

Chapter 2

Mathematical Foundations: Complex Networks and Graphs (A Review)

Abstract It is possible that the main approach to capture the global properties of complex systems is to model them as networks (graphs) whose nodes represent the units, and whose links stand for the interactions between them. This chapter is devoted to establish the main needed concepts on Graph Theory and Complex Networks we will use in building of our mathematical model.

2.1 Introduction

2.1.1 *Complex Systems and Complex Networks*

Although the concept of a network roots back to Pythagoras (fifth/sixth centuries BC) in his theory of cosmos (κᾶσμος), the first book in networks appeared in 1936 (D. König: *Theory of Finite and Infinite graphs* (see [119])). The analysis of networks had one of its most critical and exciting moments in 1999 with the discovery of new types of graphs (small-world networks and scale-free networks) so-called complex networks. From that moment, complex networks have been used for modeling different systems of the real world and have attracted considerable interest due mainly to its study has been found to be very productive in science and technology. Many complex systems of the real world can be modeled using complex networks where nodes represent the different constituents of the system and edges depict the interactions between them. Different systems such as transport networks (underground, airline networks, road networks), communication networks (computer servers, internet), biochemical networks (metabolic, protein and genomic networks), social networks, infrastructure networks (electric power grids, water supply networks) and some others (including the World Wide Web) are known to have common characteristics in their behaviour and structure [6, 7, 16, 22, 54, 62, 77, 102–104, 118].

The huge complexity of these objects stemming from their size and dynamics, requires new tools beyond classical graph theory, which have crystallized in the scientific area known as *complex network analysis*, that involves not only mathematical tools (including probability, dynamical system analysis, graph theory, matrix analysis and others), but also techniques coming from other fields (as statistical mechanics or computer sciences, to name a couple of them). So, the investigation on

complex networks has attracted an increasing amount of attention from numerous research fields due to its applications to model complex systems within the real world. In the background of the elegant and efficient descriptions of very different complex systems is the general framework of modern graph theory [70, 72] initiated by Erdős [56].

Because of this reason complex systems can be studied using non-linear mathematical models and computer modeling approaches. In fact, complex networks provide a natural alternative for representing, characterizing and modeling the structure and non-linear dynamics of all discrete complex systems.

The study of structural properties of the underlying network is very important in the understanding of the functions of a complex system as well as, for example, to quantify the strategic importance of a node (or set of nodes) in order to preserve the best functioning of the network as a whole. The improvements in computers performance in the last decades granted us the ability to analyze huge complex networks.

The model introduced throughout joins together all the research on intentional attack risk modelled from complex networks concepts and tools and from previous research based on information accessibility (its value and the anonymity level of the attacker) carried out over the last 3 years.

The original aim of this work is to locate and quantify the most susceptible elements (devices, links, accesses) within a digital communication network through the use of complex networks analysis advanced tools. It is therefore necessary to provide here an overview of key definitions and complex networks advanced tools we will use in the following chapters.

2.1.2 *Holism vs Reductionism*

Complex networks are used since they are understood to be the appropriate environment to deal with the complexity of the Intentional Risk model. In this sense, the complexity concept related to interactions is relatively new. Around 35 years ago the main scientific line/thread used to be reductionism, whose approach considers it to be enough a detailed knowledge of every system component and its fundamental laws to be able to globally understand a system. Nevertheless, it is only possible if the system is lineal so that it can be broken down into the pieces it is composed by, analyzed piece by piece and joint back together to see how the set behaves. This is what happens with, for example, a salt crystal. In spite of the fact that it is composed by a massive amount of particles, it is a system susceptible to be studied from a physical approach since all the particles present an average behavior. In any case, most of the systems are non-linear (systems with thresholds that amplify little perturbations). For instance, we can perfectly know the way a neuron works but we are far from understanding how the brain behaves to carry out its tasks related to the memory or the language. Consequently, the complexity in a system does not only lay in the number of degrees of freedom (variables that

describe the system) it has, it is also related to the non-linearity present in it and in the interactions among its components interaction structure. Complexity appears at the border between order and disorder, when a new qualitative and different from the elemental components of the system behavior emerges. For example, when a neuron receives a stimulus above a threshold, it can react in a chaotic way or in a periodic/regular way. However, if we connect just a few neurons, the result can be a synchronous electrical activity due to the synaptic interaction amongst them. Another of the features of the complexity in a system is related to the structure in which the elements it is composed by interact. For example, the nature of the structure in an average salt crystal helps to simplify the problem in the interactions among the atoms it is composed by assuming that each atom only interacts with the atoms next to it, leading into a meshed net structure (a tidy structure). A different type of simplification is applied when, for example, a star cluster is considered. In this case, we assume every star interacts with all the rest (fact that also enables a huge mathematical simplification within the interactions model).

However, several real complex systems show a different connectivity pattern that does not allow such simplifications. For example, the physical connections among computers (Internet) presents a way more complex structure than the previous ones since the number of computers an average computer of the net is connected to is not uniform. The complex networks theory allows to model this type of systems since it is, in fact, a tool that enables the representation of the interactions within a complex system through a graph where every element of the system is represented by a node (computer) and the interaction between two of its elements corresponds to a link between them which we can easily represent graphically.

2.1.3 Complex Networks and Intentional Risk

Differently from the existing security standards, intentionality complex networks aim to understand risk in a (global) holistic way in a net environment, an organization or even in a country. Analyzing the risk as a combined property in every element of the net, a way more efficient selection of effective controls is fomented. Intentionality complex networks are modelled through a graph, in which nodes are IP Addresses or IP:ports, and the edges or “links”. Among these elements appear the connections that ease the accesses. The edges are the basic elements that configure the paths or routes in the net. The main difference between the graph theory (combinatoria) and the complex networks theory lays in both: the size (smaller for the graph and bigger for the net) and in the parameters and tools employed to analyse its structures. The fact that the size is smaller for the graph and bigger for the net makes it compulsory to have extra care with the complexity of the algorithms that will be employed. Finally, it shall be highlighted that complex networks theory is based on the observation of the elements (in this case IPs and IP:port) that integrate a system and on their interactions. Once we have observed their behavior and translated it into entrances able to be registered in a data-base,

we obtain predictive models that can be applied to both: Technology (where the behavior of the system is the result of the interaction between several elements) and Biology (in the inside of the cell the elements have a defined mission they constantly repeat). It can be applied even to Social Sciences. In this last sentence, it must be highlighted that social networks, or friendship networks have turned into the paradigm of a new net type known as multiplex networks. With the Intentionality complex networks we intend to extrapolate this explicative capacity to hundreds of thousands or millions of nodes, aiming to model the intentional access risk to valuable information by potential attackers.

2.2 Basic Concepts on Graphs and Networks

2.2.1 The Origins

From a schematic point of view, a complex network (or a graph) is a mathematical object $G = (X, E)$ composed by a set of nodes or vertices $X = \{1, \dots, n\}$ that are pairwise joined by links or edges belonging to the set $E = \{\ell_1, \dots, \ell_m\}$.

This kind of representation may appear simple but, as we will see, it has an enormous potential, since some problems become simpler and more treatable if they are represented as a graph. Graph theory, the mathematical scaffold behind network science, has its origins in a problem, solved by Euler in 1736, known as the “Königsberg Bridge Problem” (Fig. 2.1). The Königsberg bridge problem asks if the seven bridges of the city of Königsberg over the river Pregel can all be traversed in

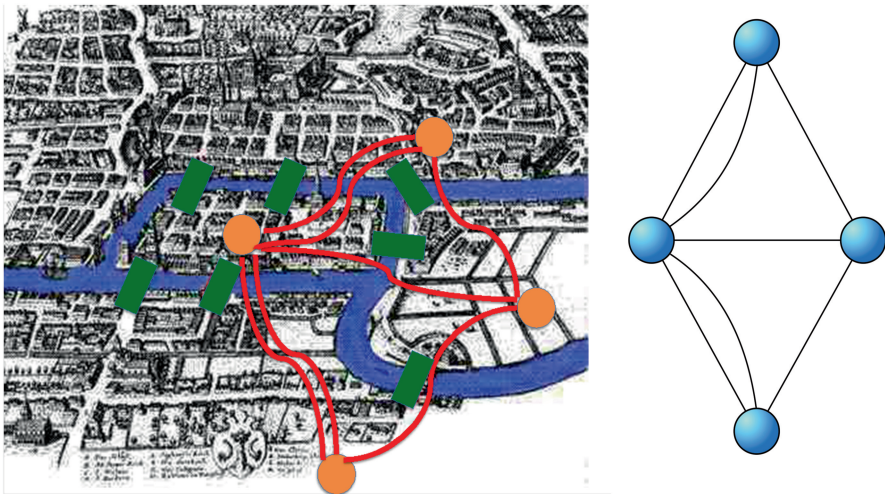


Fig. 2.1 “The Königsberg bridge problem”

a single trip without doubling back, with the additional requirement that the trip ends in the same place it began. This problem was answered in the negative by Euler, and represented the beginning of graph theory. Euler hit upon the idea of using four nodes to represent each of the four land areas separated by the river as nodes, distinguishing them with letters A, B, C, and D. Next he connected with lines each piece of land that had a bridge between them. In this way he built a graph, whose nodes were the pieces of land and links were the bridges. Once the city was represented as a graph, the Königsberg Bridge Problem can be reformulated as follows: is it possible to find a path between two nodes for which every link appears exactly once? Such path is called an Euler walk in graph theory. Now it is needed to introduce some new concepts and notations:

In a graph, two vertices are called *adjacent* (or neighbors) if they are connected by an edge. The number of neighbors of a node i , denoted by k_i , is its *node degree*. So, the degree of a node is the number of links the node shares with its neighbors and which are available for routing purposes.

