The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific
and Cultural Organization

International Atomic
Energy Agency

SMR.1656 - 10

School and Workshop on Structure and Function of Complex Networks

16 - 28 May 2005

Definitions and Methodology
Graph Structures
Scale-Invariance
The Orgins of Power Law Functions
Graph Generating Models

Guido CALDARELLI
Universita' di Roma "La Sapienza"
Gruppo Pil, Dipartimento di Fisica
Edificio Fermi
Piazzale Aldo Moro 2
00185 Roma
ITALY

Part I

Definitions and Methodology

1

INTRODUCTION TO GRAPHS

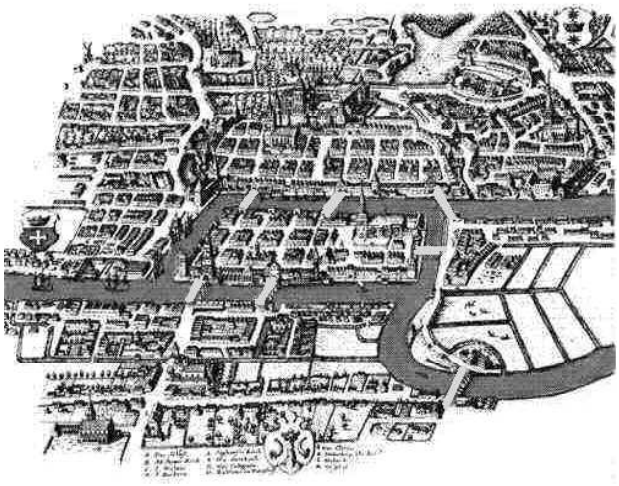

FIGURE 1.1. An ancient picture of the city of Königsberg. From MacTutor History of Mathematics archive (http://www-gap.dcs.st-and.ac.uk/∼history)

His name was Leonard Euler and he was just a country boy who had never seen Königsberg before. Still the power of mathematics helped a young Swiss mathematician in solving a puzzle about one of the most elegant cities of the $XVIII$ century. According to the current view, modern Graph Theory traces back to the mathematician Leonard Euler[1] that has been the first scientist to introduce the notion of graphs. The beginning of this story is very similar to a joke, but eventually resulted in the creation of a new branch of mathematics. Euler wanted to answer a popular question of his time. If we are in the center of the city of Königsberg (Prussian city in the first part of XVIII century) can we do a stroll by crossing only once everyone of the seven bridges shown in Fig. 1.1?

The situation was very similar to the simplified map in Fig. 1.2 apart that we didn't reproduce streets, buildings or the actual shape of the river borders and islands. A brute force solution of this problem could be summarized as follows: we start from a side, we check all the possible paths and we stop if we find the one desired. Apart the lack of elegance of such a procedure, in this way we do

[1]One of the greatest mathematicians of all times. Born the 15 April 1707 in Basel (Switzerland), he died the 18 Sept 1783 in St Petersburg (Russia). Here we discuss of his publication: *Solutio problematis ad geometriam situs pertinentis* of 1736

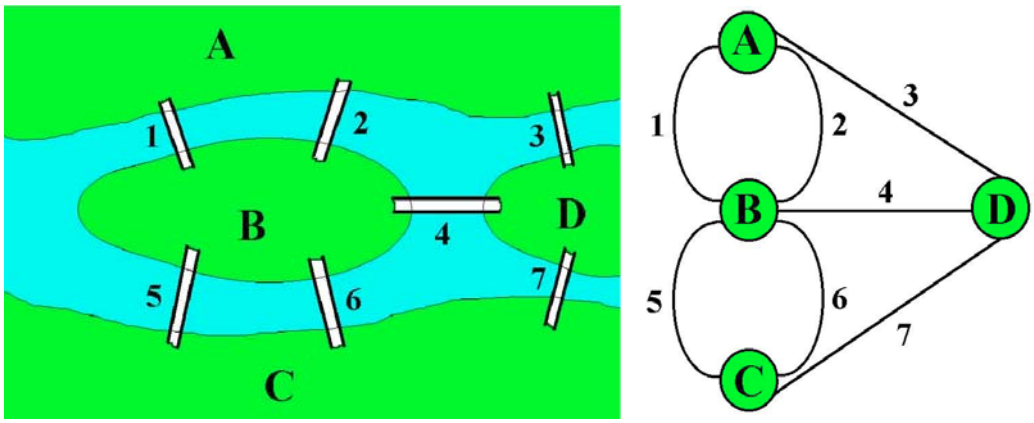

FIGURE 1.2. On the left a very schematic map of the town center of Königsberg at the time of Euler (1736). On the right the resulting graph

not provide a real solution. Indeed if we have a similar problem with a different number of bridges we simply repeat the "exact enumeration" from scratch. Even worst, if at a certain point the number of bridges is larger than seven, the possible paths become so many that it is simply impossible to proceed like that.

General solution of the problem needs the abstraction of the mathematics. The step ahead made by Euler was to condensate all the relevant information in one map of Königsberg even more simplified that the previous one. This map is shown on the right of Fig. 1.2.

Here real distances do not matter any more. Different parts of the cities (larges or small) are described by points that are called **vertices**. If they are linked (through a bridge) we draw a line (called **edge**) between them. The map of the city becomes a "graph"[2]. Through this formalism now the original problem translates in the more abstract request: "Is it possible to find a path that passes through all the edges exactly once?"

Now we have to consider if the problem is easier thanks to this formulation. The answer is a mixture of yes and no. No because "essentially" (and hopefully) the request remained the same, yes because now we restricted all the attention on these new things we called vertices. All the parts of the city are drawn in the same way (as a point). All of them are equal. Therefore a solution (if exists) must refer to some intrinsic properties of such objects.

An immediate intrinsic quantity here is the number of edges per vertex, hereafter indicated as the **degree** of the vertex. It is an integer number. Therefore it is either even or odd. If it is even (to fix ideas let us start with 2) we realize that the vertex is a crossing point. Indeed we can enter the vertex from one bridge and we exit from the other. If the number of edges is even but larger than two (i.e. $4, 6, 8$ etc.), the same argument holds. To check that, just divide the edges in couples. The degree will be a number $m = 2 * n$, where n is the number of the

[2]It is worth noting that the same trick is used today to draw the maps of the various stations in the underground of the modern cities

couples of edges. For every couple we get in through one edge and we get out from the other. The vertex is visited two, three or in general n times. Conversely, vertices with odd degree can only be starting or ending point of the path.

That is the solution of the problem! Our request to pass on every bridge exactly once can be satisfied only if the vertices with odd degree are zero (starting and ending point coincide) or two (starting and ending point do not coincide). If you think a little bit and work it out on piece of paper, you realize that you cannot have a graph with only one vertex with odd degree, so the above ones are the only two possibilities (we see later that the sum of the degrees of the various vertices is an even number, precisely it is twice the total number of edges). Now we go back to the graph in Fig. 1.2 and we discover that none of the above conditions are verified. Actually, all the four vertices have an odd degree. Therefore the path is not possible.

Starting from such problem, graph theory became more and more elaborated. Since the time of Euler many mathematicians have made important contributions to it. We do not want to provide a formal course in graph theory. Rather, we will focus here only on the basic notions allowing to study and describe scale-invariant networks. For those who would like to start a detailed study on this topic we can suggest (amongst the many resources available also in electronic form) some introductory books (Bollobás, 1979; Bollobás, 1985; Diestel, 2000; West, 2001).

In this chapter we will provide a definition of what a graph is. We also introduce a way to represent graphs through matrices of numbers. This representation will make some computations particularly easy. When graphs are of very large order (when they have many vertices) the only way to describe them is by means of statistics. We will provide here some mathematical instruments. Finally we present the probability distributions we use in the rest of the book.

1.1 Graphs, Oriented Graphs and Weighted Graphs

The various networks present in this book are different realizations of the same mathematical object known as **graph**[3].

Graph theory is a branch of mathematics (in particular geometry). To describe it there is no other choice than a rigorous way. Therefore also the order of notions has its own importance. A more traditional exposition of the basic concepts is presented in Appendix B. Here we use only part of this series of notions and I took the liberty to cut and paste from a more traditional list only the quantities strictly necessary for the purpose of this book.

[3]In general one refers to networks as any real system that can be described by means of a mathematical object called graph. Here in this book we will follow in a loose way this rule and keep the name of graph whenever talking about mathematical properties. Nevertheless sometime they will be used as synonymous

Graphs are assigned by giving the set of the vertices and the set of their connections. **Vertices** and **edges** are the technical terms used in graph theory and we use preferably them, even if some author use also site, node or link (whenever talking about graph theory we will use only the traditional notation for the sake of clarity).

Edges can have arrows or not, that is they can be crossed in one direction only (we see that this is the case of hyperlinks in a HTML document). In this case the graph is an **oriented graph**.

A further generalization is also possible, one can think that every edge has a different value assigned. In the case of transportation networks (a system of pipelines or the Internet cables) this could represent for example the maximum load allowed. Whenever this extra information is provided we deal with a **weighted graph**[4]. Further generalizations are possible and used, but for our purposes we can restrict to that.

The mathematical symbol to indicate a graph composed by n vertices and m edges is usually $G(n, m)$. These parameters n and m are not independent each other. Actually there is a bound in the number of edges we can draw. Each vertex can establish an edge with $(n-1)$ other vertices. This holds for every one of the n vertices. This give a total number of $n(n-1)$ possibilities where we counted twice the same edge. The maximum number of edges is exactly one half of that $m_{max} = n(n-1)/2$.

If the starting and ending vertices make a difference (as in the case of oriented graph) then we do not have to divide by two the above quantity. In this case the maximum number of edges is given by $n(n-1)$. Note here that under this definition, regular grids (as for example the simple cube lattice) can be considered as trivial examples of graphs.

1.1.1 *Adjacency Matrix*

The structure of the graph $G(n, m)$ can also be represented by means of a matrix. Matrices are tables of numbers very useful to solve problems of linear algebra. We assume that the reader has a basic knowledge of them. Just to refresh memory we list some of their properties in Appendix C. If needed, a good textbook must be consulted (Golub and Van Loan, 1989).

In the case of graph we introduce the **Adjacency Matrix** $\mathbf{A}(n, n)$ whose entries a_{ij} are 0 if vertices i, j are not connected and 1 otherwise. This is a somewhat extended nomenclature. Instead of listing only the edges actually drawn, we decide to write down n^2 numbers that is more than twice the maximum number of edges ($m_{max} = n(n-1)/2$) we can draw in a "simple" (i.e. non-oriented) graph. The reason of this choice is given by the fact that through this extended representation it is possible to derive analytically some results of a certain importance.

For the moment let us proceed with this representation and consider the form of this matrix. The diagonal elements represent the presence of a edge between

[4] Almost everywhere in this book the weights of an edge will be a real positive number.

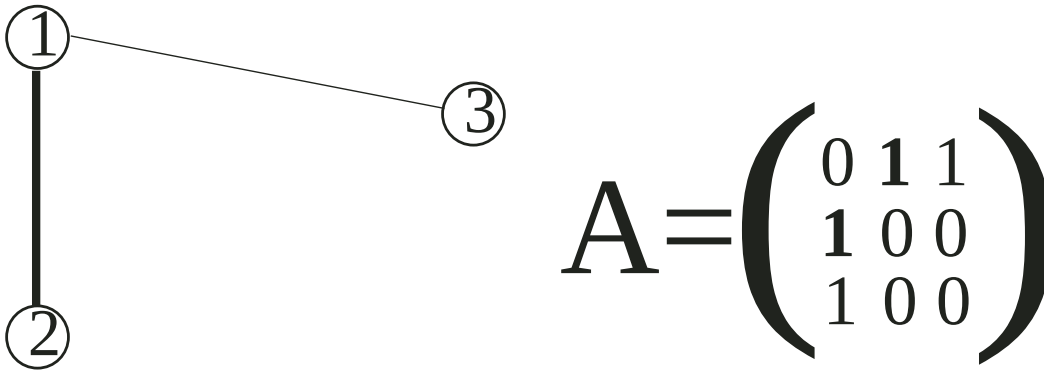

FIGURE 1.3. On the left a simple graph whose Adjacency matrix is reported on the right. Bold edge corresponds to bold entries.

a vertex and itself (whatever it could mean). Whenever differently specified we consider those entries equal to 0.

Note that this matrix is symmetric (it means $a_{ij} = a_{ji}$) only in the case of non-oriented graphs. For oriented graphs instead the elements a_{ij} are generally different from the elements a_{ji}. For example in the case of only one edge going from vertex 2 to vertex 3 we have that $a_{23} = 1$ and $a_{32} = 0$[5]

Through this representation we can easily describe the case of the weighted graphs. Now instead of giving only 1 and 0, we assign a real number (the weight) to the entry a_{ij}. We obtain then an adjacency matrix composed by real numbers for the edges present and 0 otherwise. In the following we refer to this matrix with the symbol $\mathbf{A^w}(n, n)$. Its elements will be then indicated by a^w_{ij}

1.1.2 *Quantities of interest*

As written above, readers looking for an organic list of concepts must refer to Appendix B. The following definitions are provided firstly for non-oriented and non-weighted graph. After the simple case we also provide the definition for the general case.

- The graph **order** is *the number n of its vertices.*

- The graph **size** is *the number m of its edges.*

[5]We follow here the overwhelming convention to write the oriented edges of vertex i on the i^{th} row of $\mathbf{A}$. Actually, it would be simpler to write the edges along the i^{th} column. We see that when the adjacency matrix is transformed in a transition matrix (i.e. every entry is divided by the degree of the vertex) we want $\mathbf{A^T A} = \mathbf{I}$. This result can only be obtained by writing the entries along the columns.

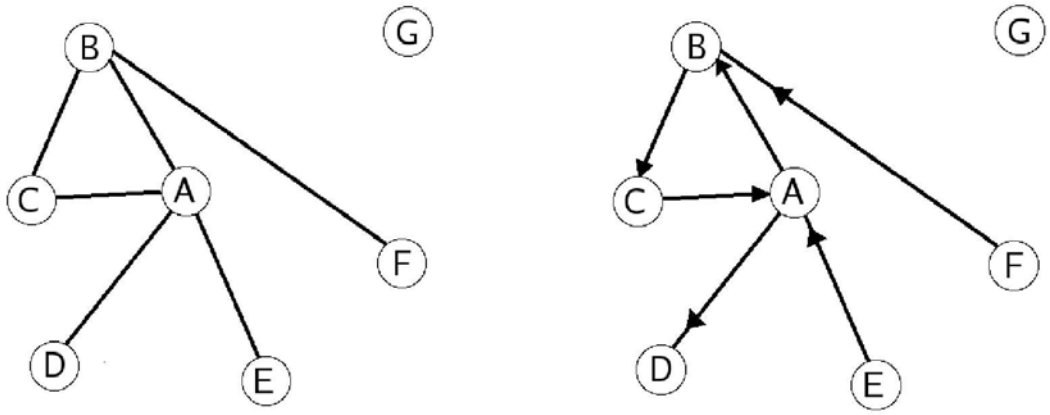

FIGURE 1.4. On the left: Graph $G(7,6)$. The **order** of the graph is 7; the **size** is 6. The **degree** of vertex A is 4, the **degree** of vertex C is 2. On the right: An oriented graph of the same size. In this case the *in-degree* of vertex A is 1, the *in-degree* of vertex C is 2 and its out-degree of vertex C is 1.

We recall that in a graph of size n we can draw a maximum number of edges given by $m_{max} = \frac{n(n-1)}{2}$. This formula is easy to understand. We have n possible starting points and for every one of them $n-1$ destinations. Repeating this procedure we count twice the path from one vertex to another.That is why we must divide by two. This factor on the denominator disappears in a oriented graph where we care about the difference between origin and destination of the edges.

Two immediate limits are present. If no edge is drawn then the graph is *empty* and it is indicated by E^n. If all the edges are drawn, the graph is *complete* and it is indicated by K^n.

- The vertex **degree** is *the number of its edges*. As mentioned before, the sum of all the degrees in the graph is twice the number of the edges in the graph. This happens because any edge contributes to the degree of the vertex origin and to the degree of vertex destination.

A compact way to compute the degree consists in running on the different columns of a fixed row in the adjacency matrix $\mathbf{A}(n,n)$ looking for all the 1's present. This means that the degree k_i of a vertex i can be computed as

$$k_i = \sum_{j=1,n} a_{ij}. \tag{1.1}$$

  - In oriented graphs this quantity splits in *in-degree* and *out-degree* for edges pointing in and out respectively.

Since the a_{ij}'s are different from the a_{ji}'s in the adjacency matrix we have that $a_{ij} = 1$ if and only if an edge goes from i to j. This means that

$$k_i^{in} = \sum_{j=1,n} a_{ji} \tag{1.2}$$

$$k_i^{out} = \sum_{j=1,n} a_{ij} \tag{1.3}$$

∗ In the above definitions all the edges count the same. This is not the case of weighted graphs.

In this case the weighted degree k_i^w of a vertex i is usually defined as $k_i^w = \sum_{j=1,n} a_{ij}^w$. Note that, with this definition one recovers the usual degree if matrix $\mathbf{A^W}$ is replaced by $\mathbf{A}$.

- The **distance** d_{ij} between two vertices i, j is *the shortest number of edges one needs to travel to get from i to j*. Therefore the neighbours of a vertex are all the vertices which are connected to that vertex by a single edge.

Using the adjacency matrix this can be written as

$$d_{ij} = min\{\sum_{k,l \in \mathcal{P}_{ij}} a_{kl}\} = min\{\sum_{k,l \in \mathcal{P}_{ij}} \frac{1}{a_{kl}}\} \tag{1.4}$$

where $\mathcal{P}_{ij}$ is a path connecting vertex i and vertex j (Note that formally both the sum the a_{ij} and that of the inverse produce the same result. This is because all the existing edges have a value of 1. For the case of "simple" graphs one or the other definitions make equally sense. In the following we see that when considering weighted graphs according to the "physical" meaning of the weight one or the other quantity have a different sense (and of course give different results).

∗ If the graph is oriented one has to follow the direction of the edges. Therefore the distances are generally larger than in the homologous non-oriented graphs.

∗ In the case of weighted graphs, instead of summing for every step a distance of 1 we can assume that the distance is related to the values of the weight. If the graph represents a distribution network as a pipeline of water, the weight can represent the section of pipe. Intuitively then one can think that two vertices related by a "strong" edge (i.e. an edge whose weight is large) are nearer than two related by a weak edge. In this case (as it could be the case of the Internet where weight reports for example the load) we define the distance as the sum of the inverse of the weights.

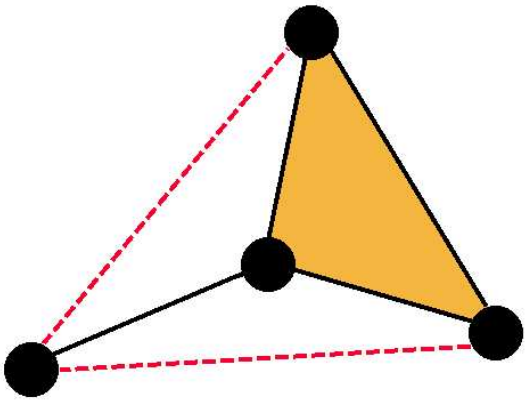

FIGURE 1.5. The clustering coefficient of the central vertex is 1/3. This is because his degree is three and its neighbours can be connected each other in three different ways. Of these possibilities (dashed line) only one is actually realized (solid line) and therefore $C = 1/3$. The three connected vertices form the coloured triangle. For that reason, sometime the clustering coefficient of a vertex is defined through the number of triangles it belongs to.

In formulas $d_{ij} = min\{\sum_{k,l\in\mathcal{P}_{ij}} \frac{1}{w_{kl}}\}$

On the other hand, if the network is an electric grid and the weight gives the resistance opposed by current flow on any edge, then the larger is the resistance the larger is the distance between two vertices. In this case the generalization of the above formula that makes more sense is given by

In formulas $d_{ij} = min\{\sum_{k,l\in\mathcal{P}_{ij}} w_{kl}\}$

- The **diameter** D of a graph is in this book *the largest distance you can find between two vertices in the graph*. Some other definition (as the average of distance) are possible.

- The **clustering coefficient** C is a basic characterization of clustering. C is given by the average fraction of pair of neighbours (of the same vertex) that are also neighbours each other. A simple example of that is shown in Fig. 1.5. In this case the central vertex has three neighbours. These can be connected in three different ways, since only one is actually present, this gives $C = 1/3$. For the empty graph E^n we have $C = 0$. The maximum value of $C = 1$ is obtained for the complete graph K^n. In general we may write the clustering coefficient as the fraction of actual edges over the possible ones between the vertices i, j, k.

Using the formalism of the adjacency matrix we have that for a vertex i

$$C_i = \frac{1}{(k_i)(k_i-1)/2} \sum_{j,k} a_{ij} a_{ik} a_{jk} \tag{1.5}$$

* If the graph is oriented, the generalization is not straightforward. It is more or less natural to consider the extension of the clustering coefficient only for the in- or out-degree, splitting therefore the contribution one has for the non-oriented graph. The problem is then to consider which direction of the edge between the neighbours has to be counted (see Fig. 1.6).

  In general one tends to join the two possible direction such that the clustering coefficient takes the form

$$C_i^{in} = \frac{1}{(k_i^{in})(k_i^{in}-1)/2} \sum_{j,k} a_{ji} a_{ki} \frac{(a_{jk}+a_{kj})}{2} \tag{1.6}$$

$$C_i^{out} = \frac{1}{(k_i^{out})(k_i^{out}-1)/2} \sum_{j,k} a_{ij} a_{ik} \frac{(a_{jk}+a_{kj})}{2} \tag{1.7}$$

* For the weighted graphs the situation is even more complicated. It is easy to generalize the numerator of the expression above, but we do not know an expression for the denominator. The point is that in a non weighted graph we can always imagine to be in condition to draw another edge if the graph is not complete. In this case the relative entry in the adjacency matrix will be invariably one. This accounts for the term $(k_i(k_i-1))/2$ giving the total number of edges one can draw. Here instead we also have to assign a weight on the missing edge! Therefore the concept of total weight of triangles is ill-defined.

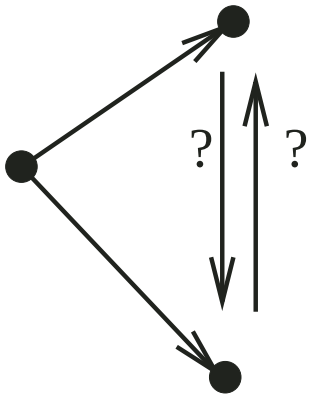

FIGURE 1.6. When dealing with oriented graphs it is not clear, how to close the triangle at the basis of clustering coefficient definition. The same problem holds for weighted networks.

Different choices are available in order to overcome such problem. Here we present this definition

$$C_i^w = \frac{1}{< a^w >^3 (k_i)(k_i-1)/2} \sum_{j,k} a_{ij}^w a_{ik}^w a_{jk}^w \qquad (1.8)$$

where the quantity $< a^w >= \frac{1}{n} \sum_{ij} a_{ij}^w$ is the average weight of an edge in the graph.
Other choices are possible especially in some real situation where fluctuations from average plays a crucial role (i.e. the average is not a representative measure of the set). A. Barrat et al. (Barrat, Barthélemy, Pastor-Satorras and Vespignani, 2004) introduced for example the following definition

$$C_i^w = \frac{1}{(k_i^w)(k_i-1)} \sum_{j,k} \frac{a_{ij}^w + a_{ik}^w}{2} \theta(a_{ij}^w)\theta(a_{ik}^w)\theta(a_{jk}^w) \qquad (1.9)$$

where $\theta(x)$ is the step function equal to 1 when the argument is larger than 0.

In general, according to the particular case one or another definition can have more sense.

One important quantity is given by *cliques*. A **bipartite clique** consists of two sets of vertices such that every vertex of one set is connected with every vertex of the other set. The same concept can be generalized for **tripartite cliques** when we have three sets. More generally cliques can be composed by a number n of sets. This quantity is usually very difficult to measure and visualize. Empirically the number of bipartite cliques $bc(m,n)$ for two sets of order m,n decays very fast (exponentially) for large m,n in almost any network of interest.

Simpler methods to describe quantitatively such systems have been introduced to compute the presence of communities.

This very schematic list closes the first part of description of graph quantities. Other definitions of centrality are possible and are used in graph theory, but for the moment we stop here.

It is important to understand that real networks are different each other. They happen to share some statistical properties with the others, but in general they display also intrinsic characteristics for which specific quantities are necessary. For example, let us focus on a very important case of study, that of the World Wide Web. Here the graph is oriented, the vertices are the html documents and edges can point in or out a certain page. Given the meaning of the graph, it is very likely (and therefore fair to assume) that vertices (i.e. web-pages) with a large number of outgoing edges are pages specifically suited (i.e. Yahoo) to reach as many other pages are possible.

On the other hand, if one page has a large number of ingoing edges, it is because its content is probably very important (i.e. an on-line newspaper). The larger the number of edges the more recognized is this importance.

This calls for a new division of vertices, valid only for oriented graphs and justified only by the specific character of the case of study (Kleinberg, 1998; Kleinberg and Lawrence, 2001)

- **hubs** are those web pages that point to a large number of authorities (i.e. they have a large number of outgoing edges).
- **authorities** are those web pages pointed by a large number of hubs (i.e. they have a large number of ingoing edges).

From that example we understand that a new set of quantities is needed on top of the general ones. Sometime they can be used for more than one system, sometime they are indeed rather particular as in the above case.

1.2 Centrality measures, Betweenness and Robustness

As we see in the following, all scale-free networks are characterized by a Degree Density Distribution whose tail goes to zero very slowly.

This means that for large values of the degree the distribution function is power law shaped, that is

$$P(k) \propto k^{-\gamma}. \tag{1.10}$$

For the moment we can consider that the above information means that the structure present few vertices (later called "hubs") with many edges and many vertices with few edges. In some sense, the vertices with the largest degree are the "most important" in the graph. This concept is particularly clear in the case of the Internet. Whenever hackers want to interrupt the service, they attack the routers with the largest number of connection. One could then fairly assume that the largest the degree the largest the importance of the vertex in the graph.

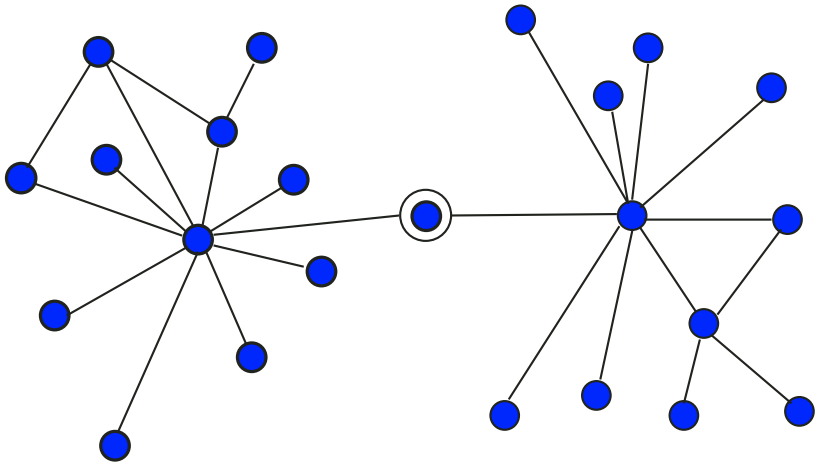

FIGURE 1.7. A particular graph where we can disconnect the system acting on the circled vertex with degree 2.

