Complex networks: Statics and Dynamics

Albert Diaz-Guilera

Departament Fisica Finamental, Universitat de Barcelona, Barcelona, SPAIN

Abstract. We present some of the results obtained during the last 8 years about complex networks. Starting with the collection of data in the form of networks or graphs, we proceed on the characterization at different scales: microscopic, macroscopic, and mesoscopic. We introduce also the basic models incorporating complexity in the pattern of connectivities. Finally we review some results on dynamical features on complex networks.

Keywords: complex networks, dynamics, synchronization
PACS: 89.75.–k,89.75.Da,05.45.Xt

1. INTRODUCTION

Complex systems show emergent properties that cannot be understood as a simple superposition of the dynamical behavior of the single units that form it. Emergent properties arise as a collective effect from the interaction between units. For most of the complex systems we can find in nature or in society the interaction patterns of the units are far from being regular or from being completely random. These two extreme patterns of connectivity between units had been the subject of analysis during the last decades, but since the pioneering work of Watts and Strogatz [1] and Barabasi and Albert [2] most of the interest turned to non-trivial patterns of interaction.

Access to large datasets and the power of the new generation of computers have enabled to generate networks from the data and to scrutinize them, finding the relevant statistical properties and the common features shared by many of them. And such common features are not the subject of particular disciplines; this new framework has received a really interdisciplinar support involving disciplines such diverse as economy, social sciences, medicine, engineering, physics, chemistry, biology, just to mention a few (see [3, 4, 5, 6, 7] for reviews, and [8, 9] for contributed essays on various topics by leading researchers).

Although in principle the interest relied on the structural properties, i.e. the topology of the interactions, in the last years a great deal of attention has turned to the dynamical properties. This dynamical properties can involve the growth of the network (in terms of nodes or links or both), the regeneration of links between the nodes, and even the dynamics of the nodes properties.

In this paper we will cover the different aspects of complex networks. Starting with an overview of networks that are found in our environment, then we follow with the characterization of the networks at different scales (microscale, macroscale, and mesoscale) and introducing some simple network models that describe the topological features. Finally we describe some dynamical models on networks.

2. NETWORKS EVERYWHERE

We can currently find structures among data that can be classified as networks in many unrelated fields. The only thing that is needed is some sort of relation between the data. Networks or graphs are formed by nodes that are connected by links or edges. What is a node and a link will a characteristic of the data. For instance in social networks nodes can be individuals and links could be any kind of relation: friendship, coauthoring, trust, ... In technological networks, as for example the Internet, the edges correspond to physical wiring between computers or routers. Or in biology, metabolic networks are formed by metabolites as nodes and the links represent the biochemical reactions. Those are just a few examples of the type of data that can be represented as a network. A very large number of publicly available repositories of huge databases are at our fingertips.

2.1. Classification

Here we present a list of examples, mainly from papers published in Physics journals, of data that has been collected in the form of network. They can roughly be classified into four categories (following [6]):

- Social networks
  Actor collaborations [10, 1]
  Boards of directors [11, 12]
  Physics and biology coauthorships [13, 14, 15]
  Email messages [16, 17, 18]
  Sexual contacts [19, 20]
  Jazz bands and musicians [21]
  Pretty Good Privacy trust network [22]
- Information networks
  World Wide Web [23, 2]
  Citation networks [24]
  Word co-occurrence [25, 26]
- Technological and transport networks
  Internet [27]
  Power grid [1]
  Software packages [28]
  Software routine calls [29]
  Electronic circuits [30]
  Airport network [31, 32]
  Railroad network [33]
- Biological networks
  Metabolic networks [34, 35, 36, 37, 38]
  Protein interactions [39, 40, 41, 42, 43]

Food webs [44, 45]
Neural networks [1, 46, 47, 48]
Genetic regulatory networks [49, 50]
Signaling networks [51]

This is by no means a complete list. Nevertheless, it attempts to be a cross section of the various lines of investigation where network analysis has been useful.

2.2. How are networks constructed from data

It is then clear that every dataset will give rise to a network in which nodes and links have very different meanings: social agents and social relationships, computers and cables, species and predator-prey relationships, neurons and synapses, web pages and hyperlinks, and so on.

The link can be directed or undirected, depending on the reciprocity of the relation. An example of directed network, in which some links can be directed, is presented in Fig. 1(top-left). Links with an arrow pointing from one node to another one are directed, whereas the links without any arrow are undirected or bidirectional, because the relation holds in the two directions.

Some of the networks that are constructed are called bipartite, those are graphs that contain nodes of two different types, with links only between unlike types. This is what happens for instance in the actors movie database with actors and movies, or with the coauthor-ship databases. In this case networks are constructed by linking those actors that appear in the same movie or those authors that participate in the same paper. An example of such network is presented in Fig. 1(bottom)[15]. This is the network obtained from the coauthorship of presentations in the Spanish Statistical Physics meetings; it has been obtained by accumulating the collaborations along all the editions of the meeting. There are a few nodes that are identified because they correspond to members of the different scientific committees; the role played by these members will be discussed later, in the community identification discussion section.

Another fact that has become very important in the last years is the weight of the edges. If one is just interested in the existence of the relation then we talk about unweighted networks. If, on the contrary, there is some measure for the strength of the relation, as for instance the number of flights or the number of passengers between airports, or the band-width between Internet routers, some weight is associated to the link, and in this case the networks are called weighted [52]. In Fig. 1(top-right) we present a network obtained from the email exchange of the Universitat Rovira i Virgili. We take each node in the original network and measure the number of steps across the e-mail network needed to reach any other node. Then we average over all the nodes in the same center and obtain average distances between centers. This average distance between the centers accounts for the weight of the link. To summarize this information in a new network of centers, we proceed as follows. First, we calculate the distance from one center A to all other centers, d_{AB}, d_{AC}, and so on. Then we compute the average distance from A to the other centers $\langle d_A \rangle$. Finally, node A (that now represents a center, not an individual) is linked to another node B if $d_{AB} < \langle d_A \rangle$. In this case, the network is directed because, in general, $d_{AB} < \langle d_A \rangle$ does not imply $d_{BA} = d_{AB} < \langle d_B \rangle$. [53]

3. CHARACTERIZING NETWORKS

3.1. Microscale

From a microscopical point of view, the interest would lie on the role played by the nodes in the overall context of the whole network. This has been the main issue for decades from the social sciences viewpoint [54]. Several measures of centrality were introduced and the special roles played by the nodes discussed. For instance, the degree of a node corresponds to its number of links or the mean distance is a measure of the average distance, measured as the shortest number of links necessary to reach one node from another, from a node to the rest of the population. Another example is the clustering coefficient of a node, which measures the fraction of links between neighbors of a given node. Finally, another interesting measure is what is called the betweenness, of a node, which corresponds to the number of shortest paths between each pair of nodes in the network that go through the reference node. In many problems related to flow or traffic in the network the betweenness is a good measure for the load of the node [55, 56].

3.2. Macroscale

On the other hand, when dealing with very large networks, the roles played by the individual nodes has not meaning at all and the interest is turned to the statistical characterization of the network at the global or macroscopic scale. Now one studies average quantities like the mean degree, the mean distance between nodes, the average clustering coefficient, the diameter of the network (measured as the maximum distance between nodes). Another statistical characterization of the network comes in terms of the distributions of degree, of load, or on the correlations.

It was the initial study of these statistical characterizations of the networks that started the big interest from the Statistical Physics community. In particular, as we will explain shortly, there were two crucial facts that could not be explained by means of known graph models: the small-world effect and the observation that the distribution of degrees followed a power law indicating that there are no characteristic scales in this distribution and hence those networks were called "scale-free" networks. In Fig. 2(right) we plot the in- and out-degree cumulative distributions [1] in the PGP web of trust of Ref. [22], as an example of power-law distribution; in this case, as a directed network, the in- and out-degrees distributions do not need to be identical.

[1] The cumulative distribution $P(k)$ is simply related to the probability density function $p(x)$ by $P(k) = \int_{-\infty}^{k} dx\, p(x)$. In particular, if $p(x)$ is a power law $p(x) \sim x^{-\alpha}$, then $P(k) \sim k^{-\alpha-1}$, and if $p(x)$ is an exponential $p(x) \sim \exp(-x/k^*)$, then $P(k) \sim \exp(-k/k^*)$.