Euler proved that any graph with n nodes and m links satisfies that $\sum_{i=1}^n k_i = 2m$. Hence, the sum of degrees is even and, consequently, the number of nodes with odd degree is even. Well, having in mind this result, it is not difficult to show that, if w is the number of nodes of G with odd degree, then if $w > 2$, then no Euler walk exists. On the other hand, if $w = 0$, there are Euler walks starting from any node. Finally, if $w = 2$ Euler walks only exist starting from one of the odd nodes. Therefore, since the underlying graph of the Königsberg Bridge Problem had four nodes with odd degree, there was no solution to the problem, that is, no Euler walk exists.

So, in this sophisticated and elegant manner, Euler gave the first example of how small changes (i.e., adding or removing one bridge) may have global consequences (in this case, the existence or not of Euler walks) (see Fig. 2.2).

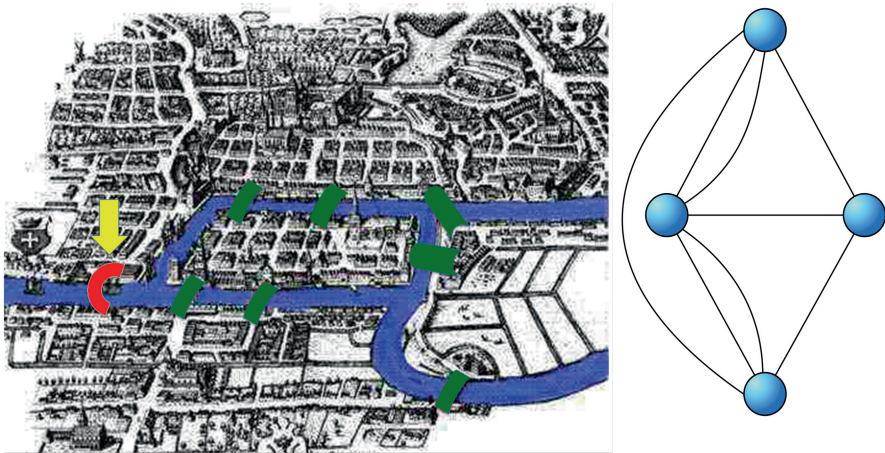


Fig. 2.2 A “solution” for Königsberg’s inhabitants

2.2.2 *Graphs vs Networks*

As we have said, a complex network (or a graph) is a mathematical object $G = (X, E)$ where $X = \{1, \dots, n\}$ is the set of nodes of G , and $E = \{\ell_1, \dots, \ell_m\}$ is the set of links (or edges) of G .

The term “complex network”, or simply “network”, often refers to real systems while the term “graph” is generally considered as the mathematical representation of a network. Some authors consider that there are some differences between the graph theory and the complex networks theory. These differences lay mainly in the size (smaller for the graph and bigger for the network) and in the parameters and the tools employed to analyze both structures. Following this idea, we will talk for example, about a graph of twelve nodes and about a network of five hundred nodes. Specifically, the fact that the size is smaller for the graphs and bigger for the networks makes it compulsory to have an extra care with the complexity of the algorithms that will be employed. In any case, in the sequel, we will use the terms “graph” and “network” interchangeably. For example, the union of two networks $G = (X, E)$ and $G' = (X', E')$ is the union of their node and edge sets: $G \cup G' = (X \cup X', E \cup E')$. When V and V' are disjoint, the union $G \cup G'$ is referred to as the disjoint union. Similarly, the intersection of two networks $G = (X, E)$ and $G' = (X', E')$ is $G \cap G' = (X \cap X', E \cap E')$. It may be noted that we will assume implicitly that in any network there are no self-loops or multiple links between its nodes (i.e., more than one link between two specific nodes).

2.2.3 *Matrices, Degrees, Link Density and Some Interesting Graph Families*

The adjacency matrix of a graph $G = (X, E)$ is a way to determine the graph completely. The *adjacency matrix* $A(G) = (a_{ij})$ of $G = (X, E)$ is defined by the conditions

$$a_{ij} = \begin{cases} 1 & \text{if } \{i, j\} \in E \\ 0 & \text{if } \{i, j\} \notin E. \end{cases} \quad (2.1)$$

Obviously, $\forall i \in \{1, \dots, n\}$ we get that $a_{ii} = 0$ (since there are not self-loops). On the other hand, the degree of a node i can be now easily calculated by the expression $\sum_{j=1}^n a_{ij}$. A node with a high degree can be considered as being well connected and a node with a relatively low degree can be considered weakly connected. In the sequel, we will denote by $\delta(G)$ the minimum node degree of the nodes of G , by $\Delta(G)$ the maximum node degree of the nodes of G and by

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^n k_i \quad (2.2)$$

the average degree calculated over all the nodes of G . Now, it is easy to get that $\langle k \rangle = \frac{2m}{n}$. The degree of an individual node and the minimum, maximum and average degree over all the nodes are standard characterization metrics in graph theory. If all the nodes of G are pairwise adjacent, then G is called *complete* and a complete network of n nodes is denoted by K_n . Note that a complete network with n nodes has exactly $\frac{n(n-1)}{2}$ edges. The graph K_3 is called a *triangle*.

Given a network $G = (X, E)$ where $X = \{1, \dots, n\}$ and $E = \{\ell_1, \dots, \ell_m\}$, it is obvious that the number of edges of G is at most $\frac{n(n-1)}{2}$. The link density of G is defined by the equation

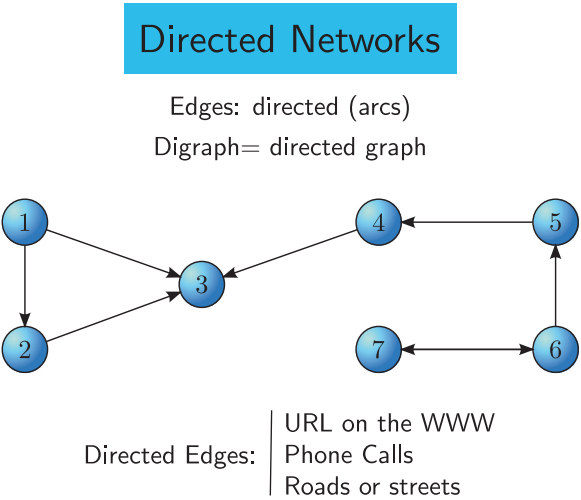
$$\Delta = \frac{2m}{n(n-1)}. \tag{2.3}$$

Obviously, the link density of a network takes values within the interval $[0, 1]$, where $\Delta \ll 1$ means the network is *sparse*, $\Delta \approx 1$ means the network is *dense*, and $\Delta = 1$ means the network is the complete network of n nodes.

A *bipartite* network is a network whose nodes can be divided into two disjoint sets U and V , $X = U \cup V$ such that every link connects a node in U to one in V . If $\forall u \in U$ and $\forall v \in V$ there is an edge connecting u and v , the bipartite network is called *complete*, and if $U = \{1, \dots, r\}$, and $V = \{1, \dots, s\}$, the complete bipartite network $G = (X, E) = (U \cup V, E)$ is denoted by $K_{r,s}$. In Fig. 2.4 some complete bipartite networks are represented.

