

Chapter 2

Networks Analysis and Beyond

This chapter is aimed to present an overview of the basic concepts underlying the use of network paradigms in data analysis.

2.1 Introduction

In real-world systems, there are plenty of phenomena occurring as a result of complex interactions among several elements. Network paradigm has been successfully introduced to describe the relationships among elements, and hence for analyzing complicated systems.

The reason for the great success of networks is primarily due to the fact that the concept of network is universal; besides the network's apparatus offers a set of intuitive tools to analyze big amounts of data; as such it can be applied to a wide range of fields, including (but not limited to) biology, mathematics, economics, etc. Furthermore, as networks can be represented as sets of nodes and edges drawn between the nodes, from the mathematical viewpoint managing networks is likewise studying graphs. In following, we are going to provide some definitions and notational conventions useful to deal with this formalism.

A graph $\mathcal{G} = (V, \mathcal{E})$ consists of a (finite) set denoted by V , and a collection $\mathcal{E} \subseteq V \times V$, of unordered pairs $\{u, v\}$ of distinct elements from V .

Each element of V is called a vertex (point, node), and each element of E is called an edge (line, link). Typically, it is assumed that self-loops, i.e., edges of the form (u, u) , for some $u \in V$, are not contained in a graph.

A sequence of connected vertexes forms a path. The number n of vertexes, (i.e., the cardinality of V), is called the order of graph and denoted by $|V| := n$. The number m of edges (the cardinality of \mathcal{E}), is called the size of graph and denoted by

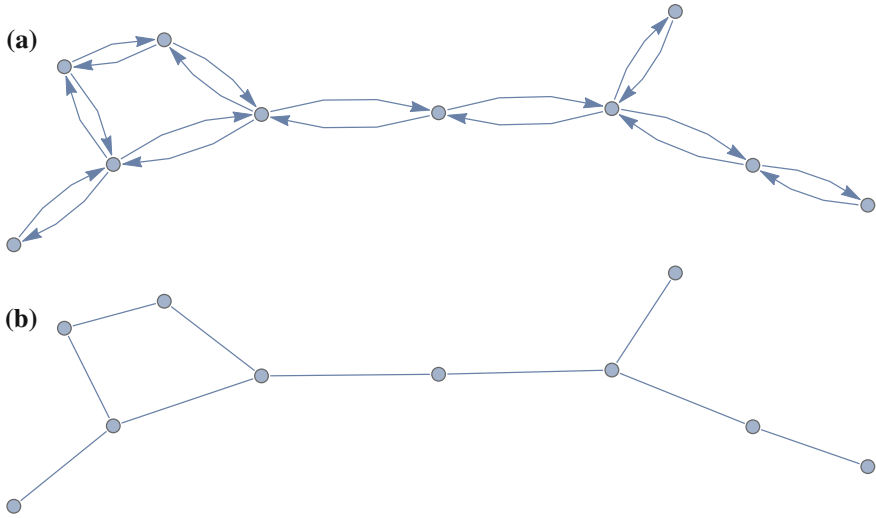


Fig. 2.1 From *left to right* directed (a) versus undirected graphs (b)

$|\mathcal{E}| := m$. The degree k_v of a vertex $v \in V$ is the number of its neighbors in the graph.

Moreover, the graph \mathcal{G} will be claimed to be

- directed, if its edge set is composed of ordered vertex (node) pairs; undirected if the edge set is composed of unordered vertex pairs: Fig. 2.1 shows an example of both types of graphs;
- simple, if it has no loops or multiple edges;
- acyclic, if there is not any possibility to loop back again from every vertex; cyclic if the contrary holds.
- connected, if there is a path in \mathcal{G} between any given pair of vertexes, otherwise it is disconnected;
- regular, if all the vertexes of \mathcal{G} have the same degree;
- complete, if every two distinct vertexes are joined by exactly one edge. The complete graph with n vertexes will be denoted by K_n : some examples of complete graph, varying the number of vertexes from 3 to 8 are provided in Fig. 2.2;
- a path, if consisting of a single path. The path graph with n vertexes will be denoted by P_n ;
- bipartite, if the vertex-set can be split into two sets in such a way that each edge of the graph joins a vertex in the first set to a vertex in the second. A complete bipartite graph is a bipartite graph in which each vertex in the first set is joined to each vertex in the second set by exactly one edge;
- a tree, if it is connected and it has no cycles. If \mathcal{G} is a connected graph, the spanning tree in \mathcal{G} will be a subgraph of \mathcal{G} which includes every vertex of \mathcal{G} and is also a tree. The minimum length spanning tree is called Minimum Spanning Tree (MST). We will turn back on it in Chap. 7, in the second part of this book.