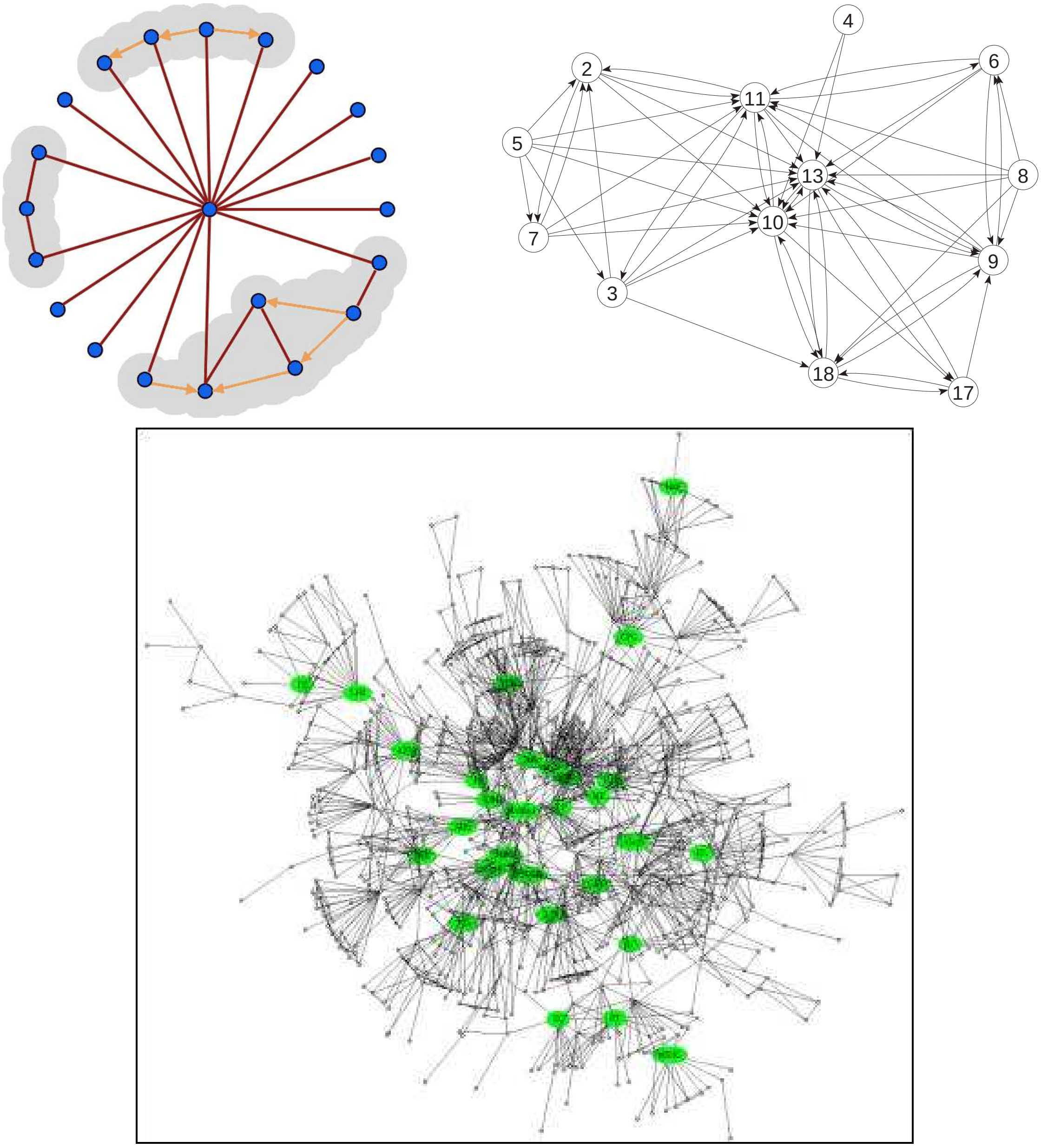

FIGURE 1. Examples of generated networks from data. Top-Left: Directed network. Nodes are colored blue. Bidirectional links are colored brown whereas unidirectional links are colored orange. The shadowed areas correspond to the robust areas of the network, i.e. they keep connected when the central node is eliminated. Top-Right: Directed and weighted network. Network obtained from the email network at Universitat Rovira i Virgili. The nodes correspond to the centers and the links are related to the distance between centers (see text). Since the distance can take real non-negative values the link, and hence the network, are weighted. Bottom: Collaboration network. Cumulative network of collaborations during Spanish Statistical Physics meetings. Green nodes correspond to members of the scientific committee [15].

Later on, different characterizations of the networks have been introduced; for instance in weighted networks, the distribution of weights is also a scale-free [52]. Also, other characterization in the large scale have appeared. For instance, the degree-degree correlation $P(k'|k)$ is the conditional probability that a link of a node with degree k is linked to a node with degree k'; if this probability depends on k we say that the node is correlated, or uncorrelated in the opposite case. In terms of this conditional probability, it is more useful to define the average degree of the nearest neighbors of nodes with degree k

$$k_{\mathrm{nn}}(k) = \sum_{k'} k' P(k'|k). \tag{1}$$

If $k_{\mathrm{nn}}(k)$ is a decreasing function of k then we say that the network is disassortative, as happens in technological or biological networks, whereas if it is a increasing function we call it assortative, as happens in many of the social networks, where clearly the meaning is that most connected nodes tend to be connected between them and less with poorly connected nodes.

4. MODELS OF NETWORKS

4.1. The random graph model of Erdös and Renyi

This is the most simple model of graph[57]. Let us consider a set of N nodes and the probability that every two nodes are connected (form a pair or a link) is p (see Fig. 3(left)). Then the expected value of the connectivity is simply:

$$\bar{k} = p(N-1). \tag{2}$$

When considering very large networks and keeping the average value fixed, the distribution of connectivities approaches a Poisson distribution with mean λ:

$$P(k) = \frac{\lambda^k e^{-k}}{k!} \tag{3}$$

which is sharply peaked at λ, as can be seen in Fig. 2(left).

4.2. The small-world model of Watts and Strogatz

In the paper by Watts and Strogatz [1] they realized that many networks in nature had a statistical behavior that could not be fitted to that of the known results up to that time: regular lattices or random graphs. On the one hand regular lattices have very large average distance between nodes, this the so called "small-world" effect, and high average clustering coefficient, due to the high interconnection between neighbors. On the other hand, random graphs have very short average distances, due to the existence of short-cuts and very low clustering due to the random uncorrelated nature of the connections. And the conclusion that the authors got from the analysis of the networks was that they had

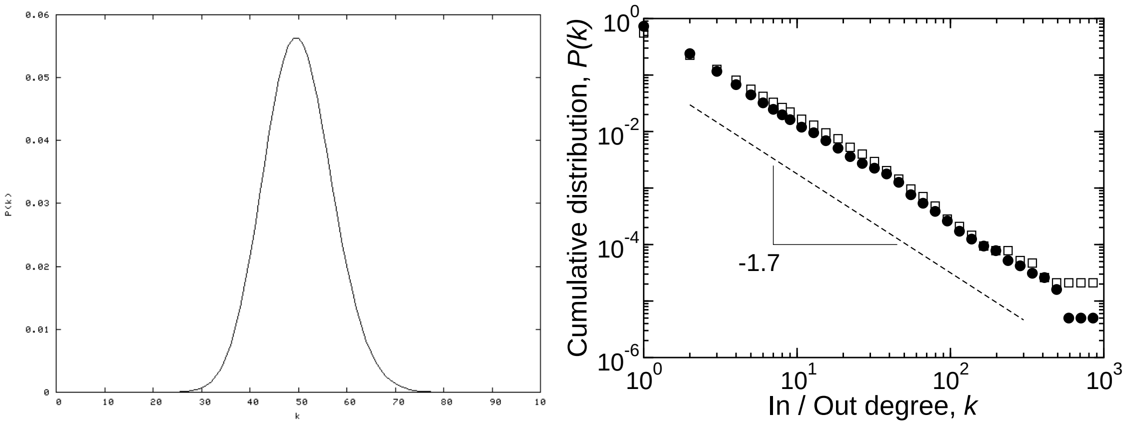

FIGURE 2. Two exemples of distribution of connectivities. Left: Poisson distribution with average 50 in a linear-linear scale. Right: Power-law cumulative distribution of in- and out-degrees (incoming and outgoing connections) in the PGP web of trust of Ref. [22].Notice that in this case the scale is log-log.