Interestingly, some special cases show that this not always true. Therefore the concept of importance (or better *centrality*) can be improved.

Sometime the situation is rather different. If the graph is the one shown in Fig. 1.7 the hacker should attack the central vertex even if it has a small degree. The notion of importance then depends upon the actual shape of the graph and upon the particular physical meaning of the services provided on the graph. For that reason different measures of **centrality** of a vertex have been presented. These concepts have been mainly used in the field of social science (Freeman, 1977). We describe them here because they can play a role in some cases of scale-free networks.

The most immediate way to define the most central vertex is to look for the vertex whose average distance from all the others is the minimum one.

That corresponds to say that centrality $c(i)$

$$c(i) = \frac{1}{\sum_{j=1,n} d_{ij}}. \tag{1.11}$$

Many graphs present a characteristic average distance (this phenomenon is called small world effect). This means that also this quantity will show a typical average scale.

Another measure of centrality comes from the dynamical properties of the graph. Whenever the graph is supposed to represent a transportation network (water, electricity, information etc.) then a flux is present on the edges of the graph. It makes then sense to ask how this flux is distributed along the vertices. We start from the simplest choice. We assume that on every edge there is a uniform load. That is to say, any edge has the same capacity of the others. Under this hypothesis, a sensible measure of centrality is given by the number of times that we cross one vertex k in going from one vertex i to another j

following the path of minimal length (distance $d(i,j)$). This quantity is called **site betweenness** $b(i)$ and in formulas is given

$$b(i) = \sum_{\substack{j,l=1,n \\ i \neq j \neq l}} \frac{\mathcal{D}_{jl}(i)}{\mathcal{D}_{jl}} \tag{1.12}$$

where $\mathcal{D}_{jl}$ is the total number of different shortest paths (distances) going from j to l and $\mathcal{D}_{jl}(k)$ is the subset of those distances passing through i. The sum runs over all pairs with $i \neq j \neq l$.

An immediate generalization of this quantity is given by the **edge betweenness** where now the number of paths considered are those passing for a certain edge.

On large graph one has to use some care in order to compute this quantity along the graph. Simple algorithms tend to increase the time of computation very easily, so that in some cases a series of methods to have an approximate value of the betweenness have been produced.[6]

A particular interest for application is to check whether a particular graph can resist to a certain number of deletion of both edges and vertices. Traditionally such deletions are divided into two classes (Albert, Jeong and Barabási, 2000; Albert, Jeong and Barabási, 2001). The *random* failure where a vertex is removed regardless its importance (degree) or centrality in the network and the *attack* where a vertex is removed with a probability related to its importance (usually the degree). The key quantity to monitor to check if the properties of the network are unaffected is the change in the value of the diameter D or of the average distance with respect to the fraction f of removed vertices.

1.3 A Simple Subset of Graphs: the Trees

There is one general case in which the networks have a particular characteristic shape. In the case of a distribution network (as for example water supply, but in principle anything), the good is delivered to all clients trying to avoid to pass on the same vertex twice. The class of graphs without closed paths are called **trees**. A more formal definition is the following:

- For our purposes we can informally define a **cycle** as *a closed path that visit only once the vertices (apart the end-vertices that coincide).*
- *A set of vertices connected each other without cycles is a* **tree**.
- *A set of disconnected trees is a* **forest**.
- For the oriented trees *the vertices with (out-)in-degree equal to one* (the peripheral vertices of the tree) are called **leaves**. Sometime it is needed to define a special vertex that is called **root**. In the case of river networks (as shown in Chapter 8) the root is the vertex (always present) for which the

[6] Actually very recently a fast and efficient algorithm to compute exactly such quantity has been presented in Ref. (Brandes, 2001). We sketch the idea of the algorithm and the structure of a routine in language C in the Appendix.

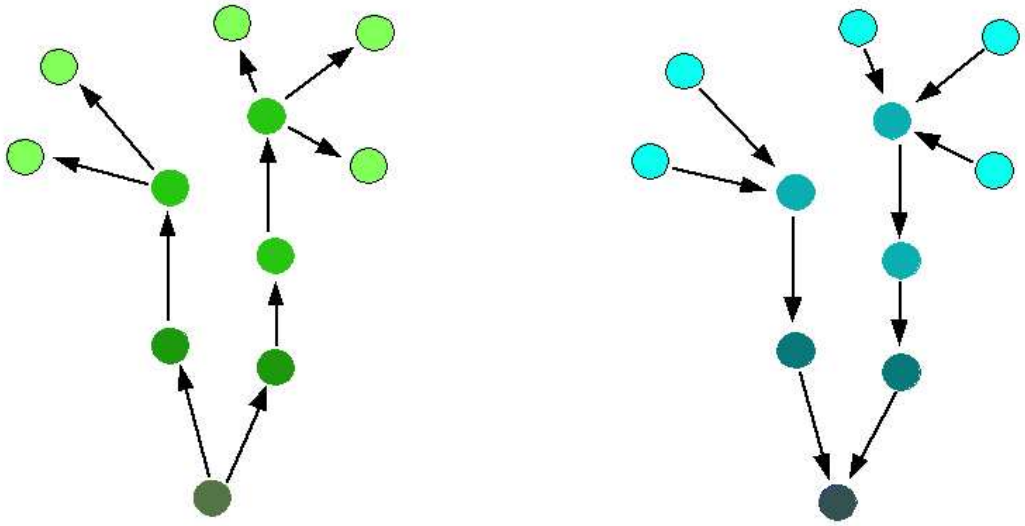

FIGURE 1.8. Two example of special vertices in a tree. On the left (as in a real tree) nutrients flow from the root (dark) to reach the leaves (light). Root and leaves are defined through their in-degree. On the right the case of river networks. Here light vertices represents the highest zone in the basin (no points uphill). The dark vertex is the outlet of the whole basin. Root and leaves are defined through the out-degree.

out-degree is zero. In food-webs instead people prefer to define as root the vertex whose in-degree is zero (see Fig. 1.8).

In a non-oriented tree there is always a path between any couple of vertices. For oriented trees instead, it is possible that some of the vertices are isolated from the others because the direction of the edges does not allow to join them. Therefore distance are generally larger as it happens in the oriented graphs. Some graph quantities as the degree or the betweenness can be computed also in the case of the trees. Instead, by construction the clustering coefficient is zero. We remember here that in the definition given in eqn (1.5) the clustering coefficient is related to the number of triangles (cycles of order 3) present in the graph. Since the tree is an a-cyclic graph, we see immediately that this quantity does not have any sense in the case of trees.

Anyway, for this sub-class of graphs we can define a new quantity given by the structure of the tree. As shown in Fig. 1.9 we have that a tree can be defined as a set of (sometime) nested sub-trees. In the picture proposed we have the sub-basins A, B, C whose order (number of vertices) is respectively $4, 5, 6$ (note that in a tree of n vertices, the total number of edges is $n-1$, so whenever talking about trees the size and order of these graphs differ only by one).

The **Probability Distribution** $P(n)$ of the size n for the various nested subtrees. According to the different systems this quantity takes several names. It is called "in-degree component" for a general oriented tree as well as *drained area* in the language of river networks. In this particular case the vertices cor-

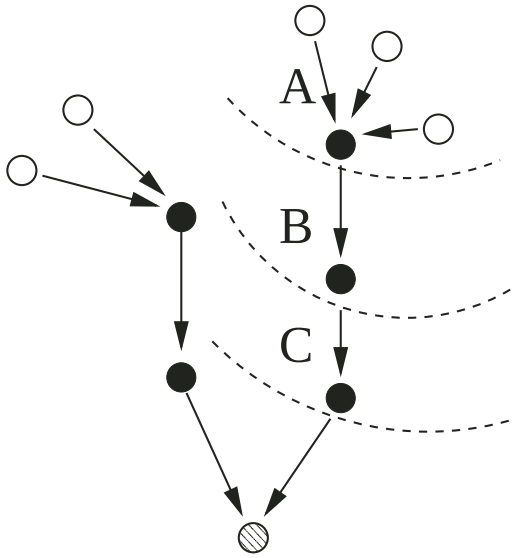

FIGURE 1.9. Three different sub-basins nested: A is composed by 4 vertices, B is composed by 5 vertices, C is composed by 6 vertices.

respond to area of Earth surface where water is collected from rainfall. The edges represent the directions along which the water flows in the system. For the present system n represents the total amount of points "uphill" in the tree.

To compute this quantity one can repeat the natural process of water collection in the basin.

- We start by identifying the root of the tree.
- We assign an oriented direction to the edges such that a path exists from any vertex to the root (in the case of River Networks we see that different height values in the basin areas makes these choices rather natural).
- We start from a leaf (light vertices on the right of Fig. 1.8) that are sub-basin of area $A = 1$.
- We follow the direction of the edges and we compute for the destination vertex how many points are uphill.

The various values of n along the tree give the frequency distribution of n. The latter one tends to the probability function $P(n)$ in the limit of infinite trials (that is to say an infinite tree).

Some of the trees we present in this book are of some importance for their functional properties, that is to say they are shaped in a particular way in order to perform a specific task (as it is the case of transportation networks). Some other trees arise naturally from the process of classification. The most intuitive example of this phenomenon is given by natural taxonomy of plants or animals. Starting from field observation generations of naturalists recursively grouped together in larger and larger groups the real species around us. Note that real species as the laurel (*Laurus nobilis*) or the domestic cat (*Felix Felix*) are the "experimental data". The larger categories in which they are grouped instead

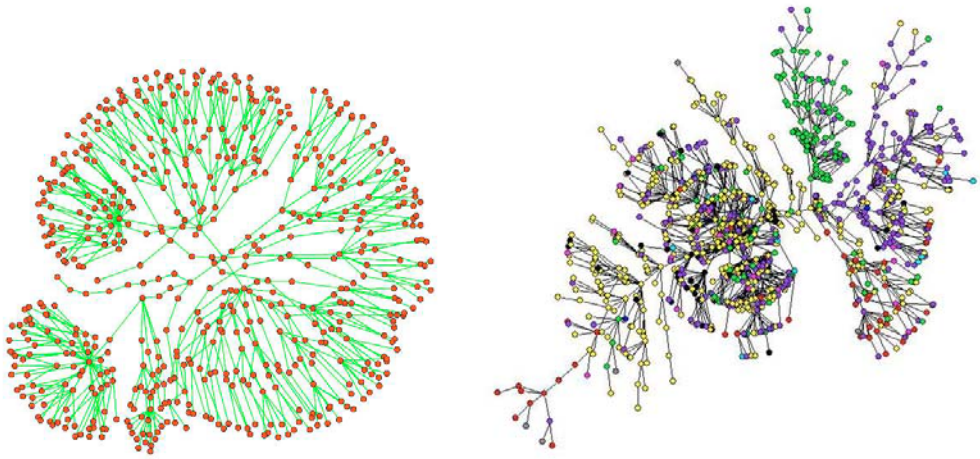

FIGURE 1.10. Left, the taxonomy tree of an Argentinian ecosystem. Right, a minimal spanning tree obtained with the procedure explained in the text. The key quantity in this latter example is the correlation in price returns of the different stocks (the vertices of the graph). This tree will be discussed in great detail in Section 13.2 of Chapter 13

are product of men activity and cannot be recorded in any field observation. These two classes of trees are intrinsically different. In one case we have vertices with a similar nature (as the houses to be connected by a water pipeline) in the other the vertices are very different each other (a vertex corresponding to an actual species and a vertex corresponding to a human classification). This fact produces some interesting feature as shown in Chapter 7. For the moment let us see how these structure can arise from classification procedures.

1.3.1 *Graphs from Classification*

Taxonomic or classification trees appears naturally in a series of different real situation, practically whenever it is crucial to organize the information. The best and more efficient structures of taxonomy are therefore one of the open problems in data mining. In any example of trees the starting point is given by the physical correlations between the agents in the systems. Agents or class of them become the vertices of the tree and the correlations become the edges.

In the case of botany or zoology, this is very easy. We start from species and we cluster them according to their morphology. Classes of species can be clustered in the same way. Step by step we form a tree composed by different layers.

We can follow this procedure in several different ways. For example we can fix from beginning the number of layers we want to use. To reduce a little bit this freedom of choice we present here another procedure called **Minimal Spanning Tree**.

To fix the ideas let us consider as an example a financial system. More particularly, let us consider the price oscillations of a stock traded in the market. Usually, the price of different stocks is correlated in such a way that different stocks might have a similar price history. For our purposes price correlation is a real number measuring the similarity of two vertices. Now one can decide to use this measure of similarity to cluster stocks in the same way as species in a taxonomy. The passages in this procedure are the following

- Define a set of entities that will be represented as vertices in a graph and define an interaction between them that will give the edges (in the specific case we have stocks as vertices and price correlation as interaction. Technically correlation can be negative so in order to have something like an edge strength it is customary to define a distance from correlation).
- Now we have a complete $n \times n$ set of measures of similarity of any of the n vertices with all the other $n - 1$ vertices. This form a complete graph with different edges strength. If requested the graph can be transformed in a non complete one. Two choices are now possible.
  - We can assign a threshold on this weight saying that only the edges with a similarity larger than the threshold are drawn.
  - Otherwise we can classify the different vertices by means of a different procedure. For example with the minimal spanning tree procedure we can obtain a tree in the following way. Rank all the similarities between different vertices. Then draw the first two vertices in the list. Then proceed on the list and draw the second pair. If in the second pair there is one vertex already present in the first one we have the situation in Fig. 1.10. Now if by drawing the third pair we close a loop then forget about this entry in the list and proceed further. Stop when all the vertices have been drawn.

1.4 Network Motifs

One possible reduction of the complexity of graphs can be done by simplifying the graph structure into basic building blocks (Milo, Shen-Orr, Itzkovitz, Kashtan, Chklovskii, and Alon, 2002). We already know some of the structures shown in Fig. 1.11. The graph number 9 is a cycle of order three, but other configuration are possible even if they are not closed paths. In general for three vertices it turns out that thirteen possible basic configuration are available (if the edges are oriented). These configurations are called **motifs** and can help in characterising the shape of the different real networks. In some lucky case the hope is that understanding the function of each motif would allow us to understand the entire network behavior. This is because certain recurring motifs can witness a particular function required to be accomplished. They can be regarded as functional units important to whatever function the network was designed or evolved to perform.

We can in principle study the various motifs arising when considering four, five or in general a larger set of vertices. It is easy to understand that this

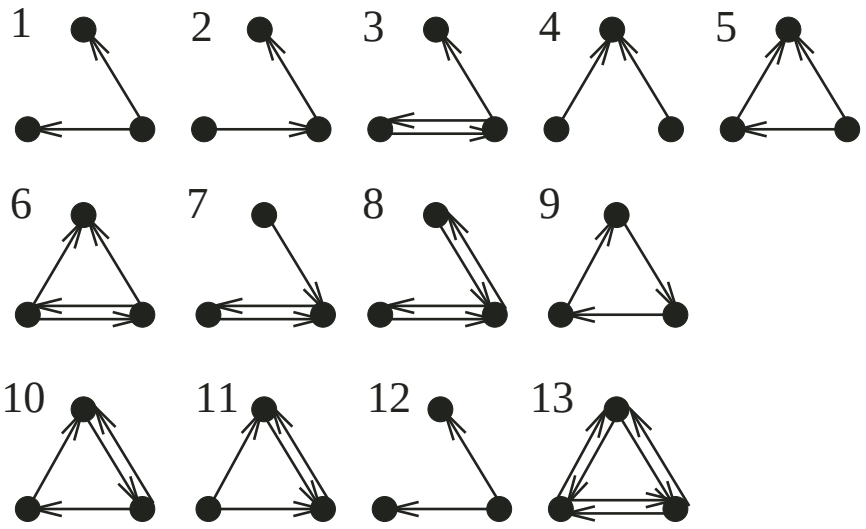

FIGURE 1.11. The basic 13 elementary motifs that can be drawn in an oriented graph of three vertices. The table made for four vertices motifs has 199 entries.

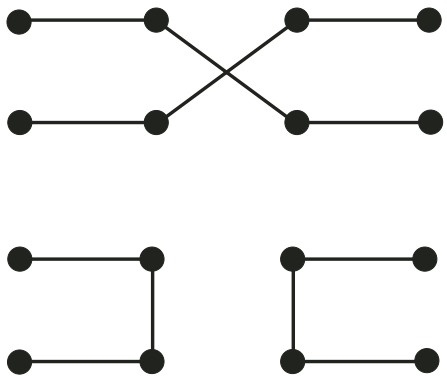

FIGURE 1.12. A possible way to rearrange edges keeping the same size, order and degree sequence.

procedure will produce an exponential number of motifs. For that reason people usually stop very soon and consider only the presence of the smallest motifs in the network. Particularly interesting is the fact that in many real networks some motifs are far more frequent than "expected". Let us see what it means "expected". We start with the graph we are studying (graph origin). After that we build a set of all the possible graphs with the same order (number of vertices), the same size (number of edges) and the same degree sequence (i.e. any vertex of the graphs in the set has the same degree it had in the original version). Edges are instead rearranged in the vertices as shown in Figure 1.12

Interestingly, in the networks analyzed (Milo, Shen-Orr, Itzkovitz, Kashtan, Chklovskii, and Alon, 2002; Mangan and Alon, 2003) some motifs are constantly repeated. In particular, their abundance is more than 10 standard deviations from the mean expected in randomized graphs. The feed-forward loop, or filter

motif, for example, is common in networks of neurons, but is relatively rare in food webs. A feed-forward loop consists of network vertices x, y and z, in which x has connections to y and z, and y also has a connection to z. In food webs, where vertices are animals, this pattern is carried out only by omnivores (x) that bot eat another animal (y) and the food (z) that that animal eats. Instead in (Milo, Shen-Orr, Itzkovitz, Kashtan, Chklovskii, and Alon, 2002) it is shown that at least in seven different food webs we can find the same two motifs. Those motifs are a chain, where one type of prey eats another, which eats another, and a diamond-shaped pattern, where one type of prey eats two others, which both eat a fourth type of prey.

Not surprisingly a relative abundance of motifs is also present in the information process network, where these modular structures can act as logic circuits performing "and" and "or" operations.

1.5 Statistical Distributions

The above graph quantities refer to the properties of a single vertex (or in the case of motifs of a small group of vertices). Since many networks of interest are composed by thousands of vertices, the above information for all the elements would be practically intractable. That is the reason that forces to give a statistical description of the system. Focusing on the degree for example, we can consider to average of the various degrees and study its mean value. Even better, we can consider *the frequency distribution* of the various values of the degrees. This means that we compute how many times (with respect to the total number of vertices N) we find one vertex with degree $1, 2, 3, ..$ etc.

It is enough to remember here, that as the number of measures becomes really large (we need infinite measurements), the values of these frequencies become very similar to their probabilities. Actually this procedure provides one of the possible non rigorous definition of probability (the limit of the frequency value when the number of observation goes to infinite). This series of frequencies (that we can call probabilities from now on, hoping in good quality measurements) is usually the first quantity checked in most of the cases of study.

Before proceeding further there is an important thing to discuss. We should consider the properties of this discrete series of values $P^d(1), P^d(2), P^d(3)$ etc. (giving the probability to find a vertex with degree $1, 2, 3...$ respectively) People instead use (almost universally) continuous function $P(k)$ where k is a real number. This means that $P(k)$ is defined also for non integer values of its argument k. In this book we also follow this approximation because it helps in obtaining some analytical results and is in general "easier" to use.

An example of this approximation is represented in Fig. 1.13 where the histogram of the various discrete probabilities is fitted through a suitable continuous function. It is important to understand that $P^d(k)$ is a **Probability Function** or **Probability distribution**. If we have $P^d(k) = 0.1$, this means that on average 10% of the times we find a degree equal to k.

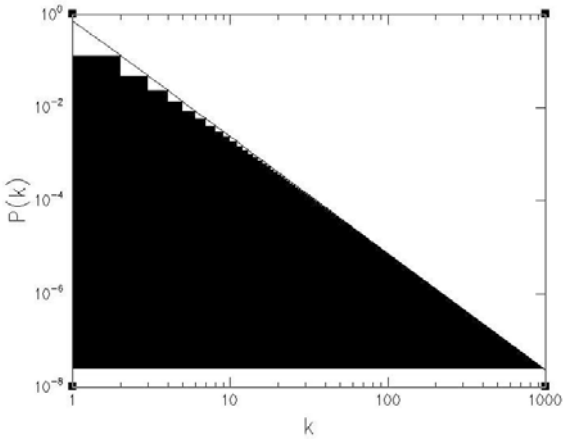

FIGURE 1.13. The histogram of a discrete probability function and its approximation through a continuous function. The relative error is $\frac{dk}{k} = \frac{1}{k}$ that in the limit of large k tends to zero.

Instead the real function $P(k)$ has not the meaning of a probability. This happens because the probability to extract exactly a precise real number (as it is now k) is always zero. The correct approach is to define a quantity that gives the probability that the degree k is within an interval dk around the expected value. This quantity is indicated as $P(k)dk$ and the $P(k)$ is called **Probability Density Function** or **Density of Probability Distribution**.

The name "density" states for the fact that one recovers the meaning of probability only by multiplying it for a suitable interval. Turning to the above example, the $P(k)$ can now assume non-zero values like for example 0.1. In this latter case this means that on average we find 10% of times a degree *whose value is between* k *and* $k + dk$.

Note also that the axes of the plot in Fig. 1.13 are in logarithmic scale. The reason of such choice of scale is related to the topic of this book. In almost all the cases of interest the histogram of the discrete probabilities $P^d(k)$ and also its fitting continuous function $P(k)$ are distributed according to a power law. This corresponds to say

$$P(k) \propto k^{-\gamma}. \tag{1.13}$$

The symbol $\propto$ means "proportional to" and we shall use this relational concept very often in this book. Probability Density and Probability Functions obey the normalization condition. That is to say we have to be sure (probability =1) that we will find at least one value of the degree whatsoever.

In this specific case this means that the frequency distribution is in fact $P(k) = Ak^{-\gamma}$. The proportionality constant A is however not important. It is fixed by the requirement that the sum of all the frequencies is equal to one. This means

$$\int P(k)dk = \int Ak^{-\gamma}dk = 1 \tag{1.14}$$

$$\rightarrow A = \frac{1}{\int k^{-\gamma}dk}. \tag{1.15}$$

Some care must be used with the extremes of integration. While the upper limit can be considered infinite, in the lower limit if $k = 0$ the integrand diverges. Therefore in computing A we have to remember to restrict ourselves to the connected part of the system, that is where $k > 1$. Whenever extremes of integration are not explicitly indicated we assume the above conditions apply.

The knowledge of the distribution function is particularly important. Through this quantity we can compute a couple of other interesting quantities. For example we can compute what is the typical (or better "mean") value that the degree assumes in the graph. This value will be indicated by $\langle k \rangle$, where the symbol $\langle ... \rangle$ indicates an average over all the possible outcomes.

A measure of the typical error we make if we assume that every vertex has degree $\langle k \rangle$ (thereby neglecting values fluctuations in our system) is given by the variance σ^2.

These two quantities are given by definition by

$$\langle k \rangle = \int kP(k)dk \tag{1.16}$$

$$\sigma^2 = \int (k - \langle k \rangle)^2 P(k)dk \tag{1.17}$$

1.6 Working with Statistical Distributions

1.6.1 *Some Important Example*

It is easy to understand that if the integral of the $P(k)$ must be equal to one and k can variate between 1 and infinite at a certain point this function $P(k)$ must decrease fast, otherwise its integral would diverge.

Amongst all the possible function obeying such requirements the real distribution probabilities are remarkably few. Here we will further restrict this choice to the ones that are more frequent.

- A **discrete** Probability distribution is the Poissonian distribution $P_P(k)$ (where k is integer) behaving as

$$P_P(k) = \frac{\mu^k e^{-k}}{k!} \tag{1.18}$$

- Whenever the argument of k the function can assume also real values we have a series of different possible distribution function. For our purpose it is enough to mention:
  the Gaussian (or normal) density function $P_G(k)$

$$f(k) = P_G(k) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(k-\mu)^2}{2\sigma^2}} \tag{1.19}$$

  The log-normal distribution

$$f(k) = P_{LN}(k) = \frac{1}{\sqrt{2\pi}\sigma k} e^{-\frac{(ln(k)-\mu)^2}{2\sigma^2}} \tag{1.20}$$

  the power law distribution $P_{pl}(k)$

$$f(k) = P_{pl}(k) = Ak^{-\gamma}. \tag{1.21}$$

The Gaussian distribution has mean μ and variance σ^2, the log-normal instead is more skewed with mean $e^{\mu+\sigma^2/2}$ and variance $e^{2\mu+\sigma^2}(e^{\sigma^2} - 1)$. power law distribution is the only one that may have no finite mean and variance.

We see in Fig. 1.14 a snapshot of the normal, the Poissonian and the power-law distributions. Note that the first two functions increase to a maximum after which they decade. The third has a smoother character always decreasing as k grows. Note also that while normal and Poissonian distribution depend upon the choice of some parameter, their qualitative behaviour does not change. Note that the area dashed in the three plot of Fig. 1.14 has a value of 1. That is another way to say that the three distributions are normalized.

Interestingly (we return on this point), while it is easy to distinguish between a Gaussian and a power law distribution, it could be difficult to spot the difference between a power law function and a log-normal function. Most of the times even eye inspection can help in determining which function is best suitable for the data. A power law distribution will look as a straight line whenever plotted on a logarithmic scale (or that is to say whenever considering the logarithms of both sides). Unfortunately, in some conditions this can be true also for the log-normal distribution.