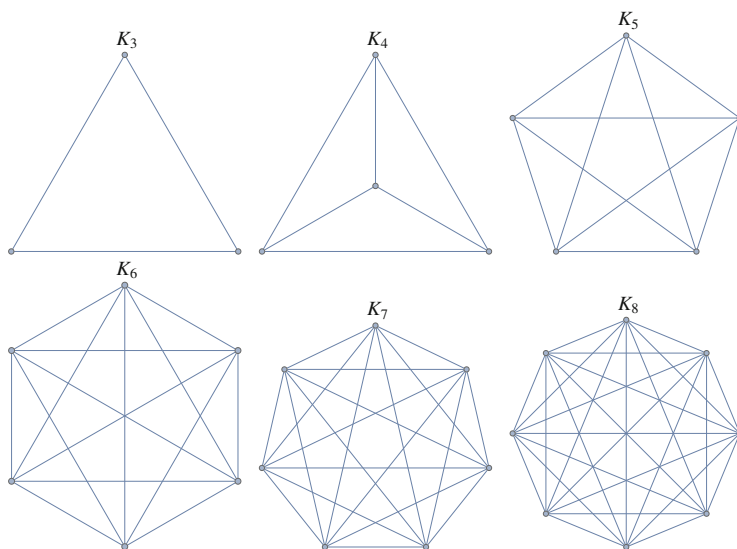


Fig. 2.2 From *top* to *bottom* and in clockwise sense; complete graphs varying the number of nodes n

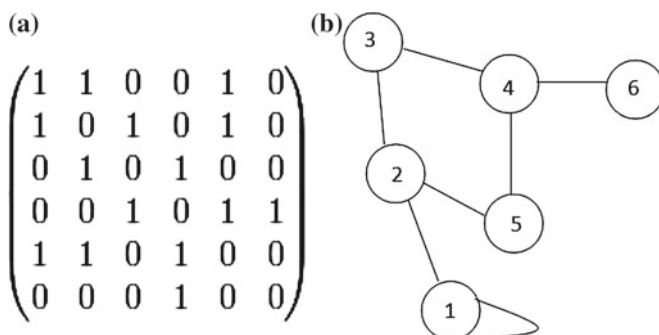


Fig. 2.3 From *left* to *right* the adjacency matrix for an undirected graph (a), and the corresponding graph (b)

Commonly, the representation of graphs passes through the building of the adjacency matrix, i.e., a matrix that marks neighbor vertexes with one, and labels with a zero those nodes that are not adjacent.

Figure 2.3 explains this idea in deeper detail. From left to right, the adjacency matrix for an undirected graph (a) and the corresponding graph (b) are represented. In the left-hand side picture, ones indicate the existence of connection among nodes, while zeroes mean no connection. The adjacency matrix provides then complete information about the graph; as a matter of fact, the provided information is self-containing and enable us, without looking at the right-hand side of Fig. 2.3, to fully

describe the graph, claiming that it has a loop in node 1, and ties between nodes 1 and 2 (2 and 1), 2 and 3, 2 and 5 (3 and 2, 5 and 2), 3 and 4 (4 and 3), 4 and 5, 4 and 6 (5 and 4, 6 and 4).

2.2 Classical Networks

The beginning of the modern networks theory dates back the 1950s and 1960s, when Paul Erdős and Alfred Rényi introduced the *random graph model* [52, 53, 54], also known as Erdős–Rényi (ER) model.

