses the number of words and the actual words following an exponential and Zipf’s probability distributions, respectively. The former distribution was measured in their data, while Zipf’s law is a well-known feature of language regarding word frequency. Simulations of the model indicated the following features, consistent with real data: transition from tree-like to scale-free networks, dynamics of changes in the clustering coefficient and path length (forming a small-world), as well as the degree distribution. Nevertheless, the model fails to reproduce the frequency of three-word motifs and the evolution of hub connectivity. An earlier study of language development appeared in [507], albeit using word co-occurrence networks.

Arbesman *et al.* [508] analyzed networks encoding the phonological structure of a language (words sounding similarly are connected) for English, Spanish, Mandarin, Hawaiian and Basque. The authors found a strong robustness to node removal in these networks. The interpretation

they gave for node removal is rather interesting, and deserves note: a node removal may occur when a person cannot remember the phonological form of a word, despite having used it before and knowing its meaning. This is called tip-of-the-tongue phenomenon. Therefore, a robustness investigation can shed light on strategies to recover from this state and also to discover which words are prone to the tip-of-the-tongue phenomenon.

The models of language development created by Dorogovtsev and Mendes [482] and Markošová [509] employ the degree-based preferential attachment rule to make connections between newly created words and words already inserted in the vocabulary. In these cases, the vocabulary is a web of words that interact in sentences, thus it is a positional (or co-occurrence) word web. Dorogovtsev and Mendes also defined that, as new nodes are created, new edges are created between old words, again using the preferential attachment rule. Markošová, instead, preferentially rewired some old edges, thus both models incorporate changes in vocabulary use. Their models fit the experimental data with a two-regime power law distribution of degrees.

Linguistic networks have been applied to several natural language processing tasks. For instance, choosing the synonym most expected in a given context can be done using a word co-occurrence network [510]. An automatic assessment of text quality was implemented using network measurements, in which correlation was established between measurements (such as degree) and text quality scores assigned by humans to essays written by high-school students in Portuguese [511]. A similar approach was adopted in [512] to assess the quality of machine translations, where the original and translated texts, or more specifically, their word-adjacency networks, were compared. The degree of structural change from source to translation was measured and correlated with translation quality. Employing a network of sentences linked using lexical co-occurrence, a set of automatic summarizers based on concepts such as shortest paths,  $k$ -cores and communities was also proposed [513]. The performance of this approach compared well with state-of-the-art summarizers for Brazilian Portuguese texts. In measurements taken from English, Bengali and Hindi, lexical networks were correlated with some issues related to the construction of spell checkers [514]. These networks were weighted, where each edge was associated with the orthographic distance between words (this distance considers the number of character substitutions, deletions or insertions necessary to transform one word into another). The authors found that the weighted clustering coefficient is correlated with the difficulty to correct non-word errors (i.e. spelling errors resulting in non-existent words). Real word errors, which are misspelled words still valid in the given language, were found to be higher in words with higher weighted degree. Finally, using a syntactic dependence network for the Romanian language, Ferrer i Cancho *et al.* [503] showed that spectral methods developed to detect community structure can be employed to group words of the same morphological class.

The papers reviewed in this section cover many of the possibilities to represent linguistic structures as networks, ranging from word-adjacency to synonym networks. Many studies on linguistic networks focus on experimental global features of the structures, while some works are devoted to model the evolution of these networks, with results agreeing with experimental data. Another important issue is the application of linguistic networks to natural language processing, for helping develop spell checkers and automatically assess text quality. As far as we know, networks using more sophisticated resources such as argument structure have not been studied yet, perhaps mainly because of the difficulty in obtaining large amounts of data. Modeling the development of language for specific individuals has also had its difficulties, since we do not know how language vocabulary or syntactic rules are stored in the brain. As for applied research, there is a great deal of possibilities unexplored, such as improving parsers and machine translation systems. The great challenge in this case is to choose the right type of network and analysis method to retrieve the desired information from interlinked structures.

## 11. Engineering

### 11.1. Transportation

Transportation networks are important for the development of a country and may be seen as indicators 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 other forms of public transport. Studying this kind of network can help one to understand the movement of people around the world and predict how diseases 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.

#### 11.1.1. 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 arc weights are given by the number of passengers of flights or number of flights itself in a day or in a week. Guimerà *et al.* [25], however, argued that the connections for the network of all airports in the world are 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.

In the worldwide airport network [25,515], 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 [25], on which a global role of cities has been developed. This measurement indicates that the vertices connecting different communities are hubs in their own community. Guimerà *et al.* [25] have 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 has been developed by Guimerà and Amaral [515], 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.

An analysis of the evolution of the Brazilian airport network [516] has shown that its structure has changed within a 12-year period. The number of airports and routes has decreased, while the betweenness centrality has increased over time. The overall behavior suggests that companies focus on concentrating the operation in the most profitable routes, increasing the number of flights and removing the less profitable ones in a dynamic way such that airports gain and lose importance over the years. One of the consequences is the possible increase in the vulnerability of the network for both random failures and targeted attacks [516].

In addition to analyzing the networks, it is important to design optimal networks for this kind of transport. Kim and Motter [204,205] analyzed networks with regard to the resource allocation, i.e. the seat occupation in the flights, and showed that airport networks have a very efficient capacity distribution, which is a consequence of the high cost of air transportation. A new route network to optimize the operational cost for the French airports network, based on graph coloring for the air traffic flow management, has been proposed in [517].

Gautreau *et al.* [518] analyzed the evolution of the US airport network from 1990 to 2000 and found out that the statistical distributions (e.g. number of connections and passengers per flight) are stationary, but at the local level there is an intense activity as a consequence of the creation and removal of connections between airports. In most cases, the properties of the new routes are almost identical to the connections removed. However, if a new connection is established between very different airports in terms of traffic, it is very volatile. The authors also proposed a model which reproduces the main features observed, including the stationary distributions and the local dynamics.

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 [519] analyzed the errors caused by an attack to the structure of the US airport network. Analogous 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 connections, the same properties are drastically altered if a few hubs are removed.

### 11.1.2. Roads and urban streets

These networks deserve special attention in our daily life since they directly influence our travel times and transport costs. Over the last years, we have witnessed an increasing number of vehicles on the roads which resulted in slower traffic flow and more frequent jams. Although possible in the past, building new roads or streets is not an easy task today because of spatial constraints imposed by the local buildings. The best solution is, therefore, to make better planning for the traffic or even to manage it in the real time so that time wasted 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 and streets 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 and street networks can be *primal* or *dual* graph [520,521]. In the former case, while intersections are nodes, the roads or streets 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 [26,522–527]. A more detailed study of this kind of networks can be found in [522], where the authors analyze 20 samples of street patterns of several world cities. With this kind of representation, it has been shown that airport networks can be obtained by the path-star transformation applied to the corresponding road networks [528], where paths with a given length and defined by random walks in the road networks give rise to star-like structures in the corresponding airport networks.

There is a limitation, however, with the primal representation, since it does not express the difficulty to find one's way in the streets [529], which can be obtained from the *dual* representation, where roads are nodes and intersections are edges. Such a representation, known as “axial mapping”, was defined in the seminal work of Hillier and Hanson [530]. Although it 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 smaller the number of road changes to reach a specific destination, the easier to find it. This kind of representation is also useful for representing rail networks [531], since most passengers are interested in the train changes necessary to reach their destinations.

Rosvall *et al.* [529] expressed the difficulty of navigation on these networks in terms of the “search information” [532]. 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 second city. Another conclusion of Rosvall *et al.* [529] is that it is preferable to replace a big number of streets with a few long, provided it reduces the number of street changes necessary to connect any two points of the network. The accessibility of places in towns and cities has also been investigated with self-avoiding random walk dynamics [524,526,527] in the primal representation, and it has been shown that the dynamics of transportation through towns and cities is strongly affected by the topology of the connections and routes. For instance, it was found that the accessibility of the less accessible regions of São Carlos in Brazil could be improved by the

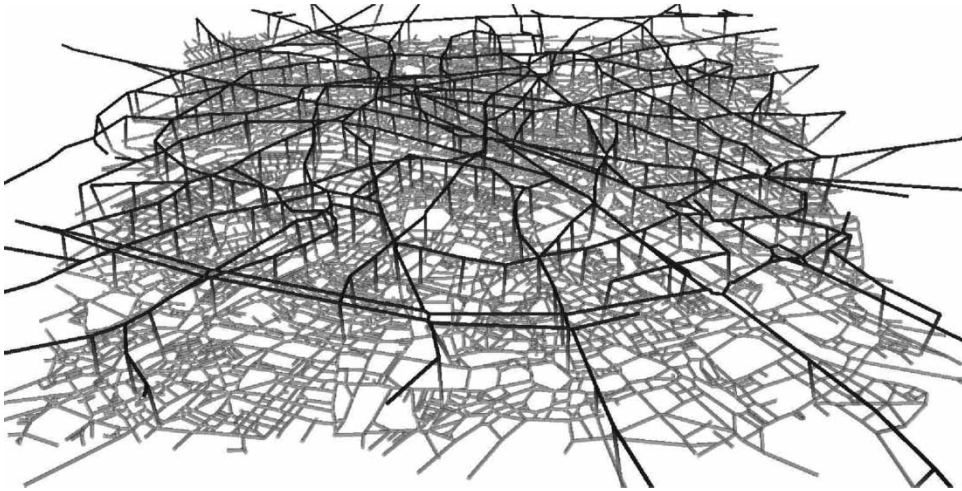


Figure 10. An illustration of Paris's streets (gray) and underground lines (black).

addition of some streets [526]. Costa *et al.* [527] also employed this measurement to compare the street networks in big cities such as London and Paris and showed that they are quite different due to the number of bridges and parks in London (i.e. some regions of this city are less accessible). The authors also showed that with the underground systems, the corresponding accessibility is increased for both cities providing a uniform access to distinct areas. Figure 10 presents a small portion of Paris's streets (green) and underground lines (black), where this effect is apparent.

Gastner and Newman [26] compared geographical networks such as an airline network, the US interstate highway network, and the Internet at the autonomous system level. They showed that these 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, e.g. the geographical and geodesic distance. The authors have also proposed a model to reproduce their findings. Another model for this type of representation is the geographic path model [533,534], where vertices are connected through paths instead of just considering the distance between them. This model is capable of reproducing many properties of the US, England and Brazilian road networks [534].

The dynamics on road networks has been studied for years, with many models proposed [535–539], ranging from macroscopic models based on the kinetic gas theory or fluid dynamics to microscopic approaches with equations for each car in the network. Schadschneider *et al.* [540] analyzed the German highway network using cellular automata as a model for the traffic flow, and showed that the simulations are faster than real time making this model suitable for traffic forecasting. It was also possible to find the bottlenecks of the highway network.

### 11.1.3. Some remarks on transportation networks

The analysis of transportation networks is important because it provides methods to investigate the economy of countries and can be used to improve their infrastructure. The main results obtained for the airport networks were that they are very efficient in terms of resource allocation, due to the expensive cost of air transportation, and tolerant to random attacks, while vulnerable to targeted ones, analogous to many scale-free networks. In the case of road networks, the best planned cities are those with low search information, i.e. it is not difficult to travel inside them, and high accessibility, i.e. distinct regions can be accessible uniformly. The majority of the models developed

for transportation networks are based on geographical and economic constraints, with the main aim of reaching optimal, robust networks. The development of better models is fundamental for achieving such results.

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 railway [531,541–543], subway [544,545] and public transportation [546,547] networks.

### 11.2. *Electric power transmission systems*

The electric power transmission system is one of the most complex man-made networks. It comprises transmission lines and several substations that include generators, i.e. the electricity power source; transmission substations, which connect high-voltage transmission lines; and load centers, which deliver the electricity to consumers [548]. 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 that affected 50 million people in the USA and Canada on 14 August 2003, and resulted in a huge loss of money (around US \$30 billion) [549]. Obviously, this kind of network is crucial for the economy of a country and deserves special attention in engineering and science.

The first studies on 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 [550–554]. 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 fails, all the power flow is redistributed to the other lines, but new failures may happen due to overflow, leading to a cascading effect. These dynamical processes were simulated [550–554], 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. For a power network of 314,123 nodes, Chassin and Posse [555] 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), and showed that the network has a radial form. The generators are at the center, with transmission substations at the middle and the load centers at the border, in the so-called “bow-tie” configuration [246,555].

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 [556–558]. Each substation was assumed to have a transmission capacity that depends on the number of shortest paths passing through it. Albert *et al.* [548], 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 [548]. 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.* [559]. It describes an electric transmission network under random perturbations, such as line or substation failures and overloads, and the operator's response to the contingency events, in 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 able not only to predict blackouts, but also to find the optimal strategy for minimizing the impact of random component failures.

Overall, the power transmission systems represent very large networks, with hundreds of thousands of vertices ranging from generators, transmission substations to load centers. This kind of network has many redundant paths to connect every generator to every load center and exhibit the “bow-tie” configuration (i.e. generators are in the center, transmission substations in the middle and the load centers, in the border). Due to the redundant paths, the power transmission systems are quite robust to random failures of generators, since the others can handle the additional power, but can have large blackouts if highly connected transmission substations fail as the other substations cannot take the additional overload. The first models developed for transmission power systems had simple topologies, aimed at simulating blackouts. More realistic models now exist which include more sophisticated topologies and are based on stochastic events.

## 12. Earth sciences

### 12.1. Earthquakes

The study of earthquake occurrence with its spatial distribution may be done through complex networks [560–564]. In general, the method used to build the networks consists in dividing the Earth surface into cells that are the nodes of the network. Two successive earthquakes (i.e without any other between them) establish a new link in the network between the cells where these shocks occurred. From the analysis of real seismic data obtained from public databases, such as the Southern California Earthquake Data Center<sup>14</sup>, it was shown that earthquake networks have a scale-free topology with coefficients depending on the tectonic plates analyzed. The hubs were found to be related to the place where a main shock occurred (i.e earthquakes with large magnitude). The mixing property was investigated using the concept of nearest-neighbor average connectivity, in which Abe and Suzuki showed that highly connected nodes are linked to each other, for both seismology networks, namely from California and Japan. It was discovered that in the growth of a seismology network after the main shock the clustering coefficient remains constant during the time  $\Delta t$ , which depends on how the network was built, typically of the order of hours [564]. After  $\Delta t$ , the clustering coefficient has a steep drop and then decays slowly according to a power law before it becomes steady again.

### 12.2. Climate networks

The extension of network theory to climate sciences has been reported by Tsonis *et al.* [41, 565, 566], where the networks represent a grid of oscillators – each of them is a dynamical system varying in some complex way. The nodes in the network represent cells of a grid of  $5^\circ$  latitude  $\times$   $5^\circ$  longitude that covers the globe. The connections between the nodes were defined through a threshold of correlation coefficient of a time series of each node. The connections were established using the database from the National Center for Environmental Prediction/ National Center for Atmospheric Research (NCEP/NCAR). Generally, data from the 500 hPa value, which indicates the height of the 500 hPa pressure level and provides a good representation of the general circulation of the atmosphere, were used for network construction. If the correlation coefficient for 500-hPa time series of two nodes exceeded 0.5, these nodes were connected. The resulting network revealed to be scale-free with small-world behavior [565].

Studying climate networks may help identify global changes and climate dynamics [565]. For instance, Tsonis and Swanson [567] constructed the El Niño and La Niña networks by taking into account the global temperature. They showed that the El Niño network is less connected and less stable than the La Niña network, because the former presents fewer links, with lower clustering coefficient and characteristic path length than the latter.

Climate networks are generally constructed by taking into account the correlation between time-series of different places worldwide, e.g. [565]. However, recently Dondges *et al.* [568,569] used nonlinear mutual information of time series analysis to establish the connections between Earth's grid points and verified that structures of high-energy flow are related to global surface ocean currents.

### 13. Basic sciences

#### 13.1. Physics

Many phenomena in physics can be modeled as networks, encompassing applications from various areas, from particle physics to astrophysics. In this section we shall describe some papers including these applications.

##### 13.1.1. Energy landscapes

Doye and Massen [570] used networks to study the configurational space of the Lennard-Jones potential. To build the network, called reaction graph, the authors took each minimum of the potential function as a node, with the surrounding of each node defining its basin of attractions [571,572]. The nodes were connected when there was a direct state transition between their corresponding minima, not considering auto-loops and multiple edges. The results indicated that the network topology is scale-free and small-world. The authors discovered a negative correlation between the degree and the potential energy of the minima, indicating that the low-energy minima act as hubs, the signature of scale-free networks. Interestingly, the scale-free nature of this network is not due to preferential attachment, but a direct consequence of the potential landscape topography, because the basins of attraction related to the low-energy minimum are larger, in general, and more likely to have surrounding transition states. The authors also suggested that the small-world effect can be useful to explain the Levinthal paradox, which states that the time of a protein to arrive at its native conformation should be 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 typical problem with  $10^{21}$  minima, algorithms based on global optimization converge, in average, after only 150 minima. Reaction graphs were already used in previous papers, such as [573], to investigate the isomers of the water dimer, and [574] to study interconnection of the protein conformal space. Gfeller *et al.* [575] also studied energy landscapes through weighted networks, achieving analytical results for low-dimensional models and discussing how the network approach is useful for the isomerization problem. The landscape of spin-glasses mapped in networks by Seyed-allaei *et al.* [576] had a Weibull distribution for the degree, instead of the power law reported previously for the Lennard-Jones potential [570]. The authors concluded that the topology of the landscape networks is not universal because it depends on the physical properties of the model system.