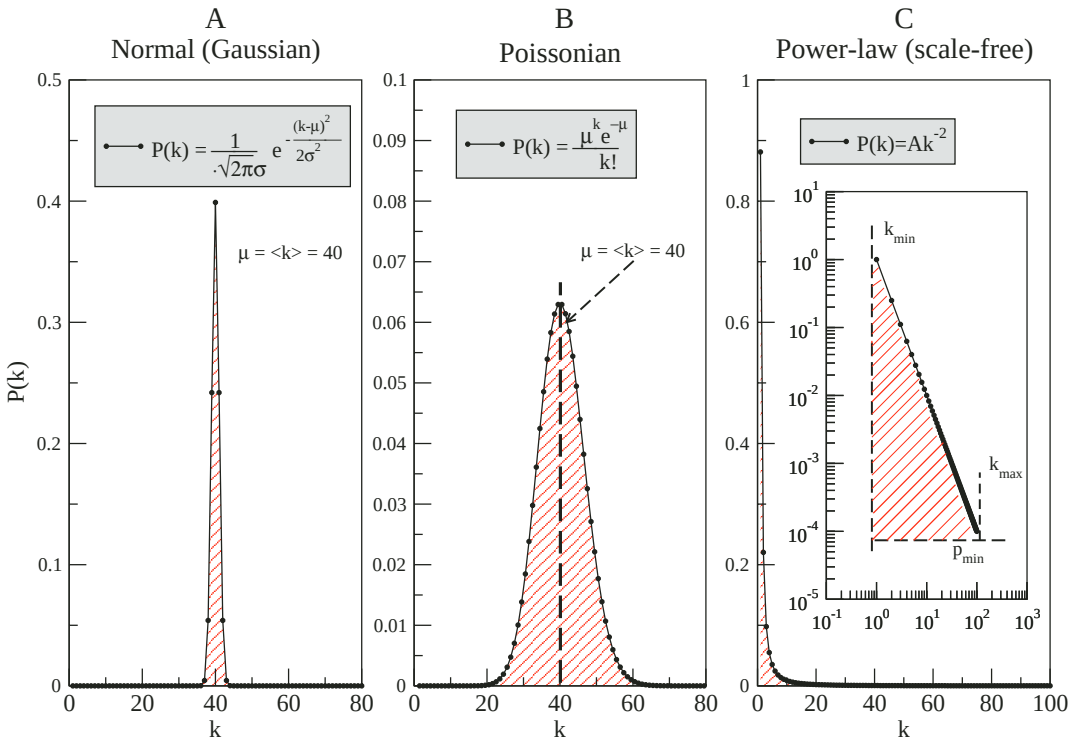

FIGURE 1.14. A The plot of a Gaussian Distribution. B The plot of a Poissonian Distribution. C The plot of a power law distribution. In the inset of C the same plot on a logarithmic scale. Note that the power law function is not defined in $k = 0$

Taking the logarithm of eqn (1.21)

$$log(f(k)) = log(A) - \gamma log(k) \tag{1.22}$$

plotting $log(f(k))$ vs $log(k)$ we have a linear relationship with slope $-\gamma$. Now making the same for eqn (1.20) we have

$$\begin{aligned} log(f(k)) &= -log(\sqrt{2\pi}\sigma) - log(k) - \frac{(log(k) - \mu)^2}{2\sigma^2} \\ &= -(\frac{\mu^2}{2\sigma^2} + log(\sqrt{2\pi}\sigma)) + (\frac{\mu}{\sigma^2} - 1)log(k) - \frac{log(k)^2}{2\sigma^2} \end{aligned} \tag{1.23}$$

where the term $-(\frac{\mu^2}{2\sigma^2} + log(2\pi\sigma))$ has the same role of the $log(A)$ in the above equation representing the constant term.

Whenever the log-normal distribution is characterized by a value $\sigma^2 >> (ln(k) - \mu)^2$ then the behaviour becomes very similar to that of a power law function.

The *apparent* slope in this case it is given by

$$\gamma_{app} = (\frac{\mu}{\sigma^2} - 1). \tag{1.24}$$

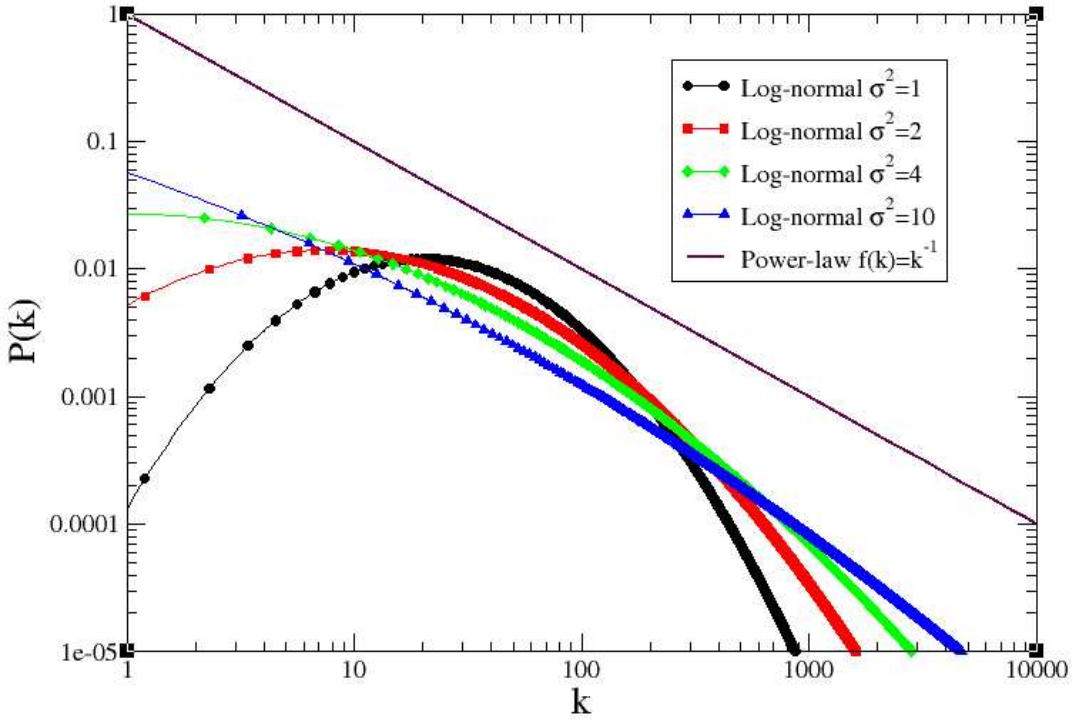

FIGURE 1.15. A comparison between log-normal distributions with different variances and a power law whose exponent is minus 1 (in this example the average μ is equal to 4, so that the very first part should decay as $k^{-0.6}$. Note that the range of k is typical of many datasets available. Therefore the true tail of a log-normal distribution could not be noticeable in some cases of study. Note also that the bending is that slow that even the intermediate region reminds the behaviour of a power law of larger slope (about -1).

Of course as we go along the tail, the logarithm of k grows and the quadratic term becomes less and less negligible. Nevertheless, since the logarithm is a slowly growing function, one may need many and many decades in order to clearly distinguish between the two functions.

This is an important point, and it is worth to discuss that for a moment. We are particularly interested in the power law because of its **scale-invariant** or **scale-free** properties. This means that if we change the scale of observation (that is to say we consider a change of scale of the kind $k' = ak$) the only distribution that maintains unaltered its analytic form is the power law. As explained in the next chapter this property can be the signature of a particular evolution of the systems as well as the signature of some correlations between the parts.

Strikingly, all the networks around us seem to have a scale-free distribution of the degree. The topic of this book is to present the experimental evidence of this fact and try to provide an explanation for such feature.

1.6.2 *Finite sizes*

Recalling the discrete nature of these function and their physical meaning we would like to stress here that even for a scale-free function (see inset of Fig. 1.14 C) a typical scale is always present. This is given by the finite size of our samples. In the specific case of networks we must have a minimum degree that is one (you cannot have less than one edge) and a maximum degree that is the size of the network (you cannot have more than $n-1$ edges). Hoping that (as in the case of the WWW) the graph is very large we can shift the upper cutoff as far as we want on the x axis. Still, we must expect a deviation from power law behaviour whenever k is similar to the size of the system.

In the case of a power law form of the $P(k)$ we obtain for the normalization condition

$$\langle k \rangle = A \int k k^{-\gamma} dk = \frac{\int k^{1-\gamma} dk}{\int k^{-\gamma} dk} \tag{1.25}$$

Since for the real world we never find one degree exactly infinite we must stop integration at a large value k_{max} (k_{max} of the order of N). The above equation becomes then

$$\langle k \rangle = \frac{\int_1^{k_{max}} k^{1-\gamma} dk}{\int_1^{k_{max}} k^{-\gamma} dk} \tag{1.26}$$

That is finite (i.e. does not grow as k_{max} grows) for $\gamma > 2$.

In order for the $P(k)$ to be a probability function it must be integrable, or in other words the A must be finite. This gives a first condition $\gamma > 1$. We see now that there are two possibilities, let us consider the result of eqn (1.26), that is

$$\langle k \rangle = \frac{\int_1^{k_{max}} k^{1-\gamma} dk}{\int_1^{k_{max}} k^{-\gamma} dk} = \frac{(1-\gamma)}{(2-\gamma)} \frac{k_{max}^{2-\gamma} - 1}{k_{max}^{1-\gamma} - 1} \tag{1.27}$$

- $\gamma < 2$ In this case the above expression eqn (1.26) grows as $k_{max}^{2-\gamma} - 1$.
- $\gamma > 2$ In this case instead the above expression as k_{max} increases, tends to the constant value $\frac{1-\gamma}{2-\gamma}$. The average $\langle k \rangle$ stays finite no matter how large is the cutoff.

An analogous behaviour holds also for the variance. This quantity remains finite whenever the exponent γ is larger than 3. It instead diverges as the cut-off k_{max} increases.

This can be seem with computation similar to those made for the variance, in this case we find

$$\sigma^2 = \int (k - \langle k \rangle)^2 P(k) dk = A \int k^{2-\gamma} dk - \langle k \rangle^2 \tag{1.28}$$

The above equation has therefore solution

$$\sigma^2 = \frac{(1-\gamma)}{(3-\gamma)} \left(\frac{k_{max}^{3-\gamma} - 1}{k_{max}^{\gamma-1} - 1} \right) - \langle k \rangle^2 \tag{1.29}$$

Again we have that in the case of $\gamma < 3$ the variance grows as

$$\sigma^2 = \frac{(1-\gamma)}{(3-\gamma)} \left(k_{max}^{3-\gamma} - 1 \right) - \langle k \rangle^2. \tag{1.30}$$

Instead for $\gamma > 3$ the variance tends to a constant value regardless the value of the cutoff k_{max}. In all these cases it makes little sense to describe the system by means of average values.

We see in the following that almost every scale-free network is characterised by a power law degree distribution with a value of γ between 2 and 3. This means that even if it would be possible to define a finite average the standard error for this value is of the order of magnitude of the size of the system.

1.7 Statistical properties of Weighted networks

As we see in the second part of the book, in most real situations the networks that one finds are weighted (Barrat, Barthélemy, Pastor-Satorras and Vespignani, 2004; Yook, Jeong, Barabási and Tu, 2001). While the generalization of the concept is straightforward, we have already seen that sometime it is not very easy to generalize the definition of the quantities, as in the case of the clustering coefficient. Whenever these quantities are defined, anyway, we are interested in their statistical distributions. Also in this case the weighted degree is in most of the case power law distributed. This is what we have already found in the "simple" network analysis, interestingly, in the case of weighted networks new scale-free relations arise.

This means that if we consider the degree k_i^w for a vertex i, where

$$k_i^w = \sum_{j=1}^{n} a_{ij}^w \tag{1.31}$$

the $P(k^w)$ is power law distributed. Incidentally note that since the value of k^w is a real number, we have a continuous probability function $P(k^w)$.

As pointed out in Ref. (Barrat, Barthélemy, Pastor-Satorras and Vespignani, 2004) also for the strength density function, we find a behaviour similar to that

of the degree. This is not surprising since the same authors measure a connection between the two quantities given by the scaling relation

$$s(k) \propto k^{\beta} \tag{1.32}$$

where now $s(k)$ is the average strength for vertices whose degree is k. In the case of absence of correlation between the two quantities one is allowed to take the averages on both sides of eqn (1.31) finding

$$s(k) = \frac{1}{2m} \sum_{i,j=1}^{N} a_{ij}^{w} k = \langle w \rangle k \tag{1.33}$$

where m is the total number of edges in the network and $\langle w \rangle$ is its average weight. In this case the value of β is equal to one. Ref. (Barrat, Barthélemy, Pastor-Satorras and Vespignani, 2004) find for their cases of study values of β between 1 and 1.5.

2

GRAPH STRUCTURES

2.1 Introduction

"And - and - what comes next?" "Oh, yes, yes, what the dickens does come next? C'est la question, ma très chère demoiselle!" As in the first lines of Thomas Mann's Buddenbrooks is there something more to say about graphs? We have seen in the previous chapter the basic quantities describing a network. Those quantities are in general directly related to the state of a single vertex (or few of them like in the case of motifs). In this chapter we instead present quantities that depend upon the shape of a large set of vertices[7]. When considering a set of vertices, they might have similar "properties" (i.e. the same neighbours), so that we can cluster them in the same "class". Sometime a whole series of vertices might be connected each other with a number of edges larger than in the rest of the graph. In the first case we have what is probably a "community" in the sense that different vertices have the same "preference" in their connections. In the second case we have a clustering of vertices. These two concepts are very closely related. A subgraph particularly clustered can witness almost in all cases the presence of a community. Communities in principle do not need to be particularly clustered, but in practice result in a clustered subgraph.

These quantities are important not only to characterise topologically the graph. Actually they can give some information both on the formation of the network and on its functionality. This is because (as is the case of Web) in most of the cases we know that the whole network is built by merging different subgraphs. In other cases the communities select only particular edges among all the possible ones. This helps in determining the traffic or the robustness of the network.

After that we also present the state of the art in the methods to compute communities and clustering. They can be divided approximately in two ways: edge detection and the spectral analysis.

Finally, we present some experimental evidence on the scale-free properties of a series of networks.

[7] Technically, also a statistical distributions of *local* properties (as the degree) is a way to characterize *globally* the shape of a graph. But we want to describe here another way to quantify these global properties.

In this chapter we study some properties of the structure of graphs. We characterize quantitatively the presence of communities and the clustering present in the structures. In this chapter we also use the matrix representation of graphs in order to compute the community structure. Finally we present some experimental evidence of the statistical properties of scale-free networks.

2.2 Correlation in Graphs, Assortative Mixing

Most of the global properties of clustering and/or community presence can be recovered from the analysis of the correlation inside the graph. Correlation here means that vertices with similar properties tend to be connected each other in the network. As usual, since the most immediate properties of a vertex is the degree we look for the presence of a correlation between vertices with similar degree. In general there is no reason to expect a particular correlation. Actually, in some situations there is a tendency for high-degree vertices to be connected to other high-degree vertices. In this case the network display what is called an **assortative mixing**. The opposite situation when high-degree vertices attach to low degree ones is referred as **disassortative mixing** (Newman, 2002).

The correct mathematical way to quantify such a measure is the *conditioned probability* $p(k_1|k_2)$ to have a vertex with degree k_1 at one side of the edge when at the other site of the edge the degree is k_2.

We have two constraints on the conditioned probability. The first one is given by normalization condition

$$\sum_{k_1} p(k_1|k_2) = 1. \tag{2.1}$$

For non oriented graphs the same quantity obeys the detailed balance distribution (Boguñá and Pastor-Satorras, 2002)

$$k_2 p(k_1|k_2)P(k_2) = k_1 p(k_2|k_1)P(k_1) \tag{2.2}$$

This balance equation simply states that the number of edges going from vertex k_1 to vertex k_2 must be equal to the number of edges going from vertex k_2 to vertex k_1.

A way to compute this quantity presented in (Pastor-Satorras and Vespignani, 2004). The authors introduce a symmetric matrix E defined in such a way that the elements $E_{k_1 k_2}$ give the number of edges between k_1 and k_2. It is easy to check that this matrix satisfies

$$\sum_{k_2} E_{k_1 k_2} = k_1 N_{k_1}$$

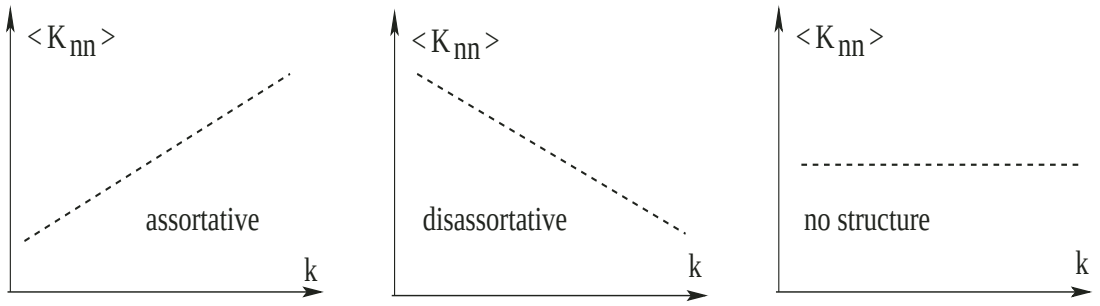

FIGURE 2.1. The three possible behaviour of the average degree of the neighbours $\langle K_{nn}\rangle$ versus the degree k of the vertex origin.

$$\sum_{k_1}\sum_{k_2} E_{k_1 k_2} = \langle k\rangle N \tag{2.3}$$

where $N_{k_i} = NP(k_i)$ is the number of vertices of degree k_i and it is given by the probability $P(k_i)$ that a vertex had degree k_i times the number of trials N. The above identities allow to define the joint probability

$$p(k_1, k_2) = \frac{E_{k_1,k_2}}{\langle k\rangle N}. \tag{2.4}$$

Note that the joint probability is a concept different from the conditioned probability. In the latter one we assume that one degree is given and check for the other. Using the joint probability we can write the probability that an edge randomly extracted connects two vertices of degree k_1 and k_2. This is given by $(2-\delta_{k_1,k_2})p(k_1,k_2)$. The above formula is essentially the joint probability with a weight that is 2 when $k_1 \neq k_2$ (because we are not interested in the order of the degree and we can exchange k_1 with k_2). The weight becomes 1 when $k_1 = k_2$ (this is the only case when δ_{k_1,k_2} is different from 0 and equal to 1.) because this indetermination is resolved.

Finally we have that

$$p(k_1|k_2) = \frac{E_{k_1,k_2}}{k_2 N_{k_2}} = \frac{\langle k\rangle p(k_1,k_2)}{k_2 P(k_2)}. \tag{2.5}$$

The above expression can be simplified in the case of no correlation between vertices. If this is the case, the term $p(k_1,k_2)$ can be written as the product $P(k_1)P(k_2)$. Therefore eqn (2.5) becomes

$$p(k_1|k_2) = \frac{1}{\langle k\rangle} k_1 P(k_1). \tag{2.6}$$

This expression has an immediate meaning: the probability that any given edge points to a vertex of degree k_1 is proportional to the density of these vertices times the number of edges.

A less rigorous but more intuitive and simple approach (because the $p(k_1|k_2)$ can be very difficult to measure) has been introduced by ref. (Pastor-Satorras, Vazquez and Vespignani, 2001). We start from a vertex whose degree is k and then compute the average degree $\langle K_{nn}\rangle$ of its neighbours. This quantity $\langle K_{nn}\rangle$ is in general a function of the degree of the vertex origin and determines the assortativity of the graph.

When plotting $\langle K_{nn}\rangle$ versus k as shown in Fig. 2.1 we can have different behaviour. If the average degree $\langle K_{nn}\rangle$ grows for large k values then big hubs are connected each other. In this case the network has assortative mixing. On the other hand, if $\langle K_{nn}\rangle$ decreases for large values of k this means that most of the edges of the large hubs are with more or less isolated nodes. In this case the network present the disassortative mixing.

2.2.1 *Assortative coefficient*

Another measure of assortativity can be obtained through the assortative coefficient r. This number is a particular case of the Pearson correlation coefficient (see 2.3.1.1). In other words is another measure of the correlation between the degrees.

Following (Callaway, Hopcroft, Kleinberg, Newman and Strogatz, 2001; Newman, 2002) we define

$$r = \frac{1}{\sigma^2} \sum_{k_1,k_2} (p(k_1|k_2) - q_{k_1} q_{k_2}) \tag{2.7}$$

where

- q_k is the normalized distribution for the "remaining degree" of vertices. Remaining degree is the degree of a vertex without the edge considered in the link. In formulas, this means that the remaining degree of vertex i is given by $k_i - 1$. The normalized distribution for such a quantity is then given by:
  $q_k = (k+1)P(k+1)/\sum_{i=1,N} iP(i)$.
- The σ^2 is the variance of the above quantity:
  $\sigma^2 = \sum_{i=1,N} k^2 q_k - (\sum_{i=1,N} k q_k)^2$

Positive values of r signal assortative mixing. Disassortativity corresponds to negative values of r. In table 2.1 we report some data analysis produced by M.E.J. Newman (Newman, 2002). From this data we note that technological and biological networks show disassortative behaviour while social networks are assortative. The reasons for such occurrence are not completely understood.

The same qualitative behaviour can be determined by considering the clustering coefficient of the neighbors of a vertex whose site is k. Also in this case the correlation between different vertices has the immediate meaning given by the assortativity.

Network	n	r
Physics Co-authorship	52909	0.363
Biology Co-authorship	1520251	0.127
Mathematics Co-authorship	253339	0.120
Film Actors Collaboration	253339	0.208
Company Directors	7673	0.276
Internet	10697	−0.189
Protein Interactions	2115	−0.156
Marine food web	134	−0.247
Little Rock Lake	92	−0.276

Table 2.1 *Order and assortative coefficient for various networks*

2.2.2 *Correlations in Weighted Graphs*

As reported in the previous chapter a weighted graph is given by assigning an individual weight to the various edges. Using the *Weighted Adjacency Matrix* $\mathbf{A^w}$ we have that the strength of edge between vertices i and j is given by the entry a^w_{ij}. As usual, if the graph is not oriented the matrix is symmetric. Entry a^w_{ij} equal to 0 means no edge between i and j.

In most of the cases the strength given by the edge weight a^w_{ij} and the degree of the end vertices k_i, k_j are correlated. A large weight (denoting for example a large amount of traffic on a Internet cable) is related to the degree of the vertices (computers) connected. To measure such correlation we study the average $< a^w_{ij} >$[8] versus the product $k_i k_j$. Also in this case we find another power law (see (Barrat, Barthélemy, Pastor-Satorras and Vespignani, 2004))

$$< a^w_{ij} > \propto (k_i k_j)^\theta. \tag{2.8}$$

If correlation exists the value of theta must be different from 0. This is the case of ref. (Barrat, Barthélemy, Pastor-Satorras and Vespignani, 2004) find a value of $\theta = 0.5 \pm 0.1$.

Interestingly the value of the exponent θ is related to the exponent β previously introduced. Indeed

$$s(k) \propto k^\beta \simeq < w > k \simeq k(kk_j)\theta \tag{2.9}$$

and therefore $\beta = 1 + \theta$.

2.3 Communities in Graphs

As already discussed sometime the clustering in a graph can be put in relation with the presence of *communities*. This is an important point, since in various real networks the presence of communities give additional information on the

[8] In this quantity the average is computed on all the vertices i, j, such that their degree is k_i, k_j.

properties of the system. In the case of the WWW, a community individuates people sharing similar interests and therefore linking each other. The knowledge of community structure helps, for example, in exploring efficiently the net.

Two different approaches are possible. The first one selects suitable subgraphs on the basis of special vertices or special edges present in the graph. For example, if we spot few vertices with many edges in common, that is a natural signature of a community. We call this method **Edge analysis**. The second choice (used mainly by graph theorists and mathematicians) is based on a method called **Spectral analysis**. This method select communities from the properties of the eigenvectors of the adjacency matrix $A(n, n)$ of the graph. In both cases we can divide further the approaches using a beautiful nomenclature introduced by M.E.J. Newman (Newman, 2004*b*). In one case we can build communities by recursive grouping of vertices (bottom up process) and this process is called *agglomerative method*. Otherwise we can operate a top down process with a recursive removal of vertices and edges. This latter process is instead a *divisive method*.

2.3.1 *Edge Analysis*

2.3.1.1 *Agglomerative methods*

Once the network is given, the *similarity* between vertices i and j can be measured according to several formulas. In almost all the case the similarity concepts can be defined by means of a suitable "distance" or vice-versa. Essentially if two vertices are similar they must be nearby.

There is no immediate choice between these various methods. Rather the study of the actual real system represented by the graph is the only way to determine which of these quantities is more sensible than the others.

- **Structural equivalence** Two vertices have a structural equivalence if they have the same set of neighbours. This strong requirement can be relaxed defining a similarity measure between the two sets of neighbours. Obviously, the smaller this similarity the greater the equivalence. Different real situations call for different measures of such a quantity. We see in Chapter 6 that a similar concept can be used for protein classification (Brun, Chevenet, Martin, Wojcik, Guénoche and Jacq, 2003; Vazquez, Flammini, Maritan and Vespignani, 2003). The simplest definition of similarity is more or less equivalent to the *Hamming distance*[9] obtained by counting the number n_{ij} of similar neighbours between vertices i and j.
  In order to have a distance properly defined (i.e. non-negative) we can consider the following quantity as presented in Refs. (Burt, 1976; Wasserman and Faust, 1994)
  $$x_{ij}^{H} = \sqrt{\sum_{k \neq i,j} (a_{ik} - a_{jk})^2}. \tag{2.10}$$

[9]named after the USA mathematician Richard Wesley Hamming (1915 - 1998) known for his work in information theory

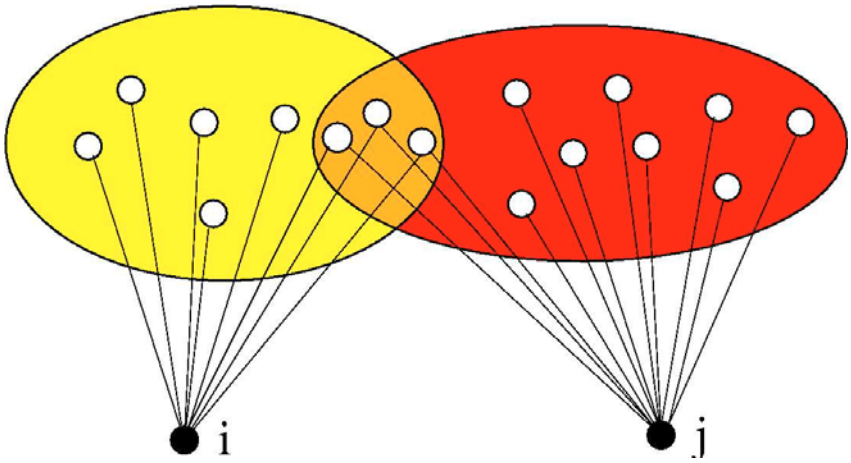

FIGURE 2.2. The Union of the two sets is composed by 16 vertices. The intersection by 3. The numerator of x_{ij}^S is given by the vertices belonging only to one or the other. In this case this number is 13. The distance x_{ij}^S between the two vertices i, j is then $\frac{16-3}{16+3} = \frac{13}{19} \simeq 0.68$.