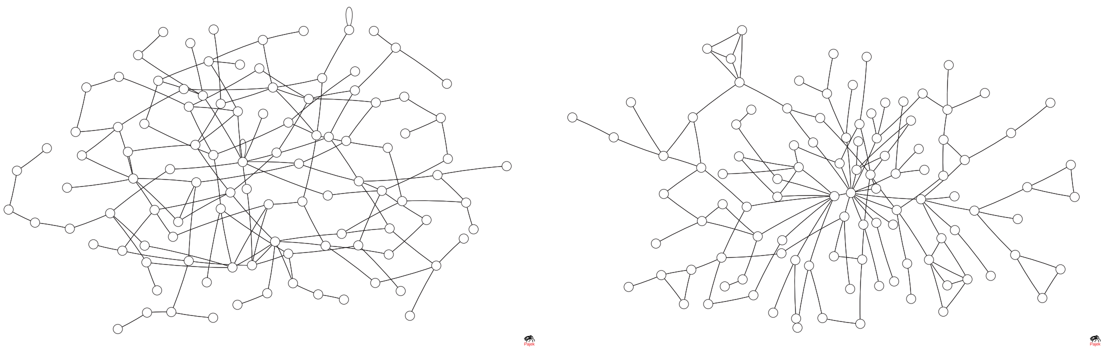

FIGURE 3. Two networks of approximately the same number (100) of nodes and links (130). Left: Erdös-Renyi Random graph. Right: Barabasi-Albert scale free network.

very short distances, like in random networks, and high clustering, like random graphs. This observation opened a completely new field of research since new models that could explain this simultaneous, and in principle opposite, behaviors were needed.

In particular, they already proposed a model, nowadays known as the Watts and Strogatz "small-world" model. To construct the network one starts with a regular one-dimensional lattice (a ring) in which all nodes are linked to their m neighbors in each direction (see Fig. 4(leftmost)); in this case the degree distribution is a delta function $\delta(k-2m)$. Then one proceeds by removing the short range links and substituting them by long-range links between two randomly chosen nodes with probability p. As can be seen in Fig. 4 by increasing p the graph loses the lattice character and resembles every time more to a random graph. Actually, with a relatively small value of p the graph acquires a short average distance between nodes without appreciably changing the clustering. In this way the observation of the simultaneous short distance and high clustering is explained by means of a very simple model.

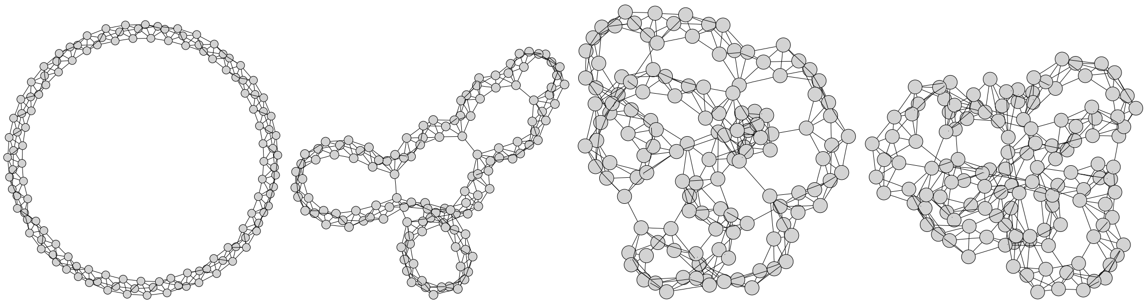

FIGURE 4. Small world model of Watts and Strogatz. Links to neighbors in the original ring are rewired with a probability 0.00, 0.05, 0.10, 0.15 (from left to right).

4.3. The scale-free model of Barabasi and Albert

Although the model introduced by Watts and Strogatz could resolve the apparent dichotomy in the observation of some regular and some random characteristics in many networks in nature, it did not change appreciably the distribution of connectivities. Starting from a delta function it rapidly evolves to a Poisson distribution for low values of p, thus the distribution of connectivities resembles that of the Erdos-Renyi model. But, just a few months later than this paper appeared, Barabasi and Albert [2] published their work in which they noticed that again many of the large networks that could be already analyzed at that time (including the Internet or the Web) showed distributions of connectivities that should be fitted to a power-law, instead of a Poisson-like as an Erdos-Renyi random graph.

In order to explain this behavior they also introduced a model, nowadays known as Barabasi-Albert model, in which there were two essential ingredients: growth and preferential attachment (this kind of attachment also gives its name to the model sometimes). On the one hand, networks are not static but are the result of a process of growing, starting from a set of a small number of completely connected nodes. On the other hand, the growth proceeds in such a way that the arriving nodes are linked preferentially to those nodes which already have more connections, as is schematically visualized in Fig. 5 (see also Fig. 3(right) for an example of such network with around 100 nodes). As can be easily interpreted from this simple rule, and also from the kind of distribution showed in Fig. 2(right), one of the main implications of this model is the existence of small fraction of highly connected nodes, named as hubs, whereas the vast majority of nodes have a very low connectivity. These hubs play a crucial role in many aspects of the network; for instance, the network is very sensitive to intentional attacks if the targets are the hubs, but is very robust under random attacks (or failures) in the case that the target is chosen at random [58, 59]. They are also important in the spreading of information or in the dynamics of synchronization, as we will see in next sections.

The finding that networks in natural or technological or social environments were scale free, showing some remarkable similarities of many other phenomena studied in the physical sciences, like critical phenomena or fractals, together with the "small-world" concept introduced by Watts and Strogatz, started the new theory of complex networks with contributions in many different fields, but with a major contribution from

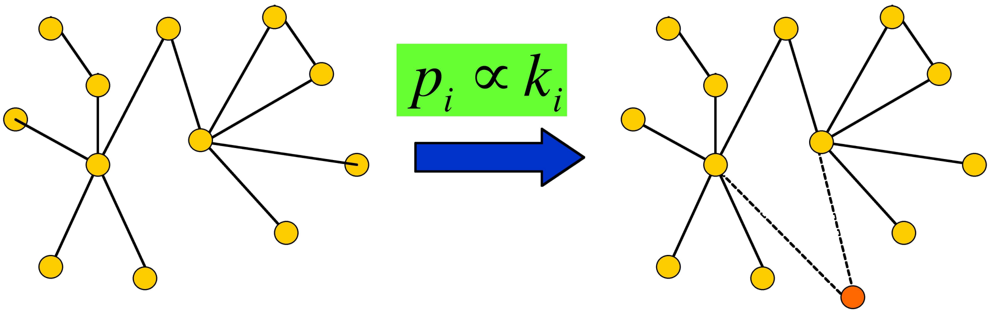

FIGURE 5. Preferential attachment rule of the Barabasi-Albert model. The arriving node is more likely to be connected to those nodes which already have more existing connections, and hence the new links correspond to the dotted lines.

the Statistical Physics community.

5. DESCRIBING THE MESOSCALE: COMMUNITIES

Clearly, in the previous sections we characterize the networks either from the microscopic or from the macroscopic point of view, but many networks show structures that are important in the intermediate scales, the mesoscale. Those structure can have different meanings depending on the origin of the network: communities in social networks[54], functional groups in biology[60], regional groups in geographically based networks, thematic clusters in the web [61, 62], and so on. Many times theses structures have an important role in their own and they have not been constructed by chance but by an ordered process of growth. For this reason identifying the communities in a network is a process from which we can gain a lot of useful information. Furthermore, dynamics is also affected by this community structure since dynamics is tightly related to the underlying topology of the network. The readers are pointed to Refs. [63, 64] for recent reviews on the subject of communities in complex networks.

Distinct modules or communities within networks can loosely be defined as subsets of nodes which are more densely linked, when compared to the rest of the network. But this is a very simple definition that cannot assure the correct identification of the groups that form the complex network.

The problem of community detection is quite challenging and has been the subject of discussion in various disciplines. A simpler version of this problem, the graph bipartitioning problem has been the topic of study in the realm of computer science for decades. In real complex networks we often have no idea how many communities we wish to discover, but in general it is more than two. This makes the process all the more costly. What is more, communities may also be hierarchical, that is communities may be further divided into sub-communities and so on [16, 21, 15, 65].

Nevertheless, many attempts to tackle these problems have been proposed recently. The proposed methods vary considerably in terms of approach and application, which makes them difficult to compare. Community identification is potentially very useful and researchers from a number of fields may be interested in using one or several of the methods for their own purposes. In [66] we review all these methods comparing their performance and their computational cost.