The model consists of n nodes joined by edges which are placed at random between pairs of vertexes; by the notation $\mathcal{G}_{n,p}$, it is meant that p is the independent probability for each possible edge to be present; on the contrary, $1 - p$ is the probability for the edges of being absent. In the case of simple networks, the expected number of edges \mathcal{E} is expressed as

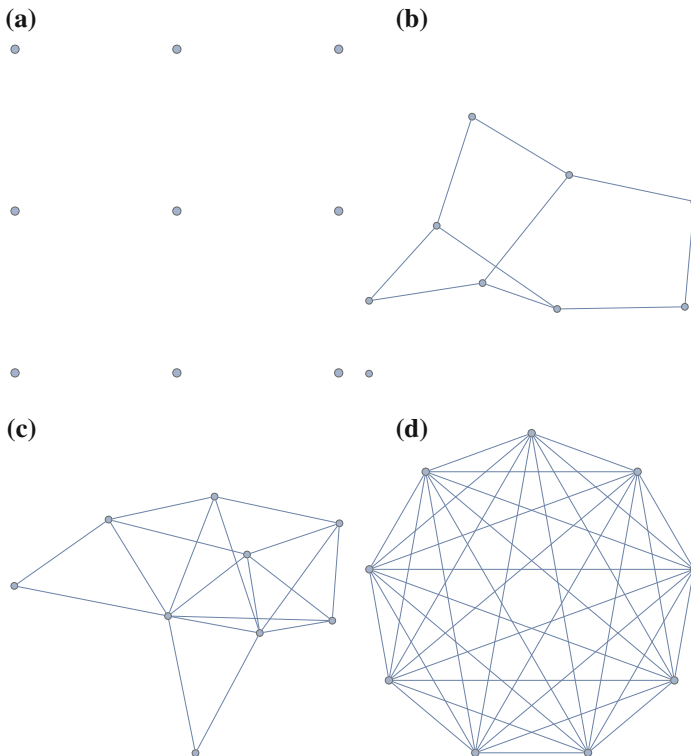


Fig. 2.4 Random graph with probability p drawn from a binomial distribution. From *top to bottom* and from *left to right*, results for $p = 0$ (a), $p = 0.4$ (b), $p = 0.6$ (c), and for $p = 1$ (d)

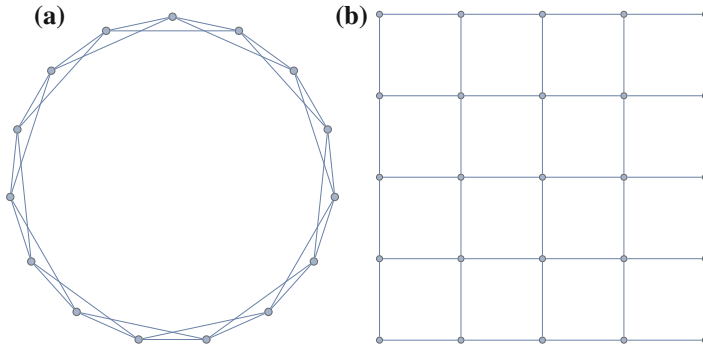


Fig. 2.5 Lattice networks: some examples. In the *left-hand side*, a one-dimensional network (a) is provided, while a rectangular two-dimensions lattice is shown in the *right-hand side* (b)

$$\mathcal{E} = p \binom{n}{2} = p \frac{n(n-1)}{2} \quad (2.1)$$

because the total number of possible edges (combinations) is $C_2^{(n)} = n(n-1)/2$.

Figure 2.4 shows some examples of the random graph model, created by assuming to draw p from a binomial distribution.¹ Networks change with the probability p .

Clearly, the literature discussed in this chapter is far from being exhaustive; the interested reader can refer to [60, 92, 125] for several examples of this approach. Note that this model has been widely employed in several research fields such as sociology, ecology, and mathematical biology because of its simplicity; however, it conflicts with real-world networks. Nevertheless, random networks are often used as benchmarks in comparison to the statistical properties observed in real-world networks, in order to test the significance of those latter.

2.3 Lattice Network

Another famous network model involves lattice networks whose examples are provided in Fig. 2.5.

Lattice networks are simply networks where nodes are arranged in a rectangular (or more generally, a regular) lattice, aimed to overcome the major drawback of the ER model. As a matter of fact, although the Erdős-Rényi random network is a nice and tractable model, it fails to capture some nonrandom aspects of real-world networks. This is, for instance, the case of patterns consisting of a small collection of nodes with a certain combinations of edges (*motifs*), the existence of nodes with much higher connectivity than the rest (*hubs*).

¹This assumption is fully straightforward, as we are going to explain later with further details in Sect. 2.4.