In general, a network $G = (X, E)$ is called *q-partite* if X admits a partition into q classes such that every edge has its ends in different classes: nodes in the same partition class must not be adjacent. Obviously, if $q = 2$ we have a bipartite network. A q -partite network in which every two nodes from different partition classes are

Fig. 2.3 Some directed networks



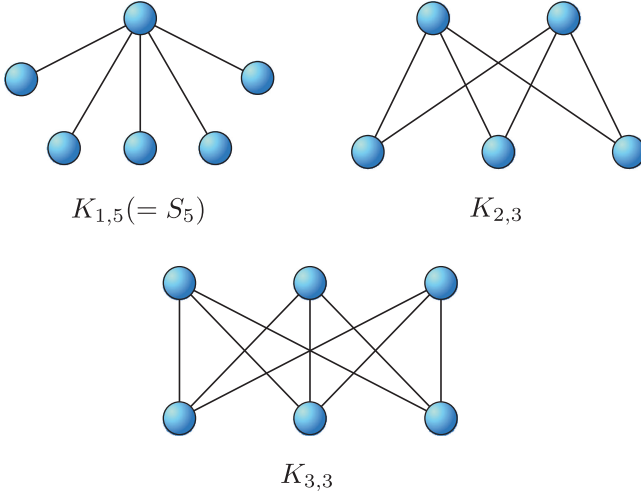


Fig. 2.4 Some examples of complete bipartite networks

adjacent is called complete. A complete bipartite network such that the set U has r elements and the set V has s elements is denoted by $K_{r,s}$. A *star* is a complete bipartite network such that one of the classes has exactly one element. A star of n nodes is denoted by S_n . A *regular* network is a network whose nodes have exactly the same degree (i.e., the same number of neighbors). A regular network with nodes of degree k is called a k -regular network.

The incidence matrix of G is a $n \times m$ matrix $I(G) = (b_{ij})$ defined by the conditions

$$b_{ij} = \begin{cases} 1 & \text{if } i \in \ell_j \\ 0 & \text{if } i \notin \ell_j. \end{cases} \quad (2.4)$$

2.2.4 Directed and Weighted Networks

So far we have implicitly assumed that the relationship between nodes is symmetric, i.e., if there exists an edge ℓ_{ij} between the nodes i and j implicitly we have assumed the existence of the edge $\ell_{ji} = \ell_{ij}$. But this situation does not apply in the general case. This may be the situation if, for example, a link represents geographical proximity (if a node A is close to another node B , then B is also close to A). However, in many cases the connections between nodes are asymmetric since the edge runs only in one direction. When this happens, we say that the link is *directed* (Fig. 2.3). Networks composed of directed edges are referred to as *directed networks* or *digraphs*. Likewise, those networks without directed edges are called *undirected networks*. Obviously, the adjacency matrix $A(G) = (a_{ij})$ of an undirected network G is symmetric, i.e., $\forall i, j \in \{1, \dots, n\}$ $a_{ij} = a_{ji}$. Nevertheless, this is not the case

of a directed network: In a directed network the adjacency matrix is not necessarily symmetric, i.e., it may happen that there exists a link between two nodes $i \rightarrow j$ such that its “inverse” $j \rightarrow i$ does not exist and, in such a case, $a_{ij} = 1$ whilst $a_{ji} = 0$.

Notice that an undirected network can be represented by a directed one having two edges between each pair of connected nodes, one in each direction.

Understandably, if the network is directed, the degree of a node is twofold. We have the so called *out-degree* of the node (i.e., the number of outgoing edges),

$$k_i^{out} = \sum_{j=1}^n a_{ij}, \quad (2.5)$$

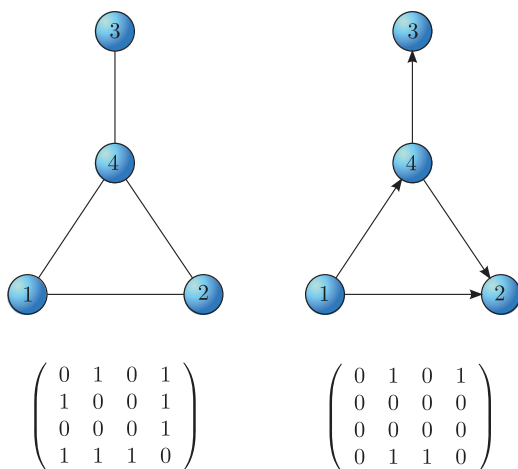
and the *in-degree* of the node (i.e., the number of ingoing edges),

$$k_i^{in} = \sum_{j=1}^n a_{ji}. \quad (2.6)$$

Hence, as in the context of directed networks the degree of a node i has two components (k_i^{in}, k_i^{out}) , the total degree of a node i is defined as $k_i = k_i^{in} + k_i^{out}$.

Edges can also carry weights to measure the capacity or the intensity of the relationship between two nodes. Examples of this situations are the existence of strong and weak ties in social networks or unequal traffic on the Internet. These networks are better described as *weighted networks*, i.e., networks in which each edge has associated a value measuring the strength of the relationship, and in these cases we consider the so called *weights matrix* $W = (w_{ij})$ whose entry w_{ij} is the weight of the edge from node i to node j . So, a weighted (directed or undirected) network is a triplet $G = (X, E, W)$ where (X, E) is a (directed or undirected) network and $W = (w_{ij})$ is the weights matrix of G (Fig. 2.5).

Fig. 2.5 Adjacency matrices of a non-directed and a directed network



2.2.5 Metric Structure, Connectedness, Geodesics and Some Other Concepts

Once we have introduced the concept of complex networks as the main object of the complex network analysis, we should give the basic parameters used in the literature in order to analyze these objects. There are plenty of mathematical functions that help to study and classify the behavior of complex network structure (see, for example [22, 61]), including metric parameters, clustering, spectral functions and dynamical parameters among many others. In the rest of this chapter we will put mainly the stress in some metric and spectral functions that we will use to analyze the structure of a complex network.

The metric structure of a complex network is related to the topological distance between nodes of the network, written in terms of walks and paths in the graph. A *walk* (of length k) in G is a non-empty alternating sequence

$$\{i_1, \ell_1, i_2, \ell_2, \dots, \ell_{k-1}, i_k\}$$

of nodes and edges such that $\ell_r = i_r, i_{r+1}$ for all $r < k$. If $i_1 = i_k$, the walk is *closed*. A *path* between two nodes is a walk through the network nodes in which each node is visited only once. A *cycle* is a closed walk that starts and ends at the same node, in which no edge is repeated. A cycle of n nodes is denoted by C_n . C_3 is a triangle. To see the difference between paths and walks, consider any two adjacent vertices not part of a cycle. Then there is exact one path between the two vertices i.e. the edge between them, but there exists an infinite number of walks between them. For example, $\{i_1, \ell_1, i_2, \ell_2, i_1, \ell_1, i_2, \ell_2, i_1\}$ is a walk from i_1 to i_2 of length 3. Specifically, paths do not allow repetition of vertices while walks do. In any case, we will say that a node j is reachable from another node i if there exists a walk connecting them.

On the other hand, by using the adjacency matrix $A(G)$, we can obtain the number of k -length walks from one node to another. More specifically, it is not difficult to check that the number of walks of length r from i to j , equals the (i, j) th entry of the r th power of the adjacency matrix $A(G)^r$.

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. A tree is a connected graph in which any two vertices are connected by exactly one path. The *length* of a path is the number of edges of that path. If $i, j \in X$ a *geodesic* between i and j is a path of the shortest length that connects i and j . The *distance* d_{ij} between i and j is the length of a geodesic between i and j . The maximum distance $D(G)$ between any two vertices in G is called the *diameter* of G . By n_{ij} we will denote the number of different geodesics that join i and j . If $v \in X$ is a node and $\ell \in E$ is a link, then $n_{ij}(v)$ and $n_{ij}(\ell)$ will denote the number of geodesics that join i and j passing through v and ℓ respectively. The *length* of a cycle is, obviously, its number of edges (or nodes). The minimum length of a cycle (contained) in a graph G is the *girth* $g(G)$ of G . The maximum length of a cycle in G is its *circumference*. If $Y \subseteq X$ is any set of nodes, then $G(Y)$ denotes the network on Y whose edges are precisely the edges

of G with both ends in Y . A network $H = (Y, F)$ is a subnetwork of $G = (X, E)$ if $Y \subseteq X$, $F \subseteq E$ and the edges in F connect nodes in Y . A *connected component* is a maximal connected subgraph of G . Two paths connecting the same pair of vertices in a network are said to be vertex-independent if they share none of the same vertices other than their starting and ending vertices. A *k-component* is a maximal subset of the vertices of a network such that every vertex in the subset is connected to every other by k independent paths. For the special cases $k = 2$, $k = 3$, the k -components are called *bi-components* and *three-components* of the network. For any given network, the k -components are nested: every three-component is a subset of a bi-component, and so forth.