#### 13.2. Chemistry

Chemical reactions were studied on systems described by networks with complex topologies. Solé and Munteanu [577] used the UMIST Database for Astrochemistry to generate networks, where 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<sub>2</sub>O. The authors suggested that the community structure in this network results from action of the most reactive free radicals of Earth atmosphere, namely Cl and OH.

#### 14. Collaboration network of the papers cited in this review

In order to have a more complete, integrative perspective of the works addressing applications of complex networks, we constructed a network of collaborations from the list of references in this 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 to filter and organize the BibTeX database. The network, which contains 1149 nodes and 5508 edges, is illustrated in Figure 11. Several features can be seen in this figure. First, the overall network is highly fragmented, indicating that the application of complex networks is still largely 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 to be known or applied by the other communities. It is felt that important and interesting advances could be achieved through a more integrated approach involving more intense collaboration between the various communities. Several of the most highly cited authors tend to have high centrality (the betweenness centrality of each author is indicated in Figure 11), with some exceptions. Almost all such authors can be found in the largest cluster, which means that important researchers on complex networks tend to collaborate with one another. The densest groups of authors result from papers involving dozens of co-authors. Nevertheless, these authors are not necessarily well connected to the main cluster.

We also analyzed the time evolution of the main features of the collaboration network since 1998 (Figure 12). Although the number of authors grows monotonically, a step-like function of average degree has been identified, which is a consequence of the publication of papers with large number of authors (Figure 12(a) – one paper with 49 authors and another with 52 authors published in 2003 and 2006, respectively). Moreover, the growth rate of the giant component is larger than that exhibited by the number of components, which means that new authors have a tendency of being connected to the largest cluster rather than forming isolated small clusters (Figure 12(b)). In Figure 12(c), we can see that, after a transient stage where the network is very small, both the assortativity and the clustering coefficient remain almost constant along the last years. The high values of assortativity indicate that in the collaboration network nodes with similar degree tend to interconnect. This feature does not favor the interaction between less connected and high profile researchers. Regarding the clustering coefficient, its high values are a direct consequence of the network construction, where cliques of co-authors are included in the network for each published paper.

#### 15. Conclusions and perspectives

The success of new areas of physics is frequently judged not only owing to their theoretical contributions, 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

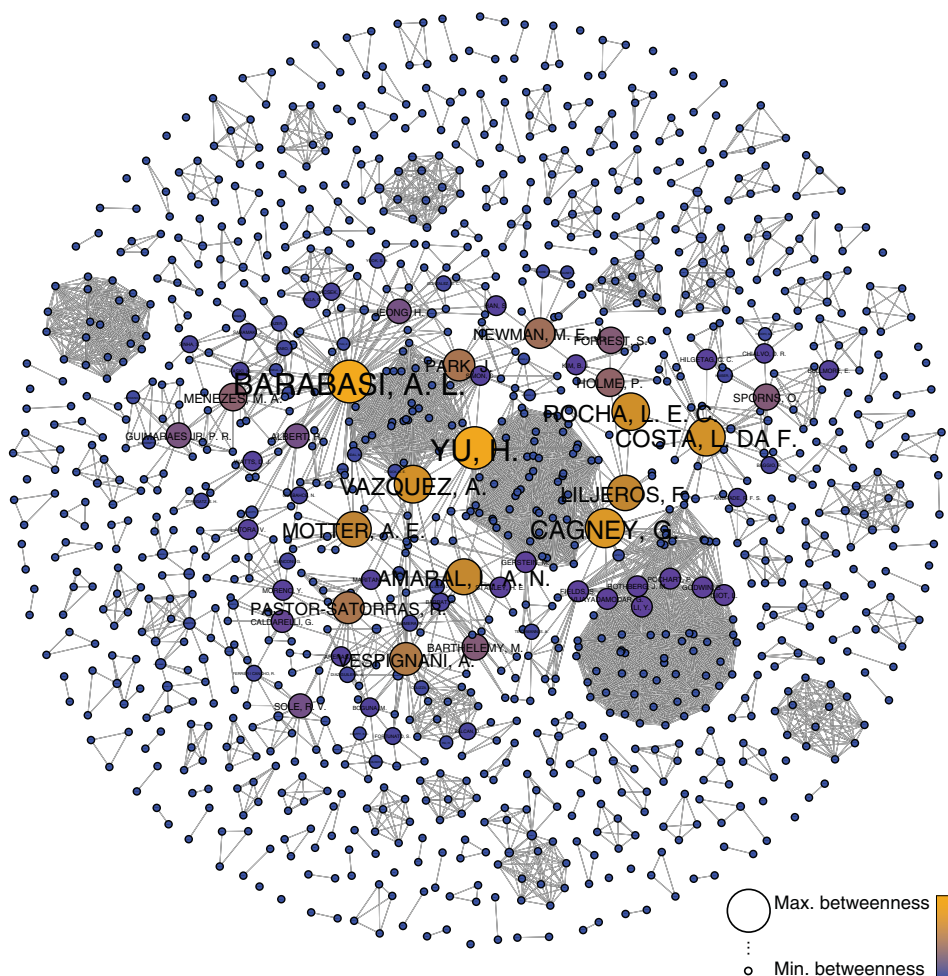


Figure 11. Collaboration network considering the papers cited in this review. Label and node sizes, as well as node colors, indicate the betweenness centrality of each author. Only authors whose centralities are large have visible names.

mechanics and graph theory, as a general and powerful theoretical framework for representing and modeling real complex systems. It has been capable of taking into account not only the connectivity structure of those systems, but also intricate dynamics (see, for instance, the surveys by Newman [31] and Boccaletti *et al.* [27]).

Judging by the large number of areas and articles reviewed in the present survey, complex networks have performed exceedingly well with respect to their application potential. Indeed, the generality of complex networks extend naturally to real-world problems in virtually any area, from neurosciences to earthquakes, encompassing at least the 11 areas considered here. With the application potential of complex networks being clearly substantiated, we conclude the present survey by conducting a global analysis of the applications reported.

Table 3 lists the 11 application areas considered here and the corresponding number of reviewed papers, the size of the networks and respective number of edges. Despite the bias implied by the choice of articles reviewed, this table still provides a representative snapshot of the state of the art in complex networks applications. The first remarkable feature in this table is the fact that sociology

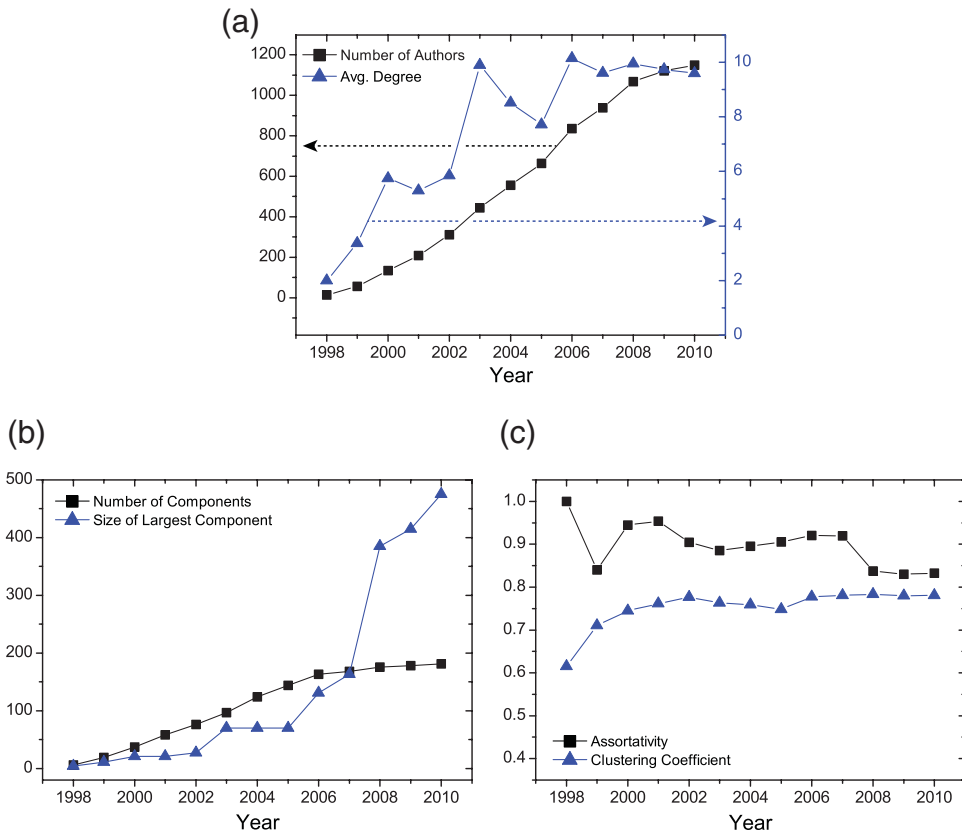


Figure 12. Properties of the collaboration network computed along each year since 1998.

is the area where complex networks have been more intensively applied (97 applications), followed by biology (83 applications). Other areas that received particular interest from complex network research include Internet, epidemiology, economy, computer science and engineering, all with at least 30 reviewed articles. The areas with the smallest number of applications (less than 10 entries) include earth and basic sciences. Moreover, most areas deal with networks of at most thousands of nodes, while for the World Wide Web, the Internet and sociology (more specifically, phone communications), networks with millions of nodes have been studied already. It is a difficult issue to identify which of these areas will remain little investigated, while others become the focus of increasing attention.

In biology, data of protein–protein interaction, metabolic relationships and genetic processes are continuously becoming more complete, allowing increasingly more accurate complex networks and paving the way to breakthroughs in prediction of biological functions as well as leading to a better understanding of evolutionary processes. Important future works in this area include the simulation of dynamics capable of reproducing biological processes such as information exchange between genes. In medicine, applications of complex networks have revealed that an intrinsic relationship has been achieved between distinct diseases, fostering the investigation of interplays between them. These networks could be integrated with other types of networks in order to obtain more comprehensive representation of the biology of whole organisms. These advances will depend on the availability of new experimental data regarding the several medical aspects

Table 3. Number of papers, network sizes and number of edges per research area considering the works cited in this review.

Research field	Papers	Number of nodes	Number of edges
Biology	83	10 – 10,000	20 – 50,000
Internet	35	112 – 1,249,747	4,334 – 28,959
World Wide Web	18	3,000 – 200,000,000	1,469,680 – 1,500,000,000
Sociology	97	20 – 50,000,000	90 – 50,000,000
Epidemiology	42	3100	17,000
Economy	30	30 – 1071	400 – 4,000
Computer Science	33	20 – 60,000	300 – 120,000
Linguistics	26	29 – 478,773	100 – 17,700,000
Engineering	50	100 – 314,000	400 – 10,000
Earth Sciences	5	4,000 – 13,000	40,000
Basic Sciences	8	4 – 4,000	5 – 61,000

such as microbiology and pharmacological interactions. Complex networks have also contributed to several advances in ecology. Especially, universal properties have been identified in different food-webs, and the effect of perturbations into ecosystems induced by alien species has been investigated. Food-webs seem to be highly resilient to extinction of species. Promising future research includes the effects of propagation of pollutants in food-webs and the simulation of more sophisticated population dynamics. Important contributions in the neuroscience field have also been reported. For instance, a strong relationship has been verified between neuronal structure and function. Some studies have shown that neural diseases and damage lead to well-defined alterations of the topology of cortical connectivity. In addition, other studies identified differences in the cerebral organization among distinct mammals. Future works would require more complete and reliable databases as well as the development of methodologies capable of relating cerebral dynamics with specific diseases and abnormalities.

Though the Internet and WWW were the subjects of the first investigations in complex networks, it is still difficult to find comprehensive public databases. These two types of networks have been acquired through the use of methods that navigate through the structures in order to map the respective topology. Therefore, the networks obtained are often incomplete and subjected to biases implied by the stochastic nature of the exploration as well as the temporary unavailability of nodes. As a matter of fact, the eventual full mapping of the WWW will yield an extremely large structure that will be difficult to be stored and analyzed, demanding state-of-the-art computing approaches.

The still incomplete social networks are being improved by using social online networks data. In particular, the online monitoring of this type of data paves the way to the understanding of global and local patterns, cultural trends, opinion formation and market strategies. Regarding scientific collaboration, it has been suggested that the most productive authors do not correspond to the hubs, but to those researchers who have greater diversity of interests. Future advances should be expected after the integration of different areas. This integration can contribute to the field of security and surveillance, which may allow the prediction of the architecture of terrorist organizations. Such a problem cannot be solved currently, but it has been shown that the weak points of such organizations are not the hubs, but the edges whose removal implies great decrease in the information propagation in the networks. Still in the social context, several studies related to communication networks have been reported. Great attention has been focused on the investigation of email networks, used in developing methods for controlling spreading of computer viruses and anti-spam filters with nearly 100% performance accuracy. Promising future works include monitoring message exchange to identify patterns of behavior and message content. Telephone

networks have also been investigated, especially regarding mobile phones. Interesting insights have been obtained on the patterns of human mobility, consequently representing important subsidies for disease control and urban planning. As most of such analysis have been performed locally, i.e. inside a single country or community, it would be interesting to obtain more global data and results.

It has been shown that scale-free networks have null epidemic threshold, which implies that any epidemic with non-null infection rate can propagate. Long range connections such as airline routes are particularly important for the outbreak onset at the global level. On the other hand, short-range connections such as highways have little effect on the global outbreaks but strongly influence the local propagation pattern. Allied to an ever growing availability of experimental data, the current models permit highly accurate predictions. Some related simulation resources are now available on the Internet. Among the strategies used, namely local, global and quasi-local, the latter has been verified to provide the most promising results.

Particularly accurate and complete financial market databases have been available for a long time. Frequently, the data are organized as an asset tree which is used to investigate the hierarchical relationships between the involved entities. However, better understanding and prediction of economy will require the development of more sophisticated models.

Some interesting results have been reported regarding the application of complex networks to software architecture. It is possible to identify proper programming practices through the study of the topological properties of the networks. It would be particularly promising to incorporate complex networks concepts and methods as an intrinsic feature in software engineering, such as in quality control and even software development.

Linguistic networks tend to be particularly complete and reliable, as a consequence of the availability of full texts as well as objective mapping methods. Several investigations have been reported concerning applications of linguistic networks, ranging from modeling language development in children to the automated text quality assessment. Further advances will require going beyond the basic semantic representations currently available, e.g. incorporating argumentative structure and cultural contexts.

The application of complex networks to transportation is important because of its economical implications. Airport networks have been found effective regarding passenger allocation. Future advances will require the integration of different types of transportation networks, such as highways, airports and railways, which may allow the optimization of the transportation system. The availability of accurate, reliable data on power grids have allowed the development of more realistic models which can be applied to identify critical parts of the grids, leading to increased robustness.

Two clear examples of complex networks applications to earth sciences are the earthquake networks and climate networks. In the first case, it is still not possible to make good predictions of earthquakes by using complex networks. However, there are clues suggesting that major topological changes in the networks are observed after earthquakes. Further advances should be achieved as more comprehensive databases become available. For the climate networks, patterns identified in these systems have been related to natural phenomena such as the El Niño, suggesting a strong relationship between the topology of climate networks and their respective dynamics. Some important issues remain unsolved, such as the definition of proper procedures for network construction as well as the selection of suitable dynamics capable of modeling climatic modifications.

The results contained in the papers reviewed, 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, 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.

### Acknowledgements

L. da F. Costa is grateful to FAPESP (proc. 05/00587-5) and CNPq (proc. 301303/06-1) for financial support. F. A. Rodrigues acknowledges FAPESP financial support (proc. 07/50633-9). P. R. Villas Boas is grateful to CNPq sponsorship (141390/2004-2). L. E. C. Rocha is grateful to CNPq and to the Swedish Research Council for financial support. L. Antikeira acknowledges FAPESP sponsorship (proc. 06/61743-7). M. P. Viana is grateful to FAPESP sponsorship (proc. 07/50882-9). We are also grateful to S. Abe, M. Ausloos, R. Baggio, Y. Bar-Yam, G. Bianconi, V. Blondel, D. Braha, R. Ferrer i Cancho, R. Criado, F. D. Fallani, S. Dos Reis, P. Holme, B. Kahng, J. Kertész, R. Lambiotte, B. Li, F. Liljeros, D. Lusseau, M. M. Babu, Y. Maeno, G. Michailidis, A. Motter, C. Oosawa, A. Penn, S. Roy, A. Sánchez, S. Sinhá, M. Small, F. Sorrentino, D. Stauffer, F. Tamarit, M. Tugrul, D. Volchenkov, Y. Xiao, and T. Zhou for comments and suggestions.