But community identification is not merely a qualitative problem; actually, the performed comparison between the different algorithms is done in terms of a quantity that measures how good a given partition is. Since communities are sometimes not perfectly defined with clear border-line separation among them, different algorithms to detect communities can give rise to slightly different partitions. Them a measure that quantifies the accuracy of the partition is welcome. A simple approach that has become widely accepted was proposed in [67]. It is based on the intuitive idea that random networks do not exhibit community structure. Let us imagine that we have an arbitrary network and an arbitrary partition of that network into n_c communities. It is then possible to define a $n_c \times n_c$ size matrix $\mathbf{e}$ where the elements e_{ij} represent the fraction of total links starting at a node in partition i and ending at a node in partition j. Then, the sum of any row (or column) of $\mathbf{e}$, $a_i = \sum_j e_{ij}$ corresponds to the fraction of links connected to i.

If the network does not exhibit community structure, or if the partitions are allocated without any regard to the underlying structure, the expected value of the fraction of links within partitions can be estimated. It is simply the probability that a link begins at a node in i, a_i, multiplied by the fraction of links that end at a node in i, a_i. So the expected number of intra-community links is just $a_i a_i$. On the other hand we know that the *real* fraction of links exclusively within a partition is e_{ii}. So, we can compare the two directly and sum over all the partitions in the graph.

$$Q \equiv \sum_i (e_{ii} - a_i^2) \tag{4}$$

This is the measure known as *modularity*, that for a very good partition approaches 1. It is important to say that the network can have a very clear community separation and then a good partition can attain a large value of the modularity.

But sometimes, we are not only interested in the best partition but in the hierarchical organization of the network in nested communities. One of the early methods of community detection, proposed by Girvan and Newman [68], consists in splitting the networks by cutting the links with the highest betweeness. In this case this procedure can be iterated up to the level of individual nodes giving rise then to a hierarchy of nested communities. The application of this procedure is very useful for the understanding on the different levels of organization in a network. We have applied this procedure to the email network of the Universitat Rovira i Virgili [16] finding that the hiercahical organization of the community structure maintains many treats of the supposed formal chart of the organization; but, at the same time, we could observe that some nodes are not placed in the supposed community. This is of course very valuable as a tool for the management of a organization [53]. Also as a tool of identifying the working communities and the respective leaders we applied the procedure to the Statistical Physics meetings network shown in Fig. 1. The network in Fig. 6 is the result of such community partition, where we can see that the green nodes, identified as the members of the scientific committees are equally distributed between the different branches and appear mainly at their tips. The former means that members have been chosen in a homogeneous way between the

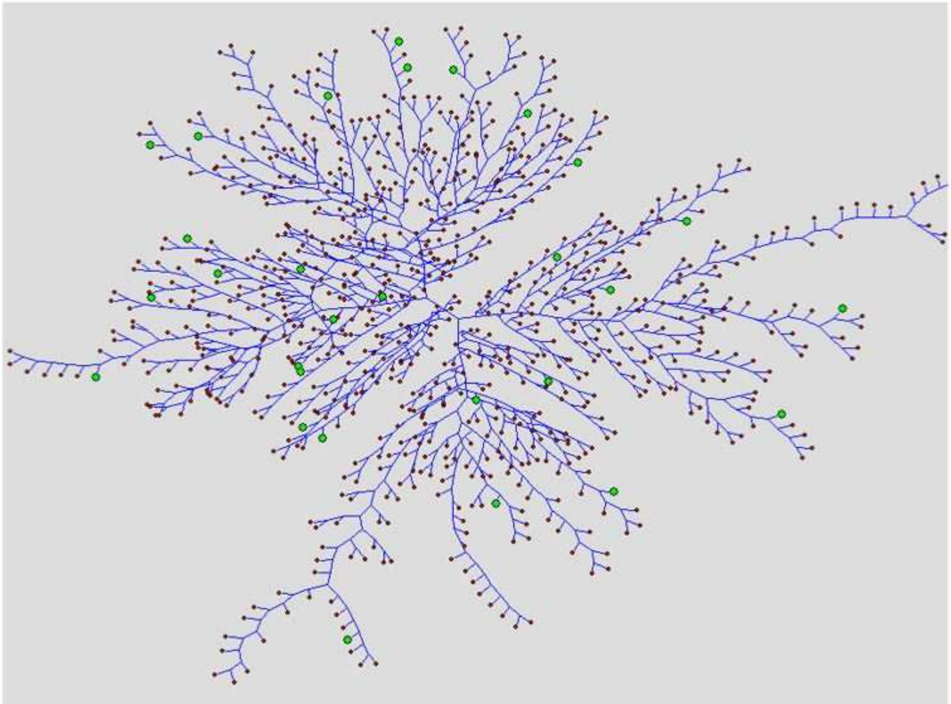

FIGURE 6. Community structure of the collaboration network in the Spanish Statistical Physics meetings. The small branches correspond to the research groups that are grouped into Universities that, at the same time, are closely grouped according to geographical proximity. The green nodes, that correspond to the members of the scientific committees appear mainly at the tips of the branches, showing their leadership in the respective groups. The homogeneous distribution of green nodes also shows that they have been chosen uniformly among the different groups.

different groups that form the Statistical Physics community and that these members are the leaders of the respective teams.

Another fact that has been obtained from this hierarchical community structure is that in many networks the distribution of community sizes also shows a power-law, indicating an underlying mechanism of auto-organization in the network and the absence of characteristic community sizes. In this way another scaling of magnitudes within communities can be analyzed and hence, in a language very familiar to physicists, networks can be classified in different universality classes [16, 15].

6. DYNAMICS ON THE NETWORK

Complex networks have become such widespread analyzed not only because of their universal topological properties, but also because the effect of the topology on the dynamics. Dynamical systems had been largely studied mainly in three different playgrounds: regular lattices, random graphs, and completely connected networks. Thus the evidence of the existence in nature and society of complex patterns of interaction again offered a large number of new possibilities to those studying the dynamical properties of complex systems. And hence, many different types of dynamics have been studied according to different patterns of connectivity. Just to mention a few in different contexts: flow of physical magnitudes or information in communication networks [56], spreading of epidemics [69, 20, 19] or rumors[70], synchronization of dynamical units (mainly oscillators) [71, 7], opinion formation [72], cultural dissemination [73], technological innovations [74], strategic games [75], Boolean dynamics in genetic networks [51], neural networks [1, 76, 77, 48].

Just to present a comprehensive view of these phenomena we will show results on two different types of dynamics: search and congestion as an example of transport in networks, and the dynamics of oscillators towards synchronization since it is a good example on how dynamics can help in elucidating some details of the topology.

6.1. Search and congestion

Concerning transport, the flow of information has been one of the mainly discussed issues. Information, in this case, can be understood as packets in a computer network [78], problems in a company that need to be solved [55], passengers in a transportation network [79]. As an example of information flow in [55] we presented a formalism that is able to cope with search and congestion simultaneously in any type of network, allowing the determination of optimal topologies. This formalism avoids the problem of simulating the dynamics of the communication process and provides a general scenario applicable to any communication process.

Let us focus on a single information packet at node i whose destination is node k. The probability for the packet to go from i to a new node j in its next movement is p^k_{ij}. In particular, $p^k_{kj} = 0 \; \forall j$ so that the packet is *removed* as soon as it arrives to its destination. This formulation is completely general, and the precise form of p^k_{ij} will depend on the search algorithm and on the connectivity matrix of the network. In particular, when the search is Markovian, p^k_{ij} does not depend on previous positions of the packet. In this case, the probability of going from i to j in n steps is given by

$$P^k_{ij}(n) = \sum_{l_1,l_2,\ldots,l_{n-1}} p^k_{il_1} p^k_{l_1 l_2} \cdots p^k_{l_{n-1} j}. \tag{5}$$

This definition allows us to compute the average number of times, b^k_{ij}, that a packet generated at i and with destination at k passes through j.

$$b^k = \sum_{n=1}^{\infty} P^k(n) = \sum_{n=1}^{\infty} \left(p^k\right)^n = (I - p^k)^{-1} p^k. \tag{6}$$

and the effective betweenness of node j, B_j, is then defined as the sum over all possible origins and destinations of the packets,

$$B_j = \sum_{i,k} b^k_{ij}. \tag{7}$$

When the search algorithm is able to find the minimum paths between nodes, the effective betweenness will coincide with the topological betweenness, β_j, as usually defined in the previous sections [80, 81].