To close this section, we think it is opportune to mention the concept of algebraic connectivity which was introduced in [60]. This measure, which depends on both the number of nodes and their respective configurations, indicates the level of connectivity in the graph [80, 92, 132]. So, the larger the algebraic connectivity is, the more difficult it is to cut the network into disconnected parts.

2.2.6 Characteristic Path Length, Efficiency and Vulnerability of a Network

The *characteristic path length* is defined as

$$L(G) = \frac{1}{n(n-1)} \sum_{j=1}^n \sum_{\substack{k=1 \\ k \neq j}}^n d_{jk} = \frac{1}{n(n-1)} \sum_{j \neq k \in X} d_{jk}. \quad (2.7)$$

This parameter give us a way of measuring the performance of a network (the bigger is $L(G)$ the smaller is its performance), but in order to use it as a measure of performance, it has a problem because of errors and attacks, networks can become disconnected. In fact, if the distance between two nodes is infinite, $L(G)$ becomes infinite. The concept of *efficiency*, introduced by Latora and Marchiori in [87] is a well defined quantity also for non-connected networks. The efficiency of a network G is defined as

$$E(G) = \frac{1}{n(n-1)} \sum_{\substack{i, j \in X \\ i \neq j}} \frac{1}{d_{ij}}. \quad (2.8)$$

The concept of network's vulnerability appears in different contexts. A series of different approaches from several branches of knowledge have been introduced to quantify the vulnerability of a complex network [1, 5, 14, 17, 20, 22, 23, 32, 40, 42, 69, 71, 75, 85, 121, 135]. In fact, what must be understood by the concept of network's vulnerability depends on the context in which that term is used [34]. For example, it is commonly understood that a structure is vulnerable if any small

damage produces disproportionately large consequences [2, 3, 12, 75, 98, 99]. In general, the concept of vulnerability in a network aims at quantifying the network security and stability under the effects of all that kind of disfunctions. In classical graph theory, the term vulnerability is related to a lack of resistance of the graph to the deletion of nodes and edges [14], i.e., how the topology of a network is affected by the removal of a finite number of links and/or nodes. This point of view matches up with another approach to the concept of vulnerability, the *structural vulnerability* which aims to study the way in which the topology of a network is affected by the removal of a finite number of links and/or nodes. In the context of classic graph theory the analysis of vulnerability is carried out by studying different versions of the connectivity of the graph [14, 72, 104]. In any case, it is important to point out that from a strictly complex networks analysis perspective, the concept of vulnerability lies on statistical graph measures and numerical models (simulation) in order to study complex (large) networks and their susceptibility to damage by errors and attacks [5, 17, 20–22, 32, 33, 40–43, 74, 83, 84, 86].

An alternative way to analyze connectivity is by considering the number of node-independent paths between two vertices or, in the same way, to the minimum number of other vertices in the network that must fail in order for those two vertices to become disconnected from one another [72, 89, 104]. In this approach, it is considered the impact of nodes and edges destruction in terms of potentially disconnecting the network. Several results related to structural vulnerability (minimal number of edges and/or nodes whose removal disconnects the network) can be found in [52, 53, 60, 65, 89], and a broader and more detailed approach to this concept can be found in [34].

2.2.7 Clustering Coefficient

The clustering coefficient of a node indicates how concentrated the neighborhood of that node is. From a sociological point of view, *clustering* measures the acquaintance property of the network, where two nodes with a common neighbour are likely to connect each other, i.e., they form a triangle [130]. Specifically, Watts and Strogatz define in [131] the local clustering coefficient C_i for a vertex i as the proportion of links between the nodes within its neighbourhood (i.e., between the neighbours of i) divided by the number of links that could possibly exist between them:

$$C_i = \frac{2t_i}{k_i(k_i - 1)} \quad (2.9)$$

i.e., the number of triangles in which node i participates normalized by the maximum possible number of such triangles, where t_i denotes the number of triangles around i (if the degree of node i is 0 or 1, we which gives us $C_i = \frac{0}{0}$, we can set $C_i = 0$). Hence $C_i = 0$ if none of the neighbours of a node are connected, and $C_i = 1$ if all of the neighbours are connected. The *clustering coefficient* $C(G)$ for the whole network G is the average of the local values C_i :

$$C(G) = \frac{1}{n} \sum_{i=1}^n C_i. \quad (2.10)$$

By definition $0 \leq C_i \leq 1$ and $0 \leq C \leq 1$.

The average of C_i taken over all nodes with a given degree k is the so-called *clustering coefficient of a connectivity class k* .

Alternatives measures of the clustering properties of a network $G = (X, E)$ are the *transitivity* T [15, 101, 102] whose value is given by the formula:

$$T(G) = \frac{3 \times (\text{number of triangles in } G)}{(\text{number of connected triples of nodes in } G)}, \quad (2.11)$$

i.e., the fraction of connected triples of nodes (triads) which also form triangles, and the *local efficiency* of G [83, 87] defined as:

$$E_{loc}(G) = \frac{1}{n} \sum_{i=1}^n E(G_i). \quad (2.12)$$

where $E(G_i)$ is the efficiency of G_i , the subnetwork formed by the neighbours of the node i . The differences between $C(G)$ and $T(G)$ are illustrated in [83, 102].

The clustering coefficient is one of the most interesting parameters used in complex networks analysis. In [61] a panoramic view of the main existing measurements of complex networks can be found.

2.2.8 Finding Out the Critical and the Most Influential Nodes: Eigenvector Centrality

The use of spectral methods in networks and graph theory have a long tradition. Eigenvector centrality is a concept often used in social network analysis and was first proposed by Bonacich [25]. Addressed more precisely, we can say that eigenvector-like centralities were introduced in sociology to measure the influence of each actor in a social group, taking into account the immediate effects, the mediative effects and the global effects of the social interaction [25], but they are also useful in many other applications such as the web search engines like Google. In this sense, it is remarkable that Google uses a similar centrality ranking technique (called PageRank) to rank the relevance of hyper-linked pages in search results [27]. In order to clarify this concept, as we will see below, the eigenvector centrality of a node is proportional to the sum of the centrality values of all its neighboring nodes, so eigenvector centrality is defined in a circular manner. In the social-network context, an important node (or person) is characterized by its connectivity to other important nodes (or people). A node with a high centrality value is a well-connected node and has a dominant influence on the surrounding network.

In an overall perspective, spectral graph theory studies the eigenvalues of matrices that embody the graph structure. One of the main objectives in spectral graph theory is to deduce structural characteristics of a graph from such eigenvalue spectra. Among other applications, these methods are used in the study of the measures of vulnerability based on the fall of connectivity. In addition, spectral analysis has a lot of applications too numerous to be collected here. For example, spectral analysis allows characterizing models of real networks [90, 113], determine communities [105], to find the edges which connect different communities and remove them in a iterative form, breaking the network into disconnected groups of nodes [10, 66] and even visualizing networks [115].

To introduce the eigenvalue spectra of a network some additional concepts are needed. The characteristic polynomial $\det(xI - A(G))$ of the adjacency matrix $A(G)$ is called the characteristic polynomial of G . The eigenvalues of $A(G)$ (i.e., the zeros of $\det(xI - A(G))$) are also called the eigenvalues of G . If μ is an eigenvalue of G , then a non-zero vector $V \in \mathbb{R}^n$ satisfying $A(G) \cdot V = \mu \cdot V$, is called an eigenvector of $A(G)$ for μ ; it is also called a μ -eigenvector. The relation $A(G) \cdot V = \mu \cdot V$ can be interpreted in the following way: if $V = (v_1, \dots, v_n)^t$, where the superscript t denotes the transpose of V (i.e., the column vector V^t) then for any node i we have that $\mu \cdot v_i = \sum_{j \sim i} v_j$, where the summation is over all neighbours j of i . If μ is an eigenvalue of G , then the set $\{V \in \mathbb{R}^n : A(G) \cdot V = \mu \cdot V\}$ is a linear subspace of $\mathbb{R}^n$, called the eigenspace of G for μ .

On the other hand, since $A(G)$ is a real symmetric matrix, it is important to point out that all the eigenvalues of G are real. Moreover, $\mathbb{R}^n$ has a basis $\{V_1, \dots, V_n\}$ of n normal eigenvectors of $A(G)$ (see, for example, [30]). The set of eigenvalues of $A(G)$:

$$\sigma(A(G)) = \{\mu_1(G), \mu_2(G), \dots, \mu_n(G)\}, \quad (2.13)$$

where

$$\mu_1(G) \leq \mu_2(G) \leq \dots \leq \mu_n(G) \quad (2.14)$$

is called the “spectrum” of G .