### Notes

1. KEGG: <http://www.genome.jp/kegg>.
2. EcoCyc: <http://ecocyc.org/>.
3. <http://navigators.com/isp.html>
4. <http://www.elsevier.com/locate/inca/505596>.
5. <http://www.cmu.edu/joss>.
6. <http://www.secondlife.com>.
7. For an extensive list of social networking websites access [http://en.wikipedia.org/wiki/List\\_of\\_social\\_networking\\_websites](http://en.wikipedia.org/wiki/List_of_social_networking_websites).
8. <http://us.imdb.com>.
9. <http://www.orgnet.com/hijackers.html>.
10. <http://www.trackingthethreat.com> and <http://www.globalsecurity.org>.
11. <http://www.census.gov>.
12. <http://www.gleamviz.org>.
13. Data collected from <http://www.intracen.org/menus/countries.htm>. and <http://www.tswoam.co.uk/world>.
14. <http://www.scecdc.scec.org/catalogs.html>.

### References

- [1] A.-L. Barabási and Z.N. Oltvai, *Nat. Rev. Genet.* 5(2) (2004), pp. 101–113.
- [2] S.N. Dorogovtsev and J.F.F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press, New York, 2003.
- [3] S. Bornholdt and H.G. Schuster, *Handbook of Graphs and Networks: From the Genome to the Internet*, Wiley-VCH GmbH KGaA, Weinheim, 2003.
- [4] P.L. Garrido and J. Marro, *Modeling Complex Systems (Seventh Granada Lectures on Computational Physics)*, AIP Conference Proceedings. American Institute of Physics, Melville, 2003.
- [5] E. Ben-Naim, H. Frauenfelder, and Z. Toroczkai, *Complex Networks*, Lecture Notes in Physics, vol. 650. Springer Verlag, New York, 2004.
- [6] M. Newman, A.-L. Barabasi, and D.J. Watts, *The Structure and Dynamics of Networks*, Princeton University Press, Princeton, NJ, 2006.
- [7] M.E.J. Newman, *Networks: An Introduction*, Oxford University Press, NY, 2010.
- [8] A.-L. Barabási, *Linked: how everything is connected to everything else and what it means for business, science, and everyday life*, Plume, 2003.



- [9] F. Chung and L. Lu, *Complex Graphs and Networks*, CBMS Regional Conference Series in Mathematics. American Mathematical Society, Providence, 2006.
- [10] P. Erdős and A. Rényi, *Bull. Int. Statist. Inst.* 38(4) (1960), pp. 343–347.
- [11] P. Erdős and A. Rényi, *Publ. Math. (Debrecen)* 6 (1959), pp. 290–297.
- [12] B. Bollobás, *Random graphs*, Cambridge Studies in Advanced Mathematics. Cambridge University Press, New York, 2001.
- [13] B. Bollobás, *Graph theory: An introductory course*, Graduate Texts in Mathematics. Springer, New York, 1990.
- [14] J.P. Scott, *Social Network Analysis: A Handbook*, 2nd ed., Sage Publications, London, 2000.
- [15] S. Wasserman and K. Faust, *Social Network Analysis*, Cambridge University Press, Cambridge, 1994.
- [16] L.C. Freeman, *The Development of Social Network Analysis: A Study in the Sociology of Science*, Empirical Press, Vancouver, 2006.
- [17] M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Comput. Commun. Rev.* 29(4) (1999), pp. 251–262.
- [18] R. Albert, H. Jeong, and A.-L. Barabási, *Nature* 401 (1999), pp. 130–131.
- [19] R. Albert, H. Jeong, and A.-L. Barabási, *Nature* 406 (2000), pp. 378–382.
- [20] H. Jeong, S.P. Mason, A.-L. Barabási, and Z.N. Oltvai, *Nature* 411(6833) (2001), pp. 41–42.
- [21] L.da F. Costa, F.A. Rodrigues, and G. Travieso, *Appl. Phys. Lett.* 89(17) (2006), p. 174101.
- [22] R. Guimerà and L.A.N. Amaral, *Nature* 433(7028) (2005), pp. 895–900.
- [23] A. Krause, K.A. Frank, D.M. Mason, R.E. Ulanowicz, and W.M. Taylor, *Nature* 426 (2003), pp. 282–285.
- [24] M.E.J. Newman, *Proc. Natl Acad. Sci. USA* 98(2) (2001), pp. 404–409.
- [25] R. Guimerà, S. Mossa, A. Turtschi, and L.A.N. Amaral, *Proc. Natl Acad. Sci. USA* 102 (2005), pp. 7794–7799.
- [26] M.T. Gastner and M.E.J. Newman, *Eur. Phys. J. B* 49(2) (2006), pp. 247–252.
- [27] S. Boccaletti, V. Latora, Y. Moreno, M. Chaves, and D.-U. Hwang, *Phys. Rep.* 424(4–5) (2006), pp. 175–308.
- [28] L.da F. Costa, F.A. Rodrigues, and G. Travieso, *Phys. Rev. E* 76(4) (2007), p. 46106.
- [29] R. Albert and A.-L. Barabási, *Reviews of Modern Physics* 74(1) (2002), pp. 47–97.
- [30] S.N. Dorogovtsev and J.F.F. Mendes, *Advances in Physics* 51(4) (2002), pp. 1079–1187.
- [31] M.E.J. Newman, *SIAM Review* 45(2) (2003), pp. 167–256.
- [32] L.da F. Costa, F.A. Rodrigues, G. Travieso, and P.R. Villas Boas, *Adv. Phys.* 56(1) (2007), pp. 167–242.
- [33] S.H. Strogatz, *Sync: The Emerging Science of Spontaneous Order*, Hyperion, 2003.
- [34] L.da F. Costa and M.S. Barbosa, *Eur. Phys. J. B* 42 (2004), pp. 573–580.
- [35] O. Sporns, D.R. Chialvo, M. Kaiser, and C.C. Hilgetag, *Trends Cogn. Sci.* 8(9) (2004), pp. 418–425.
- [36] L.da F. Costa, O. Sporns, L. Antikeira, M.G.V. Nunes, and O.N. Oliveira Jr, *Applied Physics Letters* 91 (2007), p. 054107.
- [37] L.da F. Costa and F.A. Rodrigues, *What is there between any two nodes in a complex network?* Preprint arXiv:0801.4068, 2008.
- [38] D.J. Watts and S.H. Strogatz, *Nature* 393(6684) (1998), pp. 440–442.
- [39] W. Gerstner and W.M. Kistler, *Spiking Neuron Models*, 1st ed., Cambridge University Press, 2002.
- [40] P.R. Monge and N.S. Contractor, *Theories of Communication Networks*, Oxford University Press, New York, 2003.
- [41] A.A. Tsonis, A.G. Hunt, and J.B. Elsner, *Meteorol. Atmos. Phys.* 84(3) (2003), pp. 229–242.
- [42] R.N. Mantegna, *Eur. Phys. J. B* 11 (1999), pp. 193–197.
- [43] C.J. Stam, *Neurosci. Lett.* 355(1–2) (2004), pp. 25–28.
- [44] F.S. Collins, E.D. Green, A.E. Guttmacher, and M.S. Guyer, 2003 *Nature* 422(6934) (2003), pp. 835–847.
- [45] H. Kitano, *Science* 295(5560) (2002), pp. 1662–1664.
- [46] K. Sneppen and G. Zocchi, *Physics in Molecular Biology*, Cambridge University Press, New York, 2005.
- [47] B. Vogelstein, D. Lane, and A.J. Levine, *Nature* 408 (2000), pp. 307–310.
- [48] Z.N. Oltvai and A.-L. Barabási, *Science* 298(5594) (2002), pp. 763–764.
- [49] L.A.N. Amaral and J.M. Ottino, *Eur. Phys. J. B* 38 (2004), pp. 147–162.

- [50] 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. Rillstone, K. Gandi, N.J. Thompson, G. Musso, P. St 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, *Nature* 440 (2006), pp. 637–643.
- [51] 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, *Nature* 409(6817) (2001), pp. 211–215.
- [52] 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, *Science* 302(5651) (2003), pp. 1727–1736.
- [53] E. Ravasz and A.-L. Barabási, *Phys. Rev. E* 67(2) (2003), p. 26112.
- [54] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, *Nature* 435(7043) (2005), pp. 814–818.
- [55] A.W. Rives and T. Galitski, *Proc. Natl Acad. Sci. USA* 100(3) (2003), pp. 1128–1133.
- [56] R. Albert, *J. Cell Sci.* 118(21) (2005), pp. 4947–4957.
- [57] A.E. Motter, N. Gulbahce, E. Almaas, and A.-L. Barabási, *Mol. Syst. Biol.* 4 (2008), p. 168.
- [58] S. Maslov and K. Sneppen, *Science* 296(5569) (2002), pp. 910–913.
- [59] T. Manke, L. Demetrius, and M. Vingron, *Genome Informatics* 16(1) (2005), pp. 159–63.
- [60] M.P. Joy, A. Brock, D.E. Ingber, and S. Huang, *J. Biomed. Biotechnol.* 2005(2) (2005), pp. 96–103.
- [61] S. Wuchty, Z.N. Oltvai, and A.-L. Barabási, *Nat. Genet.* 35(2) (2003), pp. 176–179.
- [62] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Science* 298(5594) (2002), p. 824.
- [63] L.H. Hartwell, J.J. Hopfield, S. Leibler, and A.W. Murray, *Nature* 402(6761) (1999), pp. C47–C52.
- [64] J.F. Poyatos and L.D. Hurst, *Genome Biol.* 5 (2004), p. R93.
- [65] I. Albert and R. Albert, *Bioinformatics* 20(18) (2004), pp. 3346–3352.
- [66] N. Williams, *Science* 272(5261) (1996), pp. 481–500.
- [67] S. Fields and O. Song, *Nature* 340(6230) (1989), pp. 245–246.
- [68] P.L. Bartel and S. Fields, *The Yeast Two-Hybrid System*, Oxford University Press US, 1997.
- [69] T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori, and Y. Sakaki, *Proc. Natl Acad. Sci. USA* 98(8) (2001), pp. 4569–4574.
- [70] P. Legrain and L. Selig, *FEBS Lett.* 480(1) (2000), pp. 32–36.
- [71] S. Mukherjee, S. Bal, and P. Saha, *Curr. Sci.* 81(5) (2001), pp. 458–464.
- [72] A.M. Lesk, *Introduction to Protein Architecture: The Structural Biology of Proteins*, Oxford University Press, Oxford, 2001.
- [73] D. Eisenberg, E.M. Marcotte, I. Xenarios, and T.O. Yeates, *Nature* 405 (2000), pp. 823–826.
- [74] I. Xenarios and D. Eisenberg, *Curr. Opin. Biotechnol.* 12(4) (2001), pp. 334–339.
- [75] R. Mrowka, A. Patzak, and H. Herzog, *Genome Res.* 11 (2001), pp. 1971–1973.
- [76] R. Saito, H. Suzuki, and Y. Hayashizaki, *Nucleic Acids Res.* 30(5) (2002), pp. 1163–1168.
- [77] P.L. Bartel, C.T. Chien, R. Sternglanz, and S. Fields, *Biotechniques* 14(6) (1993), pp. 920–924.
- [78] S. Fields and R. Sternglanz, *Trends Genet.* 10(8) (1994), pp. 286–92.
- [79] P. Uetz, L. Giot, G. Cagney, T.A. Mansfield, R.S. Judson, J.R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J.M. Rothberg, *Nature* 403(6770) (2000), pp. 623–627.
- [80] C. von Mering, R. Krause, B. Snel, M. Cornell, S.G. Oliver, S. Fields, and P. Bork, *Nature* 417(6887) (2002), pp. 399–403.
- [81] E.M. Marcotte, M. Pellegrini, H.L. Ng, D.W. Rice, T.O. Yeates, and D. Eisenberg, *Science* 285(5428) (1999), pp. 751–753.

- [82] E. Sprinzak, S. Sattath, and H. Margalit, *J. Mol. Biol.* 327(5) (2003), pp. 919–923.
- [83] H. Yu, P. Braun, M.A. Yildirim, I. Lemmens, K. Venkatesan, J. Sahalie, T. Hirozane-Kishikawa, F. Gebreab, N. Li, N. Simonis, T. Hao, J.-F. Rual, A. Dricot, A. Vazquez, R.R. Murray, C. Simon, L. Tardivo, S. Tam, N. Svrikapa, C. Fan, A.-S. de Smet, A. Motyl, M.E. Hudson, J. Park, X. Xin, M.E. Cusick, T. Moore, C. Boone, M. Snyder, F.P. Roth, A.-L. Barabási, J. Tavernier, D.E. Hill, and M. Vidal, *Science* 322(5898) (2008), pp. 104–110.
- [84] 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, *Proc. Natl Acad. Sci. USA* 96(8) (1999), pp. 4285–4288.
- [85] H. Hishigaki, K. Nakai, T. Ono, A. Tanigami, and T. Takagi, *Yeast* 18(6) (2001), pp. 523–531.
- [86] B. Schwikowski, P. Uetz, and S. Fields, *Nat. Biotechnol.* 18 (2000), pp. 1257–1261.
- [87] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, *Nat. Biotechnol.* 21(6) (2003), pp. 697–700.
- [88] K. Takemoto and C. Oosawa, *Phys. Rev. E* 72(4) (2005), p. 46116.
- [89] E. Eisenberg and E.Y. Levanon, *Phys. Rev. Lett.* 91(13) (2003), p. 138701.
- [90] A. Wagner, *Molecular Biology and Evolution* 18 (2001), pp. 1283–1292.
- [91] A. Vázquez, A. Flammini, A. Maritan, and A. Vespignani, *Complexus* 1(1) (2003), pp. 38–44.
- [92] R.V. Solé, R. Pastor-Satorras, E. Smith, and T.B. Kepler, *Adv. Complex Syst.* 5(1) (2002), pp. 43–54.
- [93] R. Pastor-Satorras, E. Smith, and R.V. Solé, *J. Theoret. Biol.* 222 (2002), pp. 199–210.
- [94] R.V. Solé and P. Fernandez, *Modularity “for free” in genome architecture?* Preprint arXiv:q-bio/0312032, 2003.
- [95] R.K. Pan and S. Sinha, *Phys. Rev. E* 76(4) (2007), p. 45103.
- [96] A. Wagner, *Proc. Biol. Sci.* 270(1514) (2003), pp. 457–466.
- [97] J. Berg, M. Lässig, and A. Wagner, *BMC Evol. Biol.* 4(1) (2004), p. 51.
- [98] C. Vogel, S.A. Teichmann, and J. Pereira-Leal, *J. Mol. Biol.* 346(1) (2005), pp. 355–65.
- [99] S.K. Ng, Z. Zhang, and S.H. Tan, *Bioinformatics* 19(8) (2003), pp. 923–929.
- [100] D. Betel, R. Isserlin, and C.W.V. Hogue, *Bioinformatics* 20 Suppl 1 (2004), pp. I55–I62.
- [101] S. Wuchty, *Mol. Biol. Evol.* 18 (2001), pp. 1694–1702.
- [102] S. Wuchty, *Proteomics* 2(12) (2002), pp. 1715–1723.
- [103] F.A. Rodrigues and L.da F. Costa, *Mol. BioSyst.* 6(1) (2009), pp. 234–240.
- [104] F.A. Rodrigues and L.da F. Costa, *Mol. BioSyst.* 5(4) (2009), pp. 385–390.
- [105] T. Manke, L. Demetrius, and M. Vingron, *J. Roy. Soc. Interface* 3(11) (2006), pp. 843–850.
- [106] G. Bagler and S. Sinha, *Bioinformatics* 23(14) (2007), p. 1760.
- [107] H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, and A.-L. Barabási, *Nature* 407(6804) (2000), pp. 651–654.
- [108] A. Wagner, *Proc. Biol. Sci.* 268(1478) (2001), pp. 1803–1810.
- [109] J.S. Edwards and B.O. Palsson, *The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities*, 2000.
- [110] E. Ravasz, A.L. Somera, D.A. Mongru, Z.N. Oltvai, and A.-L. Barabási, *Science* 297(5586) (2002), pp. 1551–1555.
- [111] L. Diambra and L.da F. Costa, *Bioinformatics* 21(20) (2005), pp. 3846–3851.
- [112] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall/CRC, Boca Raton, FL, 2006.
- [113] A. Blais and B.D. Dynlacht, *Genes Dev.* 19(13) (2005), pp. 1499–1511.
- [114] E.D. Siggia, *Curr. Opin. Genet. Dev.* 15 (2005), pp. 214–221.
- [115] S. Shen-Orr, R. Milo, S. Mangan, and U. Alon, *Nat. Genet.* 31(1) (2002), pp. 64–68.
- [116] M.M. Babu, N.M. Luscombe, L. Aravind, M. Gerstein, and S.A. Teichmann, *Curr. Opin. Struct. Biol.* 14(283) (2004), p. 15193307.
- [117] D. Balcan, A. Kabakçioğlu, M. Mungan, and A. Erzan, *PLoS ONE* 2(6) (2007), p. e501.
- [118] 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, J. Zeitlinger, E.G. Jennings, H.L. Murray, D. Benjamin Gordon, B. Ren, J.J. Wyrick, J.-B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford, and R.A. Young, *Science* 298(5594) (2002), pp. 799–804.
- [119] N. Guelzim, S. Bottani, P. Bourguin, and F. Képès, *Nat. Genet.* 31 (2002), pp. 60–63.
- [120] H.W. Ma, J. Buer, and A.P. Zeng, *BMC Bioinformatics* 5 (2004), p. 199.