A similar approach to determine the similarity of two vertices is used in (Brun, Chevenet, Martin, Wojcik, Guénoche and Jacq, 2003). They use the set S_i of vertices neighbours to i and the set S_j of vertices neighbours to j. Denoting by $N(S_{i,j})$ a function that gives the number of elements (vertices) in sets $S_{i,j}$, we have that the quantity

$$x_{ij}^S = \frac{N(S_i \cup S_j) - N(S_i \cap S_j)}{N(S_i \cup S_j) + N(S_i \cap S_j)} \tag{2.11}$$

has the desired property to be 0 when S_i and S_j coincide, growing instead when the sets differ. An example of the application of this quantity is reported in Fig. 2.2.

The above requirement can be put in the more complicated language of the adjacency matrix. The x_{ij}^S is then given by

$$x_{ij}^S = \frac{\sum_k a_{ik} + \sum_k a_{jk} - 2\sum_k a_{ik}a_{jk}}{\sum_k a_{ik} + \sum_k a_{jk}} = \frac{\sum_k a_{ik} + a_{jk} - 2a_{ik}a_{jk}}{\sum_k a_{ik} + a_{jk}} \tag{2.12}$$

- **Correlation coefficient**. In this case the distance between vertices is computed by considering the mean and the variance of the values along a row (or column) of the adjacency matrix. We use these quantities in order to compute the correlation C_{ij}^P, between vertices i, j. Correlation is different from distance. Now if the two vertices are the same (i.e. $i = j$) the value C_{ij}^P must be 1 while their distance is 0. If i, j are very different they must have a large distance while their correlation is around 0.

In formulas if mean and variance are given by

$$\mu_i = \frac{1}{n}\sum_j a_{ij} \quad \sigma_i^2 = \frac{1}{n}\sum_j (a_{ij} - \mu_i)^2 \tag{2.13}$$

the correlation coefficient[10] is defined as

$$C_{ij}^P = \frac{\frac{1}{n}\sum_k (a_{ik} - \mu_i)(a_{jk} - \mu_j)}{\sigma_i \sigma_j} \tag{2.14}$$

that is equal to 1 when $i = j$ as expected.
Note that we always used the entries of the unweighted adjacency matrix $A(n, n)$ so that the quantity μ_i corresponds to the normalized degree k_i/n of the vertex i. When the matrix is oriented μ_i gives in-/out- degree according if the sum is over the rows or the columns.

2.3.1.2 *Divisive methods: The Algorithm of Girvan and Newman* Instead of grouping together vertices, we can decide to remove from the initial graph some vertices (can be done with edges as well). If this procedure splits the graph in definite subsets, these subsets are the communities we are looking for. M. Girvan and M.E.J. Newman (Girvan and Newman, 2002) produced the following recipe. Select the edge of largest betweenness in the graph and then remove it. Edges have large betweenness if they connect parts of the graph that would be separate otherwise. This seems a rather good heuristic method in order to detect separate subsets. In other words this method selects and removes the bottleneck edges separating different communities.

Note that this method work reasonably well for sparse graphs (that is a frequent case). In this case just at the first edge removed it is clear how to cluster the vertices. In case of dense graphs, this procedure can be a little tricky. This powerful idea is the prototype of a class of edge-removing methods.

- The first and original approach (Girvan and Newman, 2002) is based on **edge-betweenness**. This quantity is computed (see Appendix E) on all the edges of the graph. Recursively the edge with the largest betweenness is removed until no edge remains. Obviously at a certain point the graph may become disconnected. This is a crucial point. We write all the vertices in one subgraph on one side. On the other side we put the other vertices. This is the first split in the dendrogram of Fig. 2.3. The determination of the various communities proceeds until another split takes place (the same procedure continues in the separate parts). If more than one edge share the same value of betweenness one can choose either to remove all of them or to select the first randomly. In this case the structure of communities

[10]This coefficient is often named after Karl Pearson. English mathematician born in London in 1857 and died in Coldharbour, Surrey, England in 1936. Founder of the journal *Biometrika*, he mainly worked in the statistical studies of populations

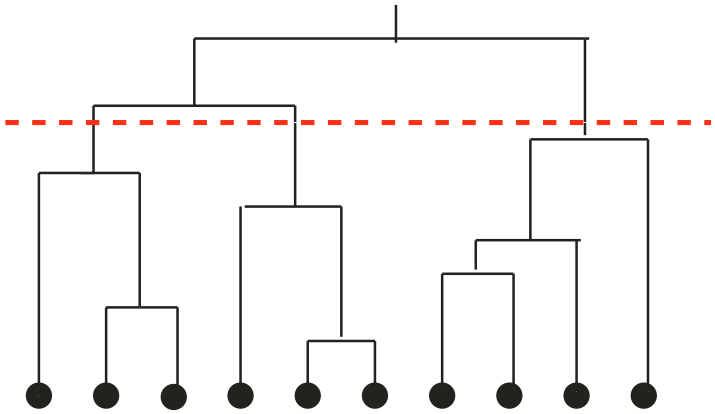

FIGURE 2.3. An example of a dendrogram computed from vertices properties. By cutting it at a certain level we obtain the communities present at that level.

can be represented by a dendrogram built in a top-down process. Starting from the root of the tree, at the first division one places on one side all the vertices in one of the community on the other all the others. This recursive splitting ends at the leaves of the tree that represent the vertices of the graph.

As discusses below, the main inconvenient of this method is that it eventually splits the graph in all the vertices, regardless the real number of communities present. To partly overcome this problem M.E.J Newman and M. Girvan (Newman and Girvan, 2003) introduced a quantity called *modularity* in order to measure how good a particular division is. If we divide the graph in g groups, we can define a matrix $\mathbf{E}$ whose entries e_{ij} give the fraction of edges that in the original graph connects group i to group j. The modularity for such division is then

$$Q = \sum_i e_{ii} - \sum_{ijk} e_{ij} e_{ki} \tag{2.15}$$

- **Approximated edge-betweenness**. Instead of computing the betweenness by considering all the possible paths through an edge, one can approximate this quantity by computing only the paths from a random selection of vertices (Tyler, Wilkinson and Huberman, 2003). With such approach we have a substantial increase of speed a stochastic measure for the communities. The number of times we find a precise vertex in the same community gives a measure of the reliability of such assignment.
- **Loops Counting Methods**. This time the edges are selected according to the number of small cycles they belong to. The idea is that edges joining communities are bottlenecks and it is unlikely they belong to a cycle. If this were the case another path would be present between the same communities (Radicchi, Castellano, Cecconi, Loreto and Parisi, 2004). Therefore by measuring the number of small cycles (in this case triangles or cycles

of order 3) one edge belongs to, we have a measure of centrality. We can expect that the lower is this number, the greater the importance of the edge in joining different communities. The key quantity is now the *edge clustering coefficient*

$$C_{ij} = \frac{z_{ij} + 1}{min(k_i - 1, k_j - 1)} \tag{2.16}$$

where i, j are two vertices of degree k_i, k_j respectively, and z_{ij} are the number of triangles an edge belongs to. The algorithm iteratively deletes the edges with lowest values of C_i and recompute this quantity for the edges left.

All the structures formed with these procedures result in a tree of classifications or *dendrogram* (from the Greek "tree-like image") as the one shown in Fig. 2.3. Technically this procedure is rather different from all taxonomy processes (described below). Whenever forming a taxonomy like for example in botany or zoology, we put together different things according to some measure of correlation they have. This is done without any reference to an underlying graph of connections (existing or not).

Here instead, the dendrogram structure comes out from a process of clustering where the correlation between the two vertices is given by the actual links between the vertices.

2.3.2 *Spectral Analysis*

This is a completely class of methods where the structure of communities is determined by the eigenvalues and eigenvectors of suitable functions of the adjacency matrix $A(i, j)$ (Hall, 1970; Seary and Richards, 1995; Kleinberg, 1999). The basic notion of linear algebra necessary to understand this approach are presented in Appendix C. We nevertheless suggest to consult a basic text for a more comprehensive background (Golub and Van Loan, 1989). Here we only mention that in most approaches, referring to undirected networks, $\mathbf{A}$ is assumed to be symmetric.

The main function introduced are the Normal matrix (actually a "normalized") and the Laplacian matrix.

2.3.2.1 *Normal Matrix*

This matrix has the same entries of the adjacency matrix, divided for the degree of the node. In linear algebra that can be written as

$$\mathbf{N} = \mathbf{K^{-1}A} \tag{2.17}$$

Where the matrix $\mathbf{K}$ is a diagonal matrix that has on the diagonal element k_{ii} the degree of vertex i. This corresponds to write $k_{ij} = \delta_{ij} \sum_{l=1}^{N} a_{il}^{w}$. Also note that matrix product is not symmetric, so that $\mathbf{K^{-1}A} \neq \mathbf{AK^{-1}}$.

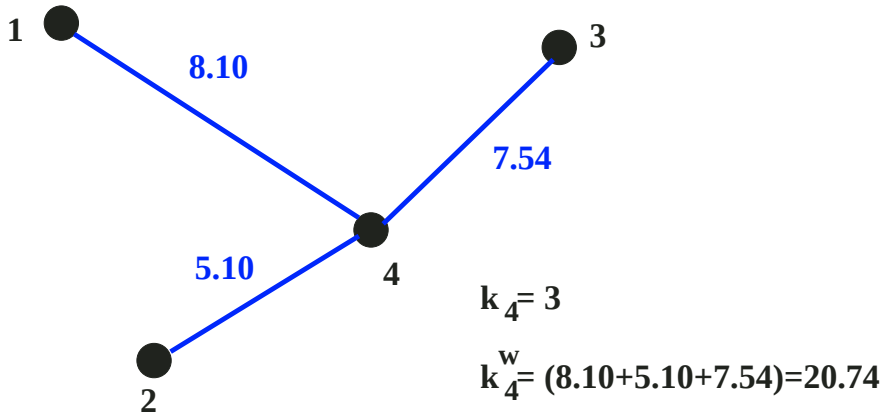

FIGURE 2.4. A simple example on a graph on which we compute $\mathbf{N}, \mathbf{N^w}, \mathbf{L}, \mathbf{L^w}$.

In practice if we are sitting on a vertex $i = 4$ with a degree 3 and this vertex is connected with vertices $1, 2$ and 3, this would correspond to have the i^{th} row of the matrix given by

$$1/3, 1/3, 1/3, 0, 0, ..., 0. \tag{2.18}$$

These entries of the matrix can be regarded as the **probabilities to pass directly** from node i to one of the neighbours. In the case of a simple graph, one assumes that any edge counts the same, so that it is equally probable to pass from one vertex to any other of the neighbours. For this probabilistic property this matrix is also known as *Transition matrix*.

Note that if the graph is oriented, the matrix $\mathbf{N^T}$ obtained from $\mathbf{N}$ by exchanging the rows with the columns corresponds to a graph with the direction of the arrows exchanged. Coming back to the probabilistic meaning, if $\mathbf{N}$ transfers from one vertex to another, we can come back using $\mathbf{N^T}$. Therefore by recursively applying N and after that N^T we must obtain the identity matrix. That is to say

$$\mathbf{N^T N = I} \tag{2.19}$$

Since the matrix product is not symmetric, all the above requirements are fulfilled if *against the standard notation* entries in the adjacency matrices are written along the columns (Servedio, 2004).

In the case of weighted graphs the probabilities can be assumed to be proportional to the weights, so also in this case the i^{th} row must give the list of these probabilities. Following the above example we can show how to specify the weights. For example we can say that the weight of edge $(1-4)$ is 8.10, the weight of edge $(2-4)$ is 5.10 and the weight of edge $(3-4)$ is 7.54 (see for example Fig. 2.4). This would give a generalized degree $k_4^w = 8.10 + 5.10 + 7.54 = 20.64$. The resulting 4^{th} row of the normal matrix

$$\mathbf{N^w} = (\mathbf{K^w})^{-\mathbf{1}}\mathbf{A^w} \tag{2.20}$$

is represented by the following entries

$$0.39, 0.25, 0.36, 0, 0, ..., 0. \tag{2.21}$$

Note that by construction (as explained by the probabilistic argument above) the sum of the entries along a row is equal to 1.

2.3.2.2 *Laplacian Matrix* Another interesting function of the Adjacency Matrix **A** is given by the Laplacian Matrix **L**. This matrix in the language of matrix operations is given by

$$\mathbf{L} = \mathbf{K} - \mathbf{A}. \tag{2.22}$$

Where **K** is the degree matrix previously defined where the non-null elements of this matrix are only on the diagonal and they are equal to the degree k_i of the vertex i.

Using our previous little example this means that the i^{th} row of the matrix **L** (for unweighted graph) would be

$$-1, -1, -1, 3, 0, ..., 0. \tag{2.23}$$

while for the weighted graph as shown in Fig. 2.4 we have

$$-8.10, -5.10, -7.54, 20.64, 0...., 0. \tag{2.24}$$

In general we can write

$$\mathbf{L} = \begin{pmatrix} k_1 & -a_{12} & .... & -a_{1n} \\ -a_{21} & k_2 & .... & -a_{2n} \\ ... & ... & .... & ... \\ -a_{n1} & -a_{n2} & .... & k_n \end{pmatrix} \tag{2.25}$$

Where in the case of weighted of networks we substitute k_i and a_{ij} with their respective correspondent k_i^w and a_{ij}^w. This produces the new weighted Laplacian matrix given by

$$\mathbf{L^w} = \mathbf{K^w} - \mathbf{A^w} \tag{2.26}$$

The reason of the name "Laplacian" given to this matrix comes from the fact that the i^{th} row of the matrix gives the value of the Laplacian operator on the vertex i.

As reported in section **??** the Laplacian operator is defined as

$$\nabla^2\phi(x,y,z) = \frac{\partial^2\phi(x,y,z)}{\partial x^2} + \frac{\partial^2\phi(x,y,z)}{\partial y^2} + \frac{\partial^2\phi(x,y,z)}{\partial z^2} \tag{2.27}$$

That is, a Laplacian Operator applied on a certain function (in this case $\phi(x,y,z)$) returns the sum of the three second derivative with respect to x, y, z.

When the function is defined only on discrete values, we use the finite differences instead of derivatives. That is

$$\frac{\partial\phi(x,y,z)}{\partial x} \rightarrow \phi(x+1,y,z) - \phi(x,y,z) = \phi(x,y,z) - \phi(x-1,y,z) \qquad (2.28)$$

In this case the Laplacian operator defined in eq.(2.27) takes the form

$$\nabla^2\phi(x,y,z) = \sum_{\xi,\eta,\zeta} \phi(\xi,\eta,\zeta) - k\phi(x,y,z) \qquad (2.29)$$

where ξ,η,ζ indicates the coordinate of one neighbor of x,y,z. k is the total number of such neighbours. By comparing this expression with any one line of the matrix $\mathbf{L}$ we immediately see that the sum of the element along the row gives exactly the Laplacian operator (with a minus sign).

2.3.3 *Eigenvectors and Communities*

Any matrix (See Appendix C for a brief explanation and (Golub and Van Loan, 1989) for a general view) is characterized by a set of quantities. These are the eigenvalues λ_i and the eigenvectors x_i that enter in the equation

$$\mathbf{A}\mathbf{x_i} = \lambda_\mathbf{i}\mathbf{x_i} \qquad (2.30)$$

A matrix of size n (n rows and n columns) has n eigenvalues and related eigenvectors. Some of these n eigenvalues (in general complex numbers) can coincide. In some particular cases (if the matrix is symmetric, if the matrix has a physical meaning (describing probability rate) etc.) some results are known on the eigenvalue structure.

In the case of the normal matrix $\mathbf{N}$ the largest eigenvalue is equal to one and it is associated to a trivial constant eigenvector. That can be easily demonstrated using the property that the sum of the element along a row in $\mathbf{N}$ is equal to one. To show that let us consider the first element of the vector $\mathbf{A}\mathbf{x_i}$ where every element of $x_i$is constant and given by $x_i = x$ is given by

$$\mathbf{A}\mathbf{x_i} = a_{11}x_1 + ... + a_{1n}x_n = a_{11}x + ... + a_{1n}x = (\sum_{i=1,n} a_{1i})x = x = x_i \qquad (2.31)$$

For the same reasons of normalisation valid for the matrix $\mathbf{N}$, also in the case of the Laplacian matrix $\mathbf{M}$ we have special eigenvalues. Formally, the existence of the null eigenvalue can be immediately proved by noticing that all rows (and thus columns) in $\mathbf{L}$ sum up to zero.

The normal matrix is so called because its row sum up to 1 (transition matrix) and thus its eigenvalues lie in the range between -1 and 1. In particular, the eigenvalue equal to 1 is always present, with a degeneracy equal to the number of connected components in the graph (similarly to the 0 eigenvalue of the Laplacian matrix), while the eigenvalue -1 is present if and only if the

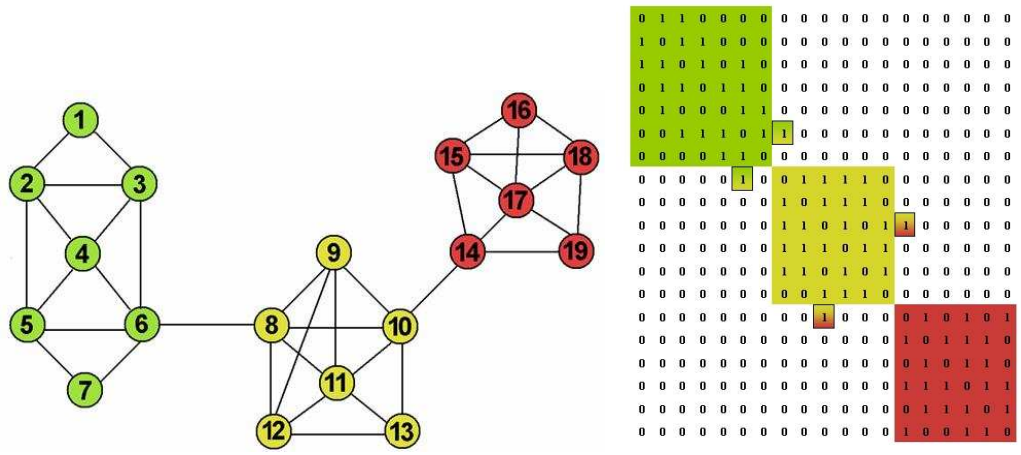

FIGURE 2.5. On the left a very simple graph, with three clear communities. On the right the aspect of the related Adjacency matrix.

graph is bipartite. Although the normal matrix *is not symmetric*, it has real eigenvalues because it possesses the same eigenvalues of the symmetric matrix $\mathbf{K^{-1/2}AK^{-1/2}}$. Because of row normalization, it can be immediately seen that the the eigenvector associated to the unity eigenvalue is the trivial eigenvector with constant components.

2.3.4 *Divisive methods*

Bisection methods

When the communities are made by separate and distinct subgraphs the adjacency matrix (as well as the other matrices) is made of distinct blocks. Every block represents the adjacency matrix for a particular subgraph. In this situation, any of the subgraphs have a constant eigenvector. The eigenvector resulting for the whole graph is given by composing the different eigenvectors for the subgraphs.

The situation remains similar if we introduce few edges to connect the subgraphs as indicated in Fig. 2.5. More generally, in a network with an m well defined communities, matrix $\mathbf{N}$ has also $(m-1)$ eigenvalues close to one. The eigenvectors associated to these first $(m-1)$ nontrivial eigenvalues, also have a characteristic structure. The components corresponding to nodes within the same cluster have very similar values x_i. This means that, as long as the partition is sufficiently sharp, the profile of each eigenvector, sorted by components, is step–like. The number of steps in the profile corresponds again the number m of communities.

Let us see how this method works by means of a simple example. Consider for example a graph made up of two disjoint complete subgraphs of order 3 and 4 as depicted in Fig. 2.6. The rows of the Laplacian matrix can be arranged in order to show the presence of these two independent blocks: one $3x3$ block and one $4x4$ block. The null eigenvalue is then present with twofold degeneracy and the corresponding vectorial space may be spanned by two orthogonal eigenvectors of

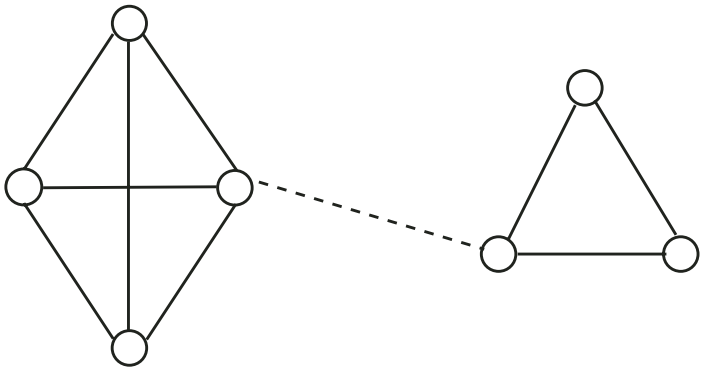

FIGURE 2.6. Undirected network employed as an example, with two disjoint complete graph of order 3 and 4. A small perturbation is then added by means of a link between them.

the type:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \tag{2.32}$$

Next we add one link (the dashed edge shown in Fig. 2.6) between two vertices lying in two disjoint components. The result is a Laplacian matrix slightly changed by a small perturbation. From matrix perturbation theory it can be shown that this perturbation removes the degeneracy of the null eigenvalue (in fact we have only one connected component in the graph now), while at the zeroth order the previous two eigenvectors change into a linear combination of them. The eigenvalues are now two. The first one is null and the other one is close to zero (because the perturbation was small). Eigenvectors are given respectively by:

$$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a \\ a \\ a \\ -b \\ -b \\ -b \\ -b \end{pmatrix} \tag{2.33}$$

with $a = 4$ and $b = 3$ such to preserve orthogonality with the trivial eigenvector. By observing at the structure of the first non trivial eigenvector we are able to discern the community structure of the graph. *Vertices are in the same community if their corresponding components in the eigenvector have the same or almost the same value.*

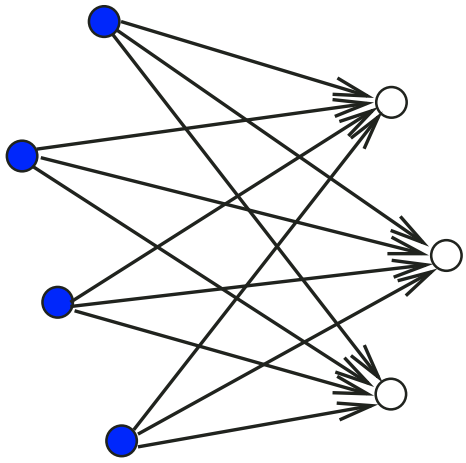

FIGURE 2.7. A very simplified example of a characteristic structure of the World Wide Web. The dark vertices are the *hubs* of the graph, the light ones are the *authorities*.

The situation remains substantially unaltered if more than 2 clear communities are present. If we have m well defined communities in the graph, the zero-th approximation is a linear combination of m vectors similar to those in eqn (2.32). The m final eigenvectors have the structure of eqn (2.33). Vertices are in the same community if their corresponding components in the m eigenvectors have a similar value.

There are m eigenvalues close to the unity, indicating that there are actually m clear communities in the graph and the following eigenvalues would be separated by these latter ones by a gap.

The method is substantially the same also with the matrix $\mathbf{N}$ except that the null eigenvalue of the Laplacian matrix is replaced by 1. Furthermore the eigenvectors in eqn (2.33) do not need to be orthogonal since the normal matrix is not symmetric.

2.3.5 *Agglomerative Methods*

- **Hubs and Authorities algorithm** For a directed network, it is possible to extract communities information from the link structure. This algorithm was proposed on empirical bases in order to find the main structures in the World Wide Web. Web pages are divided in two categories: the **hubs** and the **authorities** (see Fig. 2.7).
  Those quantities are defined in a self-organized way by the dynamics of the World Wide Web. By the creation of a link from page p to q, the author of page p increases the authority of q. So the first recipe in order to define the authority of a site would be to consider its in-degree. This is only an approximation; in the world wide web, many links are created without specific reference to the authority of a page (i.e. the hyperlinks to return to the home page). To partly overcome this problem we need to individuate at the same time also the counterpart of the authorities site, that is the

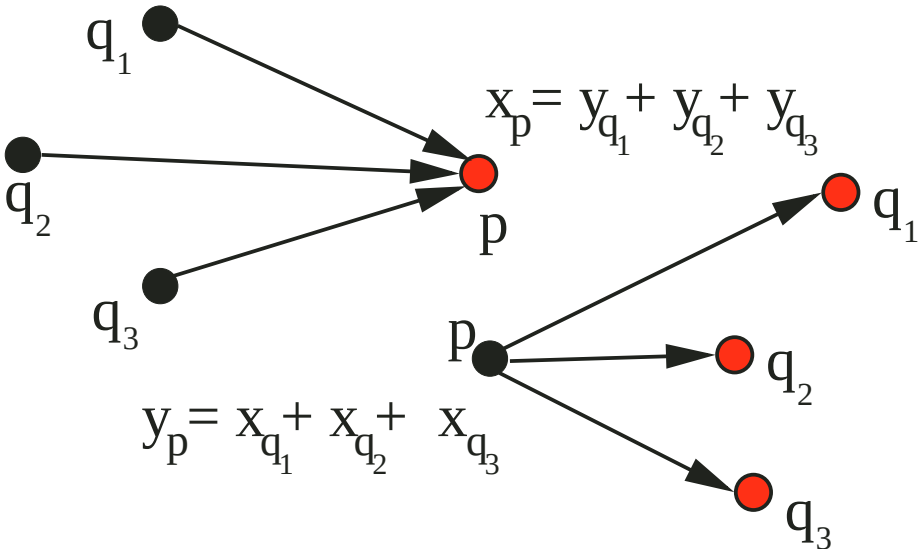

FIGURE 2.8. The definition of the hubs weight x_p and authority weight y_p

hubs. A hub is defined as a site pointing to many related authorities. It can be demonstrated that a specific relation exists between the two quantities.

The algorithm works starting with a subset of the web (obtained by means of a text searching algorithm). On that subgraph we remove the internal links and the links pointing to the webmasters. The iterative procedure starts by assigning to every vertex p (a page) a non negative authority weight x^p and a non negative hub weight y^p. The idea is to solve simultaneously all the equations in order to find the set of values of x and y satisfying the requirements.

The solution of the system is done numerically through an iteration procedure. That is, starting from fixed initial values one computes recursively the new values according to the formulas

$$\begin{aligned} x_p &= \sum_{q \to p} y_q \\ y_p &= \sum_{p \to q} x_q \end{aligned} \tag{2.34}$$

where $q \to p$ runs on all the pages q pointing to p and $p \to q$ runs on all the pages p pointing to q. Under specific mathematical conditions (verified in this case) the value of the x_p's and y_p's does not change anymore from one iteration to the other. This set of values is called "fixed point" of the iteration and gives the solution of the system of equations. From the point of view of linear algebra the solution of such problem consists in finding the set of eigenvalues and eigenvectors of the matrices AA^T and A^TA correspond to highly clustered nodes belonging to a single community.