Many authors have studied the algebraic aspects of spectral graph theory (see, for example, [18, 30, 48–50, 64, 67, 68, 79, 80, 92–95, 122]). The eigenvalue spectra of a network provide valuable information about the behaviour of many dynamical processes running within the network, but in this section we only consider the applications of spectral analysis to static networks. For example, in [49] it is shown that diameter $D(G)$ of a network satisfies $D(G) \leq r - 1$ where r is the number of distinct eigenvalues.

The largest eigenvalue of the adjacency matrix $\mu_n(G)$ is called spectral radius of G . This eigenvalue is usually denoted by $\rho(G)$. It is important to remark that for $\rho(G)$, since $A(G)$ is non-negative, there exists an eigenvector whose all entries are non-negative.

This eigenvalue of $A(G)$ has received the most attention in this context, since this quantity refers to the speed of growth of walks in the graph (the number of walks of length k is, approximately, $\rho(G)^k$) [30]. A nice and useful property is given by the following inequality [49]:

$$\sqrt{\Delta(G)} \leq \mu_n(G) = \rho(G) \leq \Delta(G), \quad (2.15)$$

where $\Delta(G)$ is the maximum node degree of G .

It is important to highlight that the spectral radius of G plays an important role in modelling virus propagation in computer networks. The smaller the largest eigenvalue, the larger the robustness of a network against the spread of viruses. In fact, the epidemic threshold in spreading viruses is proportional to $\frac{1}{\rho(G)}$ [128]. Another example that remarks the importance of $\rho(G)$ is given by the Bonacich centrality of G [25, 26, 36] (measure based on the eigenvectors associated to the spectral radius of G) generally known as *eigenvector centrality*.

The eigenvector centrality is calculated by using the adjacency matrix $A(G) = (a_{ij})$ to find central nodes in the networks. Let $V = (v_1, \dots, v_n)$ be a vector whose i th element v_i represents the centrality (normalized) measure of node i . Let $N(i)$ the set of neighbours of node i in the network G . Eigenvector centrality is defined using the following formulas:

$$v_i \propto \sum_{j \in N(i)} v_j, \quad (2.16)$$

which can be rewritten as

$$v_i \propto \sum_{j=1}^n a_{ji} v_j, \quad (2.17)$$

which can be rewritten in the form

$$A(G) \cdot v = \mu \cdot v. \quad (2.18)$$

Since $A(G)$ is a $n \times n$ symmetric matrix, it has n eigenvectors and n (not necessarily distinct) corresponding eigenvalues. The classical way to compute the eigenvalues of $A(G)$ is to find the roots of the characteristic polynomial $\det(xI - A(G))$ of $A(G)$. The principle eigenvector is the eigenvector associated to the spectral radius $\rho(G)$. After the principle eigenvector is found, its entries are sorted from highest to lowest values to determine a ranking of nodes. The most central node has the highest rank and most peripheral node has the lowest rank. So, in any network the connectivity of a node depends on the connectivity of its neighbours, and eigenvector centrality can help capture this property. It is remarkable that the main drawback of this measure is that it can only be calculated in a centralized way. To close this section it is important to remark that some new developments related to eigenvector centrality have been recently introduced in

the literature (see, for example, [112]). Moreover, there exist many interesting and recent publications related to critical element detection in a graph-theoretic setting (see, for example, [11, 123, 124, 127]). And finally, it is also remarkable that in the last decade there exists a trend towards generalizing centrality classes which may perhaps play a significant role in the context of cyber-security (see, for example, [47, 59, 78, 111, 125]).

2.2.9 Information Flow Management: Betweenness Centrality

Another approach to the study of network's centrality is based on the idea that critical nodes and links stand between others, playing the role of an intermediary in the interactions or in the communications. So, the greater the number of paths in which a node or edge participates, the higher the importance of this node or edge for the network. Thus, assuming that the interactions follow the shortest paths between two vertices, it is possible to quantify the importance of a node or an edge in terms of its betweenness centrality.

Node betweenness was first proposed by Freeman [63] in 1977 in the context of social networks. This concept has been considered more recently as an important parameter in the study of networks associated to complex systems [102]. Girvan and Newman [66] generalize this definition to edges and introduce the edge betweenness of an edge as the fraction of shortest paths between pairs of vertices that run along it.

Specifically, the betweenness centrality is related to the concentration of the geodesic structure throughout the network.

The node betweenness centrality $B(G)$ of a network G [63] is:

$$B(G) = \left(\frac{1}{n} \sum_{i \in X} b_i \right), \quad (2.19)$$

where b_i is the betweenness of the node $i \in X$ (see, for example [106, 130]) given by

$$b_i = \frac{1}{n(n-1)} \sum_{k, j \in X, i \neq j} \frac{n_{kj}(i)}{n_{kj}}, \quad (2.20)$$

where n_{kj} is the number of different geodesics that join k and j , and $n_{kj}(i)$ is the number of geodesics that join k and j passing through i .

The maximum betweenness of the network G is

$$B_{max}(G) = \max\{b_i : i \in X\}. \quad (2.21)$$

The same parameters can be defined for edges exactly in the same way as before, obtaining the edge-betweenness $B_E(G)$

$$B_E(G) = \left(\frac{1}{m} \sum_{\ell \in E} b_\ell \right). \quad (2.22)$$

where, in the same way as before, b_ℓ is the betweenness of the link $\ell \in E$ given by

$$b_\ell = \frac{1}{n(n-1)} \sum_{i,j \in X, i \neq j} \frac{n_{ij}(\ell)}{n_{ij}} \quad (2.23)$$

and the maximum edge betweenness centrality

$$B_{Emax}(G) = \max\{b_\ell : \ell \in E\}. \quad (2.24)$$

A remarkable relationship between the characteristic path length $L(G)$ and $B_E(G)$ as it is shown in [20] is the following:

$$\begin{aligned} B_E &= \frac{1}{|E|} \sum_{\ell \in E} b_\ell = \frac{1}{|E|} \sum_{\ell \in E} \left(\sum_{j,k \in X} \frac{n_{jk}(\ell)}{n_{jk}} \right) \\ &= \frac{1}{|E|} \sum_{j,k \in X} \frac{1}{n_{jk}} \left(\sum_{\ell \in E} n_{jk}(\ell) \right). \end{aligned}$$

Notice that if $\mathbf{P}_{jk}$ is the set of all geodesics joining j and k then one has

$$n_{jk}(\ell) = \sum_{g \in \mathbf{P}_{jk}} \chi_g(\ell), \quad (2.25)$$

where $\omega_g(\ell)$ is 1 if ℓ belongs to the geodesic g and 0 otherwise. Hence if $d_{j,k}$ denotes the distance between j and k in the network then

$$\begin{aligned} B_E &= \frac{1}{|E|} \sum_{j,k \in X} \frac{1}{n_{jk}} \left(\sum_{\ell \in E} n_{jk}(\ell) \right) \\ &= \frac{1}{|E|} \sum_{j,k \in X} \frac{1}{n_{jk}} \left(\sum_{g \in \mathcal{P}_{jk}} \sum_{\ell \in E} \omega_g(\ell) \right) \\ &= \frac{1}{|E|} \sum_{j,k \in X} \frac{1}{n_{jk}} \left(\sum_{g \in \mathcal{P}_{jk}} d_{j,k} \right) = \frac{n(n-1)}{|E|} L(G) \end{aligned}$$

and therefore $B_E(G)$ measures essentially the same global properties than the characteristic path length $L(G)$.

2.2.10 Degree Distributions

The *degree distribution* of a network $G = (X, E)$ is a function $P(k)$ which give us the probability of finding a node in G with degree k . In other words, $P(k)$ is referred to the fraction of nodes having degree k . So, $P(k) = \frac{N(k)}{n}$, where $N(k)$ is the number of nodes whose degree is k . Obviously, by using the degree distribution of G , we can get another expression for the average degree of G :

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^n k_i = \sum_k k \cdot P(k) = \frac{2m}{n}. \quad (2.26)$$

It is remarkable that this concept is the simplest and most studied one-node feature that we can find in a network, an it is used as the basis for network topological characterization.

Information on how the degree is distributed among the nodes of an undirected network can be obtained either by plotting $P(k)$, or by the calculation of the moments of the distribution. The ℓ -moment of $P(k)$ is defined as

$$\langle k^\ell \rangle = \sum_k k^\ell \cdot P(k). \quad (2.27)$$

So, the first moment $\langle k^1 \rangle = \langle k \rangle$ is the average degree of G , the second moment $\langle k^2 \rangle$ (degree variance) measures the fluctuation of the connectivity distribution of G , the third moment $\langle k^3 \rangle$ is the degree skewness and the fourth moment $\langle k^4 \rangle$ is the degree kurtosis.