- [121] R. Dobrin, Q.K. Beg, A.-L. Barabási, and Z.N. Oltvai, *BMC Bioinformatics* 5(1) (2004), p. 10.
- [122] H. Yu, D. Greenbaum, H. Xin Lu, X. Zhu, and M. Gerstein, *Trends Genet.* 20(6) (2004), pp. 227–231.
- [123] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, *Proc. Natl Acad. Sci. USA* 101(14) (2004), pp. 4781–4786.
- [124] K. Klemm and S. Bornholdt, *Proc. Natl Acad. Sci. USA* 102(51) (2005), pp. 18414–18419.
- [125] S. Bornholdt, *Science* 310(5747) (2005), pp. 449–451.
- [126] S.A. Teichmann and M.M. Babu, *Nat. Genet.* 36(5) (2004), pp. 492–496.
- [127] M.B. Elowitz and S. Leibler, *Nature* 403(6767) (2000), pp. 335–338.
- [128] S.A. Kauffman, *J. Theor. Biol.* 22(3) (1969), pp. 437–467.
- [129] S. Bornholdt, *Biol. Chem.* 382(9) (2001), pp. 1289–1300.
- [130] B. Drossel, *Random boolean networks*, in *Reviews of Nonlinear Dynamics and Complexity, I*, H.G. Schuster, ed., chapter 3. Wiley-VCH, Weinheim, Germany, 2009.
- [131] N.S. Holter, M. Mitra, A. Maritan, M. Cieplak, J.R. Banavar, and N.V. Fedoroff, *Proc. Natl Acad. Sci. USA* 97(15) (2000), pp. 8409–8414.
- [132] O. Alter, P.O. Brown, and D. Botstein, *Proc. Natl Acad. Sci. USA* 97(18) (2000), pp. 10101–10106.
- [133] D.S. Lee, J. Park, K.A. Kay, N.A. Christakis, Z.N. Oltvai, and A.-L. Barabási, *Proc. Natl Acad. Sci.* 105(29) (2008), pp. 9880–9885.
- [134] A.-L. Barabási, *New Engl. J. Med.* 357(4) (2007), pp. 404–407.
- [135] L.L. Chen, N. Blumm, N.A. Christakis, A.-L. Barabási, and T.S. Deisboeck, *Br. J. Cancer* 101(5) (2009), pp. 749–758.
- [136] L.da F. Costa, M.P. Viana, and M.E. Beletti, *Appl. Phys. Lett.* 88 (2006), p. 033903.
- [137] 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/0506019, 2005.
- [138] D.D. Bonchev and D. Rouvray, *Complexity in Chemistry, Biology, and Ecology*, Springer Science+Business Media, New York, 2005.
- [139] B. Drossel and A. McKane, *Handbook of Graphs and Networks: From the Genome to the Internet*, chapter Modelling food webs. Wiley-VCH GmbH & Co. KGaA, Weinheim, 2003.
- [140] J.E. Cohen, F. Briand, and C.M. Newman, *Community food webs*, Springer-Verlag, New York, 1990.
- [141] P.A. Murtaugh and J.P. Kollath, *Ecology* 78(5) (1997), pp. 1382–1387.
- [142] R.J. Williams, E.L. Berlow, J.A. Dunne, A.-L. Barabási, and N.D. Martinez, *Proc. Natl Acad. Sci. USA* 99(20) (2002), pp. 12913–12916.
- [143] I. Hanski, *Nature* 396(6706) (1998), pp. 41–49.
- [144] V. Vuorinen, M. Peltomäki, M. Rost, and M.J. Alava, *Eur. Phys. J. B* 38(2) (2004), pp. 261–268.
- [145] P.R. Guimarães Jr, M.A. Menezes, R.W. Baird, D. Lusseau, P. Guimarães, and S.F. dos Reis, *Phys. Rev. E* 76(4) (2007), p. 42901.
- [146] J. Camacho, R. Guimerá, and L.A.N. Amaral, *Phys. Rev. Lett.* 88(22) (2002), p. 228102.
- [147] C.C. Wilmers, *Trends Ecol. Evol.* 22(10) (2007), pp. 504–506.
- [148] S. Sinha, *Physica A* 346(1–2) (2005), pp. 147–153.
- [149] S. Sinha and S. Sinha, *Phys. Rev. E* 71(2) (2005), p. 20902.
- [150] D.H. Boucher, *The Biology of Mutualism: Ecology and Evolution*, Oxford University Press, Oxford, 1988.
- [151] E.L. Rezende, J.E. Lavabre, P.R. Guimarães, P. Jordano, and J. Bascompte, *Nature* 448(7156) (2007), pp. 925–928.
- [152] P.R. Guimarães Jr, V. Rico-Gray, S. Furtado dos Reis, and J.N. Thompson, *Proc. Roy. Soc. B* 273(1597) (2006), pp. 2041–2047.
- [153] J. Bascompte, P. Jordano, C.J. Melián, and J.M. Olesen, *Proc. Natl Acad. Sci. USA* 100(16) (2003), pp. 9383–9387.
- [154] P.R. Guimarães Jr, C. Sazima, S.F. Reis, and I. Sazima, *Biol. Lett.* 3(1) (2007), pp. 51–54.
- [155] L.da F. Costa, *Avalanches of activation and spikes in neuronal complex networks*, Preprint arXiv:0802.0421, 2008.
- [156] J.M. Beggs and D. Plenz, *J. Neurosci.* 23(35) (2003), p. 11167.
- [157] O. Sporns, *Complexity* 8(1) (2002), pp. 56–60.

- [158] L.da F. Costa, *Morphological complex networks: Can individual morphology determine the general connectivity and dynamics of networks?* Preprint arXiv:q-bio/0503041, 2005.
- [159] E. Bullmore and O. Sporns, *Nat. Rev. Neurosci.* 10(3) (2009), pp. 186–198.
- [160] C. Koch and G. Laurent, *Science* 284(5411) (1999), pp. 96–98.
- [161] M. Hämmäläinen, R. Hari, R.J. Ilmoniemi, J. Knuutila, and O.V. Lounasmaa, *Rev. Mod. Phys.* 65(2) (1993), pp. 413–497.
- [162] N.K. Logothetis, J. Pauls, M. Augath, T. Trinath, and A. Oeltermann, *Nature* 412(6843) (2001), pp. 150–157.
- [163] L.da F. Costa and L. Diambra, *Phys. Rev. E* 71(2) (2005), p. 21901.
- [164] L.da F. Costa, M.S. Barbosa, V. Coupez, and D. Stauffer, *Brain and Mind* 4(1) (2003), pp. 91–105.
- [165] D.J. Felleman and D.C. Van Essen, *Cereb. Cortex* 1(1) (1991), pp. 1–47.
- [166] C.C. Hilgetag, G.A. Burns, M.A. O’Neill, J.W. Scannell, and M.P. Young, *Phil. Trans. Biol. Sci.* 355(1393) (2000), pp. 91–110.
- [167] G.A. Burns and M.P. Young, *Phil. Trans. Biol. Sci.* 355(1393) (2000), pp. 55–70.
- [168] J.W. Scannell, C. Blakemore, and M.P. Young, *J. Neurosci.* 15(2) (1995), pp. 1463–1483.
- [169] P. Hagmann, L. Cammoun, X. Gigandet, R. Meuli, C.J. Honey, V.J. Wedeen, and O. Sporns, *PLoS Biol.* 6(7) (2008), p. e159.
- [170] L.da F. Costa and O. Sporns, *Eur. Phys. J. B* 48(4) (2005), pp. 567–573.
- [171] L.da F. Costa, M. Kaiser, and C.C. Hilgetag, *BMC Syst. Biol.* 1 (2007), p. 16.
- [172] L.da F. Costa and O. Sporns, *Int. J. Bifurc. Chaos Appl. Sci. Eng.* 17(7) (2007), pp. 2387–2398.
- [173] L.da F. Costa and O. Sporns, *Appl. Phys. Lett.* 89 (2006), p. 013903.
- [174] M. Barahona and L.M. Pecora, *Phys. Rev. Lett.* 89(5) (2002), p. 054101.
- [175] H. Hong, M.Y. Choi, and B.J. Kim, *Phys. Rev. E* 65(2) (2002), p. 026139.
- [176] L.F. Lago-Fernández, R. Huerta, F. Corbacho, and J.A. Sigüenza, *Phys. Rev. Lett.* 84(12) (2000), pp. 2758–2761.
- [177] B. Percha, R. Dzakpasu, M. Żochowski, and J. Parent, *Phys. Rev. E* 72(3) (2005), p. 31909.
- [178] S. Sinha, J. Saramäki, and K. Kaski, *Phys. Rev. E* 76(1) (2007), p. 15101.
- [179] S. Nadkarni and P. Jung, *Phys. Rev. Lett.* 91(26) (2003), p. 268101.
- [180] N. Chatterjee and S. Sinha, *Progr. Brain Res.* 168 (2007), pp. 145–53.
- [181] M.I. Rabinovich, P. Varona, A.I. Selverston, and H.D.I. Abarbanel, *Rev. Mod. Phys.* 78(4) (2006), pp. 1213–1265.
- [182] A.M. Aertsen, G.L. Gerstein, M.K. Habib, and G. Palm, *J. Neurophysiol.* 61(5) (1989), pp. 900–917.
- [183] F. De Vico Fallani, L. Astolfi, F. Cincotti, D. Mattia, M.G. Marciani, A. Tocci, S. Salinari, H. Witte, W. Hesse, S. Gao, A. Colosimo, and F. Babiloni, *Neuroinformatics* 6(1) (2008), pp. 23–34.
- [184] F. De Vico Fallani, L. Astolfi, F. Cincotti, D. Mattia, M.G. Marciani, S. Salinari, J. Kurths, S. Gao, A. Cichocki, A. Colosimo, and F. Babiloni, *Human Brain Mapping* 28(12) (2007), pp. 1334–46.
- [185] V.M. Eguíluz, D.R. Chialvo, G.A. Cecchi, M. Baliki, and A.V. Apkarian, *Phys. Rev. Lett.* 94(1) (2005), p. 18102.
- [186] R. Salvador, J. Suckling, C. Schwarzbauer, and E. Bullmore, *Phil. Trans. Roy. Soc. B* 360(1457) (2005), pp. 937–946.
- [187] S. Achard, R. Salvador, B. Whitcher, J. Suckling, and E. Bullmore, *J. Neurosci.* 26(1) (2006), p. 63.
- [188] C.J. Stam, B.F. Jones, G. Nolte, M. Breakspear, and P. Scheltens, *Cereb. Cortex* 17(1) (2007), p. 92.
- [189] 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, *Clin. Neurophysiol.* 117(9) (2006), pp. 2039–2049.
- [190] 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.
- [191] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge University Press, Cambridge, 2004.
- [192] G. Caldarelli, R. Marchetti, and L. Pietronero, *Europhys. Lett.* 52(4) (2000), pp. 386–391.
- [193] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, *Phys. Rev. Lett.* 87(25) (2001), p. 258701.
- [194] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. E* 65 (2002), p. 066130.

- [195] J.C. Doyle, D. Alderson, L. Li, S. Low, M. Roughan, S. Shalunov, R. Tanaka, and W. Willinger, *Proceedings of the National Academy of Sciences* 102(41) (2005), pp. 14497–14502.
- [196] P. Holme, B.J. Kim, C.N. Yoon, and S.K. Han, *Phys. Rev. E* 65(5) (2002), p. 056109.
- [197] V. Latora and M. Marchiori, *Phys. Rev. Lett.* 87(19) (2001), p. 198701.
- [198] M.A. Menezes and A.-L. Barabási, *Phys. Rev. Lett.* 92 (2004), p. 028701.
- [199] M.A. Menezes and A.-L. Barabási, *Phys. Rev. Lett.* 93 (2004), p. 068701.
- [200] J. Duch and A. Arenas, *Phys. Rev. Lett.* 96(21) (2006), p. 218702.
- [201] K. Fukuda, L.A.N. Amaral, and H.E. Stanley, *Europhys. Lett.* 62(2) (2003), pp. 189–195.
- [202] V. Rosato, L. Issacharoff, S. Meloni, D. Caligiore, and F. Tiriticco, *Physica A* 387(7) (2008), pp. 1689–1704.
- [203] S. Valverde and R.V. Solé, *Eur. Phys. J. B* 38(2) (2004), pp. 245–252.
- [204] D.-H. Kim and A.E. Motter, *New J. Phys.* 10 (2008), p. 053022.
- [205] D.-H. Kim and A.E. Motter, *J. Phys. A* 41(22) (2008), p. 224019.
- [206] S. Abe and N. Suzuki, *Europhys. Lett.* 61(6) (2003), pp. 852–855.
- [207] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, *Phys. Rev. E* 70 (2004), p. 056105.
- [208] D. Krioukov, K.C. Claffy, K. Fall, and A. Brady, *ACM SIGCOMM Comput. Commun. Rev. (CCR)* 37(3) (2007), pp. 41–52.
- [209] M. Boguñá, D. Krioukov, and K.C. Claffy, *Nature Physics* 5(1) (2009), pp. 74–80.
- [210] B. Tadić, G.J. Rodgers, and S. Thurner, *Int. J. Bifurc. Chaos* 17(7) (2007), pp. 2363–2385.
- [211] H.P. Thadakamalla, R. Albert, and S.R.T. Kumara, *New J. Phys.* 9(6) (2007), p. 190.
- [212] R. Percacci and A. Vespignani, *Eur. Phys. J. B* 32(4) (2003), pp. 411–414.
- [213] G.-Q. Zhang, G.-Q. Zhang, Q.-F. Yang, S.-Q. Cheng, and T. Zhou, *New J. Phys.* 10(12) (2008), p. 123027.
- [214] G. Bianconi, G. Caldarelli, and A. Capocci, *Phys. Rev. E* 71(6) (2005), p. 66116.
- [215] S.-H. Yook, H. Jeong, and A.-L. Barabási, *Proc. Natl Acad. Sci. USA* 99(21) (2002), pp. 13382–13386.
- [216] J. Park and M.E.J. Newman, *Phys. Rev. E* 68 (2003), pp. 026112.
- [217] S. Maslov, K. Sneppen, and A. Zaliznyak, *Physica A* 333 (2004), pp. 529–540.
- [218] V. Rosato and F. Tiriticco, *Europhys. Lett.* 66(4) (2004), pp. 471–477.
- [219] P. Holme and B.J. Kim, *Phys. Rev. E* 65 (2002), p. 026107.
- [220] S. Zhou and R.J. Mondragón, *Phys. Rev. E* 70(6) (2004), p. 066108.
- [221] S. Zhou, *Phys. Rev. E* 74(1) (2006), p. 016124.
- [222] S. Zhou, *BT Technol. J.* 24(3) (2006), pp. 108–115.
- [223] M.A. Serrano, M. Boguñá, and A. Díaz-Guilera, *Eur. Phys. J. B* 50(1–2) (2006), pp. 249–254.
- [224] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, *Proc. Natl Acad. Sci.* 101(11) (2004), pp. 3747–52.
- [225] P. Holme, J. Karlin, and S. Forrest, *ACM SIGCOMM Comput. Commun. Rev.* 38(3) (2008), pp. 5–16.
- [226] F.A. Rodrigues, P.R. Villas Boas, G. Travieso, and L.da F. Costa, *Seeking the best Internet model*, Preprint arXiv:0706.3225, 2007.
- [227] M. Piraveenan, M. Prokopenko, and A.Y. Zomaya, *Eur. Phys. J. B* 70(2) (2009), pp. 275–285.
- [228] J. Leskovec, D. Chakrabarti, J.M. Kleinberg, C. Faloutsos, and Z. Ghahramani, *J. Mach. Learn. Res.* 11 (2010), pp. 985–1042.
- [229] F. Viger, A. Barrat, L. Dall’Asta, C.-H. Zhang, and E.D. Kolaczyk, *Phys. Rev. E* 75(5) (2007), p. 056111.
- [230] C. Magnien, F. Ouédraogo, G. Valadon, and M. Latapy, *Fast dynamics in Internet topology: observations and first explanations*, *Proceedings of the Fourth International Conference on Internet Monitoring and Protection*, Venice (2009), pp. 137–142.
- [231] J.-L. Guillaume, M. Latapy, and D. Magoni, *Computer Networks* 50(16) (2006), pp. 3197–3224.
- [232] L. Dall’Asta, I. Alvarez-Hamelin, A. Barrat, A. Vázquez, and A. Vespignani, *Phys. Rev. E* 71(3) (2005), p. 036135.
- [233] V. Latora and M. Marchiori, *Chaos, Solitons and Fractals* 20(1) (2004), p. 69–75.
- [234] P. Baldi, P. Frascioni, and P. Smyth, *Modeling the Internet and the Web: Probabilistic Methods and Algorithms*, John Wiley & Sons, Chichester, 2003.