Once these quantities have been defined, we focus on the load of the network, $N(t)$, which is the number of floating packets. These floating packets are stored in the nodes that act as queues. In a general scenario where packets are generated at random and independently at each node with a probability ρ, the arrival of packets to a given node j is a Poisson process. In this simple picture, the queues are called M/M/1 in the computer science literature and the average load of the network is [82, 55]

$$\overline{N} = \sum_{j=1}^{S} \frac{\frac{\rho B_j}{S-1}}{1 - \frac{\rho B_j}{S-1}}. \tag{8}$$

There are two interesting limiting cases of equation (8). When ρ is very small, taking into account that the sum of betweennesses is proportional to the average distance, one obtains that the load is proportional to the average effective distance. On the other hand, when ρ approaches ρ_c most of the load of the network comes from the most congested node, and therefore

$$\overline{N} \approx \frac{1}{1 - \frac{\rho B^*}{S-1}} \qquad \rho \to \rho_c, \tag{9}$$

where B^* is the effective betweenness of the most central node. The last results suggest the following interesting problem: to minimize the load of a network it is necessary to minimize the effective distance between nodes if the amount of packets is small, but it is necessary to minimize the largest effective betweenness of the network if the amount of packets is large. The first is accomplished by a *star-like* network, that is, a network with one central node and all the others connected to it. The second, however, is accomplished by a very decentralized network in which all the nodes support a similar load. This behavior is similar to any system of queues provided that the communication depends only on the sender.

It is worth noting that there are only two assumptions in the calculations above. The first one has already been mentioned: the movement of the packets needs to be Markovian to define the jump probability matrices p^k. Although this is not strictly true in real communication networks—where packets are not usually allowed to go through a given node more than once—it can be seen as a first approximation [78, 83, 84]. The second assumption is that the jump probabilities p^k_{ij} do not depend on the congestion state of the network, although communication protocols sometimes try to avoid congested regions, and then $B_j = B_j(\rho)$. However, all the derivations above will still be true in a number of general situations, including situations in which the paths that the packets follow are unique, in which the routing tables are fixed, or situations in which the structure of the network is very homogeneous and thus the congestion of all the nodes is similar. Compared to situations in which packets avoid congested regions, it corresponds to the worst case scenario and thus provide bounds to more realistic scenarios in which the search algorithm interactively avoids congestion.

Equation (8) relates a dynamical variable, the load, with the topological properties of the network and the properties of the algorithm. So we have converted a dynamical

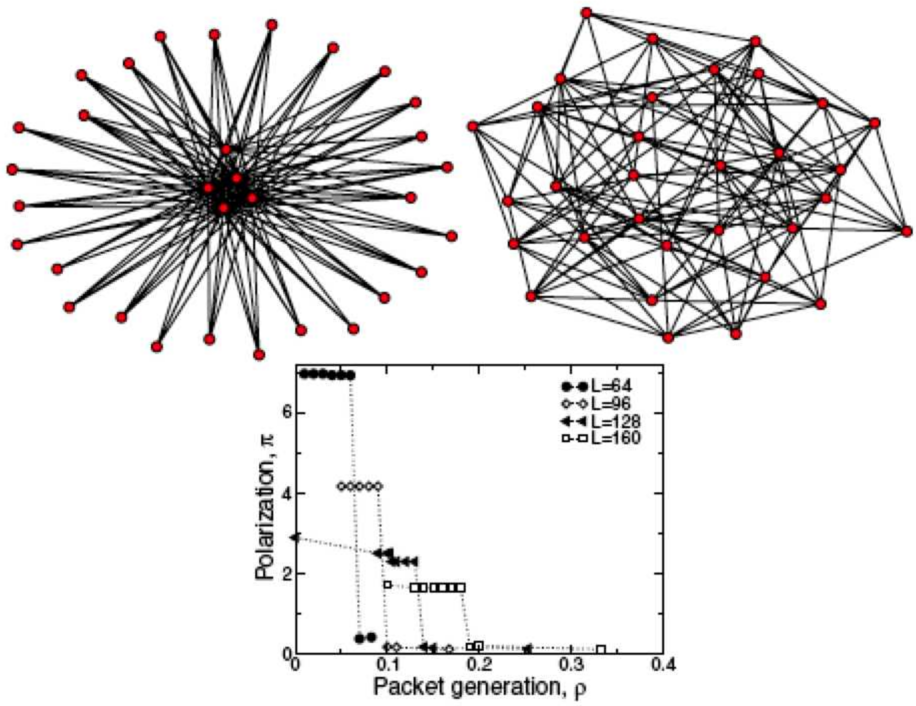

FIGURE 7. Optimal topologies for networks with $S = 32$ nodes, $L = 32$ links.

communication problem into a topological problem. Hence, the dynamical optimization procedure of finding the structure that gives the minimum load is reduced to a topological optimization procedure where the network is characterized completely by its effective betweenness distribution. In [55] we considered the problem of finding optimal structures for a purely local search, using a generalized simulated annealing procedure, as described in [85]. On the one side, we have found (see Fig. 7) that for $\rho \to 0$ the optimal network has a star-like centralized structure as expected, which corresponds to the minimization of the average effective distance between nodes. On the other extreme, for high values of ρ, the optimal structure has to minimize the maximum betweenness of the network; this is accomplished by creating a homogeneous network where all the nodes have essentially the same degree, betweenness, etc. One could expect that the transition centralized-decentralized occurs progressively. Surprisingly, the results of the optimization process reveal a completely different scenario. According to simulations, star-like configurations are optimal for $\rho < \rho^*$; at this point, the homogeneous networks that minimize B^* become optimal. Therefore there are only two type of structures that can be optimal for a local search process: star-like networks for $\rho < \rho^*$ and homogeneous networks for $\rho > \rho^*$.

Beyond the existence of both centralized and decentralized optimal networks, it is significant that the transition from one sort of networks to the other is abrupt, meaning that there are no intermediate optimal structures between total centralization and total decentralization. Our explanation of this fact is the following. Since we are con-

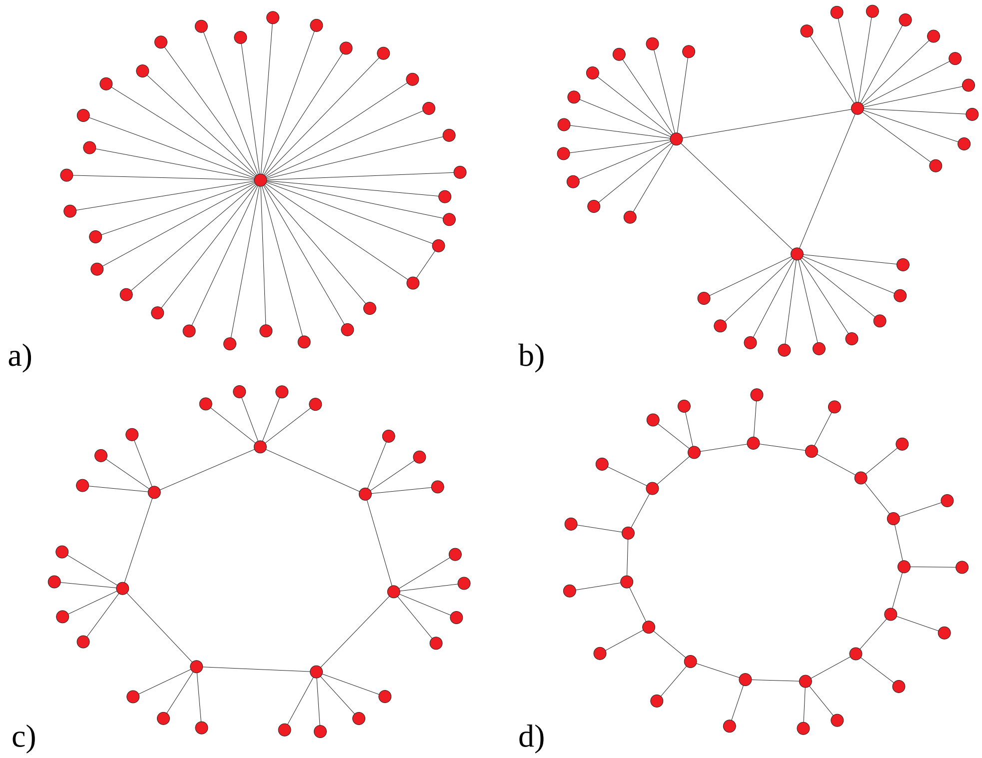

FIGURE 8. Optimal topologies for networks with $S = 32$ nodes, $L = 32$ links and global knowledge. (a) $\rho = 0.010$. (b) $\rho = 0.020$. (c) $\rho = 0.050$. (d) $\rho = 0.080$. In this case of global knowledge, the transition from centralization to decentralization seems smooth.

sidering local knowledge of the network topology, centered star-like configurations are extremely efficient in searching destinations and thus minimizing the effective distance between nodes. This explains that stars are optimal for a wide range of values of ρ, until the central node (or nodes) becomes congested. At this point, structures similar to stars will have the same problem and will be much worse regarding search; at this point, the only alternative is something completely decentralized, where the absence of congestion can compensate the dramatic increase in the effective distance between nodes. If this explanation is correct, one should be able to obtain a smooth transition from centralization to decentralization by considering global knowledge of the network, in such a way that the average effective distance (that in this case coincides with the average path length) is not much larger in an arbitrary network than in the star. Although we do not have extensive simulations in this case, Fig. 8 shows that there is some evidence to think that this is indeed the case.