Lattice networks are relatively unsuitable for hypothetical models in network analysis, because they seem to be artificial. However, they are useful when considering spatial dimensions such as distance. By assuming the regular networks in mathematical models, further, it may be easy to derive exact solutions of the models because of the regularity of lattice networks. Thus, in addition to the random network model, this model is utilized in various research fields.

However, real-world networks possess remarkable statistical properties that cannot be fully explained either by random networks or lattice networks. Over the past two decades, this issue determined the flourishing of alternative models, best tailored to fit the observable phenomena.

2.4 Scale-Free Networks

In order to discuss this kind of network, we need some preliminary definitions that we are going to introduce in the following subsection.

2.4.1 Degree Distribution

The node degree, the simplest measure of a network, is defined as the number of edges (neighbors) that a node has. A simple question might arise, regarding the way in which degree is distributed in real-world networks. Barabasi and Albert [21] answered this question by defining the degree distribution

$$P(k) = \frac{1}{n} \sum_{i=1}^n \delta(k_i - k) \quad (2.2)$$

where k_i is the degree of the i th node, and $\delta(\cdot)$ is the Kronecker's delta function. As (2.2) returns 1 when $k_i - k = 0$, and 0 otherwise, the term $\sum_{i=1}^N \delta(k_i - k)$ corresponds to the number of nodes with degree k .

In ER random networks, $P(k)$ corresponds to the probability of a node to have k edges. Since an edge is independently drawn between two given nodes with probability p , the degree distribution can be expressed as a binomial distribution.

In the case of lattice networks, the degree distribution will be peaked too, because each node has the same degree.

2.4.2 Power-Law Distribution in Real-World Networks

The degree distributions $P(k)$ of several real-world networks follows a power-law distribution [9, 49]:

$$P(k) \propto k^{-\zeta} \quad (2.3)$$

where ζ is a constant, the so-called *degree exponent* [10], with values bounded within the interval (2, 3).

A correct interpretation of the power-law distribution suggests that a few nodes integrate numerous nodes, while most of the remaining nodes do not. In such networks, the average degree is not representative, because it is common to observe vertexes (*hubs*) with a degree that greatly exceeds the average. For this reason, networks sharing this statistical property are called *scale-free networks*. Moreover, the scale-free property seems to have a strong connection to self-similarity [165], being satisfied as

$$f(Cx) = (Cx)^\alpha = C^\alpha f(x) \quad (2.4)$$

for constants C and α , where $f(x) = x^\alpha$.

2.4.3 Barabasi–Albert Model

Hubs (and hence power-law degree distribution) represent the most striking difference between a random and a scale-free network. A model of network bridging this gap was proposed and analyzed in [21], highlighting two hidden assumptions of the Erdős–Rényi model, each of which are violated in real networks, namely growth and preferential attachment.

For what is concerning the former, in fact, the random network model assumes that the number of nodes n is fixed (time invariant). Real networks, however, are the result of a growth process that leads n to continuous increase. Moving to the second feature, it has been proved that in real networks new nodes prefer to link to the more connected nodes. Due to such preferential attachment, new nodes are therefore more likely to connect to the more connected nodes than to the smaller degree nodes. Hence, the more connected nodes will acquire links at the expense of the less connected nodes, eventually turning into hubs.

The Barabasi–Albert Model (BA) attempts to reproduce those two simple mechanisms described in previous rows.

The BA model network can be generated following the steps provided below.

- (i) Define the number n_0 of starting isolated nodes, a number $0 < n_1 \leq n_0$, and the number n of desired overall nodes.
- (ii) While n_0 is lower than n , connect a new node to the n_0 isolated nodes.
- (iii) Add and connect a new node to n_1 existing nodes, which are selected with the probability given by

$$\Pi_i = \frac{k_i}{\sum_j k_j} \quad (2.5)$$

- (iv) Repeat Steps (ii) and (iii) until the network size reaches the target size n .