- [235] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt, and D.J. Weitzner, *Science* 313(5788) (2006), pp. 769–771.
- [236] S. Lawrence and C.L. Giles, *Nature* 400 (1999), pp. 107–109.
- [237] Z. Dezső, E. Almaas, A. Lukács, B. Rácz, I. Szakadát, and A.-L. Barabási, *Phys. Rev. E* 73(6) (2006), p. 066132.
- [238] B.A. Huberman, P.L.T. Pirolli, J.E. Pitkow, and R.M. Lukose, *Science* 280 (1998), pp. 95–97.
- [239] S. Brin and L. Page, *Comput. Networks ISDN Syst.* 30 (1998), pp. 107–117.
- [240] M. Thelwall, *Cybermetrics* 6/7(1) (2002/3), p. 2.
- [241] J.M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, *Lect. Notes Comput. Sci.* 1627 (1999), pp. 1–17.
- [242] S.R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, *Extracting large-scale knowledge bases from the web*, Proceedings of the 25th VLDB Conference, pp. 639–650, Edinburgh, 1999.
- [243] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal, *Stochastic models for the Web graph*, Proceedings of the 42st Annual IEEE Symposium on the Foundations of Computer Science, Redondo Beach, CA, USA, 2000, pp. 57–65.
- [244] S. Bornholdt and H. Ebel, *Phys. Rev. E* 64 (2001), p. 035104.
- [245] A.-L. Barabási and R. Albert, *Science* 286 (1999), pp. 509–512.
- [246] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, *Comput. Networks* 33 (2000), pp. 309–320.
- [247] B. Tadić, *Physica A* 293 (2001), pp. 273–284.
- [248] D.M. Pennock, G.W. Flake, S. Lawrence, E.J. Glover, and C.L. Giles, *Proc. Natl Acad. Sci. USA* 99(8) (2002), pp. 5207–5211.
- [249] F. Menczer, *Eur. Phys. J. B* 38(2) (2004), pp. 211–221.
- [250] J.S. Kong and V.P. Roychowdhury, *Physica A* 387(13) (2008), pp. 3335–3347.
- [251] J.S. Kong, N. Sarshar, and V.P. Roychowdhury, *Proc. Natl Acad. Sci. USA* 105(37) (2008), pp. 13724–13729.
- [252] A. Barrat, M. Barthélémy, and A. Vespignani, *Lect. Notes Comput. Sci.* 3243 (2004), pp. 56–67.
- [253] P.L. Krapivsky, G.J. Rodgers, and S. Redner, *Phys. Rev. Lett.* 86(23) (2001), pp. 5401–5404.
- [254] M.A. Serrano, A. Maguitman, M. Boguñá, S. Fortunato, and A. Vespignani, *ACM Trans. Web* 1(2) (2007), p. 10.
- [255] A.-L. Barabási, R. Albert, and H. Jeong, *Physica A* 281 (2000), pp. 69–77.
- [256] D. Donato, L. Laura, S. Leonardi, and S. Millozzi, *Eur. Phys. J. B* 38(2) (2004), pp. 239–243.
- [257] L.A. Adamic, *The small world web*, Proceedings of the 3rd European Conf. Research and Advanced Technology for Digital Libraries, ECDL, Lecture Notes in Computer Science, Springer Berlin/Heidelberg, 1999, pp. 443–452.
- [258] P. Ormerod and A.P. Roach, *Physica A* 339 (2004), pp. 645–652.
- [259] P. Ball, *Nature* 415 (2002), p. 371.
- [260] B. Wellman, *J. Educ. Res.* 14 (1926), pp. 126–132.
- [261] H.A. Simon, *Biometrika* 42(3/4) (1955), pp. 425–440.
- [262] D.J.S. Price, *Science* 149 (1965), pp. 510–515.
- [263] D.J.S. Price, *J. Amer. Soc. Inform. Sci.* 27(5) (1976), pp. 292–306.
- [264] L.C. Freeman, *Sociometry* 40 (1977), pp. 35–41.
- [265] L.C. Freeman, *Social Networks* 1 (1979), pp. 215–239.
- [266] S. Milgram, *Psychology Today* 2 (1967), pp. 60–67.
- [267] J. Travers and S. Milgram, *Sociometry* 32 (1969), pp. 425–443.
- [268] D.J. Watts, *Six Degrees: The Science of a Connected Age*, W.W. Norton and Company, New York, 2003.
- [269] A. Degenne and M. Forsé, *Introducing Social Networks*, Sage Publications Ltd, London, 1999.
- [270] P.J. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network Analysis*, Cambridge University Press, New York, 2005.
- [271] S.P. Borgatti, A. Mehra, D.J. Brass, and G. Labianca, *Science* 323 (2009), pp. 892–895.
- [272] C. Castellano, S. Fortunato, and V. Loreto, *Rev. Mod. Phys.* 81(2) (2009), pp. 591–646.

- [273] F. Liljeros, C.R. Edling, L.A.N. Amaral, H.E. Stanley, and Y. Aaberg, *Nature* 411(6840) (2001), pp. 907–908.
- [274] P. Cano, O. Celma, M. Koppenberger, and J.M. Buldú, *Chaos* 16 (2006), p. 013107.
- [275] T. Evans, C. Knappett and R.J. Rivers, *Physical and relational networks in the aegean bronze age*, European Conference of Complex Systems – ECCS’06, 2006.
- [276] C. Knappett, T. Evans, and R. Rivers, *Antiquity* 82(318) (2008), pp. 1009–1024.
- [277] D.J. Watts, *Nature* 445(7127) (2007), p. 489.
- [278] B. Wellman, *Science* 293 (2001), pp. 2031–2034.
- [279] J. Golbeck, *Science* 321 (2008), pp. 1640–1641.
- [280] M.C. González and A.-L. Barabási, *Nature* 3 (2007), pp. 224–225.
- [281] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, *Phys. Rev. E* 70(5) (2004), p. 56122.
- [282] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, *Phys. Rev. E* 68(6) (2003), p. 065103.
- [283] W. Bachnik, S. Szymczyk, P. Leszczynski, R. Podsiadlo, E. Rymaszewicz, L. Kurylo, D. Makowiec, and B. Bykowska, *Acta Phys. Polon. B* 36(10) (2005), pp. 3179–3191.
- [284] P. Zakharov, *Physica A* 378(2) (2007), pp. 550–560.
- [285] 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.
- [286] D. Lusseau, *Proc. Roy. Soc. B* 270 (2003), pp. 186–188.
- [287] P.S. Dodds, R. Muhamad, and D.J. Watts, *Science* 301(8) (2003), pp. 827–829.
- [288] D. Liben-Nowell and J.M. Kleinberg, *Proc. Natl Acad. Sci. USA* 105(12) (2008), pp. 46333–4638.
- [289] M.E.J. Newman, S. Forrest, and J. Balthrop, *Phys. Rev. E* 66(3) (2002), p. 035101.
- [290] M. Szell, R. Lambiotte, and S. Thurner, *Proc. Natl Acad. Sci. USA* 107(31) (2010), pp. 13636–13641.
- [291] D.M. Boyd and N.B. Ellison, *J. Computer-Mediated Commun.* 13 (2007), p. 11.
- [292] Y.Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong, *Analysis of topological characteristics of huge online social networking services*, Proceedings of the 16th International Conference on World Wide Web, ACM Press, 2007.
- [293] H. Hu and X. Wang, *Phys. Lett. A* 373 (2009), pp. 1105–1110.
- [294] H. Hu and X. Wang, *Europhys. Lett.* 86 (2009), p. 18003.
- [295] S. Han and B.J. Kim, *Physica A* 387 (2008), pp. 5946–5951.
- [296] T. Hogg and G. Szabo, *Europhys. Lett.* 86 (2009), p. 38003.
- [297] P. Bearman, J. Moody, and K. Stovel, *Amer. J. Sociol.* 110 (2004), pp. 44–91.
- [298] M.C. González, P.G. Lind, and H.J. Herrmann, *Eur. Phys. J. B* 49(3) (2006), pp. 371–376.
- [299] B.F. Blasio, A. Svensson, and F. Liljeros, *Proc. Natl Acad. Sci. USA* 104(26) (2007), pp. 10762–10767.
- [300] T. Britton, M.K. Nordvik, and F. Liljeros, *Theor. Popul. Biol.* 72(3) (2007), pp. 389–399.
- [301] L.E.C. Rocha, F. Liljeros, and P. Holme, *Proc. Natl Acad. Sci. USA* 107(13) (2010), pp. 5706–5711.
- [302] J.H. Jones and M.S. Handcock, *Proceedings of the Royal Society B* 270(1520) (2003), pp. 1123–1128.
- [303] J.H. Jones and M.S. Handcock, *Nature* 423 (2003), pp. 605–606.
- [304] F. Liljeros, *Physica A* 338 (2004), pp. 238–245.
- [305] C. Fichtenberg, J. Jennings, T. Glass, and J. Ellen, *J. Urban Health* 87(2) (2010), pp. 225–235.
- [306] S. Lozano, A. Arenas, and A. Sanchez, *PLoS ONE* 3(4) (2008), p. e1892.
- [307] J.-C. Wang and C.-C. Chiu, *Expert Systems with Applications* 34(3) (2008), pp. 1666–1679.
- [308] W. Yuan, D. Guan, Y.-K. Lee, S. Lee, and S. Jin Hur, *Knowledge-Based Systems* 23(3) (2010), pp. 232–238.
- [309] L.A.N. Amaral, A. Scala, M. Barthélemy, and H.E. Stanley, *Proc. Natl Acad. Sci. USA* 97(21) (2000), pp. 11149–11152.
- [310] 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, *Physica A* 360(2) (2006), pp. 599–616.
- [311] Y.-H. Eom, C. Jeon, H. Jeong, and B. Kahng, *Phys. Rev. E* 77(5) (2008), p. 056105.
- [312] J. Park and M.E.J. Newman, *J. Statist. Mech.* 10 (2005), p. P10014.
- [313] J. Moody, *American Sociological Review* 69 (2004), pp. 213–238.
- [314] A.-L. Barabási, H. Jeong, Z. Néda, E. Ravasz, A. Schubert, and T. Vicsek, *Physica A* 311(3–4) (2002), pp. 590–614.
- [315] M.E.J. Newman, *Phys. Rev. E* 64(1) (2001), p. 16131.



- [316] A. Arenas, L. Danon, A. Díaz-Guilera, P.M. Gleiser, and R. Guimerà, *Eur. Phys. J. B* 38(2) (2004), pp. 373–380.
- [317] M. Li, Y. Fan, J. Chen, L. Gao, Z. Di, and J. Wu, *Physica A* 350 (2005), pp. 643–656.
- [318] M.E.J. Newman, *Phys. Rev. E* 74(3) (2006), pp. 036104.
- [319] K. Börner, L. Dall'Asta, W. Ke, and A. Vespignani, *Complexity* 10(4) (2005), pp. 57–67.
- [320] M. Ausloos and R. Lambiotte, *Eur. Phys. J. B* 57(1) (2007), pp. 89–94.
- [321] C. Cotta and J.J. Merelo, *The complex network of evolutionary computation authors: an initial study*, Preprint arXiv:physics/0507196, 2005.
- [322] G. Palla, A.-L. Barabási, and T. Vicsek, *Nature* 446(7136) (2007), pp. 664–667.
- [323] M.E.J. Newman, *Phys. Rev. E* 64(1) (2001), p. 16132.
- [324] Y. Fan, M. Li, J. Chen, L. Gao, Z. Di, and J. Wu, *Int. J. Mod. Phys. B* 18(17–19) (2004), pp. 2505–2512.
- [325] M. Wang, G. Yu, and D. Yu, *Physica A* 387 (2008), pp. 4692–4698.
- [326] M. Wang, G. Yu, and D. Yu, *Physica A* 388(19) (2009), pp. 4273–4276.
- [327] L. Bornmann and H.-D. Daniel, *J. Documentation* 64 (2008), pp. 45–80.
- [328] A. Vázquez, *Statistics of citation networks*, Preprint arXiv:cond-mat/0105031, 2001.
- [329] C. Chen and D. Hicks, *Scientometrics* 59(2) (2004), pp. 199–211.
- [330] S. Redner, *Eur. Phys. J. B* 4 (1998), pp. 131–134.
- [331] K.B. Hajra and P. Sen, *Physica A* 346 (2005), pp. 44–48.
- [332] S. Bilke and C. Peterson, *Phys. Rev. E* 64 (2001), p. 036106.
- [333] C. Tsallis and M.P. Albuquerque, *Eur. Phys. J. B* 13 (2000), 777–780.
- [334] F. Radicchi, S. Fortunato, and C. Castellano, *Proc. Natl Acad. Sci. USA* 105(45) (2008), pp. 17268–17272.
- [335] A.F.J. van Raan, *Scientometrics* 63(3) (2005), pp. 549–566.
- [336] I. Hellsten, R. Lambiotte, A. Schamhorst, and M. Ausloos, *Scientometrics* 72 (2007), pp. 469–486.
- [337] M. Ausloos, R. Lambiotte, A. Schamhorst, and I. Hellsten, *Int. J. Mod. Phys. C* 19(3) (2008), pp. 371–384.
- [338] R. Lambiotte and M. Ausloos, *Phys. Rev. E* 72 (2005), p. 066107.
- [339] R. Lambiotte and M. Ausloos, *Eur. Phys. J. B* 50 (2006), pp. 183–188.
- [340] P.M. Gleiser and L. Danon, *Advances in Complex Systems* 6(4) (2003), pp. 565–573.
- [341] R.D. Smith, *J. Statist. Mech.* 2006(2) (2006), p. P02006.
- [342] A. Clauset, M.E.J. Newman, and C. Moore, *Phys. Rev. E* 70 (2004), p. 066111.
- [343] J. Park, O. Celma, M. Koppenberger, P. Cano, and J.M. Buldú, *Int. J. Bifurc. Chaos* 17(7) (2007), pp. 2281–2288.
- [344] B.H. Erickson, *Social Forces* 60 (1981), p. 188.
- [345] M.K. Sparrow, *Social Networks* 13 (1991), p. 251.
- [346] W.E. Baker and R.R. Faulkner, *American Sociological Review* 58 (1993), p. 837.
- [347] V.E. Krebs, *Connections* 24(3) (2002), pp. 43–52.
- [348] V.E. Krebs, *First Monday* 7(4) (2002).
- [349] D. Muhammad and A. Hussain, *J. Softw.* 2(6) (2007), pp. 22–29.
- [350] Y. Maeno and Y. Ohsawa, *Node discovery in a networked organization*, Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, San Antonio, TX, USA, 2009, pp. 3522–3527.
- [351] Y. Maeno and Y. Ohsawa, *Int. J. Services Sci.* 2(2) (2009), pp. 125–141.
- [352] U.K. Will, N. Memon, and P. Karampelas, *Detecting new trends in terrorist networks*, 2010 International Conference on Advances in Social Networks Analysis and Mining, Odense, 2010, pp. 435–440.
- [353] P.S. Keila and D.B. Skillicorn, *Comput. Math. Organ. Theory* 11(3) (2005), pp. 183–199.
- [354] H. Ebel, L.I. Mielsch, and S. Bornholdt, *Phys. Rev. E* 66(3) (2002), p. 035103.
- [355] J.R. Tyler, D.M. Wilkinson, and B.A. Huberman, *Email as spectroscopy: Automated discovery of community structure within organizations*, in *Communities and Technologies*, M. Huysman, E. Wenger, and V. Wulf, eds., Springer, New York, 2003, pp. 81–96.
- [356] J. Wang and P. De Wilde, *Phys. Rev. E* 70(6) (2004), p. 066121.