6.2. Dynamics towards synchronization

Physicists have largely studied the dynamics of complex biological systems, and in particular the paradigmatic analysis of large populations of coupled oscillators [86, 87, 88]. The connection between the study of synchronization processes and complex networks is interesting by itself. This synchronization phenomena as many others e.g. asian fireflies flashing at unison, pacemaker cells in the heart oscillating in harmony, etc. have been mainly described under the mean field hypothesis that assumes that all oscillators behave identically and interact with the rest of the population. Recently, the emergence of synchronization phenomena in complex networks has been shown to be closely related to the underlying topology of interactions [89] beyond the macroscopic description.

One of the most successful attempts to understand synchronization phenomena was due to Kuramoto [88], who analyzed a model of phase oscillators coupled through the sine of their phase differences. The model is rich enough to display a large variety of synchronization patterns and sufficiently flexible to be adapted to many different contexts [90]. The Kuramoto model consists of a population of N coupled phase oscillators where the phase of the i-th unit, denoted by $\theta_i(t)$. Here we consider a simplified dynamics in which all units have the same frequency, that can be set to zero without loss of generality. Thus we have

$$\frac{d\theta_i}{dt} = \sum_j K_{ij} \sin(\theta_j - \theta_i) \quad i = 1, ..., N \tag{10}$$

where K_{ij} describes the coupling between units. In absence of noise the only attractor of the dynamics is the complete synchronization, $\theta_i = \theta,\ \forall i$.

Originally, this model had been studied in networks which are complete, but recently these studies have been extended to systems where the patterns of connections is local but non-trivial [7]. In this context the interest concerns not the final synchronized state in itself but the route to the attractor. In particular, it has been shown [7] that high densely interconnected sets of oscillators (motifs) synchronize more easily that those with sparse connections. This scenario suggests that for a complex network with a non-trivial connectivity pattern, starting from random initial conditions, those highly interconnected units forming local clusters will synchronize first and then, in a sequential process, larger and larger spatial structures also will do it up to the final state where the whole population should have the same phase. This process occurs at different time scales if a clear community structure exists. Thus, the dynamical route towards the global attractor reveals different topological structures, presumably those which represent communities. Therefore, it is the complete dynamical process what unveils the whole organization at all scales, from the microscale at a very early stages up to the macroscale at the end of the time evolution. On the contrary, those systems endowed with a regular topological structure displays a trivial dynamics with a single time scale for synchronization.

We have analyzed the dynamics towards synchronization in computer-generated graphs with community structure. For this reason, we define a local order parameter measuring the average of the correlation between pairs of oscillators

$$\rho_{ij}(t) = < cos(\theta_i(t) - \theta_j(t)) > \tag{11}$$

where the brackets stand for the average over initial random phases. The main advantage of this approach is that it allows to trace the time evolution of pairs of oscillators and therefore to identify compact clusters reminiscent of the existence of communities.

The paradigmatic model of network with a well defined community structure that has been used as a benchmark for different community detection algorithms [66], was proposed by Girvan and Newman [68]. In that model the authors construct a network of 128 nodes as a set of 4 communities, each one formed by 32 nodes. Fixing the mean number of links per node at a value of 16, the parameter describing the sharpness of the community distribution is z_{in}, the average number of links within the community. In Fig. 9 we show the time evolution of one of these networks, $z_{in} = 15$ and hence a very clearly defined community structure, averaging over random initial phases.

Dealing only with topological information we can, from the connectivity matrix, construct the Laplacian matrix and compute their eigenvalue spectrum. This spectrum gives information on the time scales involved in the dynamical process. We plot the eigenvalues spectrum of this matrix in the following way: in the horizontal axis we represent the inverse of the eigenvalue, which in a dynamical process accounts for the time, and in the vertical axis we represent the index of the eigenvalue which accounts for the number of groups along the dynamics. This picture is useful because it can be compared with the way groups (clusters or communities) are formed along the synchronization process, obtaining a very striking similarity, meaning that these eigenvalues control the formation of the synchronized communities. We also plot, for completion, the dendogram of the synchronization process (Fig. 9c): In this picture we show how the groups merge according to the synchronization dynamics along time (vertical axis). Finally we also plot (d) the relative time to achieve synchronization for each pair of oscillators. This synchronization is understood as a correlation being larger than some threshold value. The characterization is completely independent of the threshold, as is shown in [91], since it only changes the absolute time scale not the relative one. Nodes are ordered in the same way than in the picture of the dendogram just to get toghether those nodes that synchronize earlier.

In this way we have been able to relate topology, in terms of the eigenvalue spectrum of the Laplacian matrix, with dynamics, in terms of the appearance of synchronized groups of oscillators. Topologically these groups correspond to the communities, but there can be some cases where communities are not so well defined and this informations keep being useful. There can be some occasions where synchronized groups of oscillators do not fit exactly with topological communities. Synchronization is a global dynamical process that can identify the relevant structures (perhaps hierarchical) along its evolution. Also the effect of hubs in the dynamical evolution is interesting, since hubs are sometimes above the community structure.

For more information on this issue the reader is pointed to [91, 92] and to the website `http://www.ffn.ub.es/albert/synchro.html`.

7. CONCLUSIONS AND OPEN PROBLEMS

Complex patterns of interactions are so often found in any natural, technological or social environment that it has been widely accepted that new tools are needed. From Statistical Physics, many valuable existing tools have been applied to this new emergent field. Researchers in many different subjects are generating new repositories of data, very large networks are generated and these tools need new implementations. A network