This can be shown arranging in a vector $\mathbf{x}$ the various x_p for every vertex p. Similarly, the various y_p can be arranged in a vector $\mathbf{y}$ Then the relation between $\mathbf{x}$ and $\mathbf{y}$ can be written as

$$\mathbf{x} = \mathbf{A^T y} \text{ or } \mathbf{y} = \mathbf{A x} \tag{2.35}$$

substituting one equation in the other we simply find

$$\mathbf{x} = (\mathbf{A^T A})\mathbf{x} \text{ and } \mathbf{y} = (\mathbf{A A^T})\mathbf{y} \tag{2.36}$$

Therefore the solution of the conditions required is represented by the eigenvectors of the matrix operator $\mathbf{A^T A}$ for $\mathbf{x}$. Similarly the eigenvalues of $\mathbf{A A^T}$ give the solution for vector $\mathbf{y}$.

Such algorithm efficiently detects the main communities, even when these are not sharply defined. However, it becomes computationally heavy when one is interested in minor communities, which correspond to smaller eigenvalues.

- **Weighted algorithm** The study of the eigenvectors profiles and the eigenvalues has practical use only when a clear partition exists, which is rarely the case. In most common occurrences, the number of nodes is too large and the separation between the different communities is rather smooth. Thus communities cannot be simply detected by looking at the first non-trivial eigenvector. One possibility is to combine information from the first few eigenvectors, and to extract the community structure from correlations between the same components in different eigenvectors (Capocci, Servedio, Caldarelli and Colaiori, 2004). This method is then very similar to the agglomerative methods already seen in the case of edge analysis. In this case the various correlation measures are computed on eigenvector components rather than directly on vertices.

2.3.6 *Minimisation and communities*

The eigenvalue problem can be reformulated in the form of a suitable constrained minimisation problem. We consider the most general case and focus on the weighted adjacency matrix $\mathbf{A^w}$, whose elements a_{ij}^w give the strength of the edge between i and j. Consider the following constrained optimization problem: Let us consider the following function

$$z(\mathbf{x}) = \frac{1}{2} \sum_{i,j=1}^{S} (x_i - x_j)^2 a_{ij}^w \,, \tag{2.37}$$

where x_i are values assigned to the nodes, with some constraint on the vector $\mathbf{x}$, expressed by

$$\sum_{i,j=1}^{S} x_i x_j m_{ij} = 1 \,, \tag{2.38}$$

where m_{ij} are elements of a given symmetric matrix $\mathbf{M}$.

The stationary points of z over all $\mathbf{x}$ subject to the constraint (2.38) are the solutions of

$$(\mathbf{K^w} - \mathbf{A^w})\mathbf{x} = \mu\mathbf{M}\mathbf{x}\,, \tag{2.39}$$

where $\mathbf{K^w}$ is the weighted degree matrix, $\mathbf{A^w}$ is the weighted adjacency matrix and μ is a Lagrange multiplier.

Different choices of the constraint M leads to different eigenvalues problems: for example:

- choosing $\mathbf{M} = \mathbf{K^w}$ leads to the eigenvalues problem

$$(\mathbf{K^w})^{-\mathbf{1}}\mathbf{A^w}\mathbf{x} = (\mathbf{1} - \mathbf{2}\mu)\mathbf{x}. \tag{2.40}$$

- while $\mathbf{M} = \mathbf{1}$ leads to

$$(\mathbf{K^w} - \mathbf{A^w})\mathbf{x} = \mu\mathbf{x}. \tag{2.41}$$

Thus $\mathbf{M} = \mathbf{K^w}$ and $\mathbf{M} = \mathbf{1}$, corresponds to the eigenvalue problem for the (generalized) Normal and Laplacian matrix respectively.

Thus, solving the eigenvalue problem is equivalent to minimizing the function in eqn (2.37) with the constraint written in eqn (2.38). The x_i's are eigenvectors components. The absolute minimum corresponds to the trivial eigenvector, which is constant. The other stationary points correspond to eigenvectors where components associated to well connected nodes assume similar values.

2.4 Optimized Graphs: Cost function and Transport

Sometime (and this is the most interesting case) the graph is a structure suitably designed in order to accomplish some specific functions. This class of graphs are called "economic" graphs. Trees are usually very good candidates as economic graphs. Indeed in a tree we connect n vertices with the minimum possible number of edges $(n-1)$. Whenever there is a cost to pay to form an edge, trees are the most economic choice.

The concept of economic graph can be related to "dynamics" as a transport along the edges of the graph. One immediate example of a transportation network could be represented by a pipeline connecting different houses. Houses are the vertices of the graph, while the pipes are the edges. Other examples can be cities connected by railways or houses served by electricity cables. But probably the most important and interesting case is just inside ourselves. It is represented by the network composed by our cells fed through the blood vessels.

We show now that in this case this transportation network is **optimised** in the sense that is the best possible to accomplish its function. More specifically it delivers the nutrients to the body **with the minimum possible amount of matter transported**. Not only blood vessels but also river networks and plants channels deliver their contents *in the most efficient way*. We report here some of the work done by Banavar et al. (Banavar, Maritan and Rinaldo, 1999; Banavar, Maritan and Rinaldo, 2000) on transport optimisation. The results hold for a

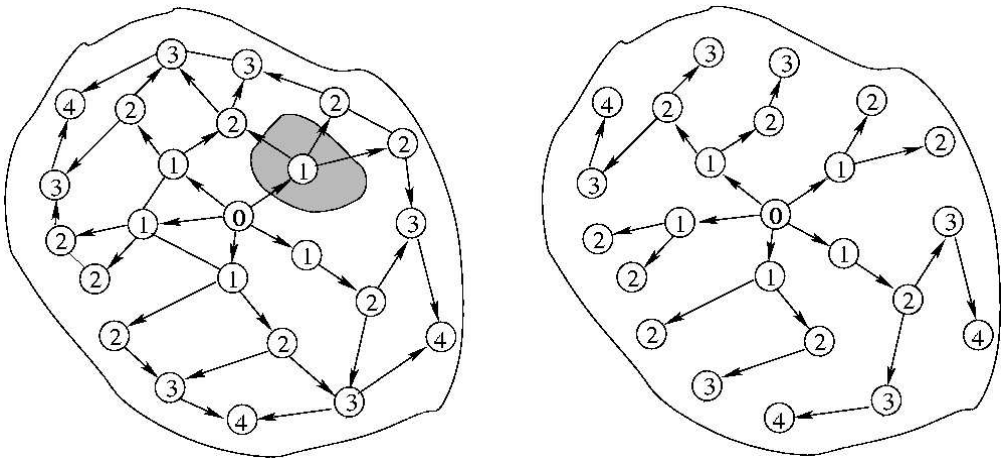

FIGURE 2.9. Left a fully connected network (for example a portion of a body with its blood vessels). The gray portion indicates the region served by the central vertex. On the right the tree obtained by measuring the distance from supplying vertex. This distance is indicated in the centre of the vertex.

lattice of any Euclidean dimension D and we try to extend this result for true graphs. The physical motivation for that research was aimed to explain the so called Kleiber relation that states that the body mass M of every animal grows with its metabolic rate with a specific exponent[11]

$$B \propto M^{3/4} \tag{2.42}$$

The value of this exponent is always the same for various species. This behaviour is called "universality" and the explanation of such universality is currently under debate (West, Brown and Enquist, 1999; Banavar, Maritan and Rinaldo, 1999). Here we do not enter this topic in detail, but we rather focus only on the transportation properties of a network as found in (Banavar, Maritan and Rinaldo, 1999; Banavar, Maritan and Rinaldo, 2000).

As far as we are concerned we consider here the various species as a volume (the body) served by a network (blood vessels). The key point is to select one source in the network (that can be regarded as the root of the transportation tree) from which the flow proceed. An animal body is described as a system of length L embedded in a space with Euclidean dimension $D = 3$. This latter value can change, the dimension of the embedding space is 2 for a river network since water flows on a surface. It is almost 2 for a thin leaf that is the space within where the nutrient flow. We have that every vertex v out of the total ones L^D is supplied with a flow F_v at a constant rate.

It can be shown (we shall not do that here) that the metabolic rate of the organism is given by

[11]For our purposes here we can think of the metabolic rate as the number of calories an organism needs to survive.

$$B = \sum_v F_v \tag{2.43}$$

This quantity B grows as L^D when the size L grows. This is because B is the product of a nearly constant term F_v times the number L^D of the vertices.

We can also compute the total volume of blood C used in the organism (this is the counterpart of M in eqn (2.43). If we denote by I_e the flow of nutrient on every edge we have that this quantity C is given by

$$C = \sum_e I_e \tag{2.44}$$

where the sum runs on every edge e. The result of a general theorem (Banavar, Maritan and Rinaldo, 1999) is that this quantity C grows with L as L^{D+1} if the graph is optimised, and as much as L^{2D} if the graph is not.

We recover the Kleiber relation only in the case of optimised transfer

$$\left.\begin{array}{l} B \propto L^D \\ C = M \propto L^{D+1} \end{array}\right\} \rightarrow B = M^{\frac{D}{D+1}} \tag{2.45}$$

That in the case of $D = 3$ gives exactly the exponent $3/4$.

We do not demonstrate here that $C \propto L^D$ is the best choice. We only note that in order to feed every one of the L^D the blood has to travel at least the mean distance (of order L). This gives exactly the exponent $D+1$ claimed to be the optimal one.

2.4.1 *The constant flux*

Let us consider a very specific case. The flow I_e is constant to every "leaf" vertex (those with only one edge). Without loss of generality we can consider this I_e unitary. It is easy to check that by recursion the role of the metabolic rate is played by the area of the basin introduced in Section 1.3 of the previous chapter.

$$A_v = \sum_i A_i \tag{2.46}$$

This is because summing on all the unitary fluxes corresponds to sum on all the points in the subbasin. We have already seen this quantity in the section devoted to trees and we saw that one possible global measure for the shape of such quantity is given by the probability distribution $P(A)$.

Consider now the case of a subtree of area A. For every one of them we can compute the quantity C. In the case of constant and unitary flux, the total volume of blood (or water or whatever) is simply given by the sum of the values A in the (sub)basin.

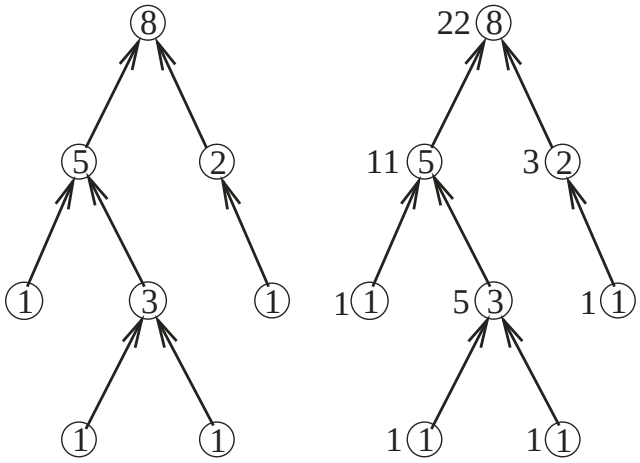

FIGURE 2.10. Values of A (left) and $C(A)$ (right) on a simple tree. Consider the blue vertex in this tree. On the right for the computation of $C(A)$ we sum all the values in the subtree, as reported on the left. This results in a value $5 = 3 + 1 + 1$.

In formulas we can therefore write

$$C(A) = \sum_{i \in A} A_i \tag{2.47}$$

where the sum runs on all the vertices i (of area A_i) in the sub-basin of area A.

The real meaning of these quantities is related to the particular case of study of river networks and biology where they appear. As shown in the next chapters, the real river networks are all characterised by the same shape of $P(A)$. Similarly, biological and other natural systems are characterised by the same $C(A)$.

2.5 The properties of scale-free networks

Until now we have considered the technical structure of graphs, and how to extract the quantities of interest for our analysis. This introduction has been necessary to present the real topic of the book, that is the recent discovery that almost every graph that you can find in nature has similar characteristic. Regardless the area (biology, computer science, physics, geology, social systems, finance and economics) all these structures display the same statistical property (Barabási, 2002; Buchanan, 2002; Dorogovtsev and Mendes, 2003; Pastor-Satorras and Vespignani, 2004). The Probability Distribution for the degree is power law distributed.

The first striking experimental evidence is that the same mathematical form holds in the different cases of study. This sort of "universality" (holding only qualitatively) means that the shape of degree distribution remains constant. As we are going to present in the next chapter a power law (scale-invariant) relationship is a very specific request. This means that the system is the same at

	Network	**Order**	$\langle k \rangle$	ℓ	C	γ
1	Internet	3700 - 10500	3.6 - 4.1	3.7	0.21 - 0.29	2.1
2	WWW	$2 \cdot 10^8$	15	16	–	2.1–??
3	Movie actors	225, 226	61	3.65	0.79	2.3
4a	Sex. Net (Males)	2810	22.63	–	–	1.6
4b	Sex. Net (Females)	2810	6.3	–	–	2.1
5	Co-authorship	56627	173	4	0.726	1.2
6	Protein Int. Yeast	1870	2.395	–	–	2.4
7	Protein Int. Drosophila	3039	2.40	9.4	–	1.26
8	Stock ownership	240	2.67	18.7	0.08	2.
9	Word Occurrence	460902	70.13	2.67	0.437	2.7
10	Coliseum flora	282	14	2.65	0.28	2.6

Table 2.2 *The general characteristics of several real networks. For each network we indicated the number of nodes, the average degree $\langle k \rangle$, the average path length ℓ and the clustering coefficient C. As a general notice the value of the size of the networks E (i.e. the number of edges) can be found from the definition of average degree. The average degree is by definition two times the number of edges E divided by the number of vertices. This gives $E = N\langle k \rangle/2$. In case of oriented networks, we have in-degree and out-degree and one edge contribute only to one of those in one vertex, so that the factor two is absent. The average length ℓ is always taken in the largest connected component of the graph.*

any level one looks at it. This could be in some cases the signature of a peculiar and interesting past history of the system. In other cases this could be also the result of the statistical noise. In this section we start by presenting the data evidence. In the next two chapters we enter in detail in the study of power laws.

Universality is also present in the statistics of other quantities even if they are not power law distributed. For example in a large variety of cases the distribution $P(d)$ of distance d is not scale-invariant. Rather the distribution is peaked around small values like $4, 5, 6$. This effect is known as **Small World effect** since in the social graphs where vertices represent individuals, a little number of relationships (edges) can connect two parts of the graph (Milgram, 1967). Similarly to the scale-free properties, the small world effect has been tested carefully in a variety of different situations. It appears in Internet, in the WWW, in the network of sexual contacts between individuals (Liljeros, Edling, Amaral, Stanley and Aberg, 2001), as well as in the network of co-authorship where different scientists are linked if they write a paper together (Newman, 2001*a*; Newman, 2001*b*).

Data set **1** (Vázquez, Pastor-Satorras and Vespignani, 2002) refer to the Internet as described by the Autonomous Systems. In Data set **2** (Broder, Kumar, Maghoul, Raghavan, Rajalopagan, R. Stata and Wiener, 2000) we have a $\ell = 6$, if orientation is removed from the links; the diameter is as large as 500 in the oriented case. As for the degree we have a clear power law for the in-degree and an unclear behaviour for the out-degree. Data set **3** (Barabási and Albert,

1999) have been collected from Internet Movie Database. Data set **4** (Liljeros, Edling, Amaral, Stanley and Aberg, 2001) present probably a bias in the distribution shape, so that the values of γ as well as the average number of partners is very different (statistically it must be the same when the size of sample goes to infinite). Figures on the number of partners are computed from the plot of the degree distributions. Amongst the possible and various datasets on co-authorship we report in dataset **5** a collection of papers from high energy physics (Newman, 2001*a*; Newman, 2001*b*). Dataset **6** refer to one of the many analyses (Jeong, B.Tombor, Albert, Oltvai and Barabási, 2000; Jeong, Mason, Barabási and Oltvai, 2002) made on Protein Interaction Network of *Saccharomyces Cerevisiae*. The fitting function proposed has quite a number of free parameters being in the form $P(k) =$. As for the data set **7** the only analysis we are aware of has been published in December 2003 (Giot, Bader, Brouwer, Chaudhuri, Kuang, Li, Hao, Ooi, Godwin, Vitols, Vijayadamodar, Pochart, Machineni, Welsh, Kong, Zerhusen, Malcolm, Varrone, Collis, Minto, Burgess, McDaniel, Stimpson, Spriggs, Williams, Neurath, Ioime, Agee, Voss, Furtak, Renzulli, Aanensen, Carrolla, Bickelhaupt, Lazovatsky, DaSilva, Zhong, Stanyon, Finley, White, Braverman, Jarvie, Gold, Leach, Knight, Shimkets, McKenna, Chant and Rothberg, 2003). Protein interaction of fruit fly *Drosophyla Melanogaster* were detected with high or lower confidence. Here we present data from high-confidence interactions. Data set **8** were collected on the Italian Stock Exchange (Garlaschelli, Battiston, Castri, Servedio and Caldarelli, 2005). The Word co-occurrence network (Ferrer i Cancho and Solé, 2001*b*) in dataset **9** is one of the many way to define a word network. Finally data set **10** refers to taxonomic trees obtained from plants collection (Cartozo, Garlaschelli, Ricotta, Barthélemy and Caldarelli, 2005).

3

SCALE-INVARIANCE

FIGURE 3.1. A fish made of other fishes (Circle Limit II 1959), the concept of self-similarity in the genius of Michael Cornelius Escher.

Many years later, as I faced the editor of this book I was to remember that distant afternoon when my tutor took me to discover fractals. Even if this episode is rather distant in time, I still remember the feeling of surprise and emotion in front of these objects. This personal experience could maybe explain the success of fractals in modern science.

It is almost impossible to present here a detailed analysis of the various scale-invariant systems appearing in nature. Therefore we present here only a brief overview of the successes achieved by using the concept of scale invariance. Many of the same ideas can be applied directly on self-similar networks (and that is the reason of such chapter) while other case are more specific.

Anyway, all the examples clarify the general interest in the self-similar properties of the networks.

The study of scale invariance has a long tradition. Among the first fields where this property has been analysed we have the theory of critical phenomena (Stanley, 1992), the percolation (Stauffer and Aharony, 1971) and the fractal geometry (Mandelbrot, 1975). The first two topics are familiar probably only to physicists, while fractals have certainly a larger audience. Fractal geometry is

related to the activity of the mathematician Benoit Mandelbrot[12] (Mandelbrot, 1975; Mandelbrot, 1982). The basic importance of such activity is the understanding that fractals represent an usual outcome of systems evolution and not an horrible paradox (as believed by some of the mathematicians who developed them in the 19^{th} Century). Starting the publication of his book "Les Objects fractals" in 1975 (Mandelbrot, 1975), a great activity started in various scientific fields. More and more phenomena have been described by means of this new formalism.

One of the first example considered by Mandelbrot was the price fluctuations of cotton in commodities market. The future price cannot be obtained with arbitrary precision from the past series. Still these series have some form of regularity. Indeed, the curves for daily, weekly and monthly price fluctuations are statistically similar. This example is particularly interesting since clarifies the degree of prediction that can be expected from fractal analysis. Actual prediction is impossible, but we can measure the probability of some future events. The fact that the same statistical features are found on different time scales is a typical signature of fractal behaviour. Also in the case of coastline lengths we find fractal objects. Strictly speaking, we always miss parts of the profile even if they are as small as the grains of the sand. Regardless the scale of observation we will find the same complexity whenever looking more carefully to the object.

Actually fractal behaviour might refer to different properties. In some systems the scale-free structure is in the shape. In this class the fractal shape can be "robust" as is the case of the branched patterns of dendrite growth, fractures and electric breakdowns. We say robust because these phenomena happen for a various range of external conditions. In the same class we have that other systems are more "fragile". Fragile in the sense that they arise after a very precise tuning of some physical quantity. This is the case of percolation and critical phenomena.

A completely different class of phenomena is verified when the scale-invariance is not geometrical but it is related to the dynamics or to the evolution. For example the time activity of one system could display a self-similar behaviour. Eye-inspection would not help now. The only possible signature of fractal behaviour is the mathematical form (power law) of the time series. This happens for earthquakes, avalanches and species extinctions. Technically those power laws regards the form of the probability distributions of energy released, duration and size of avalanches respectively.

Finally the self-similarity can be present in the way the different parts of a system interact each other. This is the case of self-similar graphs and the power law scaling appears in the distribution of topological quantities like the number of interaction per part of the system. To detect when and if this happens is the main topic of this book.

[12]Born in Warsaw in 1924 he moved to France in 1936. He is currently Sterling Professor of Mathematical Sciences Mathematics Department Yale University and IBM Fellow Emeritus at TJ Watson Research Center

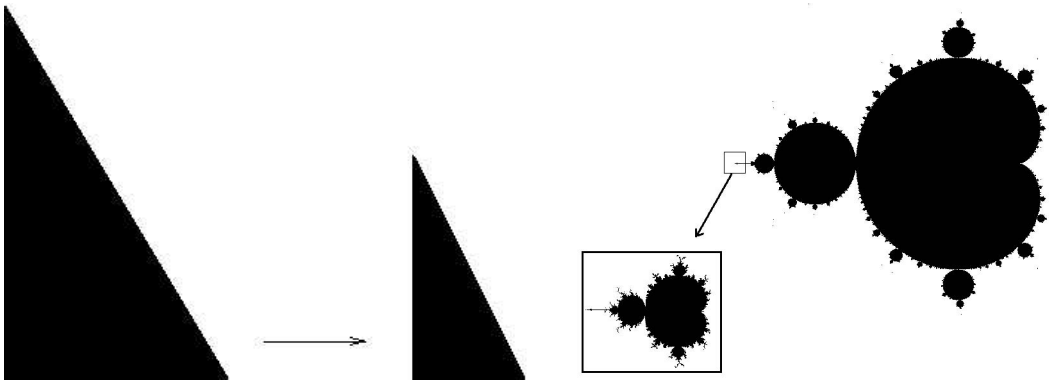

FIGURE 3.2. The concept of self-similarity. On the light two similar triangles. On the right the Mandelbrot set. In the latter some subset are similar to the whole.

Here in this chapter we present some experimental evidence of scale-free structures. The most immediate of them, are geometrical objects that result from mechanical or electrical breakdown. They are "traditional" fractals. Other scale-free relations present in Nature regard completely different quantities as probability distributions or time activity. All of them are characterised by the same shape and are sometime are related each other or are caused by similar reasons. We show how to measure fractal dimension and more generally how to plot power laws.

3.1 Geometrical Scale-invariance: Fractals

Self-similarity describes invariance of some physical objects with respect to a change of the scale of observation. From very basic geometry, we know that the technical term "similarity" is introduced to describe objects (like for example two triangles) whose respective edges are proportional. Whenever this happens, we know that by enlarging the smaller objects by the factor of proportionality we obtain a perfect copy of the second one.

Some other systems are very complicated and still similar to themselves. This means that a little part of them when enlarged is a copy of the whole. It is then natural to refer to such systems as *self-similar*. This is not yet a mathematical definition for such entities. Empirically, we can define as self-similar (hereafter *"fractal"*) all the objects that look the same, squeezing or enlarging the scale of observation. Probably the most famous object of this kind is the Mandelbrot set shown in Fig. 3.2. The shape of the set is determined by exploring which set of complex numbers has the property to keep a certain function finite[13] The region

[13]Technically, the Mandelbrot set is the region of the complex space corresponding to the numbers c's such that the succession $z_n = z_{n-1}^2 + c$, starting from $z_0 = 0$ remains finite.

FIGURE 3.3. Starting on the left we have three steps of construction of the Sierpinski gasket. After an infinite number of steps it looks like the last picture on the right (actually only five steps of iteration).

of the complex space formed by such numbers form this incredibly complicated lines of separation. Using this idea, we now try to build a fractal object trying to provide step by step a more rigorous characterization of it.

3.1.1 *Fractals by iteration*

The procedure is shown in Fig. 3.3. We start with a triangle whose size is L; we then divide it into four parts and we cut the central one. This produces the second object in the figure. Now we iterate such recipe for all the three surviving triangles obtaining the third object in Fig. 3.3. It is easy to check that any of the subset with size $1/2$ of the system resembles the original one at scale 1[14]. By abstraction we can think to repeat infinitely this procedure obtaining an object that would look very similar to the last object in Fig. 3.3 (actually only 5 iterations).

This object is known as Sierpinski Gasket[15] and we now characterise it in a more precise way. The striking difference with respect to the starting triangle is the regular presence of empty regions between the black ones. This strange distribution of filled and empty regions makes impossible to measure the area (how much space is filled) of the objects. As first notation, we see that the area of the compact triangle is an over-estimate the area of Sierpinski Gasket. Then we must find quantity that in standard triangles is larger than in fractals. To find the right measure we must generalize the usual concepts.

To measure something means to compare it with a sample object called "unit of measure". When we say that a building is ten meters high we mean that we need to put ten times one meter over another in order to have the same height. The same procedure holds also for surfaces: an area of 10 square kilometers is covered by 10 replicas of a large square of 1 Km per side (even if this is not the smartest way to actually make such a measure). A non trivial question arises if we want to know how many unit measures of surface we need in order to cover a Sierpinski gasket. For the first iteration we only need three triangles of

[14]Technically that is true provided one cuts off the scales of observation. I.e. considering only sizes greater than $1/8$ and smaller than 1

[15]named after the mathematician Waclaw Sierpinski born in Warsaw in the 1882 and died in the same place in 1969

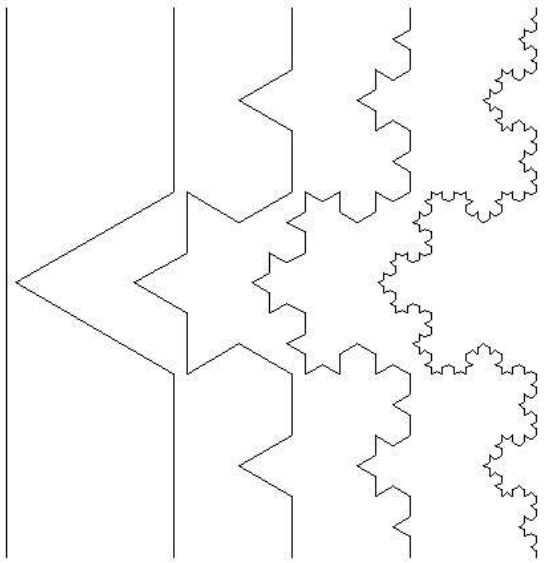

FIGURE 3.4. The Koch curve, obtained by infinite recursion of kinks production in the middle part of the segment.

side 1/2 (Note that for a whole triangle we would need four of them.) For the second iteration we need nine triangles of side 1/4 (Note again that for the whole triangle we would need sixteen of them). In general for the compact triangle the number of triangles needed grows quadratically as we reduce the size.