- [357] P. Pollner, G. Palla, D. Abel, A. Vicsek, I. J. Farkas, I. Derenyi, and T. Vicsek, *Physica A* 387 (2008), pp. 4959–4966.
- [358] D. Braha and Y. Bar-Yam, *Complexity* 12(2) (2006), pp. 59–63.
- [359] C.C. Zou and D.W. Gong, *Proceedings of the 13th International Conference on Computer Communications and Networks*, 2004, pp. 409–414.
- [360] J.P. Eckmann, E. Moses, and D. Sergi, *Proc. Natl Acad. Sci. USA* 101(40) (2004), pp. 14333–14337.
- [361] R.D. Malmgren, D.B. Stouffer, A.E. Motter, and L.A.N. Amaral, *Proc. Natl Acad. Sci. USA* 105(47) (2008), pp. 18153–18158.
- [362] 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.
- [363] J. Abello, P. Pardalos, and M.G.C. Resende, *On maximum clique problems in very large graphs*, *Dimacs Series in Discrete Mathematics and Theoretical Computer Science*, (1999), pp. 119–130.
- [364] 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*, *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*, New York, NY, USA, 2006, ACM, pp. 435–444.
- [365] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M.A. Menezes, K. Kaski, A.-L. Barabási, and J. Kertész, *New J. Phys.* 9(6) (2007), p. 179.
- [366] J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, *Proc. Natl Acad. Sci. USA* 104(18) (2007), pp. 7332–7336.
- [367] M. Granovetter, *Amer. J. Sociol.* 78(6) (1973), p. 1360.
- [368] R. Lambiotte, V.D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. Van Dooren, *Physica A* 387 (2008), pp. 5317–5325.
- [369] V.D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *J. Statist. Mech.* 2008(10) (2008), p. 10008.
- [370] Y. Xia, C.K. Tse, W.M. Tam, F.C.M. Lau, and M. Small, *Phys. Rev. E* 72(2) (2005), p. 026116.
- [371] M.C. González, C.A. Hidalgo, and A.-L. Barabási, *Nature* 453(7196) (2008), pp. 779–782.
- [372] N. Eagle, A.S. Pentland, and D. Lazer, *Proc. Natl Acad. Sci. USA* 106(36) (2009), pp. 15274–15278.
- [373] D. Brockmann, L. Hufnagel, and T. Geisel, *Nature* 439(7075) (2006), pp. 462–465.
- [374] P. Wang, M.C. Gonzalez, C.A. Hidalgo, and A.-L. Barabási, *Science* 324(5930) (2009), pp. 1071–1076.
- [375] J.O. Kephart and S.R. White, *Directed-graph epidemiological models of computer viruses*, *IEEE Computer Society Symposium on Research in Security and Privacy* (1991), pp. 343–359.
- [376] A. Kleczkowski and B.T. Grenfell, *Physica A* 274(1–2) (1999), pp. 355–360.
- [377] C. Moore and M.E.J. Newman, *Phys. Rev. E* 61(5) (2000), pp. 05678–05682.
- [378] M. Kuperman and G. Abramson, *Phys. Rev. Lett.* 86(13) (2001), pp. 2909–2912.
- [379] M. Small, P. Shi, and C.K. Tse, *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* E87-A(9) (2004), pp. 2379–2386.
- [380] O. Shanker and T. Hogg, *Mod. Phys. Lett. B* 23(10) (2009), pp. 1249–1262.
- [381] X. Li and X.A. Wang, *Int. J. Syst. Sci.* 38(5) (2007), pp. 401–411.
- [382] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* 63(6) (2001), pp. 066117–066124.
- [383] M. Boguñá and R. Pastor-Satorras, *Phys. Rev. E* 66(4) (2002), pp. 047104–047107.
- [384] D. Volchenkov, L. Volchenkova, and P. Blanchard, *Phys. Rev. E* 66(4) (2002), pp. 046137–046146.
- [385] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, *Phys. Rev. Lett.* 90(2) (2003), pp. 028701–028703.
- [386] Y. Moreno, R. Pastor-Satorras, and A. Vespignani, *Eur. Phys. J. B* 26(4) (2002), pp. 521–529.
- [387] M.E.J. Newman, *Phys. Rev. E* 66(1) (2002), pp. 016128–016139.
- [388] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* 65(3) (2002), pp. 035108–035111.
- [389] X.-F. Wang and G. Chen, *Circuits Systems Mag. IEEE* 3(1) (2003), pp. 6–20.
- [390] F. Liljeros, C.R. Edling, H.E. Stanley, Y. Aberg, and L.A.N. Amaral, *Nature* 423(6940) (2003), pp. 605–606.
- [391] T. Zhou, Z.-Q. Fu, and B.-H. Wang, *Progr. Nat. Sci.* 16(5) (2006), pp. 452–457.
- [392] Z. Dezsö and A.-L. Barabási, *Phys. Rev. E* 65(5) (2002), pp. 055103–055106.
- [393] R. Pastor-Satorras and A. Vespignani, *Phys. Rev. E* 65(3) (2002), pp. 036104–036111.

- [394] R. Pastor-Satorras and A. Vespignani, *Epidemics and immunization in scale-free networks*, in *Handbook of Graphs and Networks: From the Genome to the Internet*, S. Bornholdt and H.G. Schuster, eds., Wiley-VCH GmbH & Co. KGaA, Weinheim, 2003.
- [395] R. Cohen, S. Havlin, and D. ben Avraham, *Phys. Rev. Lett.* 91(24) (2003), pp. 247901–247904.
- [396] J. Gómez-Gardeñes, P. Echenique, and Y. Moreno, *Eur. Phys. J. B* 49(2) (2006), pp. 259–264.
- [397] W.-P. Guo, X. Li, and X.-F. Wang, *Physica A* 380 (2007), pp. 684–690.
- [398] Y. Hayashi, M. Minoura, and J. Matsukubo, *Phys. Rev. E* 69(1) (2004), pp. 016112–016119.
- [399] M. He, B. Wang, X. Xiao, and L. Sun, *Int. J. Mod. Phys. C* 18(9) (2007), pp. 1453–1458.
- [400] L.A. Meyers, M.E.J. Newman, M. Martin, and S. Schrag, *Emerg. Infect. Dis.* 9(2) (2001), pp. 204–210.
- [401] P.M.A. Sloot, S.V. Ivanov, A.V. Boukhanovsky, D.A.M.C. van de Vijver, and C.A.B. Boucher, *Int. J. Comput. Math.* 85(8) (2008), pp. 1175–1187.
- [402] L.M. Sander, C.P. Warren, I. Sokolov, C. Simon, and J. Koopman, *Math. Biosci.* 180(1) (2002), pp. 293–305.
- [403] L.M. Sander, C.P. Warren, and I.M. Sokolov, *Physica A* 325(1–2) (2003), pp. 1–8.
- [404] M. Small and C.K. Tse, *Physica A* 351(2–4) (2005), pp. 499–411.
- [405] M. Small, C.K. Tse, and D.M. Walker, *Physica D* 215(2) (2006), pp. 146–158.
- [406] M. Small, D.M. Walker, and C.K. Tse, *Phys. Rev. Lett.* 99(18) (2007), p. 188702.
- [407] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical processes on complex networks*, Cambridge University Press, 2008.
- [408] V. Colizza and A. Vespignani, *Phys. Rev. Lett.* 99 (2007), p. 148701.
- [409] V. Colizza, A. Barrat, M. Barthélemy, A.-J. Valleron, and A. Vespignani, *PLoS Med.* 4 (2007), p. e13.
- [410] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, *BMC Medicine* 5(1) (2007), p. 34.
- [411] D. Balcan, V. Colizza, B. Gonçalves, H. Hu, J.J. Ramasco, and A. Vespignani, *Proc. Natl Acad. Sci.* 106 (2009), pp. 21484–21489.
- [412] P. Bajardi, C. Poletto, D. Balcan, H. Hu, B. Gonçalves, J.J. Ramasco, D. Paolotti, N. Perra, M. Tizzoni, W.V. den Broeck, V. Colizza, and A. Vespignani, *Emerging Health Threats J.* 2 (2009), p. e11.
- [413] D. Balcan, B. Gonçalves, H. Hu, J.J. Ramasco, V. Colizza, and A. Vespignani, *J. Comput. Sci.* 1 (2010), p. 132.
- [414] S. Goyal, *Connections: An Introduction to the Economics of Networks*, Princeton University Press, Princeton, NJ, 2007.
- [415] F. Schweitzer, G. Fagiolo, D. Sornette, F. Vega-Redondo, and D.R. White, *Adv. Complex Systems* 12(4) (2009), pp. 407–442.
- [416] M.A. Serrano and M. Boguñá, *Phys. Rev. E* 68(1) (2003), pp. 015101–015104.
- [417] Y.L. Li, J.T. Bi, and H.J. Sun, *Physica A* 387(23) (2008), pp. 5852–5856.
- [418] I. Tzekinaa, K. Danthi, and D.N. Rockmore, *Eur. Phys. J. B* 63 (2008), p. 541.
- [419] G. Fagiolo, J. Reyes, and S. Schiavo, *Phys. Rev. E* 79(3) (2009), p. 036115.
- [420] A. Ausloos and R. Lambiotte, *Physica A* 382(1) (2006), pp. 16–21.
- [421] J. Miśkiewicz and M. Ausloos, *Practical Fruits of Econophysics* 5 (2006), pp. 312–316.
- [422] M. Gligor and A. Ausloos, *Eur. Phys. J. B* 63(4) (2008), pp. 533–539.
- [423] E. Bompard and Y. Ma, *IEEE Transactions on Power Systems* 23(4) (2008), pp. 1590–1600.
- [424] X. Li, Y.Y. Jin, and G. Chen, *Physica A* 343(–) (2004), pp. 573–582.
- [425] A.Z. Górski, S. Drożdż, J. Kwapien, and P. Oświęcimka, *Acta Phys. Polon. B* 37(11) (2006), pp. 2897–2906.
- [426] M.J. Naylor, L.C. Rose, and B.J. Moyle, *Physica A* 382(1) (2007), pp. 199–208.
- [427] J.S. Andrade Jr, D.M. Bezerra, J. Ribeiro Filho, and A.A. Moreira, *Physica A* 360(2) (2006), pp. 637–643.
- [428] M. Ferrary and M. Granovetter, *Economy and Society* 38(2) (2009), pp. 326–359.
- [429] Z.-Q. Jiang, W.-X. Zhou, B. Xu, and W.-K. Yuan, *AIChE J.* 53(2) (2007), pp. 423–428.
- [430] W. Souma, Y. Fujiwara, and H. Aoyama, *Small-world effects in wealth distribution*, Preprint arXiv:cond-mat/0108482, 2001.
- [431] T. Di Matteo, T. Aste, and S.T. Hyde, *Exchanges in complex networks: income and wealth distributions*, in *The Physics of Complex Systems (New Advances and Perspectives)*, F. Mallamace and H.E. Stanley, eds, IOS Press, Amsterdam, 2004, pp. 435–442.

- [432] M.-B. Hu, R. Jiang, Y.-H. Wu, R. Wang, and Q.-S. Wu, *Physica A* 387(23) (2008), pp. 5862–5867.
- [433] R. Baggio, *Information Technology and Tourism* 8(1) (2006), pp. 15–29.
- [434] R. Baggio, *Tourism Analysis* 13(1) (2008), pp. 1–20.
- [435] N. Scott, C. Cooper, and R. Baggio, *Use of network analysis in tourism research*, Proceedings of the Advances in Tourism Marketing Conference (ATMC), Valencia, Spain, September 2007.
- [436] N. Scott, C. Cooper, and R. Baggio, *Ann. Tourism Res.* 35(1) (2008), pp. 169–188.
- [437] R. Baggio and M.A. Corigliano, *Tourism destinations: a network analysis of the Web space*, Proceedings of the Advances in Tourism Marketing Conference (ATMC), Valencia, Spain, September 2007.
- [438] N. Scott, C. Cooper, and R. Baggio, *Network Analysis and Tourism: From Theory to Practice*, Channel View, Clevedon, UK, 2008.
- [439] L.da F. Costa and R. Baggio, *Physica A* 388(19) (2009), pp. 4286–4296.
- [440] G. Bonanno, G. Caldarelli, F. Lillo, and R.N. Mantegna, *Phys. Rev. E* 68(4) (2003), p. 046130.
- [441] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Phys. Rev. E* 68(5) (2003), p. 056110.
- [442] J.-P. Onnela, A. Chakraborti, K. Kaski, J. Kertész, and A. Kanto, *Physica Scripta* T106 (2003), pp. 48–54.
- [443] J.-P. Onnela, K. Kaski, and J. Kertész, *Eur. Phys. J. B* 38(2) (2004), pp. 353–362.
- [444] C. Eom, G. Oh, and S. Kim, *J. Korean Phys. Soc.* 53(6) (2008), pp. 3837–3841.
- [445] R.K. Pan and S. Sinha, *Phys. Rev. E* 76(4) (2007), p. 046116.
- [446] T. Heimo, J. Saramäki, J.-P. Onnela, and K. Kaski, *Physica A* 383(1) (2007), pp. 147–151.
- [447] T. Heimo, J.M. Kumpula, K. Kaski, and J. Saramaki, *J. Statist. Mech.* 2008(8) (2008), p. P08007.
- [448] A. Garas and P. Argyrakis, *Europhys. Lett.* 86(2) (2009), p. 28005.
- [449] T. Heimo, K. Kaski, and J. Saramäki, *Physica A* 388(2–3) (2009), pp. 145–156.
- [450] M. Bolgarian and R. Raeli, *Physica A* 389(18) (2010), pp. 3822–3827.
- [451] A. Garas, P. Argyrakis, and S. Havlin, *Eur. Phys. J. B* 63(2) (2008), pp. 265–271.
- [452] O. Hein, M. Schwind, and M. Spiwoks, *J. Econom. Interaction and Coordination* 3(1) (2008), pp. 59–71.
- [453] J. Jiang, W. Li, and X. Cai, *Physica A* 387(2–3) (2008), pp. 528–536.
- [454] Y. Kim, H.-J. Kim, and S.-H. Yook, *Phys. Rev. E* 78(3) (2008), p. 036115.
- [455] A.V. Aho and J.D. Ullman, *Foundations of computer science*, Computer Science Press, Inc. New York, NY, USA, 1992.
- [456] P.J. Denning, *Computer science: The discipline*, Encyclopedia of Computer Science, 2000.
- [457] M.A. Bishop and M. Bishop, *Computer Security: Art and Science*, Addison-Wesley Professional, 2003.
- [458] S. Valverde, R. Ferrer i Cancho, and R.V. Solé, *Europhys. Lett.* 60(4) (2002), pp. 512–517.
- [459] A.P.S. de Moura, Y.-C. Lai, and A.E. Motter, *Phys. Rev. E* 68(1) (2003), p. 017102.
- [460] S. Valverde and R.V. Solé, *Hierarchical small worlds in software architecture*, Preprint cond-mat/0307278, 2003.
- [461] H. Zhang, H. Zhao, W. Cai, J. Liu, and W. Zhou, *J. Supercomputing* 53(2) (2010), pp. 352–369.
- [462] N. LaBelle and E. Wallingford, *Inter-package dependency networks in open-source software*, Preprint arXiv:cs/0411096, 2006.
- [463] C.R. Myers, *Phys. Rev. E* 68(4) (2003), pp. 046116–046130.
- [464] D. Challet and A. Lombardoni, *Phys. Rev. E* 70(4) (2004), pp. 046109–046113.
- [465] X. Zheng, D. Zeng, H. Li, and F. Wang, *Physica A* 387(24) (2008), pp. 6190–6200.
- [466] Y. Ma, K. He, and D. Du, *A qualitative method for measuring the structural complexity of software systems based on complex networks*, Software Engineering Conference, 2005. APSEC '05. 12th Asia-Pacific, 2005, pp. 257–263.
- [467] S. Jenkins and S.R. Kirk, *Inform. Sci.* 177(12) (2007), pp. 2587–2601.
- [468] K.-Y. Cai and B.-B. Yin, *Inform. Sci.* 179(12) (2009), pp. 1903–1928.
- [469] A. Iamnitchi, M. Ripeanu, and I. Foster, *Data-sharing relationships in the Web*, Proceedings of the Twelfth International World Wide Web Conference (WWW12), 2003.