On the other hand, if the network is directed, the degree of a node is twofold. We have the so called *out-degree* of the node (i.e., the number of outgoing edges),

$$k_i^{out} = \sum_{j=1}^n a_{ij}, \quad (2.28)$$

and the *in-degree* of the node (i.e., the number of ingoing edges),

$$k_i^{in} = \sum_{j=1}^n a_{ji}. \quad (2.29)$$

Consequently, in the directed case, we need two distributions $P(k^{out})$ and $P(k^{in})$ to describe the network.

2.3 Some Interesting Complex Networks Models

2.3.1 Random Networks

Paul Erdős (1913–1996) is indeed the most prolific mathematician in history after Euler. He published around 1500 articles in his lifetime about many branches of science: graph theory, probability theory, set theory, classical analysis, approximation theory and number theory. Standing out, especially, the development of Ramsey theory and the application of the probabilistic method. However, Erdős is famous by his eccentricities: all his life was travelling between scientific conferences and the homes of colleagues all over the world. Erdős and his colleague Alfred Rényi started working on graphs to understand the structure of social networks. In other words, Erdős and Rényi, having in mind the original motivation of analyzing, by means of probabilistic methods, the properties of networks as a function of the increasing number of random connections, were the first to study this kind of mathematical objects. To tackle this issue, they focused on the so-called *random graphs* [55], in which the existence of an edge between a pair of nodes has probability p . This implies that, given a network with n nodes, the average degree is:

$$\langle k \rangle = p \cdot (n - 1) \propto p \cdot n \quad (2.30)$$

and the number of random connections between them is

$$\langle L \rangle = \frac{1}{2} p \cdot n \cdot (n - 1) = \frac{1}{2} \langle k \rangle \cdot n. \quad (2.31)$$

Now, given a random graph with $n \gg 1$ nodes and average degree $\langle k \rangle = p \cdot (n - 1)$ fixed, the probability of a node having degree k is

$$P(k) = \binom{n}{k} p^k (1 - p)^{n-k} \simeq e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}, \quad (2.32)$$

where the approximation becomes exact when $n \rightarrow \infty$ but $\langle k \rangle$ remains fixed. In other words, the degrees of a random graph follow a Poisson distribution [22, 24, 129].

As it is known, a Poisson distribution has a well defined mean value indicating that all nodes have “essentially” the same number of links. Nodes on either sides of the peak may have many more or fewer links than the average node, but these nodes are extremely rare since the distribution rapidly diminishes for values far from the mean.

In random graphs, the probability that two neighbours of a node are connected is the probability that two randomly chosen nodes are linked. Then, the clustering coefficient is simply

$$\bar{C} = p \cong \frac{\langle k \rangle}{n} \ll 1, \quad (2.33)$$

which means that large scale random networks have no clustering in general.

On the other hand, if the average degree in a random graph is $\langle k \rangle$, every node has approximately $\langle k \rangle$ neighbours. Then, since each neighbour has in turn other $\langle k \rangle$ neighbours, every node has (approximately) $\langle k \rangle^2$ second neighbours. Extending this argument, the number of links n_ℓ needed to reach all nodes in the network can be roughly estimated by the condition:

$$\langle k \rangle^{n_\ell} \sim n, \quad (2.34)$$

and, in that circumstances,

$$\ell \sim \frac{\log(n)}{\log(\langle k \rangle)}, \quad (2.35)$$

which implies that the average distance in a random network is rather small, even for very large networks. It is remarkable that we can carry out this calculation because the clustering is small. In a network with large clustering a node i does not have $\langle k \rangle^2$ second neighbours, since many of those second neighbours are also themselves neighbours of i . As a consequence, the number of second nodes will be (possibly) much smaller.

2.3.2 *Small-World Model*

In the last years, several models of complex networks have been proposed after the pioneering idea of random graph model of Erdős-Rényi [55, 56] such as the small-world model of Watts and Strogatz [131]. The main reason for this was the discovery that real networks like the Internet graph have characteristics which are not explained by uniformly random connectivity. Specifically, the *small-world property* indicates that the network diameter is much smaller than the number of nodes or, in other words, most vertices can be reached from the others through a small number of edges, like in social networks. The term *small world* was introduced by Watts and Strogatz [131] in their study of various real-world networks, such as the network of Hollywood movie actors. In [131] the average distance (or characteristic path length), which measures global distances in a network, and the clustering coefficient, which is a measure of “cliquishness” of neighbourhoods in a network, are introduced. Watts and Strogatz [131] proposed a *small-world* model based on a rewiring procedure of the edges of a regular network with a probability p . The Watts-Strogatz algorithm starts with a regular network with n nodes arranged in a ring. Each edge is randomly rewired with probability p . Varying p from $p = 0$ to $p = 1$ the transition between the regular network and randomness can be monitored and, by for example, for a small rewiring probability the network clustering coefficient scarcely differs from its initial value, but the characteristic path length downfalls

rapidly and is of the same order as that of random networks. So, Watts and Strogatz created the first model that conciliated the existence of a large clustering with a small diameter or characteristic path length.

In [8] diverse real-world networks are studied, arising three classes of small-world networks, namely *scale-free*, *broad scale* and *single scale* networks.

2.3.3 Scale-Free Networks

In many real world networks the degree distribution does not follow a Poisson-like distribution (for instance the world-wide web, electric power grid, network of world airports), but instead does follow a power law, i.e., $P(k) = ak^{-\gamma}$ where a is a constant and γ is a positive exponent (this exponent, empirically varies between two and three for the majority of the real world networks). Having a $P(k)$ that has a decaying tail in the power law means that the vast majority of nodes have low degree and that there exist few nodes, the so-called *hubs*, that have an extremely high connectivity. A network with degree power law distribution is called *scale-free*. This model was introduced in 1999 by Barabási and collaborators [4, 13] in 1999, after verifying that the degree distribution of some complex systems follows “power laws” instead of being Poisson-like distribution and, additionally, many systems are strongly clustered with a big number of short paths between the nodes, i.e., they obey the small world property.

Scale-free networks usually emerge in the context of a growing network in which new nodes prefer to connect to the highly connected nodes in the network. When there are constraints limiting the addition of new edges, like aging of the nodes or cost of adding edges to the nodes or the limited capacity of a node, then the broad-scale or single-scale networks appears [8].

One of the most frequently used graphical methods of identifying power-law probability distributions using random samples are log-log plots. This method consists of plotting the logarithm of an estimator of the probability that a particular number of the distribution occurs versus the logarithm of that particular number. Usually, this estimator is the proportion of times that the number occurs in the data set. If the points in the plot tend to “converge” to a straight line for large numbers in the x axis, then the researcher concludes that the distribution has a power-law tail.

A power-law function can be expressed as a polynomial $p(x) = ax^{-\gamma}$ where a and γ are constants and γ is called the power-law exponent. A power law distribution has no peak at its average value and is a relatively slow decreasing function, but the main property of power laws is their scale invariance, i.e., if we substitute the argument x by the same argument multiplied by a scaling factor c we get a proportionate scaling of the function itself, i.e., $p(cx) = a(cx)^{-\gamma} = c^{-\gamma}p(x) \propto p(x)$, which means that they are proportional and therefore it preserves the shape of the function itself. Moreover, by taking logarithms the following linear relation is obtained $\log p(x) = \log a - \gamma \log x$ (Fig. 2.6).