More formally, if $N(\epsilon)$ is the number of triangles of size $1/\epsilon$, then the quantity

$$D = \frac{lnN(\epsilon)}{ln1/\epsilon} \quad (= \frac{ln4}{ln2} \text{ or } \frac{ln16}{ln4}) \tag{3.1}$$

is constant as ϵ changes and it is equal to 2. This is a very complicated way to say that a triangle has "Euclidean" dimension $D = 2$. Things are getting more interesting by considering the fractal Sierpinski Gasket. By using the previous formula we obtain

$$D = \frac{lnN(\epsilon)}{ln1/\epsilon} \quad (= \frac{ln3}{ln2} \text{ or } \frac{ln9}{ln4}). \tag{3.2}$$

Again the value of D does not depends on the particular ϵ, but now D is equal to $ln(3)/ln(2) \simeq 1.58496..$

This quantity called **"fractal dimension"** measures the difference of compactness of a Sierpinski gasket with respect to a regular triangle. D it is lower than 2 because the Sierpinski gasket it is less dense than compact figures. D is also larger than 1 and that is reasonable because the gasket is denser than a line. This is only an example, different objects can be produced in a similar fashion. Here we produced a fractal by iteration, forming more and more empty region in a compact truly two-dimensional triangle. We can produce fractals with the opposite procedure. That is to say we can kink a single segment, a truly one-dimensional object. In this way it becomes clear the etymology of the word "fractal". It comes from the Latin word *fractus* i.e. broken. Starting with a line

FIGURE 3.5. A rough representation of the coast of Norway, even on such coarse grained scale it is evident the fractal nature of the line

segment, we divide it into three different pieces and replace the inner one with a reversed V shape as shown in Fig. 3.4. As usual we must imagine how the system looks like after an infinite series of iteration. The final shape is somewhat similar to the last picture in Fig. 3.4. By computing now the value of the fractal dimension we find a value of

$$D = \frac{lnN(\epsilon)}{ln1/\epsilon} = \frac{ln4}{ln3} = 1.2618.. \tag{3.3}$$

Denser than the value expected for a simple linear object.

3.1.2 *Statistical Fractals, Scaling and Self-Affinity*

Passing from mathematics to applied sciences the situation changes a little bit. Real data do not look so regular and nice as the pictures presented until now. No iterative formula is given for the cases of study. Therefore, all the relative information on fractal dimension must be extracted from the data. In many text-books the difference between toy objects like Sierpinski gasket and real fractals is sometime indicated as the difference between "Deterministic Fractals" and "Statistical Fractals" respectively. That is not completely correct since you can have statistically randomness also in the toy pictures as a Sierpinski gasket.

The situation is particularly clear by considering the coastline problem. In Fig. 3.5 we present a plot of the Norway coastlines. It resembles for its kinky shape a fractal of the type of the Koch curve. Yet we cannot expect that for any small division exactly the same recursion takes place. The most probable case is that **on average** the number of boxes needed to cover the structure is peaked around a mean value having sometime oscillation on smaller or larger values. Since we are interested in the limit behaviour when the linear size of the box

tends to zero we have to infer the limit behaviour of this averaged quantity (see next section).

Another important extension of self-similarity is given by the concept of self-affinity. Affine transformation in usual geometry, mean that you apply different magnifications on different directions. Eventually this result in stirring or shrinking along a certain direction more than the other. This could result in transforming a square in a rectangular object. Exactly as the self-similar objects are figures that are similar to one of their subset, self-affine objects are "affine" to their subsets. In the following we will have different example of quantities whose magnification properties are different along different directions. We will indicate by saying that they have different "scaling" properties with respect to the unit measure we are considering. This means that if the object is doubled along one direction it could be three time longer along another. The two different direction also have a peculiar "scaling" one each other.

A typical curve that shows self-affine characteristic is a typical price curve of a stock. This curve is determined by plotting on the y-axis the price of a certain stock at time t reported on the x-axis. Since the function is single valued (it means that for any value of time t there must be one and only one price) oscillations are allowed on the y-axis. In this case the concept of fractal dimension is no longer appropriate. Rather the behaviour of oscillations with respect to time evolution makes more sense. We will show in the following sections how power laws are intimately related to fractals. In that perspective we can say that peculiar scaling of quantities is still a signature of some fractal behaviour.

3.2 Measuring the Fractal Dimension

Having in mind that in the real world data the self-similarity can be expected to hold only in a statistical sense (because at a certain scale you may have little fluctuations in the size of the empty regions with respect to another scale) we present here two of the most used methods to compute the value of the fractal dimension.

The first one i.e. the *"Box-counting"* comes immediately from the mathematical definition of fractal dimension. It can be applied whenever we deal with a picture or a spatial grid of data. Let us consider the case of Fig.3.6, where a satellite image of a picture of a wildfire is given. The method counts the number of units of measure (in this case boxes) needed to cover the structure. Of course two limit situation are present. Boxes much larger than the whole set are uninteresting, the structure is in any case inside. This is called the "upper cut-off". will not give sensible information, because the definition will be so poor that we cannot distinguish anything of the structure. This is the reason why the fractal dimension is defined in the limit of the linear size of this units of measure going to 0. Anyway there is also a lower cut-off given by the image resolution, so that is not sensible to consider units smaller than a pixel of the image. Then one simply plots the number of boxes occupied by the structure when the size of this boxes

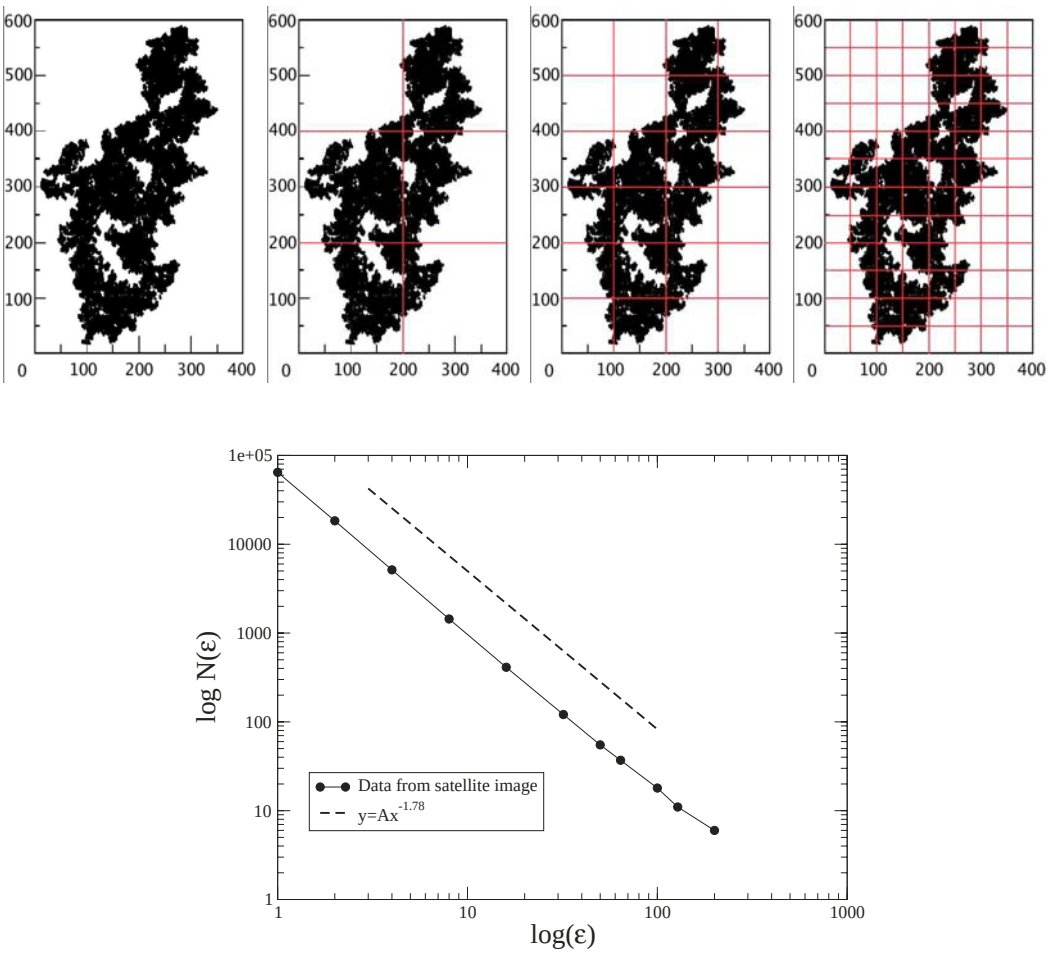

FIGURE 3.6. Above first steps of box-counting for a satellite image of a wildfire. Every black pixel is an area of vegetation destroyed. Below the plot of $N(\epsilon)$ versus ϵ.

passes from one (large as the whole set) to the size of the pixel of the image. Different stages of this procedure are presented in Fig.3.6

3.3 A home-made fractal, the mass-length relationship

In some of the cases (as in the case of three dimensional objects) a picture of the fractal set could not be available. To overcome such problem, one can use the assumption that the density of mass on the fractal objects does not change, so that fluctuations in the mass are only given by the geometry of the sample. In such a way since for a compact object holds $M = \rho L^D$ (where M is the mass and L its linear size) we expect the same relation to hold also for a fractal (where D is a non integer number). Therefore by measuring an object and weighing its mass it is possible to compute its fractal dimension. This method is called *"mass-length relationship"* and we can illustrate it with a little experiment.

Let us take two sheets of paper in the format A4. A good quality printing paper weighs $80g/m^2$, so that an A4 ($210mm \times 297mm$) weighs $4.9896g$. Take one of the sheets and wrap it in a little ball. After this procedure it does not

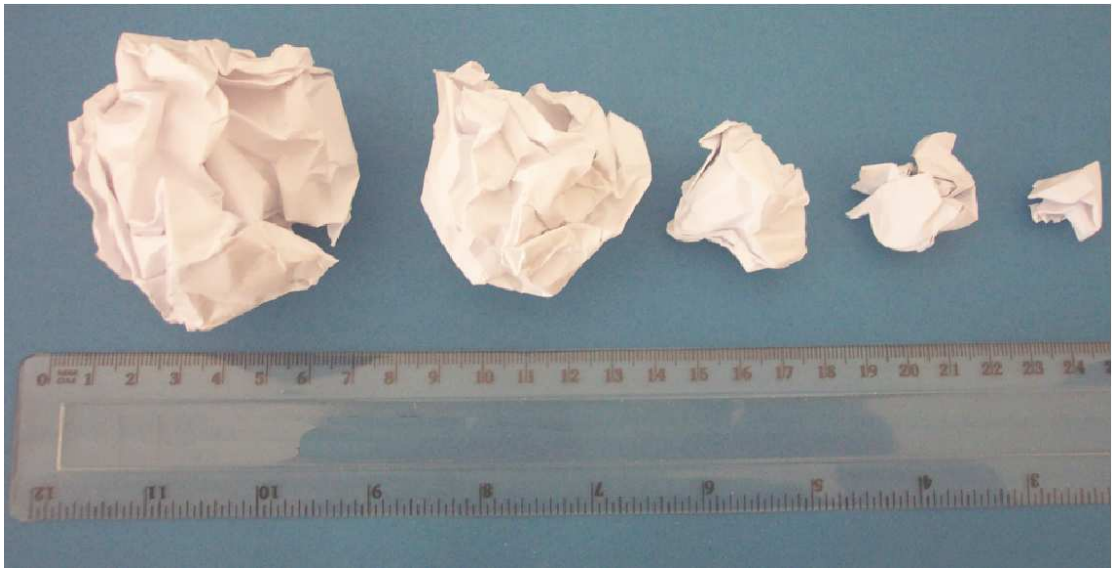

FIGURE 3.7. One realization of the experiment. A ruler in centimeters and inches is available below.

look nice as before but still weighs about 5 grams. Then take the other sheet and divide it in two. Now wrap one of the parts. It will form a ball of about 2.5 grams. Divide again in two the parts left. Wrap one of the two pieces. It will form a little ball of about 1.2 grams. Continue like that for a couple of times. Now you are left with $5, 6$ balls of different mass. If you now take a ruler and you try to determine the linear size of the objects you will obtain very likely the values reported on the left of Figure 3.7. Technically the best way to proceed is to determine the diameter D from which we compute the radius R. Note that in principle we could also plot directly the mass versus the diameter, since radius and diameter differ only by a constant factor $(1/2)^D$. Technically, using radius we can reduce the size of error bars in our measurements. Generally though, the error bars are better reduced by means of accurate observations and more statistics; nevertheless for the sake of this little experiment, even with a rather rough estimate the plot of points on the right part of Fig.3.8 looks quite nice.

3.4 Scale Invariance and Power Laws

The mathematical form of self-similarity is represented by power laws. A power law is a special case of function $y = f(x)$. Whenever the $f(x)$ can be represented as a power of the x then the relation between y and x is a power law. Simple linear function $y = x$ and quadratic function $y = x^2$ are the most immediate example of power laws. Even if very simple, they are not at all trivial. Elastic force grows with the distance r from equilibrium as a power law since $F_{el} \propto r$. Gravitational and Electrostatic forces $F_{G,E}$ decay with distance as a power law with exponent -2 since in both cases $F \propto 1/r^2$ (i.e. $F \propto r^{-2}$).

In the case of fractals we have just seen that their geometry can be identified by considering the number of boxes $N(\epsilon)$ of linear size $1/\epsilon$ needed to cover the

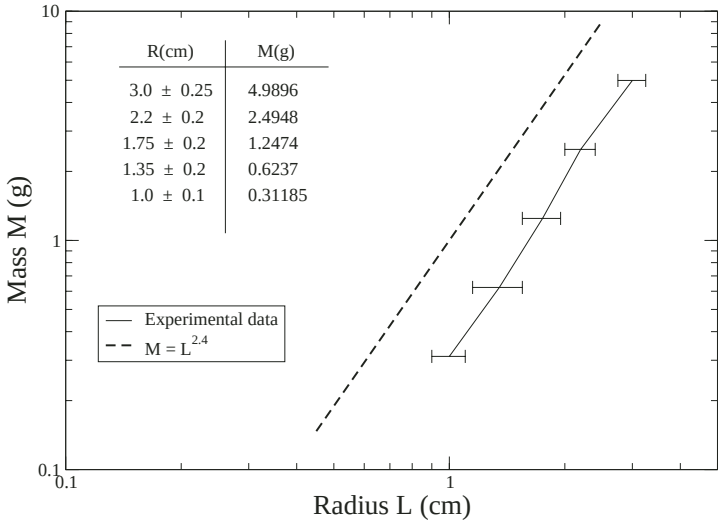

FIGURE 3.8. The mass-length relationship for the wrapped sheets allow to determine the fractal dimension of these objects.

structure. In particular,

$$N(\epsilon) \propto (1/\epsilon)^D \equiv N(\epsilon) \propto \epsilon^{-D}. \tag{3.4}$$

That is, the number of boxes is a power law of the linear size whose exponent is D. We already know that D is called the *fractal dimension*. The above formula is a way to define D, but this quantity is usually measured by means of another power law relation. This is the above mentioned mass-length relationship

$$M \propto L^D \tag{3.5}$$

As we will see in the following there is a large series of physical situations where the objects are geometrically compact. Still their dynamics or their time evolution obey a self-similar behaviour. In all these case we have power law distributions for such quantities (very often with non integer exponents).

3.5 Plotting a power law

Once defined a suitable way to measure the fractal dimension we now turn our attention on the procedure of fitting the experimental data in order to determine the quantities of interest. As already clarified, these quantities as the fractal dimension have the form of a power law. This means that whenever we put the data on a logarithmic plot the shape we must expect is that of a straight line as shown in Fig. 3.9.

Let us consider now the case shown in Fig. 3.10 **A**. This plot is very noisy at the end of the distribution. This is a very general feature; very likely this is what we must expect from a real experiment. That is not surprising when using logarithms. The feature of a logarithmic plot (consider for example the base 10 logarithm) is that the same space is given on the x-axis between 10 and 100 and

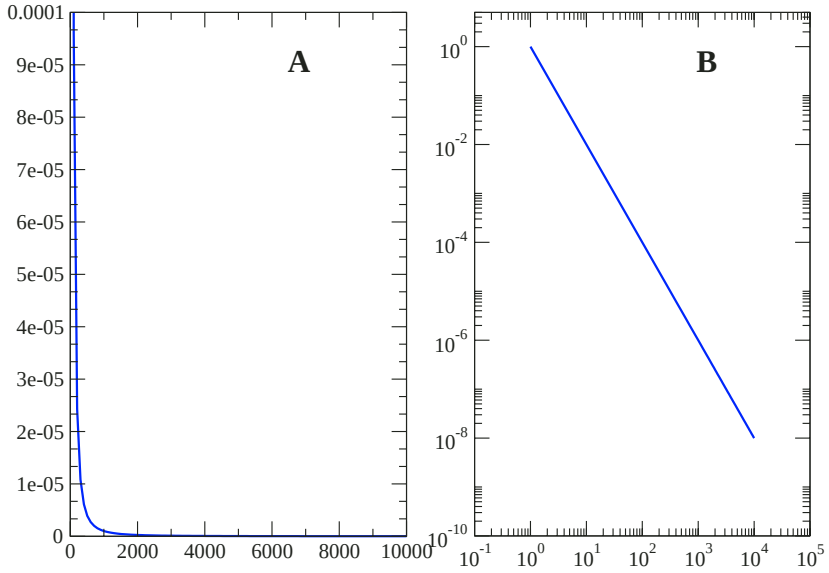

FIGURE 3.9. The plot of the function $y = x^{-2}$ in linear (**A**) and logarithmic (**B**) axes. The scale is different in the two plots, because otherwise on the left (linear scale) the curve would appear indistinguishable from the axes.

between 1000 and 10000. This happens because in both cases we have a change of 1 in the exponent, and with logarithm axes we measure exactly these exponent changes (we have $10 = 10^1$ and $100 = 10^2$ so that the exponent passes from 1 to 2; in the latter case we pass from $1000 = 10^3$ to 10000 that is 10^4, again with a change of 1). If we realize N experiments, very likely we test in a very accurate manner the interval between 10 and 100 if N is large enough. This is less likely for the following decades (like the one between 1000 and 10000 where 9000 different results are possible) even if N is rather large. In general if N is kept fixed (as it is always the case) we must expect a very noisy behaviour on the tail of the plot.

Different solutions are possible to overcome this problem. We show them in Fig. 3.10 **B** and Fig. 3.10 **C**. In the first case we applied the method of "binning" while in the second case we used a cumulative distribution.

These methods both reduce statistical noise using the elementary property that the sum of fluctuations from average must be zero[16] Therefore when summing the data, fluctuations average out and we recover the average. Before proceeding further explaining these two methods, let us note an important difference with Fig. 3.9. In Fig. 3.10 we simulate the values of an experiment. Therefore we cannot have values of y lower than 1 (we cannot measure one event 0.3 times). Dividing the number of observation by the total number of trials we have a lower limit for the y given by $1/N$. This cut-off will play a role of a certain importance in the following.

[16]This is very simple to see, in formula the sum of fluctuations from average is given by $\sum_{i=1,N}(x_i - \langle x \rangle) = N\langle x \rangle - N\langle x \rangle = 0$

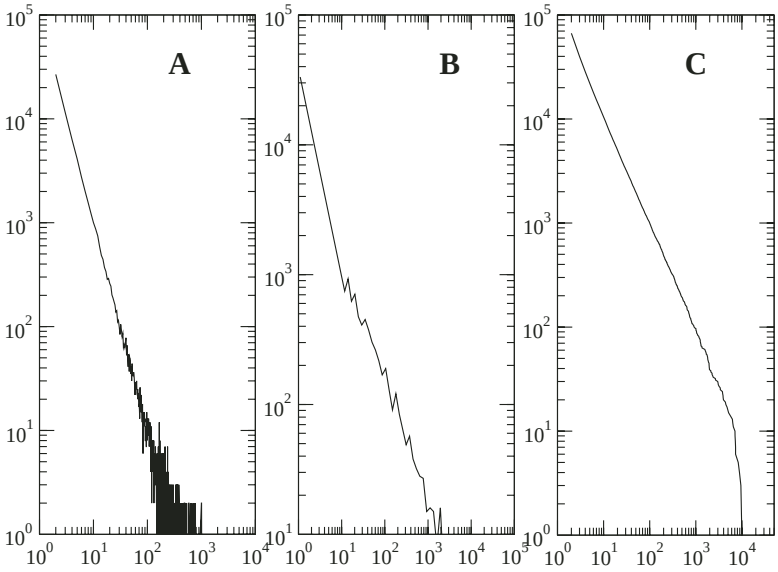

FIGURE 3.10. **A** The plot of a computer generated (with imposed power law shape) frequency distribution of random numbers. **B** The plot of the same Frequency distribution after a binning with sample of size exponentially growing. The first bin is between 1 and 1.2 (not shown), the second between 1.2 and 1.2^2. In general $n-th$ bin collects data between 1.2^{n-1} and 1.2^n. **C** The cumulative distribution of the same data. Note the change in the slope.

3.5.1 *Binning procedure*

In this method the averaging is made by dividing the x-axis in interval and summing the data within the intervals called bins.

- We divide the whole range of x in bins;
- we assign an average value for the x of the bin and an average value for the y relative to the bin;
- finally we plot them.

For example, we can take the frequency of all the numbers between 1 and 10. The bin is 10 units wide and $\langle x \rangle$ is 5. The sum of all the frequency divided by 10 will be the average value of the y. If the size of bins is constant we have a big problem. For example in the decade between 100000 and one million, the points would be so many to be hardly recognizable. Furthermore, as noticed above, as the x grows, more and more trials would be necessary to test any single bin of length 10.

The correct procedure is to use a *logarithmic binning*. For example take the size of the first bin of two units (i.e. average all the point between 1 and 3), then the second bin will have a size of four units (i.e. between 3 and 7), the third one of eight units (i.e. between 7 and 15) and so on. In this case the size of the bin is a power of 2 (i.e. $2^1, 2^2, 2^3$....) but the basis can be any number larger than 1. In particular in Fig. 3.10 **B** we used a value of the bin given by powers of 1.2. The drawbacks of this procedure (Newman, 2004*c*) are the following: this

method does not completely reduce the noise and that the choice of the most appropriate bin must be successive tentatives. For a noisy and not extremely large set of data bins too wide will cut off any interesting behaviour in the tail of the distribution. On the other hand, bins too narrow will not help in averaging out the fluctuations.

3.5.2 *Cumulative distribution*

In this method, the averaging out of the fluctuations is made on all the data set. Instead of asking the probability that a certain value x (or in the case of a continuous formulation a value between x and $x+dx$) appears in the experiment, we focus on the probability $P^{>}(x)$ that the outcome is *larger* than x.

In formulas this correspond to consider a new probability distribution $P'(x)$

$$P^{>}(x) = \int_x^{\infty} P(x)dx. \tag{3.6}$$

In this case if the $P(x)$ is a power law of the kind $P(x) = Ax^{-\gamma}$ this gives again a power law

$$P^{>}(x) = \int_x^{\infty} P(x)dx = \int_x^{\infty} Ax^{-\gamma}dx = \frac{A}{\gamma-1}x^{-\gamma-1}. \tag{3.7}$$

Results are shown in Fig. 3.10 **C** where most of the noise has been greatly reduced and the slope of the distribution changed according to eqn (3.7). Note that despite the widespread use of this method one can sometimes experience catastrophic results.

- If the exponent γ is near to one, the integral does not behave like a power law but rather like a logarithm.
- More frequently, in almost all the cases of study the upper limit of integration is rather far from infinite and it is represented by a value of x_{max}. The correct integration of equation 3.7 would then give

$$P^{>}(x) = \int_x^{x_{max}} Ax^{-\gamma}dx = \frac{A}{\gamma-1}(x^{-\gamma-1} - x_{max}^{-\gamma-1}). \tag{3.8}$$

  Now the $P^{>}(x)$ will follow a power law more and more closely as the value of x_{max} tends to infinite. If this is not the case, a likely effect would be a bending of the distribution resulting in a very difficult estimate of the value of gamma.

3.6 Scale-Invariance in Natural Sciences

The most impressive and known family of fractal structures in natural sciences is given by the class of fractals related to the deposition phenomena. One of the first examples of this class was studied by Brady and Ball (Brady and Ball, 1984). They considered a particular form of growth of crystals in their laboratory. In this phenomenon (called electro-deposition), a crystalline deposit of metal grows

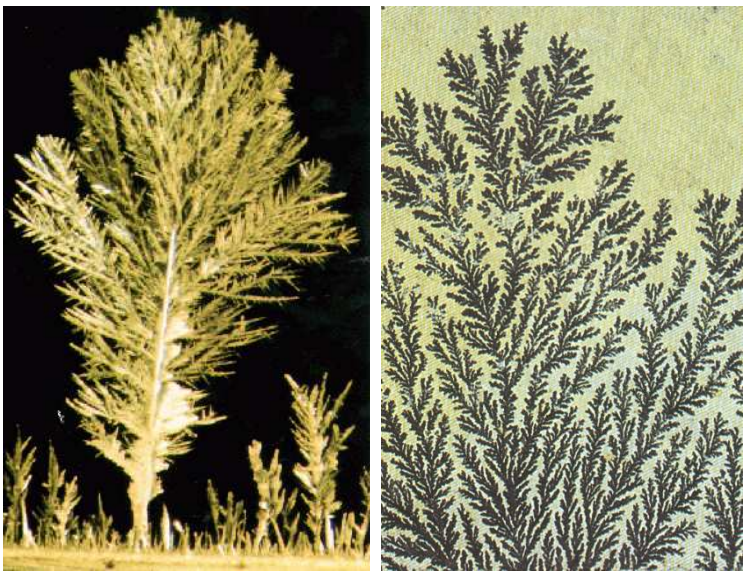

FIGURE 3.11. On the left a three dimensional crystal of copper made by electro-deposition. On the right a two dimensional mineral dendrite of Manganese Oxide (Guyon and Stanley, 1991).

on one of the electrodes put in a solution of metal ions. This process takes place when a difference of electric potential is applied to the electrodes. When the difference of electric potential is reasonably low, ions can arrange their way to the electrode and the deposition is rather ordinate. In these conditions we have an highly packed sample. On the other hand when an higher potential is applied, the growth is quicker and such optimal packing cannot be obtained. Samples grown in the latter regime are like those presented on the left of Fig. 3.11. They show a characteristic branching pattern very similar to that obtained in a model called of "Diffusion Limited Aggregation"(see Appendix **??**).

Even minerals present the same branched pattern (see Fig. 3.11 on the right). In this case the process is generated by diffusion of Manganese ions in the cracks of the rock. Whenever Manganese ions enter in contact with Oxide ions of the rock they form a dark compound (MnO) that precipitates if the concentration is larger than a certain threshold. Presence of the crystallized compound triggers new precipitation as in the DLA irreversible attachment, thereby forming similar patterns (Chopard, Herrmann and Vicsek, 1991).