- [470] A. Iamnitchi, M. Ripeanu, and I. Foster, *Small-world file-sharing communities*, Proceedings of the Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2004), 2 (2004), pp. 952–963.
- [471] N. Leibowitz, M. Ripeanu, and A. Wierzbicki, *Deconstructing the Kazaa network*, Proceedings of the Third IEEE Workshop on Internet Applications (WIAPP 2003), 2003, pp. 112–120.
- [472] A. Iamnitchi, M. Ripeanu, and I. Foster, *Lect. Notes Comput. Sci.* 2429 (2002), pp. 232–241.
- [473] M. Ripeanu, I. Foster, and A. Iamnitchi, *IEEE Internet Comput.* 6 (2002), pp. 50–57.
- [474] J.-L. Guillaume and S. Le-Blond, *Statistical properties of exchanges in P2P systems*, Proceedings of the International Conference on Parallel and Distributed Processing Techniques and Applications, vol. 3, 2004, pp. 1423–1428,
- [475] R. Ferrer i Cancho, C. Janssen, and R.V. Solé, *Phys. Rev. E* 64(4) (2001), pp. 046119–046123.
- [476] A.-L. Barabási, M.A. Menezes, S. Balensiefer, and J. Brockman, *Eur. Phys. J. B* 38(2) (2004), pp. 169–175.
- [477] C. Teuscher, *On irregular interconnect fabrics for self-assembled nanoscale electronics*, Boston, MA, USA, June 2006. 2nd IEEE International Workshop on Default and Fault Tolerant Nanoscale Architectures, NANOARCH’06.
- [478] R.V. Solé, B. Corominas-Murtra, S. Valverde, and L. Steels, *Complexity* 15(6) (2010), pp. 20–26.
- [479] 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.
- [480] B. Pang and L. Lee, *A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts*, Proceedings of the ACL 2004, pp. 271–278.
- [481] R. Mihalcea, *Random walks on text structures*, Proceedings of CICLing-2006, Lecture Notes in Computer Science, vol. 3878, February 2006, pp. 249–262.
- [482] S.N. Dorogovtsev and J.F.F. Mendes, *Proc. Roy. Soc. Lond. B* 268(1485) (2001), pp. 2603–2606.
- [483] R. Ferrer i Cancho and R.V. Solé, *Proc. Roy. Soc. Lond. B* 268(1482) (2001), pp. 2261–2265.
- [484] R. Ferrer i Cancho, O. Riordan, and B. Bollobás, *Proc. Roy. Soc. B* 272(1562) (2005), pp. 561–565.
- [485] R. Ferrer i Cancho, *BioSystems* 84(3) (2006), pp. 242–253.
- [486] J. Borge-Holthoefer and A. Arenas, *Entropy* 12(5) (2010), pp. 1264–1302.
- [487] V. Loreto and L. Steels, *Nature Physics* 3(11) (2007), pp. 758–760.
- [488] A.J. Holanda, I.T. Pisa, O. Kinouchi, A.S. Martinez, and E.E.S. Ruiz, *Physica A* 344 (2004), pp. 530–536.
- [489] M. Steyvers and J.B. Tenenbaum, *Cogn. Sci.* 29 (2005), pp. 41–78.
- [490] A.E. Motter, A.P.S. Moura, Y.-C. Lai, and P. Dasgupta, *Phys. Rev. E* 65 (2002), p. 065102.
- [491] D. Storri, A. Bombaci, and H. Bingol, *Cross comparison of synonym graphs in a multi linguistic context*, 22nd International Symposium on Computer and Information Sciences, Ankara, Turkey 2007, pp. 1–7.
- [492] M. Sigman and G.A. Cecchi, *Proc. Natl Acad. Sci. USA* 99(3) (2002), pp. 1742–1747.
- [493] M.L. Doerfel and S.L. Connaughton, *J. Amer. Soc. Inform. Sci. Technol.* 60(1) (2009), pp. 201–218.
- [494] L.da F. Costa, *Int. J. Mod. Phys. C* 15 (2004), pp. 371–379.
- [495] S.M.G. Caldeira, T.C. Petit Lobão, R.F.S. Andrade, A. Neme, and J.G.V. Miranda, *Eur. Phys. J. B* 49 (2006), pp. 523–529.
- [496] A.P. Masucci and G.J. Rodgers, *Phys. Rev. E* 74 (2006), p. 026102.
- [497] S. Zhou, G. Hu, Z. Zhang, and J. Guan, *Physica A* 387(12) (2008), pp. 3039–3047.
- [498] C. Bedogne’ and G.J. Rodgers, *Physica A* 387(27) (2008), pp. 6863–6868.
- [499] A.P. Masucci and G.J. Rodgers, *Adv. Complex Syst.* 12(1) (2009), pp. 113–129.
- [500] L. Sheng and C. Li, *Physica A* 388(12) (2009), pp. 2561–2570.
- [501] R.M. Roxas and G. Tapang, *Int. J. Mod. Phys. C* 21(4) (2010), pp. 503–512.
- [502] R. Ferrer i Cancho, R.V. Solé, and R. Köhler, *Phys. Rev. E* 69 (2004), p. 051915.
- [503] R. Ferrer i Cancho, A. Capocci, and G. Caldarelli, *Int. J. Bifurc. Chaos* 17(7) (2007), pp. 2453–2463.
- [504] H. Liu, *Physica A* 387(12) (2008), pp. 3048–3058.
- [505] H. Liu and F. Hu, *Europhys. Lett.* 83(1) (2008), p. 18002.
- [506] B. Corominas-Murtra, S. Valverde, and R.V. Solé, *Adv. Complex Syst.* 12(3) (2009), pp. 371–392.
- [507] J. Ke and Y. Yao, *J. Quant. Linguist.* 15(1) (2008), pp. 70–99.
- [508] S. Arbesman, S.H. Strogatz, and M.S. Vitevitch, *Int.J. Bifurc. Chaos* 20(3) (2010), pp. 679–685.

- [509] M. Markošová, *Physica A* 387(2–3) (2008), pp. 661–666.
- [510] P. Edmonds, *Choosing the word most typical in context using a lexical co-occurrence network*, Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics, 1997, pp. 507–509.
- [511] L. Antiqueira, M.G.V. Nunes, O.N. Oliveira Jr, and L.da F. Costa, *Physica A* 373 (2007), pp. 811–820.
- [512] D.R. Amancio, L. Antiqueira, T.A.S. Pardo, L.da F. Costa, O.N. Oliveira Jr, and M.G.V. Nunes, *Int. J. Mod. Phys. C* 19(4) (2008), pp. 583–598.
- [513] L. Antiqueira, O.N. Oliveira Jr, L.da F. Costa, and M.G.V. Nunes, *Information Sciences* 179(5) (2009), pp. 584–599.
- [514] M. Choudhury, M. Thomas, A. Mukherjee, A. Basu, and N. Ganguly, *How difficult is it to develop a perfect spell-checker? a cross-linguistic analysis through complex network approach*, Proceedings of the Second Workshop on TextGraphs: Graph-Based Algorithms for Natural Language Processing, 2007, pp. 81–88.
- [515] R. Guimerà and L.A.N. Amaral, *Eur. Phys. J. B* 38(2) (2004), pp. 381–385.
- [516] L.E.C. Rocha, *J. Statist. Mech.* 2009(4) (2009), p. P04020.
- [517] N. Barnier and P. Brisset, *Ann. Oper. Res.* 130(1–4) (2004), pp. 163–178.
- [518] A. Gautreau, A. Barrat, and M. Barthélemy, *Proc. Natl Acad. Sci. USA* 106(22) (2009), pp. 8847–8852.
- [519] L.P. Chi and X. Cai, *Int. J. Mod. Phys. B* 18(17–19) (2004), pp. 2394–2400.
- [520] S. Porta, P. Crucitti, and V. Latora, *Physica A* 369(2) (2006), pp. 853–866.
- [521] V. Kalapala, V. Sanwalani, A. Clauset, and C. Moore, *Phys. Rev. E* 73(2) (2006), p. 026130.
- [522] A. Cardillo, S. Scellato, V. Latora, and S. Porta, *Phys. Rev. E* 73(6) (2006), p. 066107.
- [523] B.A.N. Travençolo and L.da F. Costa, *J. Phys. A* 41(22) (2008), pp. 224004–224004.
- [524] B.A.N. Travençolo and L.da F. Costa, *Phys. Lett. A* 373(1) (2008), pp. 89–95.
- [525] B.A.N. Travençolo, M.P. Viana, and L.da F. Costa, *New J. Phys.* 11(6) (2009), pp. 063019–063036.
- [526] B.A.N. Travençolo and L.da F. Costa, *Outward accessibility in urban street networks: Characterization and improvements*, Preprint arXiv:0802.3665, 2008.
- [527] L.da F. Costa, B.A.N. Travençolo, M.P. Viana, and E. Strano, *Europhys. Lett.* 91 (2010), p. 18003.
- [528] L.da F. Costa, F.A. Rodrigues, and P.R. Villas Boas, *Int. J. Bifurc. Chaos* 20(3) (2010), pp. 795–804.
- [529] M. Rosvall, A. Trusina, P. Minnhagen, and K. Sneppen, *Phys. Rev. Lett.* 94(2) (2005), p. 028701.
- [530] B. Hillier and J. Hanson, *The social logic of space*, Cambridge University Press, Cambridge, 1984.
- [531] P. Sen, S. Dasgupta, A. Chatterjee, P.A. Sreeram, G. Mukherjee, and S.S. Manna, *Phys. Rev. E* 67(3) (2003), p. 036106.
- [532] K. Sneppen, A. Trusina, and M. Rosvall, *Europhys. Lett.* 69(5) (2005), pp. 853–859.
- [533] P.R. Villas Boas, F.A. Rodrigues, and L.da F. Costa, *Modeling highway networks with path-geographical transformations*, Complex Networks: Results of the 1st International Workshop on Complex Networks (CompleNet 2009), Springer, 2009, p. 115.
- [534] P.R. Villas Boas, F.A. Rodrigues, and L.da F. Costa, *Phys. Lett. A* 374(1) (2009), pp. 22–27.
- [535] K. Nagel and M. Schreckenberg, *J. Phys. I France* 2(2) (1992), pp. 2221–2229.
- [536] W. Knospe, L. Santen, A. Schadschneider, and M. Schreckenberg, *J. Phys. A* 33(48) (2000), pp. L477–L485.
- [537] D. Helbing, *Reviews of Modern Physics* 73(4) (2001), pp. 1067–1141.
- [538] A. Schadschneider, *Physica A* 313(1–2) (2002), pp. 153–187.
- [539] K. Nagel and J. Esser, *Annual Reviews of Computational Physics VII* 7 (2000), pp. 151–202.
- [540] A. Schadschneider, W. Knospe, L. Santen, and M. Schreckenberg, *Physica A* 346(1–2) (2005), pp. 165–173.
- [541] K.A. Seaton and L.M. Hackett, *Physica A* 339(3–4) (2004), pp. 635–644.
- [542] K.P. Li, Z.Y. Gao, and B.H. Mao, *Int. J. Mod. Phys. C* 17(9) (2006), pp. 1339–1347.
- [543] W. Li and X. Cai, *Physica A* 382(2) (2007), pp. 693–703.
- [544] V. Latora and M. Marchiori, *Physica A* 314(1–4) (2002), pp. 109–113.
- [545] A.P. Masucci and G.J. Rodgers, *Physica A* 387(14) (2008), pp. 3781–3788.
- [546] J. Sienkiewicz and J.A. Holyst, *Acta Phys. Polon. B* 36(5) (2005), pp. 1771.
- [547] J. Sienkiewicz and J.A. Holyst, *Phys. Rev. E* 72(4) (2005), p. 046127.
- [548] R. Albert, I. Albert, and G.L. Nakarado, *Phys. Rev. E* 69(2) (2004), p. 025103.

- [549] 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*, Proceedings of the International Conference on Communications, Circuits and Systems, 4, 2006.
- [550] I. Dobson, B.A. Carreras, V.E. Lynch, and D.E. Newman, *An initial model for complex dynamics in electric power system blackouts*, Proceedings of the 34th Annual Hawaii International Conference on System Sciences, (2001), pp. 710–718.
- [551] B.A. Carreras, V.E. Lynch, M. Sachtjen, I. Dobson, and D. Newman, *Modeling blackout dynamics in power transmission networks with simple structure*, HICSS '01: Proceedings of the 34th Annual Hawaii International Conference on System Sciences (HICSS-34), p. 2018, Washington, DC, USA, 2001. IEEE Computer Society, Silver Spring, MD.
- [552] B.A. Carreras, V.E. Lynch, I. Dobson, and D.E. Newman, *Chaos* 12(4) (2002), pp. 985–994.
- [553] B.A. Carreras, V.E. Lynch, I. Dobson, and D.E. Newman, *Chaos* 14(3) (2004), pp. 643–652.
- [554] B.A. Carreras, D.E. Newman, I. Dobson, and A.B. Poole, *IEEE Trans. Circuits and Systems I: Regular Papers* 51(9) (2004), pp. 1733–1740.
- [555] D.P. Chassin and C. Posse, *Physica A* 355(2–4) (2005), pp. 667–677.
- [556] A.E. Motter and Y.-C. Lai, *Phys. Rev. E* 66(6) (2002), p. 065102.
- [557] P. Crucitti, V. Latora, and M. Marchiori, *Phys. Rev. E* 69(4) (2004), p. 045104.
- [558] R. Kinney, P. Crucitti, R. Albert, and V. Latora, *Eur. Phys. J. B* 46(1) (2005), pp. 101–107.
- [559] M. Anghel, K.A. Werley, and A.E. Motter, *Stochastic model for power grid dynamics*, HICSS '07: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, p. 113, Washington, DC, USA, 2007. IEEE Computer Society.
- [560] S. Abe and N. Suzuki, *Europhys. Lett.* 65(4) (2004), pp. 581–586.
- [561] S. Abe and N. Suzuki, *Physica A* 337(1–2) (2004), pp. 357–362.
- [562] S. Abe and N. Suzuki, *Nonlinear Processes Geophys.* 13(2) (2006), pp. 145–150.
- [563] S. Abe and N. Suzuki, *Phys. Rev. E* 74(2) (2006), pp. 026113–026117.
- [564] S. Abe and N. Suzuki, *Eur. Phys. J. B* 59(1) (2007), pp. 93–97.
- [565] A.A. Tsonis, K.L. Swanson, and P.J. Roebber, *Bull. Amer. Meteorol. Soc.* 87(5) (2006), pp. 585–596.
- [566] A.A. Tsonis, J.B. Elsner, A.G. Hunt, and T.H. Jagger, *Geophys. Res. Lett.* 32(9) (2005), p. L09701.
- [567] A.A. Tsonis and K.L. Swanson, *Phys. Rev. Lett.* 100(22) (2008), p. 228502.
- [568] J.F. Donges, Y. Zou, N. Marwan, and J. Kurths, *Eur. Phys. J.-Special Top.* 174(1) (2009), pp. 157–179.
- [569] J.F. Donges, Y. Zou, N. Marwan, and J. Kurths, *Europhys. Lett.* 87 (2009), p. 48007.
- [570] J.P.K. Doye and C.P. Massen, *J. Chem. Phys.* 122(8) (2005), p. 084105.
- [571] J.P.K. Doye, *Phys. Rev. Lett.* 88(23) (2002), p. 238701.
- [572] T. Carmi, S. Havlin, C. Song, K. Wang, and H.A. Makse, *J. Phys. A* 42(10) (2009), p. 105101.
- [573] T. Taketsugu and D.J. Wales, *Mol. Phys.* 100(17) (2002), pp. 2793–2806.
- [574] F. Rao and A. Caffisch, *J. Mol. Biol.* 342(1) (2004), pp. 299–306.
- [575] D. Gfeller, P. De Los Rios, A. Caffisch, and F. Rao, *Proc. Natl Acad. Sci. USA* 104(6) (2007), pp. 1817–1822.
- [576] H. Seyed-allaei, H. Seyed-allaei, and M.R. Ejtehadi, *Phys. Rev. E* 77(3) (2008), p. 031105.
- [577] R.V. Solé and A. Munteanu, *Europhys. Lett.* 68(2) (2004), pp. 170–176.
- [578] R. Diestel, *Graph Theory*, Springer, Berlin, 2000.
- [579] D.B. West, *Introduction to Graph Theory*, Prentice Hall, Englewood Cliffs, NJ, 2001.
- [580] B. Wang, H. Tang, C. Guo, and Z. Xiu, *Physica A* 363(2) (2006), pp. 591–596.
- [581] G. Caldarelli, R. Pastor-Satorras, and A. Vespignani, *Eur. Phys. J. B* 38(2) (2004), pp. 183–186.
- [582] H.D. Rozenfeld, J.E. Kirk, and E.M. Boltt, *J. Phys. A* 38 (2005), p. 4589.
- [583] G. Bianconi and N. Gulbahce, *J. Phys. A* 41 (2008), p. 224003.
- [584] G. Bianconi and M. Marsili, *J. Statist. Mech.* 2005 (2005), p. P06005.
- [585] G. Bianconi and M. Marsili, *Phys. Rev. E* 73(6) (2006), p. 66127.
- [586] D.J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness*, Princeton University Press, Princeton, NJ, 2003.
- [587] M.E.J. Newman, *Social Networks* 27(1) (2005), p. 39–54.
- [588] L.da F. Costa, *Phys. Rev. Lett.* 93(9) (2004), p. 098702.
- [589] L.da F. Costa and F. Silva, *J. Statist. Phys.* 125 (2006), p. 841–872.

- [590] L.da F. Costa and L.E.C. Rocha, *Eur. Phys. J. B* 50 (2006), p. 237–242.
- [591] L.da F. Costa and R.F.S. Andrade, *New J. Phys.* 9 (2007), p. 311.
- [592] J.J. McAuley, L.da F. Costa, and T.S. Caetano, *Appl. Phys. Lett.* 91 (2007), p. 084103.
- [593] V. Colizza, A. Flammini, M.A. Serrano, and A. Vespignani, *Nat. Phys.* 2 (2006), pp. 110–115.
- [594] S. Zhou and R.J. Mondragón, *Commun. Lett. IEEE* 8(3) (2004), pp. 180–182.
- [595] M.E.J. Newman, *Phys. Rev. Lett.* 89(20) (2002), p. 208701.
- [596] L. Page, S. Brin, R. Motwani, and T. Winograd, *The PageRank citation ranking: Bringing order to the Web*, Technical report, Stanford Digital Library Technologies Project, 1998, p. 17.
- [597] M.E.J. Newman, *Eur. Phys. J. B* 38 (2004), pp. 321–330.
- [598] M. Girvan and M.E.J. Newman, *Proc. Natl Acad. Sci. USA* 99(12) (2002), pp. 7821–7826.
- [599] L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, *J. Statist. Mech.* 2005(9) (2005), p. P09008.
- [600] J. Bagrow, E. Bollt, and L.da F. Costa, *Network structure revealed by short cycles*, Preprint arXiv:cond-mat/0612502, 2006.
- [601] S.B. Seidman, *Social Networks* 5(3) (1983), pp. 269–287.
- [602] S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes, *Phys. Rev. Lett.* 96(4) (2006), p. 040601.
- [603] I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, *Adv. Neural Inform. Process. Syst.* 18 (2006), p. 41.
- [604] B. Bollobás, *Modern graph theory*, Graduate texts in mathematics. Springer, New York, 1998.
- [605] J.D.J. Han, D. Dupuy, N. Bertin, M.E. Cusick, and M. Vidal, *Nat. Biotechnol.* 23(7) (2005), pp. 839–844.
- [606] R.M. May, C. Wiuf, and M.P.H. Stumpf, *Proc. Natl Acad. Sci. USA* 102(12) (2005), pp. 4221–4224.
- [607] P.R. Villas Boas, F.A. Rodrigues, G. Travieso, and L.da F. Costa, *J. Statist. Mech.* 2010(3) (2010), p. P03009.