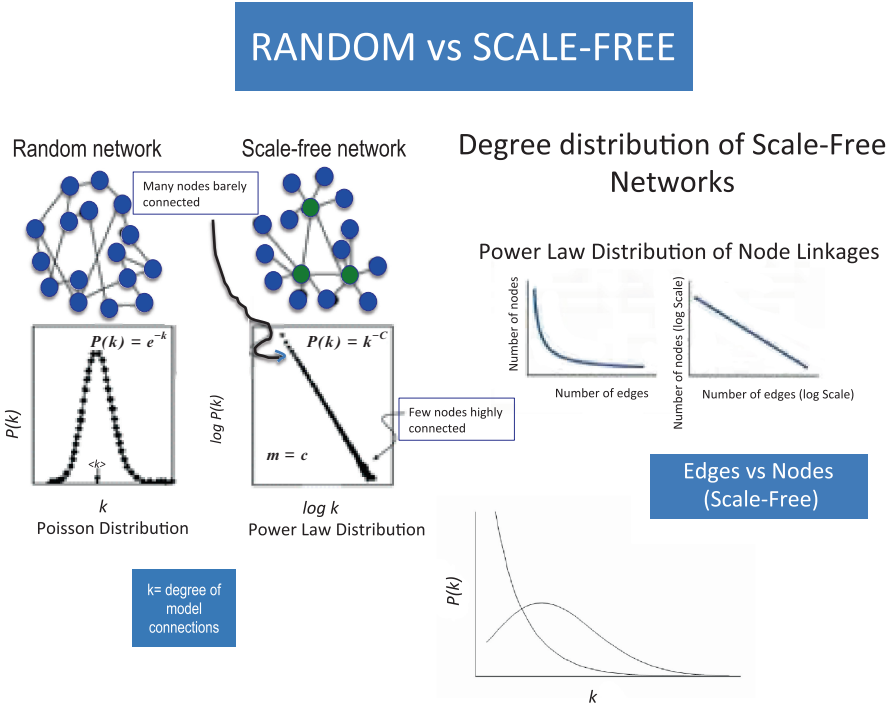


Fig. 2.6 A visual comparison between random and scale-free networks

The model proposed by Barabási and Albert [13] is based on two observed facts in real networks: networks expand continuously by the addition of new vertices, and new vertices attach preferentially to sites that are already well connected. The model starts with a small number of nodes at step $t = 0$, and at every time step a new node is connected to a number of nodes of the existing graph. The probability of the new node to be connected to an existing node depends on the degree of that node, in the sense that nodes with higher degree have stronger ability to grab links added to the network.

As we have said, a typical value for the degree power-law exponent in the majority of real networks is $2 \leq \gamma \leq 3$. The Barabási-Albert model produces a degree power-law distribution with exponent $\gamma = 3$, meanwhile the Watts-Strogatz and the Erdős-Rényi follow a Poisson distribution.

Related to this, it is important to introduce the concept of *network's assortativity* as another fundamental property of networks. This feature measures the degree-degree correlation between nodes. In assortative networks, most edges connect nodes that exhibit similar degrees (nodes aristocracy). On the other hand, *disassortative networks* are such that high-degree nodes are connected to low-degree nodes. For a survey of all the previous network properties and more we refer the reader to [22].

2.4 New Approaches and Developments of Interest for Our Model

2.4.1 *When Edges are More Important than Nodes: Line Graph and Related Concepts*

The choice of the proper network representation for a given problem and its associated tools may determine our ability to use network theory successfully. In some cases there is a natural representation of the problem using networks. In other cases, the chosen representation can help us to solve the posed problem. For example, sometimes it seems appropriate to give more importance to the edges of a network than to its nodes in the context of certain networks and graphs. An example of this comes from urbanism where sometimes the streets of a city are represented by the nodes of the graph, and the intersections between them as the links [44–46, 110]. Distribution networks constitute another example of this situation. The appropriate object to support this point of view is the line graph $L(G)$ (also called dual graph) as it has been shown in the context of urbanism above and transport networks [9, 126] or urban traffic [76].

To study this kind of problems the concept of line graph is introduced in a natural way.

The line graph associated to $G = (V, E)$ is the network $L(G) = (E, L)$ whose set of nodes is the initial set of edges of the graph G , with the assumption that two such nodes ℓ and ℓ' are connected by the edge $\{\ell, \ell'\}$ if on the initial graph G the edges ℓ and ℓ' share some node. So, the line graph of the “claw graph” $K_{1,3}$ is the triangle C_3 , the line graph of the cycle C_n is itself, and the line graph of the star graph S_n is the complete graph K_{n-1} .

In any case, it is remarkable that line graphs have been studied for more than 80 years, although this concept has been rediscovered several times throughout this period of time, with different names such as the adjoint graph, derived graph or edge-to-vertex dual [73]. The first time this concept appears in the literature was in 1932 [133].

Of course, many properties of a graph G that depend only on adjacency between edges may be translated into equivalent properties in $L(G)$ that depend on adjacency between vertices. For instance, the line graph of a connected graph is connected since if G is connected, it contains a path connecting any two of its edges, which translates into a path in $L(G)$ containing any two of the vertices of $L(G)$. However, a graph G that has some isolated vertices, and is therefore disconnected, may nevertheless have a connected line graph. More relevant is the following property: If a graph G has an Euler cycle, that is, if G is connected and has an even number of edges at each vertex, then the line graph of G is Hamiltonian, i.e., a graph possessing a Hamiltonian cycle (a Hamiltonian cycle in graph G is a cycle that passes through all its nodes exactly once). At this point it is remarkable that there are no known (non-trivial) conditions that would be necessary and sufficient for the existence of a Hamiltonian cycle in a graph.

On the other hand, line graph has been considered for networks (graphs with many nodes and edges) only in a reduced number of studies and applications. Particularly, networks are commonly used in the analysis of urban and territorial cases, for example, if we are looking for the places or streets inside a city that are more important—or more frequented—than others due to their geographic situation. As it was shown in [110], the study of centrality in complex networks within the context of urban design based on the primal graph representation give us similar results than the corresponding analysis made using its associated line graph.

A natural question then arises: given G and $L(G)$ and some certain parameter (such as efficiency, centrality, clustering, betweenness, etc.) can we estimate the value of this parameter on G (respectively, on $L(G)$) if we know its value on $L(G)$ (respectively, on G)? As we will see, in some cases the answer is “Yes”. The point is that sometimes it may be easier to work with $L(G)$ instead of G , or conversely.

The bipartite network $B(G)$ associated to G is defined by $B(G) = (X \cup E, E(B(G)))$ whose adjacency matrix is given by

$$A(B(G)) = \left(\begin{array}{c|c} 0 & I(G) \\ \hline I(G)^t & 0 \end{array} \right), \quad (2.36)$$

where $I(G)$ is the incidence matrix of G . It is shown that

$$A(B(G))^2 = \left(\begin{array}{c|c} A(G) + gr & 0 \\ \hline 0 & A(G^*) + 2I_m \end{array} \right), \quad (2.37)$$

where $A(G) + gr = I_G I_G^t$ denotes the matrix obtained by adding to $A(G)$ the diagonal matrix (b_{ii}) where (b_{ii}) is the degree of the vertex v_i , and $L(G)$ denotes the line (or dual) network associated to G [70, p. 26]. Observe that the equality $I_G^t I_G = A_{L(G)} + 2I_m$, where I_m is the identity matrix in $\mathbb{R}^m$, trivially holds. On the other hand, as it is shown in [36], if we know the Bonacich centrality $c(L(G))$, we can recover $c(B(G))$ and reciprocally. If, in addition, G is regular then each of the three centralities can be recovered from any of the other.

On the other hand, it is not difficult to get the relationships between the metrics in G , $L(G)$ and $B(G)$ established in the following propositions (see [70, p. 302] if needed):

Proposition 2.1. *Let i, j and i', j' be two pair of nodes in $G = (V, E)$ joined respectively by the edges $\ell = \{i, j\}$ $\ell' = \{i', j'\}$ and such that $\ell \neq \ell'$. Then, the distance in $L(G) = (E, L)$ between the edges ℓ and ℓ' is $d_{L(G)}(\ell, \ell') = 1 + d_G(\{i, j\}, \{i', j'\})$, where d_G is the Hausdorff distance between the sets $\{i, j\}$ and $\{i', j'\}$, i.e.*

$$d_{L(G)}(\ell, \ell') = 1 + \min\{d_G(i, i'), d_G(i, j'), d_G(j, i'), d_G(j, j')\}. \quad (2.38)$$

Proposition 2.2. *Let $B(G)$ the bipartite graph associated to $G = (V, E)$. Then for every $i, j, k, i', j' \in V$*

- (i) $d_{B(E)}(i, j) = 2d_G(i, j)$,
- (ii) $d_{B(G)}(i, \{j, k\}) = 1 + 2 \min\{d_G(i, j), d_G(i, k)\}$,
- (iii) $d_{B(E)}(\{i, j\}, \{i', j'\}) = 2d_{L(G)}(\{i, j\}, \{i', j'\})$.