Whenever the surface of the structure is smooth, all the portion of it are suitable candidates to receive further growth. When little bumps start to form, this "degeneracy" is removed. These bumps attract further growth in an accelerating process giving rise to branches. When branches appear, they restrict the possible trajectories of next walkers. The net effect is that growth happens only on the tips (see the red portion of Fig. 4.2). This basic mechanism takes place in a variety of different physical scenarios. When considering electric discharge for example, the situation is very similar. A dielectric is a material that usually does not conduce electricity. When the field applied exceeds a certain threshold

FIGURE 3.12. On the left the viscous fingering effect made by injecting air in a glycerin solution. On the right growth of bacteria colonies (Guyon and Stanley, 1991).

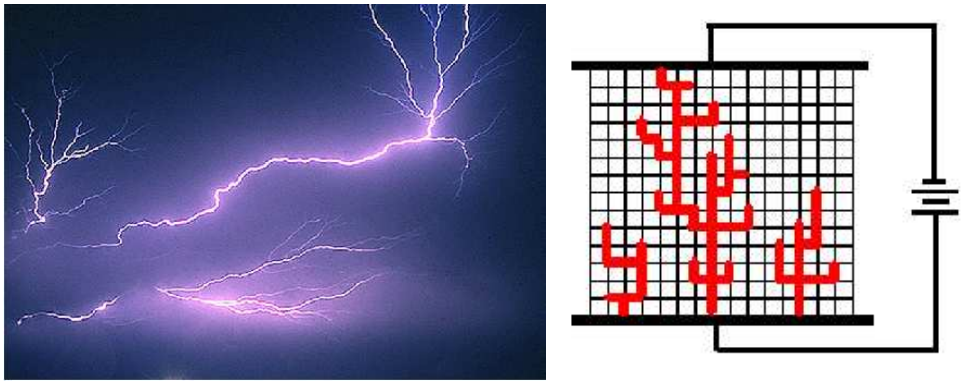

FIGURE 3.13. On the left the familiar pattern of a lightning (picture taken from http://teslamania.delete.org/frames/lichtenbergs.html), on the right the computer realization of a Dielectric Breakdown Model.

this property is no longer valid. The dielectric breaks up and a discharge flow through it. This is the phenomenon of lightnings. The electric field is so strong to attract the electrons out of the atoms in the atmosphere. Assuming that breakdown takes place with a probability proportional to the strength of the applied electric field (Niemeyer, Pietronero and Wiesmann, 1984), one can reproduce the statistical features of real lightnings as shown in Fig. 3.13.

In this case the instability given by the first breakdowns, changes the future history of the pattern, since the growth is irreversible. The same happens when the driving field is represented by concentration pressure. The tip instability is called Saffman-Taylor instability and it is at the basis of the viscous fingering phenomenon. Finally even bacteria growth can be described in such a way. In this case the gradient is given by the concentration of food (Matsushita and Fukiwara, 1993).

On the same spirit we can approach the problem of fractures. Whenever a solid object is subjected to an external load, it has a change of shape. The situation is complicated by the vectorial nature of the problem. As it is evident

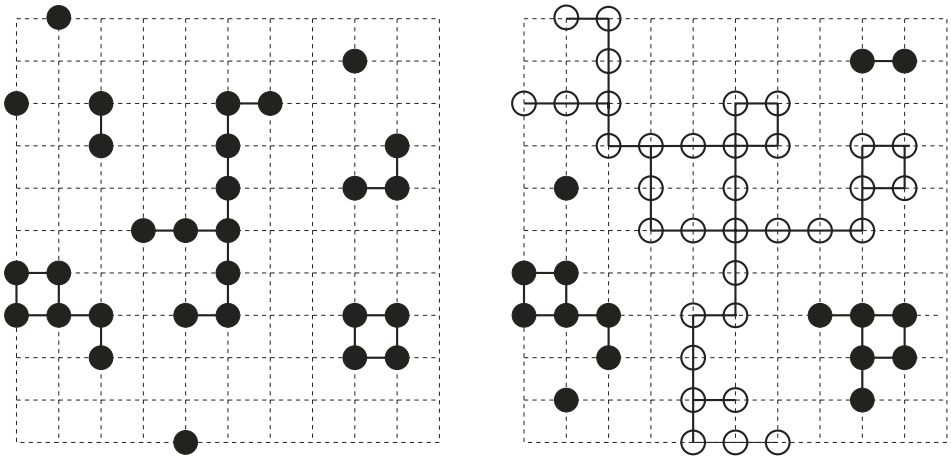

FIGURE 3.14. On the left an example of a system where nodes are drawn with a probability lower than the critical one p_c. On the right at the critical value of percolation, new sites are drawn and a percolating cluster with fractal properties (that in white) appears in the system.

in the case of a rubber string, an elongation (i.e. a deformation along its length) is very often[17] coupled by a diameter reduction (i.e. a deformation orthogonal to the force applied). When the force applied is too large, the solid eventually breaks apart. Very often the line of fracture and the part damaged show a fractal shape.

According to the particular character of the material the breakdown can happen when the material is already damaged. In this case the deformation are permanent. On the other hand other materials may break in their elastic regime where the deformations disappear when the external load is removed. A large activity is devoted to characterisation and modelling of these various cases and we do not enter in the details of the fracture theory. Nevertheless by means of very simplified models of breakdown we can spot a behaviour similar to the electric breakdown. Also in the case of elastic deformations, the load on the parts of the material are strongest on the neighborhood of the cracks. Also in this case the future breakdown proceeds from former fractures as in the case of the DLA-like structures.

3.6.1 *Scale Invariant systems in Physics: Percolation and Critical Phenomena*

3.6.1.1 *Percolation* Another way used by nature to produce fractal structures is *percolation*. This model is one of the most ancient and elegant models of fractal growth. It has been introduced to explain phenomenon of fluid penetration in a porous medium. The idea is that the various points of the lattice can be occupied or not with a certain probability p. When $p = 0$ no points are occupied, if $p = 1$ instead Interesting phenomena happen for intermediate values of p. When this occupancy probability grows from 0 to 1 small clusters of points appear in the

[17] with the notable exception of corks that can be stretched with no side deformation

FIGURE 3.15. On the left a wildfire originated in Valle del Biferno (Italy) in August 1988 destroying 56 square Km of vegetation. On the right a wildfire on Mount Penteli that in July 1995 destroyed 60 square Km of vegetation.

system. At a precise value $p = p_c$ depending on the particular lattice considered, a giant fractal cluster spans the system from one side to the opposite one. A self-organised version of this model of growth has been introduced more recently (Wilkinson and Willemsen, 1983) and it is called *Invasion Percolation*. It was introduced in order to explain the patterns observed when pushing water in a medium containing oil. The medium is represented by a series of links between the sites of the lattice. Every link has assigned a random variable describing its diameter. Since the capillarity effect, the invasion of water proceeds by selecting the bond with the smallest diameter on the boundary of the region invaded by water. This model self-organizes in a critical state forming a fractal percolation cluster in the steady state. This features is particularly interesting since self-organisation could represent a possible explanation for the ubiquity of fractal structures.

There are many real situations where percolation models can be used to describe experimental results. Here we report the study of wildfires from satellite images. From these pictures it is easy to note that after a wildfire the area of vegetation burnt as a typical fractal shape (see Fig. 3.15). Various explanations have been provided for such phenomenon, most of them closely related to the idea of percolation.

Those particular data sets shown in Fig. 3.15, consist of Landsat TM satellite imagery (30m $\times$30m ground resolution) of wildfires, acquired respectively: over the Biferno valley (Italy) in August 1988 and over the mount Penteli (Greece) in July 1995. In all the cases the image was acquired a few days after fire. The burnt surfaces were respectively 58 and 60 square Kilometers. Bands TM3 (red), TM4 (near infrared) and TM5 (mid infrared) of the post-fire sub-scene are classified using an unsupervised algorithm and 8 *classes*. This means that in the above three bands any pixel of the image is characterized by a value related to the luminosity of that area. By clustering in *classes* those values one can describe different type of soil, and in particular the absence or presence of vegetation. In particular, the maps of post-fire areas have been transformed into binary

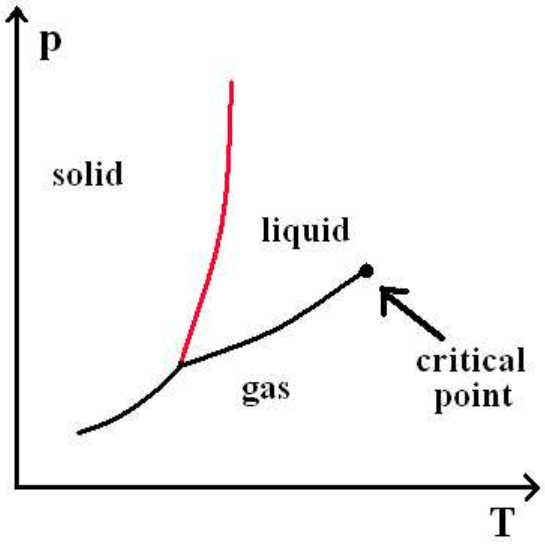

FIGURE 3.16. A typical phase diagram for a pure substance, the red line is giving the separation between liquid and solid at the various pressures and temperatures. Notably water has a different (and much more complicated) phase diagram. As an example in a certain range of the parameters the slope of this line is negative rather than positive, so that at fixed temperature an increase of pressure liquefies the ice (this phenomenon allows to skiing and skating).

maps where black corresponds to burned areas. Time evolution of wildfires is better described by a Dynamical version of percolation in which the probability of ignition decades with time so that at the beginning the fire will grow almost in a compact way leaving a fractal boundary at the end of the activity.

3.6.1.2 *Critical Phenomena* Percolation is not the only process that form a fractal by a careful tuning of a physical quantity. Traditionally this is the case of critical phenomena in Thermodynamics. Consider a typical pure (for our purpose let us restrict to only one type of molecule of a simple chemical element) chemical substance. Then plot on a graph the values of pressure and temperature at which this substance changes its phase. This means when it becomes solid from liquid (solidification) or vapor from liquid (evaporation) or vapor from solid (sublimation)[18]. In general, all the information can be reported in a chart like the one presented in Fig. 3.16.

The fact that the vapor-liquid curve stops means that if we choose the right path we can go from one phase to another without having a phase transition. This is a peculiar situation. In the standard cases (liquid-solid or liquid-vapor below the critical point) the passage between one phase and the other happens

[18]These transitions can happen at the same temperature and pressure also on the other direction (in this latter case they take respectively the names of melting, condensation, deposition)

FIGURE 3.17. On the left the data collected for N earthquakes from 1700 to 2005. On the right the spatial distribution of epicenters in ...

somehow abruptly (we have a jump in some thermodynamical quantities like the specific heat). As the critical point is approached instead, the system somehow adjust itself on a microscopic level. At temperatures T nearby the one of the critical point T_c the quantities like the specific heat, the compressibility etc. are power-laws of the quantity $(T - T_c)$. Regardless the precise substance or variable involved many systems approaching the critical present a similar behaviour.

3.6.2 *Scale-invariance in Time Activity: Avalanches and Extinctions*

There are some physical situations (as in the case of networks) where the systems of interest do not follow the fractal geometry but we still have self-similar behaviour. This is the situation of system whose time evolution is characterized by bursts of activity separated by long periods of quiescence. In such cases the *dynamics* of the system is characterized by series of causally connected events called "avalanches".

Two examples are biological evolution of species or distribution of earthquakes. In these case we find again a power law[19]. In the case of biological species the statistics is given by the number of species created from parent species. In the case of earthquakes, the statistics is considered by collecting all the earthquakes with a similar magnitude. A plot of the frequency observed for earthquakes of a certain magnitude is shown in Fig. 3.17 This behaviour is well reproduced by the law of Gutenberg and Richter (Gutenberg and Richter, 1956)

$$P \propto M^{-0.6} \tag{3.9}$$

This power-law means that no characteristic scale for earthquakes must be expected. Furthermore, after a major event, a series of little other earthquakes (aftershocks) are often present. This is a typical example of an avalanche and in many case also the distribution of the time activity for such avalanches is a power law. A plausible simple explanation for the above behaviour has not been produced yet even if many models are successful in reproducing the Gutenberg-Richter.

Coming back to avalanches, it has been thought that this peculiar dynamics is at the basis of species differentiation between plants and animals. From fossil records it appears that during particular moments of the Earth history as the Permian (291 millions of years ago) or the Cambrian (570 millions of years ago) many different species appear. In other periods a similar size of extinctions was present. This led to the formulation of *Punctuated Equilibrium* Theory made by Eldredge and Gould (Eldredge and Gould, 1972) where authors suppose that

[19]Note that in the case of earthquakes this is coupled with a fractal spatial distribution of epicenters

evolution takes place by means of intermittent bursts of activity rather by means of progressive changes.

Many models have been introduced inspired to these ideas. Amongst them the most famous is probably the Bak and Sneppen model (Bak and Sneppen, 1993). Here a food chain in an ecosystem is represented by a series of species arranged on a one dimensional lattice. A species i is represented by a real number η_i giving its fitness. Species with a low fitness are unlikely to resist in the ecosystem. In the model the species with the lowest value is removed together with its prey and its predator. Their place is taken by three new species with randomly extracted fitness values. This elementary process represents an avalanche of extinctions of order 3. If one of the newcomer also becomes extinct then the avalanche lasts for another time step. The process continues with new species entering the system. The size of these avalanches shows no particular time scale, and big extinctions event that sweep all the system are separated by long periods of small activity.

3.7 Scale-invariance in Economics and Social Science

As above mentioned, one of the first example of fractal geometry provided by Benoit Mandelbrot was related to commodity prices. In particular a chart of the different prices of cotton in a trading period looked similar to another chart with a different time resolution.

Actually, power laws in economics were introduced even before, by the seminal and pioneering work of Vilfredo Pareto[20]. Pareto noticed that in a variety of different societies regardless countries or times (Pareto, 1897) the distribution of incomes and wealth follows what is called now *Pareto's law*

$$N(X > x) \propto x^{-m} \tag{3.10}$$

where $N(X > x)$ is the number of income earners whose income is larger than x.

A similar law can be derived in a completely different context. In linguistic for example, one can be interested in the frequency of use of different words in a text. Also in this case scale-invariance appears under the form of a power law. This result called *Zipf's law* (Zipf, 1949) after the name of George Kingsley Zipf[21] states that

$$f \propto r^{-b} \tag{3.11}$$

where f is the frequency and r is the rank of the word (the most used, the second more general, the third etc..). Since b is in most of the cases equal to one we have that the frequency of a word is roughly proportional to its rank. Zipf's law is far more general than linguistic, the above case applies to cities population to and to. More generally, Zipf's law and Pareto's laws are two ways to represent the common formation of power laws in Nature. Indeed whenever we deal with a

[20]Italian economist Vilfredo Pareto was born in Paris in 1848 and died in 1923 in Céligny.

[21]Harvard professor of linguistic

power law density distribution both Zipf's and Pareto's laws apply. We can link the value of the exponents m and b (for a very clear derivation of the following see Ref. (Adamic, 2002)). The idea is that when an object whose rank is r has frequency y, it means that r words appear more than y times exactly. This is the prediction of Pareto's law, apart the fact that axes x and y are inverted. So that

$$f \propto r^{-b} \rightarrow r \propto f^{-1/b} \tag{3.12}$$

Now the frequency f is equivalent to incomes x as rank r is equivalent to the number of people with income larger than x and therefore we can link the two exponents

$$1/b = m \tag{3.13}$$

In the original formulation of the Pareto's law, we are considering the cumulative distribution of incomes rather than the density. As we recall in the Appendix of Basic Statistics the first is the integral of the second one. This means that the frequency with which we have persons with income exactly (in the case of continuous functions we need a dx) is

$$N(X = x) \propto x^{-m-1} = x^{-a}. \tag{3.14}$$

So when considering the probability functions instead of the cumulative we have the exponent relation

$$1 + 1/b = a \tag{3.15}$$

The validity of the above laws can be easily tested for two particular cases of study in the case of Zipf's law we report the analysis of the text of the Odyssey and Iliad in this case the different forms one word can have in ancient Greek (declination for nouns and adjectives and conjugation for verbs) have been considered as different words. English has a much less limited number of different forms a single word can have but nevertheless the same texts present a similar frequency distribution. By using the data set provided by Forbes company (http://www.forbes.com) we can test the law of Pareto for the 100 wealthiest persons in the world in year $2000 - 2004$. Apart the first 10 or so that seem to deviate from the expected behaviour the others behave with a rather nice power law behaviour as shown in Fig. 3.18.

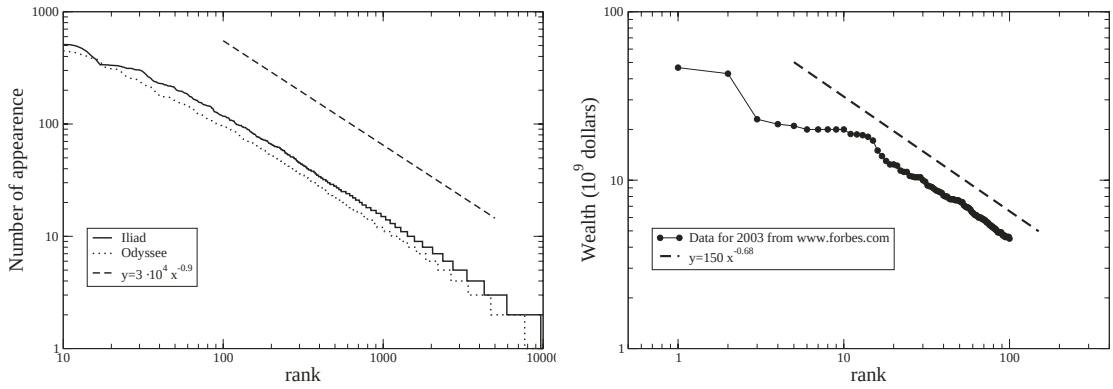

FIGURE 3.18. On the left the Zipf's law for the Greek versions of Odyssey and Iliad. On the right the Income Distribution for the first 100 wealthiest persons in the world. We use here the Zipf' form of the Pareto distribution for a comparison between the two phenomena.

4

THE ORIGINS OF POWER LAW FUNCTIONS

4.1 Introduction

All self-similar systems are power law alike. Each power law we see in Nature is self-similar in its own way. The aim of this chapter is to present the reasons for which things as different as the scale-free networks, the geometrical fractals and avalanche phenomena behave all as power laws. These reasons are various, and quite surprisingly amongst them there is also randomness. This mean that difference of properties between the parts of the system in certain conditions can produce power laws. In summary, these are the mechanisms we are going to present in this chapter.

- *Diffusion processes.* Things tend to mix, so that particle of one kind diffuse in particles of other kind, exactly as heat tend to be transferred from one body to another. The mechanism of diffusion is a powerful power law generator as we see in the following.
- *Minimisation.* Some recent trends in the research focussed on the idea that self-similarity of any kind could be related to a common evolutionary process. The common mathematical form of power laws characterise several different quantities arising from nature evolution. This is in some sense related to other theories predicting that self-similarity arises from an unknown minimisation principle (i.e. some kind of "energy" is minimised through fractals).
- *Dynamical Evolution* Self-similarity can be related to a peculiar dynamic evolution of the system. This explanation is known as Self-Organised Criticality (Bak, Tang and Wiesenfeld, 1987). This method is actually an explanation of the way in which a system could develop a fractal shape. SOC does not clarify why the steady state is power-law and therefore the question remains open. Nevertheless, this concept had quite a success in statistical physics and biology and it is likely to play a role also in the study of fractal networks. Very likely, this dynamical evolution can be driven by minimisation principles and therefore this ingredient is closely related to the above one.
- *Multiplicative Processes* For the importance of the above questions it is necessary to point out that sometimes the power laws *are not* the signature of something complicate and interesting going on in the system. Actually, in many many cases the power laws come out from very simple and "uninteresting" processes. Strictly speaking there is nothing "uninteresting" in research. We only intend to say that in most of the cases one expects

normal distributions and not power laws. This is because the central limit theorem[22], provides an excellent explanation for the ubiquity of normally distributed quantities. We show in this chapter that by using this theorem we find that in multiplicative processes (the value of a variable at time t is a percentage of its value at previous time $t-1$) we must expect variables log-normally distributed. The log-normal distribution is awfully similar to a power law one and therefore in most of the cases, due to the small size we can easily confuse them (see for example see section1.6.1 in Chapter1).

The latter case produce false power laws that look like the real ones. Actually, we can obtain also real power laws from exponential distributions if we introduce thresholds or we sample the data in a peculiar way (Reed and Hughes, 2002).

Here we present the mechanism and models that produce power laws. We start from the very traditional ones as the random walk. We then present a brief expositions of minimisation principles. The main part of the chapter is devoted in the discussion of multiplicative processes. This phenomenon either produces true power laws or log-normal distributions. In both case this is one of the most important mechanism in the formation of scale-free networks.

4.2 Random Walk, Laplace equation and Fractals

The problem of the random walk was originated by the experimental evidence about the motion of particles produced by the bothanist Robert Brown[23] (Brown, 1828). The simplest possible model (see for example (Reichl, 1980)) is given by a particle starting in $y = 0$ and constrained on a one-dimensional line (it can move only to the left or to the right). If no particular preference is given the particle moves on the left or to the right with equal probabilities $p_L = p_R = 0.5$. After some timesteps (as shown in Fig. 4.1 the walker starts to wander around. The trajectories remains centered around the position $x = 0$ but the average distance from this position grows with time. The net displacement m (it is an integer number) after N steps will be given by composing all the moves n_L on the left (let us assume they are negative) and all the moves n_R to the right (let us assume they are positive). That is

$$\begin{aligned} n_R - n_L &= m \\ n_R + n_L &= N \end{aligned} \tag{4.1}$$

[22]If we sum together different stochastic variable we obtain a new variable that is normally distributed. This holds whatever is the distribution of the original variables, provided their variance is finite.

[23]Born in Montrose (Scotland) 1773 and died in London in 1858. His accurate observations started the problem of the motion of microscopic particles. The current explanation has been given in the framework of kinetic theory by A. Einstein.

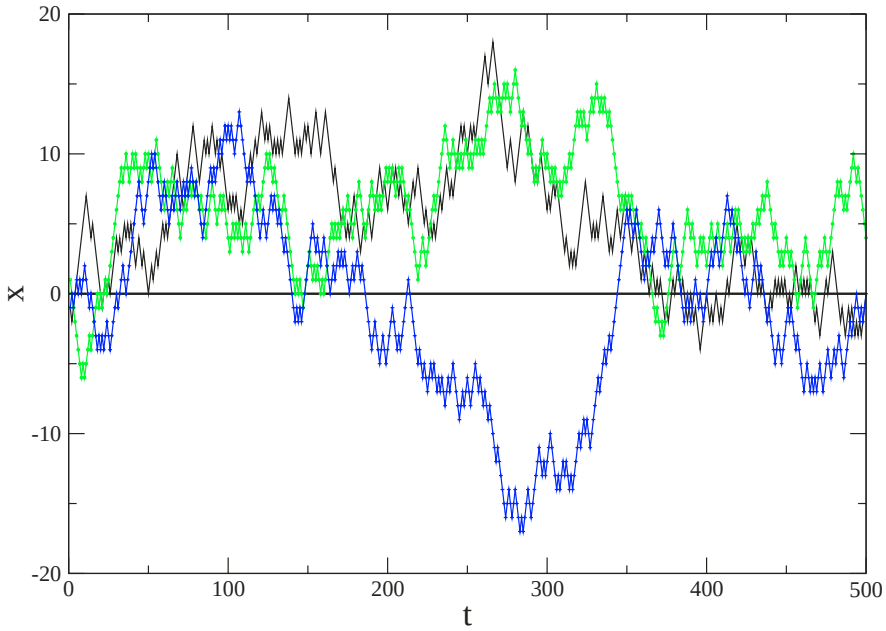

FIGURE 4.1. Three different realizations of a one-dimensional random walk made of 500 steps of unitary length.

For a large series of steps the probability distribution for the displacement will be given by

$$P(m) = \frac{2}{\sqrt{\pi}N} e^{-\frac{m^2}{2N}} \tag{4.2}$$

This is only a mathematical simplified model, in order to better describe the property of the brownian motion we must define a similar model of walk but in three dimension and with a variable step length. Let us proceed along this direction and let us start by varying the step length. In the case of one step of length l (on average), the net displacement is now given by the real number $x = ml$. Now let us also assume that the portion Δx of trajectory observed is much larger than the typical step size l (this is the case of Brownian motion where l is the intermolecular free path, while Δx can be of the order of the millimeters) For a large number of steps the probability that the particle is in the interval $x, x + \Delta x$ is given by

$$P(x) = \frac{1}{\sqrt{2\pi N l^2}} e^{-\frac{x^2}{2Nl^2}}. \tag{4.3}$$

In three dimensions this formula does not change very much apart the fact that x represent now a three dimensional vector (and apart the normalization constant). So that a particle moving of Brownian motion is characterised by a Gaussian distribution of the displacement around the mean. Surprisingly a small modification of this process is one of the most studied fractal generator.

4.2.1 *Diffusion Limited Aggregation*

This become clear coming back again to the microscopic dynamics of Brownian motion. To study this motion in the eighties was introduced a cellular automaton called Diffusion Limited Aggregation (DLA) (Witten and Sander, 1981). Cellular automata are a bit like a board game like the chess. You have the elements of the game (the chessboard, the pieces like the queen, the rooks etc.) and rules of the game. In DLA the chessboard is made by a grid $L \times L$ (it is possible to relax this assumption and take any portion of space as chessboard). The pieces are particles that occupy the sites of this grid. In the middle of the grid we fix a particle ("seed").

The game then starts following these rules:

- **Rule A: Birth of a particle.** Particles are added on one of the sites on the boundary of the grid. It will start to move as prescribed by the following rule **B**.
- **Rule B: Life of a particle.** Particle move from their starting node by means of random walk. For random walk we intend that the particle has no memory of the past path. In any point of the walk the particle can (with the same probability) proceed straight, come back or turn left or right. Hereafter this particle will be also called "random-walker".
- **Rule C: Death of a Particle.** This means that whenever the random walker in its walk passes nearby a particle deposited, it sticks on it. At the very beginning it will eventually sticks on the seed. When this happens a new particle is added following rule **A**.
- **Rule D End of game** Stop when the arrest point of a particle is near (i.e. $L/4$) to any of the boundaries.

At this point the name of the model is clear, we have a particle that has a motion of *diffusion*. This diffusion is *limited* by the boundary of the grid and above all by the *aggregation* of particles. For the first time in this chapter we note that coupling thresholds and randomness in a dynamical process we can drive a self-similar behaviour. These idea will appear again and again during this chapter and it is probably the main ingredient in the formation of self-similar objects.