Fig. 2.6 An example of Barabasi–Albert network model (a) and the corresponding power-law degree distribution (b)

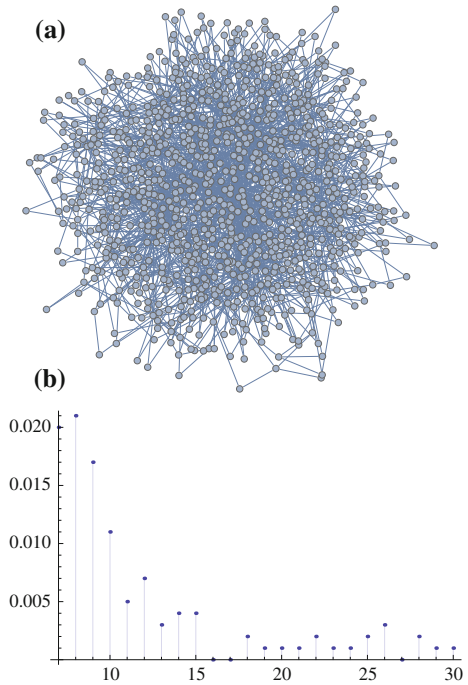


Figure 2.6 shows an example of Barabasi–Albert network model and the corresponding power-law degree distribution.

2.5 The Configuration Model

As seen in Sect. 2.4, the degree distribution is a handy tool for exploring properties of networks. An interesting point, however, concerns the discussion of the extent to which the degree distribution captures certain aspects of the network. In particular, adding minimal structure beyond a degree distribution could allow the exploration of the consequences of having a large spread of degrees, including hubs that have a much larger degree than most nodes.

The configuration model [25, 114] makes possible to generate a network model that has exactly a fixed degree distribution; as such, it allows to generate networks with the same degree distribution as a given network.

In order to understand the model, a preliminary definition can be given.

Definition 2.1 The degree sequence of a graph $\mathcal{G} = (V, \mathcal{E})$ is the sequence of degrees of vertexes V written in nonincreasing order.

For example, going back to the graph illustrated in Fig. 2.2, its degree sequence is $(3, 2, 2, 2, 2, 1)$. Note that not all nonincreasing sequences of nonnegative integers can be realized as degree sequences of simple graphs. When it is possible, these sequences are called graphical sequences.

Definition 2.2 A stub or half hedge is a hedge connected only on one side, while the other remains free.

Thanks to those definitions, the configuration model algorithm can be summarized as follows:

- (i) Define a sequence (d_1, d_2, \dots, d_n) such that $d_1 \geq d_2 \geq \dots \geq d_n$, and $\sum_{i=1}^n d_i$ is even.
- (ii) Create vertexes $V = \{1, 2, \dots, n\}$, and assign them stubs or half edges according to the sequence (d_1, d_2, \dots, d_n) .
- (iii) Pick any two stubs uniformly at random, and connect their free ends; these two stubs became one edge.
- (iv) Repeat Step (iii) until no free stubs are left.

Observe that the algorithm allows both loops (created when picking two stubs from the same vertex) and multiple edges (created when picking pairs of stubs from the same pairs of vertexes). A variant without multiple edges can be found in [34, 35].

2.6 Small-World Networks

Small-world (SM) networks, according to Watts and Strogatz [178], are a class of networks that are highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. As a result, those networks, also known as Watts–Strogatz (WS) networks, have unique properties of regional specialization with efficient information transfer. Interestingly, in those networks the distance between a given node pair is known to be surprisingly small, although the network size is very large.

An intuitive example of SM organization is provided by social networks, in which cliques or clusters of friends are interconnected but each person is really only five or six people away from anyone else. This property is referred to as the *small-world property*, and was originally known as the *six degrees of separation* in sociology.

The WS network can be built in four steps:

- (i) Start with a one-dimensional lattice with n nodes.
- (ii) Select in a clockwise sense a node and the edge connecting it to its nearest neighbor.
- (iii) Rewire the edges with the probability p , as given in (2.5), and elect at random a new target node.
- (iv) Repeat Steps (ii) and (iii) until one lap is completed.