## Appendix. Basic concepts

A *graph* or *undirected graph*  $G$  is an ordered pair  $G = (\mathcal{N}, \mathcal{E})$ , formed by a set  $\mathcal{N} \equiv \{n_1, n_2, \dots, n_N\}$  of *vertices* (or nodes, or points) and a set  $\mathcal{E} \equiv \{e_1, e_2, \dots, e_E\}$  of *edges* (or lines, or links)  $e_k = \{n_i, n_j\}$  that connect the vertices [12,578,579]. Here only simple graphs (graphs without multiple edges between a given pair of vertices) without self-connections are considered. In the Physics literature, a graph has also been called a network [32].<sup>15</sup> When the edges between pairs of vertices have direction, the graph is said to be a *directed graph*. In this case, the graph can be represented by  $G^\rightarrow$ , which is an ordered pair  $G^\rightarrow = (\mathcal{N}, \mathcal{E}^\rightarrow)$ , 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  $n_j$ , then  $n_j$  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  $n_{w0}e_{w0} \dots e_{wk-1}n_{wk}$  of vertices and edges in  $G$  such that  $e_{wi} = \{n_{wi}, n_{wi+1}\}$  for all  $i < k$ . If  $n_{w0} = n_{wk}$ , 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, in which no vertex is repeated. A graph  $G^* = (\mathcal{N}^*, \mathcal{E}^*)$  is a *subgraph* of  $G = (\mathcal{N}, \mathcal{E})$  if  $\mathcal{N}^* \subseteq \mathcal{N}$ ,  $\mathcal{E}^* \subseteq \mathcal{E}$  and the edges in  $\mathcal{E}^*$  connect nodes in  $\mathcal{N}^*$ . 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*. If the network is disconnected, it is formed by a set of *connected components*, each one being a maximal connected subgraph of the network. In the case of directed graphs, a similar notion is that of *strongly connected components*, sets of nodes in which each vertex can be reached from all other vertices by following the directed edges.

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  $\mathcal{W} \equiv \{w_1, w_2, \dots, w_E\}$  weights, i.e. real numbers attached to the edges, where  $w_i$  is the weight associated with  $e_i$ . 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  $G^{w\rightarrow}$  [32].

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



Graphs can be represented, among others, by using *incidence 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. The list may also have a third element which represents the intensity of the connection. In the latter case, the graph is represented by a matrix  $A$  whose elements  $a_{ij}$  are equal to 1 whenever there is an edge connecting nodes  $i$  to node  $j$ , 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  $j$ . From the weight matrix, an adjacency matrix can be derived through a thresholding operation,  $A = \delta_T(W)$ , that associates for each element  $w_{ij}$  whose value is larger than a threshold  $T$  the value  $a_{ij} = 1$  (0 is otherwise assigned) [32].

In order to characterize and represent complex networks, many measurements have been developed [2,32]. The most traditional ones include 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 number of connections. For undirected networks, using the adjacency matrix,

$$k_i = \sum_{j=1}^N a_{ij}. \quad (\text{A1})$$

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

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i. \quad (\text{A2})$$

If the network is directed, it is possible to define, for each node  $i$ , its *in-degree* (the number of incoming arcs),  $k_i^{\text{in}} = \sum_{j=1}^N a_{ji}$ , and *out-degree* (the number of outgoing arcs),  $k_i^{\text{out}} = \sum_{j=1}^N a_{ij}$ , as well as the corresponding averages considering the whole network. The total degree of a node  $i$  is defined as  $k_i = k_i^{\text{in}} + k_i^{\text{out}}$ .

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 ( $P(k) \sim k^{-\gamma}$ ) for many real-world networks, as discussed in this review. For directed networks, there are two distributions, for incoming links,  $P(k^{\text{in}})$ , and outgoing links,  $P(k^{\text{out}})$ . We can also define an *information entropy* [580] related to the degree distribution, which provides insights into the heterogeneity of the connectivity pattern in a given network. The information entropy is expressed as

$$H = - \sum_{k=1}^{\infty} P(k) \log P(k). \quad (\text{A3})$$

The *clustering coefficient* is related to the presence of triangles (cycles of order 3) in the network [38]. The 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,

$$c_i = \frac{2e_i}{k_i(k_i - 1)} = \frac{\sum_{j=1}^N \sum_{m=1}^N a_{ij} a_{jm} a_{mi}}{k_i(k_i - 1)}. \quad (\text{A4})$$

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

$$\langle c \rangle = \frac{1}{N} \sum_{i=1}^N c_i. \quad (\text{A5})$$

It is important to note that, in order to avoid anomalous values in the above definition of clustering coefficient, there is an alternative expression, which is written as

$$\langle c \rangle = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}}. \quad (\text{A6})$$

Loops of different orders in networks have been studied [214,581–584]. The number of loops of order  $h = 3$  can be calculated by [214]

$$N_3 = \frac{1}{6} \sum_i (a^3)_{ii}. \quad (\text{A7})$$

For order  $h = 4$  and  $h = 5$ ,

$$N_4 = \frac{1}{8} \left[ \sum_i (a^4)_{ii} - 2 \sum_i (a^2)_{ii} (a^2)_{ii} + \sum_i (a^2)_{ii} \right], \quad (\text{A8})$$

$$N_5 = \frac{1}{10} \left[ \sum_i (a^5)_{ii} - 5 \sum_i (a^2)_{ii} (a^3)_{ii} + 5 \sum_i (a^3)_{ii} \right]. \quad (\text{A9})$$

Bianconi and Marsili [585] showed how to determine the number of triangles in a two-vertex correlated network in terms of the maximal eigenvalue and the corresponding eigenvector of the average adjacency matrix of the network ensemble. They also introduced an analytic approach to estimate the average number of loops in random scale-free networks [584].

The length of a path connecting the vertices  $i$  and  $j$  is given by the number of edges along that path. A *shortest path* (or *geodesic path*) between vertices  $i$  and  $j$  is any of the paths connecting these two nodes whose length is minimal [586]. For the whole network, it is possible to represent the geodesic distances by a distance matrix  $D$ , in which each entry  $d_{ij}$  represents the length of the shortest paths between the nodes  $i$  and  $j$ , or  $\infty$  if they are disconnected. The *average shortest path length* is then given by

$$\ell = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N d_{ij}, \quad (\text{A10})$$

where the sum considers  $i \neq j$  and disregards pairs that are not connected. This measurement is sometimes called *diameter*. Other authors consider the network diameter as the maximum value of the entries of the matrix  $D$ . Still regarding the distances in networks, we can define the *global efficiency* [197] as a measure to evaluate the transport efficiency of a network. The mathematical expression of the global efficiency is

$$E = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{d_{ij}}. \quad (\text{A11})$$

At the local level, the importance of nodes is also quantified by the so-called centrality measurements. In particular, we have the *betweenness centrality* [264], which measures the importance of a given node by counting the number of shortest paths that pass through this node, as expressed in the following:

$$bc_i = \sum_{m=1}^N \sum_{n=1}^N \frac{\sigma_{m,i,n}}{\sigma_{m,n}}, \quad (\text{A12})$$

where  $\sigma_{m,n}$  is the number of shortest paths from  $m$  to  $n$ , and  $\sigma_{m,i,n}$  is the number of these paths which pass through  $i$ . The *closeness centrality* [587] is another centrality measurement commonly used in complex

network studies. It is related to the distance between one node and all other nodes, i.e.

$$cc_i = \sum_{j=1}^N \frac{1}{d_{ij}}. \quad (A13)$$

All the measurements discussed above can be extended to weighted networks, in which case the *node strength* is defined with the weight matrix  $W$  [224],

$$s_i = \sum_{j=1}^N w_{ij}. \quad (A14)$$

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

$$\langle s \rangle = \frac{1}{N} \sum_{i=1}^N s_i. \quad (A15)$$

The *weighted clustering coefficient* of a node  $i$  can be defined as [224],

$$C_i^w = \frac{1}{s_i(k_i - 1)} \sum_{j>k} \frac{w_{ij} + w_{ik}}{2} a_{ij} a_{ik} a_{jk}, \quad (A16)$$

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

$$\langle C^w \rangle = \frac{1}{N} \sum_i C_i^w. \quad (A17)$$

The average shortest path length for weighted networks is determined similarly as in Equation (A10), considering the weight of the edges.

Hierarchical or concentric measurements, systematized by Costa *et al.* [170,588–592], generalize the traditional respective measurements in order to characterize the neighborhood of a node from purely local (first neighborhood) to completely global (most distant vertices) scales. These measurements are defined in terms of *n-rings*, which are composed by all vertices distant  $n$  edges from a given node. The *n*-ring is also known as *n-neighborhood* (see Figure A1 for an example of distinct neighborhoods of a node).

The hierarchical measurements are extensions of the traditional local ones where successive levels of neighborhoods of a given node are taken into account. For the next definitions, only unweighted and undirected measurements will be considered. The *hierarchical degree* of a node  $i$ ,  $k_n(i)$ , is the number of edges connecting ring  $R_n$  to ring  $R_{n+1}$ . The *hierarchical clustering coefficient* of a node  $i$ ,  $c_n(i)$ , is the number of edges connecting all nodes of ring  $n$  divided by the total number of possible edges between these nodes.

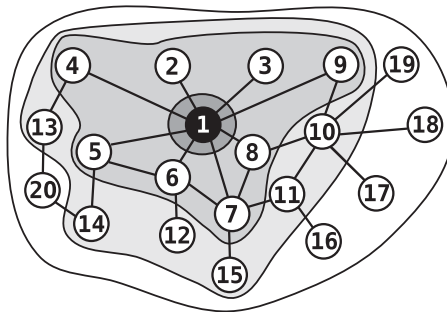


Figure A1. An illustration of the hierarchical neighborhoods of a reference node (black). The first neighborhood is composed by nodes 2–9, the second, by nodes 10–15, and the third, by nodes 16–20. The hierarchical or concentric measurements of level  $n$  are defined according to these neighborhoods.

In addition to the extension of traditional measurements, others can be defined by the inter-relation between subsequent rings. The *convergence ratio* of node  $i$ ,  $C_n(i)$ , is given by

$$C_n(i) = \frac{k_n}{\mathcal{N}(R_{n+1})}, \quad (\text{A18})$$

where  $\mathcal{N}$  represents the number of nodes of the ring  $R_{n+1}$ . The *divergence ratio* of node  $i$ ,  $D_n(i)$ , is defined as

$$D_n(i) = \frac{\mathcal{N}(R_{n+1})}{k_n}. \quad (\text{A19})$$

Several complex networks exhibit a tendency of hubs to be well connected to each other. Such a trend is known as the rich-club effect and is observed, for instance, in science where researchers tend to form collaborative groups and publish together [593]. In principle, high-degree nodes, also known as rich nodes, are likely to form tight and interconnected subgraphs, the rich-clubs. In order to quantify this phenomenon, Zhou and Mondragon [594] introduced the *rich-club coefficient* of degree  $k$  which is defined as,

$$\phi(k) = \frac{1}{|\mathcal{R}(k)|(|\mathcal{R}(k)| - 1)} \sum_{i,j \in \mathcal{R}(k)} a_{ij}, \quad (\text{A20})$$

where  $\mathcal{R}(k) = \{v \in \mathcal{N}(G) | k_v > k\}$  is the set of nodes of  $G$  with degree greater than  $k$ . It can be noticed that this measurement is similar to the clustering coefficient, however the above sum corresponds to twice the number of edges between the nodes in the club.

Another important structural aspect of complex network characterization is the analysis of how vertices with different degrees are interconnected. The degree correlation can be determined from the Pearson correlation coefficient of the degrees at both ends of the edges [595], which yields the *assortativity* coefficient:

$$r = \frac{(1/M) \sum_{j>i} k_i k_j a_{ij} - \left[ (1/M) \sum_{j>i} \frac{1}{2} (k_i + k_j) a_{ij} \right]^2}{(1/M) \sum_{j>i} \frac{1}{2} (k_i^2 + k_j^2) a_{ij} - \left[ (1/M) \sum_{j>i} \frac{1}{2} (k_i + k_j) a_{ij} \right]^2}, \quad (\text{A21})$$

where  $M$  is the total number of edges and  $-1 \leq r \leq 1$ . 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 those with few connections); for  $r = 0$  there is no correlation between node degrees, and the network is called non-assortative.

The *PageRank* is a measure of how likely it is to visit a node while “surfing” the graph, i.e. visiting nodes that are neighbors and sometimes performing jumps to other non-adjacent nodes. Brin and Page [239] devised PageRank as a random surfer model which, in the equilibrium, gives the probability of visits to a given node

$$PR_i = (1 - d) + d \sum_{j=1}^N \frac{PR_j}{k_j^{\text{out}}} \quad (\text{A22})$$

where  $d$  is called damping factor, the probability of jumping to a random node in the next time step. Intuitively speaking, the PageRank of a node tends to be high if other nodes that point to it have high PageRank. An iterative algorithm of fast convergence can be used to compute PageRank [596]. Figure 9 contains an example of a graph and its respective PageRank distribution.

The *search information* quantifies how difficult it is to send a message from node  $a$  to  $b$  [532]. It is assumed that messages take only shortest paths, and the probability of choosing a shortest path  $p(a, b)$  from  $a$  to  $b$  in relation to all paths connecting these nodes is denoted by  $P[p(a, b)]$ . Then the search information is computed as follows:

$$S(a \rightarrow b) = -\log_2 \sum_{p(a,b)} P[p(a, b)], \quad (\text{A23})$$

where the summation takes into consideration all different paths of shortest length linking nodes  $a$  and  $b$ . The higher the search information is, the more difficult it is to send a message from  $a$  to  $b$ .

The ability of the nodes to interact with other nodes can also be evaluated by the recent proposed measure *accessibility* [527]. This measure can be understood as a generalization of the classical concept of degree, which takes into account the transitions probability from a given node to all other nodes of the network. The mathematical expression of the accessibility is given by

$$A_i(H) = \exp \left\{ - \sum_{j=1}^N P_{ij}(H) \log P_{ij}(H) \right\}, \quad (\text{A24})$$

where  $P_{ij}(H)$  is the transition probability to reach node  $j$  after  $H$  steps, departing from  $i$  and performing a random walk (or other dynamics).  $A_i(H)$  denotes the accessibility of node  $i$  for step  $H$ . Nodes with homogeneous transition probabilities have high values of accessibility, while heterogeneous access implies lower accessibility values.

The measurements 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* or *modules*, are common in many real networks, formed by sets of nodes densely connected among themselves while being sparsely connected to the remainder of the network [597,598]. Communities play an important role in network structure, evolution and dynamics, defining modular topologies. Unfortunately, their identification is an NP-complete problem [599], so that many heuristic algorithms have been proposed for this purpose [32, 597,599].

Depending on the application, subgraphs can be essential to characterize network structures. For instance, modular structures in networks can be associated with different functions, such as scientific collaboration areas. In addition to communities, other types of subgraphs are found in complex networks, such as motifs [62] and cycles [600]. *Motifs* are subgraphs that appear more frequently in real networks than could be statistically expected [62,115] (see Figure 8). Another important subgraph in complex networks is the so-called *k-core* [601], which is the largest subgraph of a given graph whose nodes have degree at least  $k$  – an example is shown in Figure A2. In various complex network studies, researchers have employed *k*-cores [602], including as a tool for network visualization [603]. A *tree* is a connected acyclic graph, and is an important concept borrowed from graph theory [604]. A related concept is the *spanning tree*, which is a subgraph including all nodes of a graph without cycles (Figure A3). For weighted networks, the *minimum spanning tree* is a spanning tree formed with minimal sum of edge weights.

A fundamental issue related to network measurements 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 not much sensitive to perturbations [605,606]. Such measurements must reflect the differences in distinct networks structures [607]. This analysis is crucial for the theory of complex networks and constitutes a promising research field.

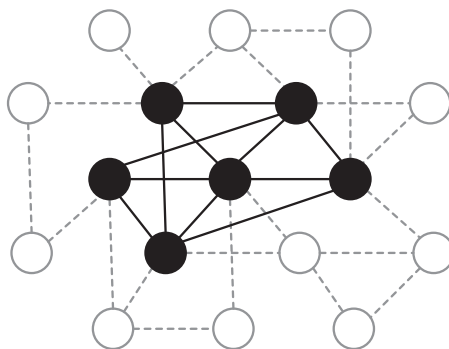


Figure A2. A graph and its 3-core in black. Nodes in the 3-core have at least degree 3 considering only the connections among them. To obtain a 3-core, it suffices to remove all nodes with degree less than 3 in a recursive manner, until only nodes with degree at least 3 remain. This is why in this example there are nodes with degree 3 or 4 outside the core.

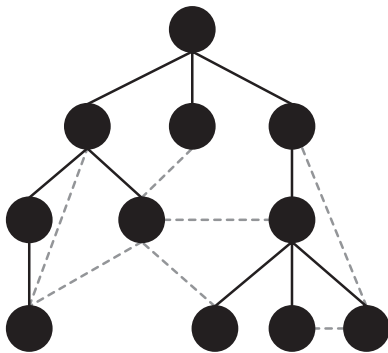


Figure A3. A graph and one of its spanning trees. The dashed edges are those that need to be removed in order to obtain the spanning tree.