Now, having in mind that the adjacency matrix of a directed network (digraph) $\vec{G} = (V, E)$, also denoted by $A(\vec{G})$, is the $n \times n$ dimensional $(0, 1)$ -matrix $A(\vec{G}) = (a_{ij})$ determined by the conditions:

$$a_{ij} = \begin{cases} 1 & \text{if } (i, j) \in E \\ 0 & \text{if } (i, j) \notin E. \end{cases} \quad (2.39)$$

and the bipartite network $B(\mathbf{G})$ associated to $\mathbf{G}$ is defined by $B(\mathbf{G}) = (V \cup E, E(B(\mathbf{G})))$ whose adjacency matrix is given by

$$A_{B(\vec{G})} = \left(\begin{array}{c|c} 0 & H_{\vec{G}} \\ \hline T_{\vec{G}}^t & 0 \end{array} \right), \quad (2.40)$$

where $H = H_{\vec{G}}$ is the incidence matrix of heads of $\vec{G}$ defined by

$$H_{ij} = \begin{cases} 1 & \text{if } \ell_j = (i, -) \\ 0 & \text{otherwise,} \end{cases} \quad (2.41)$$

and $T = T_{\vec{G}}$ is the incidence matrix of tails of $\vec{G}$ defined by

$$T_{ij} = \begin{cases} 1 & \text{if } \ell_j = (-, i) \\ 0 & \text{otherwise} \end{cases} \quad (2.42)$$

Definition 2.1. Given a graph, G , the line graph associated to G , denoted by $L(G)$, is the graph whose vertices are the edges of G , while (ℓ_i, ℓ_j) is an edge in $L(G)$ if ℓ_i and ℓ_j share a node in G (G is now the primal graph of $L(G)$).

It is easily checked that

$$(A_{B(G)})^2 = \left(\begin{array}{c|c} A_G + D & 0 \\ \hline 0 & A_{L(G)} + 2I_m \end{array} \right), \quad (2.43)$$

where $A_G + D$ denotes the matrix obtained by adding to A_G the diagonal matrix $D = b_{ij}$ and b_{ii} is the degree of the vertex i . Whenever there is no risk of confusion A_G will simply be denoted by A .

Also

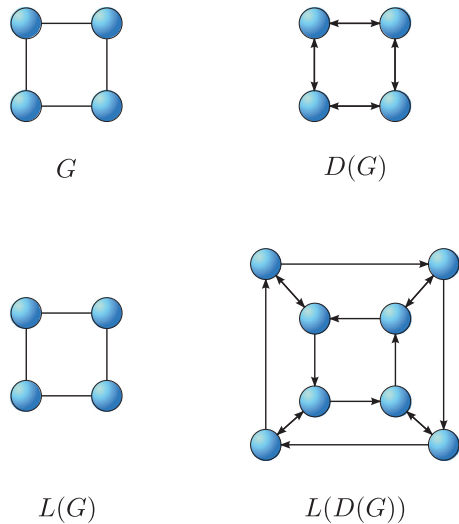
$$(A_{B(\vec{G})})^2 = \left(\begin{array}{c|c} A_{\vec{G}} & 0 \\ \hline 0 & A_{\vec{L}(\vec{G})} \end{array} \right) \quad (2.44)$$

where $\vec{L}(\vec{G})$ denotes the directed line network (line digraph) associated to the directed network $\vec{G}$, i.e., the digraph $\vec{L}(\vec{G})$ whose vertices are the arcs $E(\vec{G})$ of $\vec{G}$, while (e, f) is an arc in $\vec{L}(\vec{G})$ if the end of e coincides with the origin of f (Fig. 2.7).

Remarkably $(A_{B(\vec{L}(\vec{G}))})^2$ is simpler than $(A_{B(G)})^2$.

If now G is an undirected graph, we will denote by $D(G)$ the associated symmetric digraph obtained by replacing each edge of G by an arc pair in which the two arcs are inverse to each other. Observe that $A_G = A_{D(G)}$. Using this idea, there is an alternative definition for the line graph associated with G that has received relatively little attention: the directed line graph $\vec{L}(D(G))$. In any case the networks we are working on are mainly directed networks and therefore the linegraph we will refer in the next chapters is $\vec{L}(\vec{G})$. A more detailed and in-depth work about this topic may be found in [36, 37].

Fig. 2.7 An example of the line-graph of an undirected network (on the *left*) and for a directed network (on the *right*)



2.4.2 Multilayer Networks

In the last few years network scientists have directed their attention to the multiplex character of real-world systems, and explicitly considered the multi-layered nature of networks (see, for example, [19, 38, 81]). In [19] a comprehensive review of both structural and dynamical organization of a network made of diverse relationships (layers) between its constituents may be found. That review covers several relevant issues, from a full redefinition of the basic structural measures, to understanding how the multilayer nature of the network affects processes and dynamics. Complex systems incorporate multiple channels of connectivity and communication. *Multilayer networks* incorporate explicitly multiple channels of connectivity and constitute the natural environment to describe systems interconnected through different categories of connections in such a way that each channel (relationship, activity, category) will be represented by a layer and the same node or entity may have a different kinds of connections (different set of neighbors in each layer). For instance, in social networks, one can consider several types of different actors' relationships: friendship, vicinity, membership of the same cultural society, etc. In such a case, the adequate representation is a multilayer network, where nodes interact through multiple layers of links to properly encompass those topological properties of heterogeneous-type systems which couldn't be captured by the classical single-layer network representation.

Following [19], a *multilayer network* is a pair $\mathcal{M} = (\{G_1, \dots, G_L\}, \{E_{ij} \subset X_i \times X_j, i, j \in \{1, \dots, L\}, i \neq j\})$ where $\{G_1, \dots, G_L\}$ is a family of (directed or undirected, weighted or un-weighted) graphs $G_k = (X_k, E_k)$ (called layers of $\mathcal{M}$) and each $E_{ij} \subset X_i \times X_j, i \neq j$ is the set of interconnections between nodes of the layers G_i and G_j . Note that if $L = 1$ the multilayer network $\mathcal{M}$ is a classic complex network that we call a *monolayer network*.

If $\mathcal{M}$ is a multilayer network, there are two monolayer networks associated to $\mathcal{M}$. The *projection network* of $\mathcal{M}$ is the graph $\text{proj}(\mathcal{M}) = (X, E)$ where

$$X = \bigcup_{k=1}^L X_k, \quad E = \left(\bigcup_{k=1}^L E_k \right) \cup \left(\bigcup_{1 \leq i \neq j \leq L} E_{ij} \right). \quad (2.45)$$

Note that the projection network $\text{proj}(\mathcal{M})$ is the *monolayered vision* of the multilayer network $\mathcal{M}$. In addition to the projection network, the *interconnection network* of $\mathcal{M}$ is the graph $\text{inter}(\mathcal{M}) = (Y, F)$, where $Y = \{\ell_1, \ell_L\}$ and

$$F = \{(\ell_i, \ell_j); E_{ij} \neq \emptyset\}. \quad (2.46)$$

This monolayer network $\text{inter}(\mathcal{M})$ takes into account the (interlayer) connection between layers in the multiplex network $\mathcal{M}$, since each node of $\text{inter}(\mathcal{M})$ represents each layer of $\mathcal{M}$ and two layers are connected in $\text{inter}(\mathcal{M})$ if there are interlayer connection between the corresponding layers of $\mathcal{M}$.

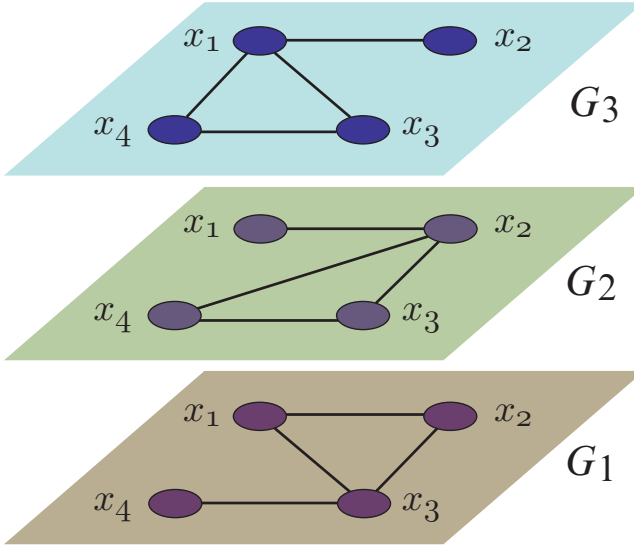


Fig. 2.8 A multiplex network with three layers

It is important to remark that the concept of multilayer network extends some objects introduced in the literature, specifically, a *multiplex network* is a special type of multilayer network in where $X_1 = X_2 = \dots = X_k = X$ and the only possible type of inter-layer connections are ones in which a given node is connected to its counterpart nodes in the other layers, i.e., $E_{ij} = \{(x, x); x \in X\}$ (Fig. 2.8).

In [19, 81] we can find a more detailed and in-depth review about these new concepts and their related topics.

Intentional Risk Management through Complex
Networks Analysis

Chapela, V.; Criado, R.; Moral, S.; Romance, M.

2015, XV, 126 p. 34 illus. in color., Softcover

ISBN: 978-3-319-26421-9