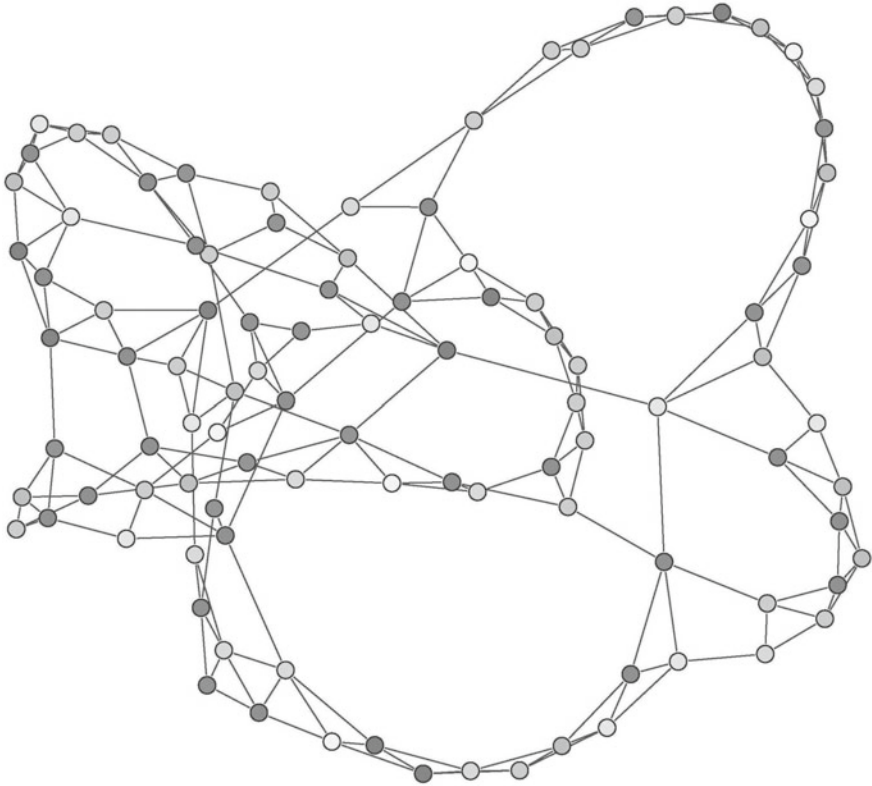


Fig. 2.7 An example of small-world network model

An example of small-world network is provided in Fig. 2.7.

The model generates networks similar to lattice networks when $p = 0$, and close to random networks when $p = 1$. In practice, the WS model includes the random network and the lattice network models as special cases; in this sense, we can claim that WS network expresses the transition from lattice networks to random networks.

2.7 Measuring the Robustness of Networks

We are now going to provide some insights on the main tools employed in order to assess networks features; more specialized indexes will be then introduced when necessary in the second part of the book, in the chapters focusing on practical applications.

2.7.1 Average Shortest Path Length

The distance between a node pair can be measured using the average shortest path length of a network, which is defined as

$$ASPL = \frac{1}{n(n-1)} \sum_{i=1}^n \sum_{j=1}^n d(i, j) \quad (2.6)$$

where $d(i, j)$ indicates the shortest path length between nodes i and j , with $d(i, i) = 0$, and $d(i, j) \rightarrow +\infty$, if there is no shortest path between nodes i and j . Thus, the average shortest path length is only calculated in connected networks, in which there are shortest paths between all node pairs.

2.7.2 Clustering Coefficients

Clustering coefficients help to determine the level of organization inside the network.

A first measure concerns the density evaluation among neighbors for each network node

$$CC_i = \frac{E_i}{\binom{k_i}{2}} = \frac{2E_i}{k_i(k_i - 1)} \quad (2.7)$$

where E_i is the number of edges among the neighbors of node i , and k as the degree of the i th node. Note that CC_i is defined as the ratio of the number of edges among the neighbors to the number of all possible connections among the neighbors.

The overall tendency of clustering can be then measured by the average clustering coefficient $C = \frac{1}{n} \sum_{i=1}^n CC_i$, where n is the number of nodes in the network. A high average clustering coefficient implies that the network is clustered.

2.7.3 Hierarchical Modularity

Hierarchical modularity relates to the possibility for a system to exhibit modularity on several topological scales. For many years, researchers have been fascinated by the ubiquity of modularity and hierarchical modularity across social, technological, and biological systems.

One of the earliest and most influential ideas was formulated by Simon [163, 164], who argued that a “nearly decomposable” system built of multiple, sparsely interconnected modules allow faster adaptation or evolution of the system in response to changing environmental conditions.