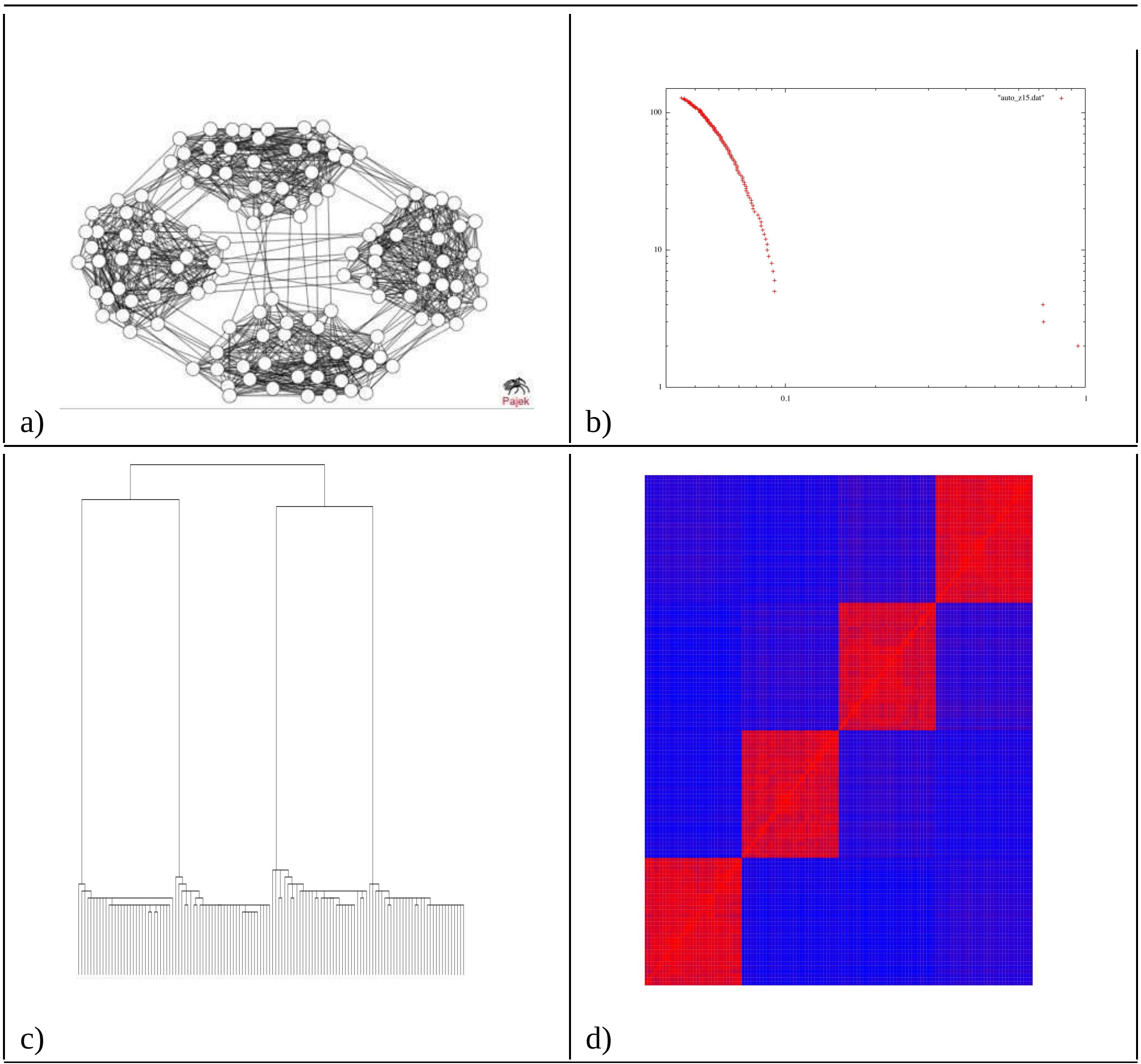

FIGURE 9. Synchronization process in a network with a homogeneous distribution of communities. a) the network structure; b) eigenvalue spectrum; c) dendogram of the community merging; d) time needed for each pair of oscillators to synchronize. Red for shorter times, blue for larger times.

is not just a collection of nodes and binary relationships between those nodes. Nodes and links can be anything, depending on the considered data, but nodes can have weights, links can have weights as well, and hence new theories have appeared to deal with this additional degree of complexity.

Usually, networks are characterized either from the microscopic level or from the macroscopic level. From a microscopic point of view we are mainly interested in node properties: degree, different measures of centrality, clustering, and so. However, from a macroscopic point of view we deal with statistical properties of the set of nodes and/or links; which are the distributions of connectivities, of load, of distances, and how the different measures are, on average, correlated. These characterizations enable to classify the networks into different universality classes, which is quite common in physics grounds. We also know that in many problems in physics we have descriptions that are scale invariant and hence we can move from the microscopic to the macroscopic scale. Here we have reviewed some concepts and methods in the intermediate scale, the mesoscale, where the definition and identification of communities or functional groups play a crucial role. Up to now, there has been a large amount of work on methods of community identification. Which are the most efficient in terms of accuracy or which are the more economic in terms of computer resources needed. These properties have also turned out to show some degree of universality.

Nevertheless, this identification based solely on topological properties needs to be related with the exact relations between the nodes of the different groups. Nodes can belong topologically to a given group but their functionality can be quite different. Understanding these relations, why topological communities are or are not related with functional groups, social communities, or some sort of thematic clusters, is still one of the open problems related with the mesoscale properties. Another interesting point that needs more clarification is the community structure at different scales, why are they ordered in some kind of hierarchical or nested way and their relation again with some ordering in this scales that can be related with some dynamical properties of the processes taking place on the network. This hierarchical structure goes far beyond many of the current methods to identify community partitions in networks; all this methods try to find the optimal value of a kind of cost function, called modularity, which is a property of the network and of the partition, then the best partition is that with the highest modularity, but there can be partitions that, even with a high value of modularity, are very unlikely from a physical point of view. Hence a proper understanding of the precise location and the neighboring areas in the partition space of special configurations can be of great help in understanding the functionality of networks.

But, at the same time, networks are not formed by static objects; nodes (social agents, computers, companies, ...) evolve in time and they can change their status and this evolution is strongly correlated with the evolution of the links (social relationships, hard rewirings, new business strategies, ...), All these new evolving, rewiring, updating, growing, removing, ... open many new problems that will be faced in the next future. Also, as stated in the previous paragraph, we need a proper understanding of the topologies and its relation with the dynamics of the node properties. We have presented here just two examples on how the topological structure affects dynamics. First, a problem of transport in which the nodes are agents that process and deliver information that has to arrive to the right destination. Here we have found the characteristics of the optimal network depending on the external load. Second, the time evolution of synchronized populations of oscillators shows a striking degree of community ordering that reflects the topological structure; furthermore, we have highlighted the relations between topological properties of the connectivity matrix with dynamical properties of the synchronization. This is just to get a glance on the wide applicability of these ideas in physical, economical, social, biological, or even engineering problems.

In any case, we are dealing with a subject, Complex Networks, that is very young, but that in such a short period of time has given so many relevant contributions (in the form of reviews, technical books, popularization books, ...) that we have to think that the future has just started and many new players are welcome to the ground.

ACKNOWLEDGMENTS

The author gratefully acknowledges fruitful and enlightening discussions with L.A.N. Amaral, A. Arenas, M. Boguña, L. Danon, J. Duch, X. Guardiola, R. Guimera, M. Llas, C.J. Perez-Vicente. I would like also to thank L. Danon and J. Duch for providing some of the data and figures.

REFERENCES

1. D. J. Watts, and S. Strogaz, *Nature* **393**, 440–442 (1998).
2. A. L. Barabási, and R. Albert, *Science* **286**, 509–512 (1999).
3. S. H. Strogatz, *Nature* **410**, 268–276 (2001).
4. A. L. Barabási, and R. Albert, *Review of Modern Physics* **74**, 47–97 (2002).
5. S. Dorogovtsev, and J. F. F. Mendes, *Advances in Physics* **51**, 1079–1187 (2002).
6. M. E. J. Newman, *SIAM Review* **45**, 167–256 (2003).
7. S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Physics Reports* **424**, 175–308 (2006).
8. S. Bornholdt, and H. G. Schuster, editors, *Handbook of Graphs and Networks - From the Genome to the Internet*, Wiley-VCH, Berlin, 2002.
9. R. Pastor-Satorras, M. Rubí, and A. Díaz-Guilera, editors, *Statistical Mechanics of Complex Networks*, Springer, 2003.
10. L. Amaral, A. Scala, M. Barthelemy, and H. Stanley, *Proceedings of the National Academy of Sciences, USA* **97**, 11149–11152 (2000).
11. G. F. Davis, M. Yoo, and W. E. Baker, *preprint, University of Michigan Business School* (2001).
12. M. E. J. Newman, S. Strogatz, and D. J. Watts, *Physical Review E* **64**, 026118 (2001).
13. M. E. J. Newman, *Physical Review E* **64**, 016132 (2001).
14. M. E. J. Newman, *Proceedings of the National Academy of Sciences, USA* **98**, 404–409 (2001).
15. A. Arenas, L. Danon, A. Diaz-Guilera, P. M. Gleiser, and R. Guimerà, *European Physical Journal B* **38**, 373–380 (2004).
16. R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas, *Physical Review E* **68**, 065103 (2003).
17. H. Ebel, L. I. Mielsch, and S. Bornholdt, *Physical Review E* **66**, 035103 (2002).
18. M. E. J. Newman, S. Forrest, and J. Balthrop, *Physical Review E* **66**, 035101 (2002).
19. F. Liljeros, C. R. Edling, and L. A. N. Amaral, *Microbes and Infections* **5**, 189–196 (2003).
20. F. Liljeros, C. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Aberg, *Nature* **411**, 907–908 (2001).
21. P. Gleiser, and L. Danon, *Advances in Complex Systems* **6**, 565–573 (2003).
22. X. Guardiola, R. Guimerà, A. Arenas, A. Diaz-Guilera, D. Streib, and L. Amaral, *preprint* pp. cond–mat/0206240 (2002).
23. R. Albert, H. Jeong, and A.-L. B. ., *Nature* **401**, 130 (1999).
24. S. Redner, *European Physical Journal B* **4**, 131–134 (1998).
25. S. N. Dorogovtsev, and J. F. F. Mendes, *Proceedings of the Royal Society, London B* **268**, 2603–2608 (2001).
26. R. F. i Cancho, and R. Solé, *Proceedings of the Royal Society London B* **268**, 2261–2265 (2001).
27. M. Faloutsos, P. Faloutsos, and C. Faloutsos, *Comp. Comm. Rev.* **29**, 251–262 (1999).
28. M. E. J. Newman, *Physical Review E* **67**, 026126 (2003).
29. S. Valverde, R. F. i Cancho, and R. Solé, *Europhysics Letters* **60**, 512–517 (2002).
30. R. F. i Cancho, C. Janssen, and R. Solé, *Physical Review E* **64**, 046119 (2001).
31. R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, *Proceedings of the National Academy of Sciences, USA* **102**, 7794–7799 (2005).
32. V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani, *Proceedings of the National Academy of Sciences, USA* **103**, 2015 (2006).
33. P. Sen, S. Dasgupta, A. Chatterjee, P. A. Sreeram, G. Mukherjee, and S. S. Manna, *Physical Review E* **67**, 036106 (2003).