Modular systems can evolve by change in one module at a time, or by duplication and mutation of modules, without risking loss of function in modules that are already well adapted. Well-adapted modules thus represent stable intermediate states such that further evolution of other modules does not jeopardize function of the entire system. This robustness is a major advantage for any system evolving under changing or competitive selection criteria, and this may explain the widespread prevalence of modular architectures across a very wide range of information processing systems.

This statistical property is defined as

$$HC(k) = \frac{\sum_{i=1}^n CC_i \times \delta(k_i - k)}{\sum_{i=1}^n \delta(k_i - k)} \quad (2.8)$$

where CC_i is the clustering coefficient of node i as defined in (2.8), and $\delta(\cdot)$ is the Dirac's delta, as previously defined in Sect. 2.4.1: [142] found that the degree-dependent clustering coefficient follows a power-law function in several real-world networks.

2.7.4 Assortativity

The relationship among the degrees in a connected node pair is very interesting. Since real-world networks are nonrandom, we can expect this relationship to be significant. However, the degree distribution only involves the degree of each node. Thus, we need an alternative measure for characterizing such a relationship of degrees.

To characterize the relationship between node degrees, [124] proposed the assortative coefficient

$$\rho = \frac{4k_i k_j - (k_i + k_j)^2}{2(k_i^2 + k_j^2) - (k_i + k_j)^2} \quad (2.9)$$

where k_i and k_j are the degrees of two nodes at the ends of an edge. This is simply the Pearson correlation coefficient of degrees between a connected node pair, and it lies in the range $[-1, 1]$. The relationship between the assortative coefficient and network structures can be described as follows:

- For $\rho > 0$, the network shows assortativity.
- For $\rho = 0$, there is no correlation between the degrees in a connected node pair, that is, such networks are randomly constructed.
- For $\rho < 0$, the network shows disassortativity, in which low-degree nodes tend to connect to high-degree nodes.

It is generally known that social and technological networks exhibit assortativity and that biological and ecological networks exhibit disassortativity.

2.7.5 Degree Correlation

Assortativity relates to another important measure known as degree correlation. The degree correlation characterizes the expected degree of the neighbors of a node with degree k

$$\bar{k}_{nn}(k) = \frac{\sum_{i=1}^N \zeta_i \delta(k_i - k)}{\sum_{i=1}^N \delta(k_i - k)} \quad (2.10)$$

where ζ_i denotes the average nearest-neighbor degree, being:

$$\zeta_i = \frac{1}{k_i} \sum_{h \in V(i)} k_h, \quad (2.11)$$

and $V(i)$ corresponds to the set of neighbors of node i , that is, the positive and negative degree correlations indicate assortativity and disassortativity, respectively.

2.8 Centrality Measures

In networks theory, the centrality refers to indicators which identify the most important vertexes within a graph. Several measures of centrality have been proposed, at present. We simply enumerate the most important ones to the extent of our study.

The *degree centrality* is the simplest centrality measure, as for the generic node i is defined as

$$C_D(i) = \frac{k_i}{n - 1} \quad (2.12)$$

where n is the network size (i.e., the total number of nodes). Since this centrality is essentially similar to the node degree, this is widely used in network analysis.

The *closeness centrality* [62] is based on the shortest path length between nodes i and j , and it is expressed as

$$C_C(i) = \frac{n-1}{\sum_{j=1, j \neq i}^n \text{dist}(i, j)} \quad (2.13)$$

where $\text{dist}(i, j)$ is the distance between nodes i and j .

The *betweenness centrality* [62] is based on the shortest path between nodes

$$C_B(i) = \sum_{s \neq t \neq i} \frac{sp_{st}(i)}{sp_{st}} \quad (2.14)$$

where $sp_{st}(i)$ and sp_{st} are, respectively, the number of shortest paths between nodes s and t , on which node i is located, and the number of shortest paths between nodes s and t . For normalization, the betweenness centrality is finally divided by the maximum value.

Finally, the *Katz centrality* [88] measures the influence of a node in a network

$$C_{Katz}(i) = \sum_{k=1}^{+\infty} \sum_{j=1}^n \beta^k (A^k)_{ji} \quad (2.15)$$

where A is an adjacency matrix, and β is an attenuation constant.



<http://www.springer.com/978-3-319-21439-9>

Computational Intelligence Paradigms in Economic and
Financial Decision Making

Resta, M.

2016, XX, 172 p., Hardcover

ISBN: 978-3-319-21439-9