34. H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási, *Nature* **407**, 651–654 (2000).
35. S. M. Gomez, S. H. Lo, and A. Rzhetsky, *Genetics* **159**, 1291–1298 (2001).
36. O. Ebenhoh, and R. Heinrich, *Bulletin of Mathematical Biology* **65**, 323–57 (2003).
37. S. Schuster, T. Pfeiffer, F. M. I. Koch, and T. Dandekar., *Bioinformatics* **18**, 351ñ61 (2002).
38. A. Wagner, and D. A. Fell, *Proceedings of the Royal Society London B* **268**, 1803–10 (2001).
39. H. Jeong, S. Mason, A. L. Barabási, and Z. N. Oltvai, *Nature* **411**, 41–42 (2001).
40. D. S. Goldberg, and F. P. Roth, *Proceedings of the National Academy of Sciences, USA* **100**, 4372–76 (2003).
41. M. Vendruscolo, N. V. Dokholyan, E. Paci, and M. Karplus, *Physical Review E* **65**, 061910 (2002).
42. A. Wagner, *Molecular Biology and Evolution* **18**, 1283–92 (2001).
43. S. Wuchty, *Proteomics* **2**, 1715–23 (2002).
44. J. A. Dunne, R. J. Williams, and N. D. Martinez, *Proceedings of the National Academy of Sciences, USA* **99**, 12917–22 (2002).
45. J. M. Montoya, and R. V. Solé, *Journal of Theoretical Biology* **214**, 405–12 (2002).
46. S. Morita, K. Oshio, Y. Osana, Y. Funabashi, K. Oka, and K. K. ., *Physica A* **298**, 553–61 (2001).
47. O. Shefi, I. Golding, R. Segev, E. B.-J. E, and A. Ayali, *Physical Review E* **66**, 021905 (2002).
48. V. M. Eguíluz, D. Chialvo, G. Cecchi, M. Baliki, and A. Apkarian, *Physical Review Letters* **92**, 028102 (2005).
49. A. Bhan, D. J. Galas, and T. G. Dewey, *Bioinformatics* **18**, 1486–1493 (2002).
50. N. Guelzim, S. Bottani, P. Bourgine, and F. Kepes, *Nature Genetics* **31**, 60–63 (2002).
51. L. A. N. Amaral, A. Díaz-Guilera, A. A. Moreira, A. L. Goldberger, and L. A. Lipsitz, *Proceedings of the National Academy of Science* **101**, 15551–15555 (2004).
52. A. Barrat, M. Barthelemy, R. Pastor-Satorras, and A. Vespignani, *Proceedings of the National Academy of Science* **101**, 3747 (2004).
53. R. Guimerà, L. Danon, A. Arenas, A. Díaz-Guilera, and F. Giralt, *Journal of Economic Behavior and Organization* (2007).
54. S. Wasserman, and K. Faust, *Social Network Analysis, Methods and Applications*, Cambridge University Press, 1994.
55. R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, *Physical Review Letters* **89**, 248701 (2002).
56. B. Tadic, G. Rodgers, and S. Thurner, *preprint* (2006).
57. P. Erdos, and A. Renyi, *Publ. Math. Debrecen* **6**, 290–297 (1959).
58. R. Albert, H. Jeong, and A.-L. Barabasi, *Nature* **406**, 378 (2000).
59. R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, *Physical Review Letters* **85**, 4626 (2000).
60. H. Zhou, and R. Lipowsky, *preprint* (2005).
61. G. W. Flake, S. Lawrence, C. L. Giles, and F. M. Coetzee, *IEEE Computer* **35**, 66 – 71 (2002).
62. J.-P. Eckmann, and E. Moses, *Proceedings of the National Academy of Sciences, USA* **99**, 5825–5829 (2002).
63. M. E. J. Newman, *European Physical Journal B* **38**, 321–330 (2004).
64. L. Danon, J. Duch, A. Arenas, and A. Díaz-Guilera, *COSIN project*, World Scientific, 2005, chap. Community structure identification.
65. M. E. J. Newman, *Physical Review E* **69**, 066133 (2004).
66. L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas, *J. Stat. Mech* p. P09008 (2005).
67. M. E. J. Newman, and M. Girvan, *Physical Review E* **69**, 026113 (2004).
68. M. Girvan, and M. E. J. Newman, *Proceedings of the National Academy of Sciences USA* **99**, 7821–7826 (2002).
69. R. Pastor-Satorras, and A. Vespignani, *Physical Review Letters* **86**, 3200–3203 (2001).
70. D. H. Zanette, *Physical Review E* **64**, 050901 (2001).
71. L. Donetti, P. I. Hurtado, and M. A. Muñoz, *Physical Review Letters* **95**, 188701 (2005).
72. F. A. Rodrigues, and L. da F. Costa, *International Journal of Modern Physics C* **16**, 1785–1792 (2005).
73. K. Klemm, V. M. Eguíluz, R. Toral, and M. San Miguel, *Physical Review E* **67**, 026120 (2003).
74. M. Llas, P. M. Gleiser, A. Díaz-Guilera, and C. J. Pérez, *Physica A* **326**, 567–577 (2003).
75. H. Ebel, and S. Bornholdt, *Physical Review E* **66**, 056118 (2002).
76. L. F. Lago-Fernández, R. Huerta, F. Corbacho, and J. A. Sigüenza, *Physical Review Letters* **84**, 2758–2761 (2000).

77. M. Aldana, and H. Larralde, *Physical Review E* **70**, 066130 (2004).
78. T. Ohira, and R. Sawatari, *Physical Review E* **58**, 193 (1998).
79. M. Barthelemy, and A. Flammini, Optimal traffic networks (2006).
80. L. C. Freeman, *Sociometry* **40**, 35–41 (1977).
81. M. E. J. Newman, *Physical Review E* **64**, 016133 (2001).
82. O. Allen, *Probability, Statistics and Queueing Theory with Computer Science Application*, Academic Press, New York, 2nd edition,, 1990.
83. A. Arenas, A. Diaz-Guilera, and R. Guimera, *Physical Review Letters* **86**, 3196–3199 (2001).
84. R. Sole, and S. Valverde, *Physica A* **289**, 595–605 (2001).
85. C. Tsallis, and D. A. Stariolo, *Annual Rev. Comp. Phys. II*, World Sci. Singapore, 1994.
86. A. Winfree, *The geometry of biological time*, Springer, 2001.
87. S. H. Strogatz, *Sync: The Emerging Science of Spontaneous Order*, Hyperion, 2003.
88. Y. Kuramoto, *Chemical oscillations, waves, and turbulence*, Dover, 2003.
89. F. M. Atay, T. Biyikoglu, and J. Jost., *IEEE Trans. Circuits and Systems* **53** (2006).
90. J. A. Acebrón, L. L. Bonilla, C. J. Pérez Vicente, F. Ritort, and R. Spigler, *Reviews of Modern Physics* **77**, 137–185 (2005).
91. A. Arenas, A. Diaz-Guilera, and C. J. Perez-Vicente, *Physical Review Letters* **96**, 114102 (2006).
92. A. Arenas, A. Diaz-Guilera, and C. J. Perez-Vicente, *Physica D* **96**, (submitted) (2007).