Despite the apparent simplicity of this model it gives rise to the very complicated object shown in Fig. 4.2. This model of fractal growth is the prototype of most of the deposition and corrosion processes observed experimentally.

A model of growth related to the Diffusion Limited Aggregation (DLA) is the Dieletric Breakdown Model (DBM) introduced in 1984 by Niemeyer, Pietronero and Wiessman (Niemeyer, Pietronero and Wiesmann, 1984). The basic idea is to describe formation of structures like the natural lightnings or the artificial ones as the Lichtenberg figures[24]. These structures are formed by discharging a high voltage inside an insulating material (dielectric). Insulating material do not allow electric current to flow (i.e. electrons are bound to their atoms). If

[24]These are named after the German physicists Georg Christoph Lichtenberg (1742 – 1799)

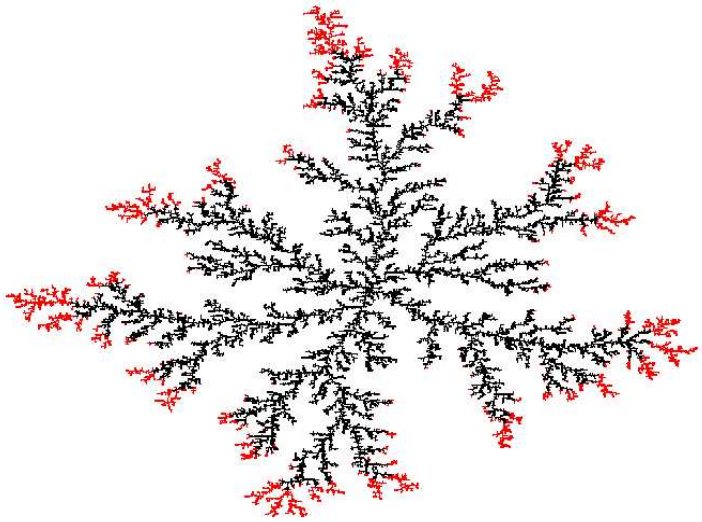

FIGURE 4.2. A cluster grown according to the model of Diffusion Limited Aggregation (DLA). For this specific cluster, more than 30000 different walkers have been sequentially sent from the boundary of the system. The last 5000 walkers have been drawn in lighter colour to show that time evolution of the object happens by growth on the tips. This means that the inner part is stable i.e. it will not be filled by successive growth.

the applied electrostatic potential grows above a threshold, this field is strong enough to remove the electrons from the atoms and a discharge takes place. An example of these figures is shown in Fig. 4.3 where a beam of accelerated electron is shot on a transparent acrylic material.

This process is described by one of the Maxwell's equations (Jackson, 1998) relating the divergence of the electric field to the electrostatic charge ρ. This equation tells us that (in standard units of Meter, Kilogram and Seconds, i.e. MKS)

$$\nabla \mathbf{E} = \frac{\rho}{\epsilon_\mathbf{0}} \tag{4.4}$$

where ϵ_0 is the permittivity of free space. By using the definition of electrostatic potential we can put this equation in the form of a Poisson equation

$$\nabla^2 \phi(x, y, z) = \frac{\rho}{\epsilon_0}. \tag{4.5}$$

In the case of dielectric the total charges in the medium are 0 so that the above equation takes the name of Laplace equation becoming

$$\nabla^2 \phi(x, y, z) = 0. \tag{4.6}$$

Maybe the symbols described in the above equations can be rather obscure to some of the readers. The above equation only states to compute and sum the second derivative of the function $\phi(x, y, z)$.

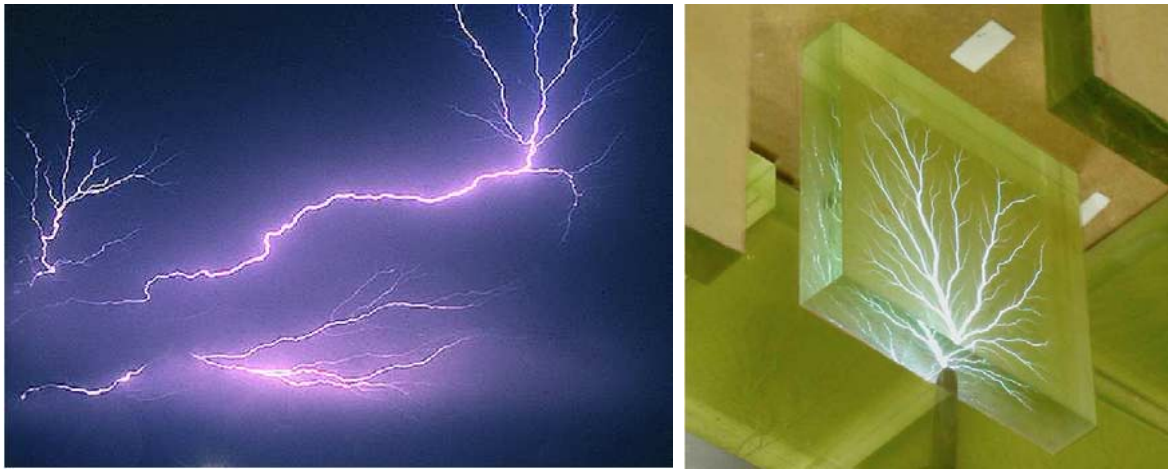

FIGURE 4.3. On the left a picture of a lightning, on the right a Lichtenberg picture. These pictures are taken from http://teslamania.delete.org/frames/lichtenbergs.html

$$\nabla^2\phi(x,y,z) = \frac{\partial^2\phi(x,y,z)}{\partial x^2} + \frac{\partial^2\phi(x,y,z)}{\partial y^2} + \frac{\partial^2\phi(x,y,z)}{\partial z^2} = 0 \tag{4.7}$$

Let us now consider the medium as a regular square lattice, so that the physical quantities are defined only on the vertices of a grid. Let us now compute the value of the electrostatic potential on these points when an external electric field is applied. For every point with coordinates (x,y,z) it must be $\nabla^2\phi(x,y,z) = 0$, but since the coordinates are now integer numbers we have that derivatives can be computed as finite differences.

$$\frac{\partial\phi(x,y,z)}{\partial x} = \phi(x+1,y,z) - \phi(x,y,z) \tag{4.8}$$

or alternatively

$$\frac{\partial\phi(x,y,z)}{\partial x} = \phi(x,y,z) - \phi(x-1,y,z) \tag{4.9}$$

By applying firstly eqn (4.8) (*forward* finite difference) and then eqn (4.9) (*backward* finite difference) we obtain that

$$\frac{\partial^2\phi(x,y,z)}{\partial x^2} = \frac{\partial}{\partial x}\frac{\partial\phi(x,y,z)}{\partial x} = \phi(x+1,y,z) + \phi(x-1,y,z) - 2\phi(x,y,z) \tag{4.10}$$

Therefore the condition $\nabla^2\phi(x,y,z) = 0$ now reads

$$\phi(x,y,z) = \frac{1}{6}\sum\phi(\xi,\eta,\zeta), \tag{4.11}$$

where (ξ,η,ζ) is one of the six neighbours of (x,y,z). That is to say, the value of the electrostatic potential in every point must be the average of the values in the neighborhood (therefore the name "harmonic" potential). This is a very

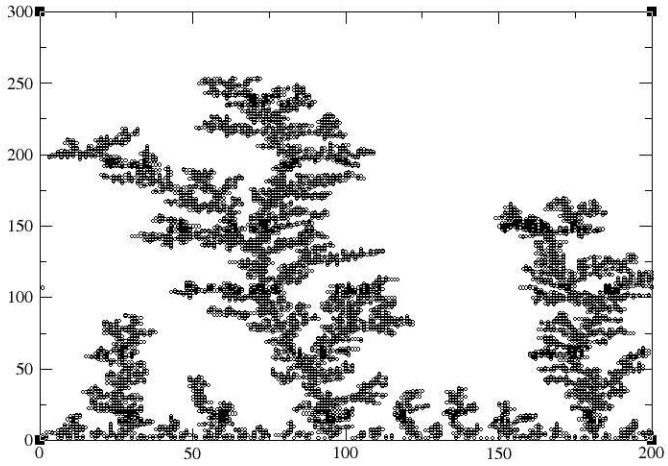

FIGURE 4.4. A pattern of DBM grown in cylindrical geometry.

complicated way to find that at rest the electrostatic field penetrates the medium forming planes of different strength from on electrode to the other.

Situation becomes more interesting when breakdown occurs. We model the breakdown stochastically. This means that we break out the bonds with a probability proportional to their electrostatic field. When bonds are burnt, the sites involved drops at potential 0. Therefore with this new boundary condition we need to recompute the values of ϕ on the sites left. Step by step the structure grows, and the steady state of this process is presented in Fig. 4.4.

4.3 Power laws from Minimisation principles

One idea presented in order to explain the onset of self-similarity is the conjecture of *"feasible optimality"*. Self-similar structures tend to be a shape that minimizes some cost function or some generalized potential present in the system. This can be better explained through the following example: consider the problem of delivering water from the source to a series of different users. In this case a reasonable cost function is to require the shortest possible number of pipelines. Sometime this requirement is complicated by other requests as to link all the clients in such a way to have them as near as possible to the source. As presented in Fig. 4.5 we see that a good compromise between user benefits and global optimisation is given by a fractal structure

In this class of phenomena the fractal shape does not correspond to the absolute minimum of this potential. Absolute minimum is often obtained only with a very precise and particular shape. This configuration is only one out of the infinite possible ones. On top of that, this minimum configuration is usually very difficult to reach during the evolution of the system. Fractal appear because self-similar configuration with similar statistical properties correspond to relative minima and occupy a relatively large region of configuration space. Therefore they are not only minimising structures, but are also more easily accessible by system evolution.

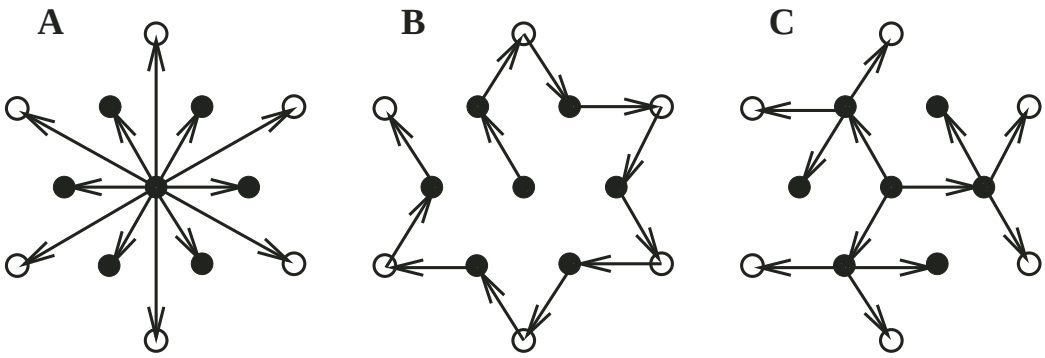

FIGURE 4.5. Let us assume that this simple series of graphs represents a water pipeline starting from the source to the various houses. Pipes have a linear cost (also, a section cost that we shall not consider in this simple example). To save resources one would like to minimize the total pipe length. On the other hand any user wants to be as near as possible to the source; for example to reduce failure risks or to have fresher water. Three classes of solutions are possible from the interplay of these two requests of mimisation, one global and the other local. A) On the left an egoistic approach. Everyone connects to the source. Maximum of local benefit (everyone is directly connected) but poor global optimisation. If the edge of the triangular lattice measures 1 we need about 16.39 (i.e. $6(1+\sqrt{3})$) pipeline units. B) In the centre one minimum of social costs. We need only 12 pipeline units, but distance from source for the unlucky last user can be as large as the system size. C) Right a self-similar structure that is a good compromise between the A and B. Distance increases slowly and pipeline units needed are still 12.

Let us explain this statement through a specific example relative to the river network case. As presented in Chapter 8 there is an hypothesis according which river networks sculpted the landscape in such a way to minimise the dissipation of total gravitational energy. In Fig. 4.6 we can see both a fractal relative minimum and the configuration for the absolute minimum. The first picture represents a computer evolution of a spanning tree designed in order to minimize such quantity. After a long time the numerical analysis remains trapped in the local fractal minimum. The second picture is the shape of the absolute minimum represented by a very regular structure known as Peano curve.

4.3.1 *Self-Organised Criticality*

Closely related to this idea of evolution according to minimisation there is the concept of *Self-Organised Criticality* (Bak, Tang and Wiesenfeld, 1987). In this process, the steady state of the evolution for a dissipative dynamical system is fractal. In order to better explain how this happens it was introduced a toy model inspired to the evolution of the piles of sand. In this model a series of sites can host different grains of sand until a certain threshold. At the threshold the site becomes "critical" and gives the sand to the neighbours. At this point, these

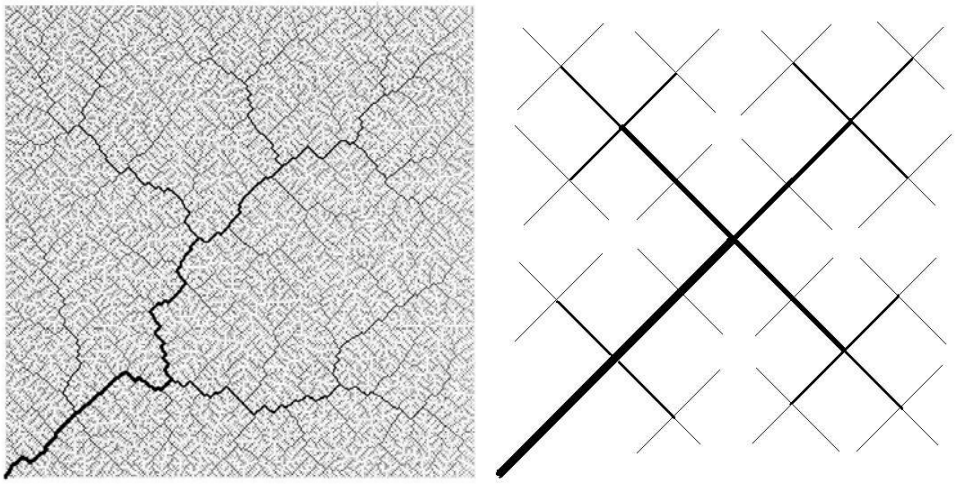

FIGURE 4.6. Two kinds of optimization. On the left a "feasible" configuration, on the right the absolute minimum of total gravitational energy dissipation

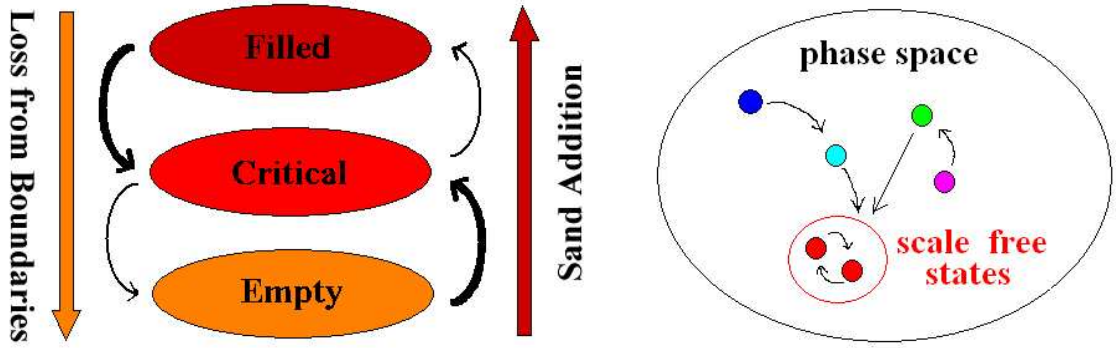

FIGURE 4.7. The basic mechanism at the basis of self-organized criticality in sandpile models.

latter ones can become critical and this elementary process can create a large scale avalanche. "Criticality" (i.e. a power law distributed series of avalanches) is maintained by a feedback mechanism ensuring that whenever the system is full of sand no more grain can be added (since you have to wait for the activity to stop). On the other hand, whenever the system is empty, no activity occurs, and then you have a lot of grain additions in a short time interval. As a result, the amount of sand tends to oscillate always around a mean value. This is not a minimisation mechanism, but rather a condition of dynamical equilibrium.

4.3.2 *Optimisation of the Entropy*

The example used by Mandelbrot refers to the distribution of words. Consider that in a language you have a series of n words and the cost attached to the use of j-word is C_j. The cost of the transmission is given by the letter (even space) used for the word, so that if d is the alphabet size, it is natural to think that $C_j \propto log_d(j)$. The average cost per word can be defined as $C = \sum_{j=1,N} C_j p_j$.

Suppose that language evolved in such a way to minimize the unit transmission cost. The probability that a word j is used in a transmission is p_j. In the case of linguistic systems the role of the entropy is played by the information transmitted in a text. We can define the entropy H as

$$H = -\sum_{j=1,N} p_j log_2(p_j) \tag{4.12}$$

Optimisation of languages in this case means that the maximum amount of information is transmitted with the minimum length of the word. At least loosely, it makes sense to assume that languages evolved in order to transfer as much information possible with the minimum cost. By imposing that the quantity C_j/H is the minimum, we can determine the form of the distribution of p_j's. In formulas

$$min(C_j/H) \rightarrow \frac{d}{dp_j}(C_j/H) = 0 \tag{4.13}$$

To ensure that this is really a minimum, we should also check the second derivative. We do not do that here, but we ensure that the value of p_j for which the first derivative is zero is really a minimum. This value is

$$\frac{d}{dp_j}(C_j/H) = \frac{C_j H - C log_2(e p_j)}{H^2} \rightarrow p_j = 2^{-HC_j/C}/e \tag{4.14}$$

If we remember our previous assumption that $C \propto log_d(j)$ we get a power law for the p_j.

4.4 Multiplicative Processes and Normal Distribution

So far we considered what is the most exciting part, that is to say the "complex" origins about the ubiquity of scale-invariant systems. Here we show that instead some system only appear to be scale-invariant but they are not (Goldstein, Morris and Yen, 1982). Some other systems are truly scale-invariant but for reasons that are not at all "complex".

These situations are characteristic of the system evolving according to the so-called **multiplicative process**. We do not want to enter in the debate if data observed can be best fit by power law or log-normal variables. Here it is enough to note that the mechanism of multiplicative process is probably the most immediate model for fat-tail pheonemena in nature since it naturally produces both. Many textbooks and scientific papers deal with this topic. A beautiful and complete description of the various approaches in different fields can be found in (Mitzenmacher, 2004) and references therein. Also ref. (Newman, 2004*c*) can be consulted on that.

Suppose you have an evolution process, where for an example an organism transforms itself in time. As a general statement the state S_t at time t will be determined by the previous states and by the external conditions. This consideration does not help in predicting what will be the evolution. This prediction becomes possible in a more restricted case.

This is the case of the **multiplicative processes** where the effect of the external shock can be written as in the following transition rule:

$$S_t = \epsilon_t S_{t-1}. \tag{4.15}$$

In other words the state at time t is proportional to the state at time $t-1$. In biology this could represent the fact that the growth of an organism is ruled by its body mass at the previous step. In the case of a city growth (Zanette and Manrubia, 1997; Manrubia and Zanette, 1998; Gabais, 1999) this equation states that the population at a certain time step is proportional to what it previously was. In both cases the proportionality constant is given by the factor ϵ_t that can change its value at any time step.

Turning back to eq.(4.15) we can immediately see that we the variable S_t is determined by the product of the various ϵ_τ where τ is between 0 and t

$$S_t = \epsilon_t S_{t-1} = \epsilon_t \epsilon_{t-1} S_{t-2} = \epsilon_t \epsilon_{t-1} \epsilon_{t-2} ... \epsilon_2 \epsilon_1 S_0. \tag{4.16}$$

At this point the distribution probability of the state is related to that of the ϵ's.

As we see the relation is rather peculiar; regardless the precise form of the distribution for the variables ϵ_τ the S_t is log-normal distributed[25]. To show that we can rewrite eq.(4.16) by taking the logarithms of both sides

$$log(S_t) = log(S_0) + \sum_{\tau=1,t} log(\epsilon_\tau). \tag{4.17}$$

In this way the product of the epsilon's transforms in the sum of the logarithms.

This sum of the logarithms of the ϵ_τ (under very mild conditions) is a normal distributed variable regardless the distribtion of the ϵ_τ. This result comes from the application of the so-called "central limit theorem". For the central limit theorem, in certain very general hypothesis the sum of random variables is a new stochastic variable normally distributed. As a result if $log(S_t)$ is normally distributed, the variable S_t is log-normally distributed.

This very simple mechanism has been rediscovered and explained over and over many times since the definition of log-normal distributions in 1879 (McAlister, 1879). As reported in (Mitzenmacher, 2004) this ideas traces back at least to the economist Gibrat (Gibrat, 1930; Gibrat, 1931) who uses essentially this model under the name of proportionate effect. On a different language a somewhat similar idea is introduced at the beginning of XX century for biological problems (Kapteyn, 1903).

[25] We remind the shape of log-normal function is given by

$$f(k) = P_{LN}(k) = \frac{1}{\sqrt{2\pi}\sigma k} e^{-\frac{(ln(k)-\mu)^2}{2\sigma^2}}$$

All these log-normal distribution, as explained in Chapter 1 can be easily confused with power laws. This happens whenever the log-normal distribution is studied in a range of k for which $\sigma^2 >> (ln(k) - \mu)^2$. On top of that this mechanism trigger also the formation of true power laws as shown in the next subsection.

4.4.1 *Power laws out of multiplicative process*

A tiny modification of the above mechanism, namely the introduction of a threshold, has a dramatic effect in the results (see Section 4.2.1). In particular *when multiplicative processes are coupled with fixed thresholds we have production of power laws distributions.*

The derivation here (as reported in (Mitzenmacher, 2004)) of the demonstration is essentially that followed by the economist Champernowne (Champernowne, 1953) who wanted to recover Pareto's law in a simplified model of economy. Pareto's law states that if we rank the people in a society according to their income we obtain a power law distribution.

In the simple model of the economy we use for the demonstration we assume that there is a minimum income possible and that is m. The poorest people in this artificial society are in class 1 and have an income between m and γm. Going to the upper class (number 2) we have people with income between γm and $\gamma^2 m$. More generally a class j will rank indivuduals whose income is between $\gamma^{j-1} m$ and $\gamma^j m$. Therefore we have a multiplicative process in which we move from one class to the upper one changing the income of an amount γ. The threshold is in the basal class where people cannot have an income lower than m.

In this economy we assume that individuals can do a change of class going to the upper class or to the lower class with a probability p_+ and p_- that remain constant along the different classes i. This means that the probability to pass from the poorest class 1 to class 2 is p_+ and the same value characterise the passage from class $n-1$ to the richest one n.

We can assume to fix ideas that $p_+ = 1/3$ and $p_- = 2/3$ and $\gamma = 2$[26]. Therefore every change of class correspond to doubling or halving the income. In this case the equilibrium probability to be in class j is simply $1/2^j$. Henceforth the probability to be in a class larger than (or equal to) j is $1/2^{j-1}$.

$$P(X > 2^{j-1} m) = 1/2^{j-1} \rightarrow P(X \geq x) = m/x. \tag{4.18}$$

This multiplicative process therefore results in a power-law distribution.

4.4.2 *Combination of Exponentials*

We have already seen that the (Gaussian) normal distribution is widespread in various phenomena, it is interesting to see that another way to produce power laws is related to a simple idea. Consider a process that grows exponentially and consider that we can observe it only at certain random times as explained

[26] When considering class 1 the $p_- = 2/3$ gives the probability to stay in this lowest class.

in Ref. (Reed and Hughes, 2002). The latter situation is particularly clear for geophysical processes like earthquakes where only the recent data started to be recordered sistematically. The distribution of these observed state behaves as a power law, even if the distribution in all the possible cases is not.

Let us see the simplest example of an exponential process of growth where a variable (as example the size of a population in an environment with infinite resources) grows exponentially with time $x(t) = e^{\mu t}$. Individuals (or whatever the x represents) have also an extinction probability given by $P(t) = e^{-\nu t}$ where $\nu > 1$. The state $X = e^{\mu T}$ where T is extracted from the $p(t)$ is power law distributed with distribution

$$f_X(x) = \left(\frac{\nu}{\mu}\right) x^{-\frac{\nu}{\mu}-1} \tag{4.19}$$

This is only a particular case where two exponential concurr to form a power law. Let us see the general case as clearly and beautifully exposed in (Newman, 2004*c*). If the quantity y is exponentially distributed

$$P(y) = e^{\alpha y} \tag{4.20}$$

and we are interested in another quantity x given by

$$x = e^{\beta y} \rightarrow y = \frac{1}{\beta} lnx. \tag{4.21}$$

we find that this latter quantity x is distributed according the power law function shown in eq.(4.19).

Let us see why: a basic property of distribution of probabilities is that the probability is a measurable quantity and it must be recovered regardless the distribution used, so that from conservation of probability ($P(x)dx = P(y)dy$) it follows that the distribution of x's is power-law.

In the above case from

$$P(x)dx = P(y)dy \tag{4.22}$$

it follows

$$P(x) = P(y)\frac{dy}{dx} = \frac{e^{\alpha y}}{\beta x} = \frac{x^{\alpha/\beta}}{\beta x} = \frac{1}{\beta} x^{\alpha/\beta - 1} \tag{4.23}$$

that is exactly eq.(4.19) where $\alpha = -\nu$.

The application of that mechanism is very frequent and certainly plays a role in some of the topic treated in the following.

4.5 Preferential Attachment, S. Matthew Effect, Richer gets Richer

This is the most successful mechanism adopted so far in the study of growing network. Interestingly, the idea that we are going to explain has been independently rediscovered several times in different fields and ages. Precisely for such

reason it has also been called with several names. To name a few of them we have *Yule Process, San Matthew effect, Richer gets richer, Preferential Attachment.*

In the community there is some agreement (Mitzenmacher, 2004; Newman, 2004*c*) on the fact that the first one to present this idea has been G. Yule (Yule, 1925) in order to explain the relative abundance of species and genera in biological taxonomic trees. As we are going to see in Chapter 7 when considering set of biological species we have a classification (taxonomical) tree with scale-free properties. Various explanation can be provided, none of which is totally repodrucing the shape of the data. The null hypothesis consists in considering that set of species comes out from evolution. Therefore starting fromn one parent species we obtain a new one that will likely to be grouped in the same genus. Every now and then though, speciated species (the new one) can be as different from the parent one that can form a new genus on their own (or be grouped in an existing different one). Probability to speciate will be larger for genera already large, since mutation rate is constant for any individual.

This explanation allow us to focus on the two ingredients of the model. Firstly you have to invoke a certain a priori dynamics (hereafter called **growth**). Secondly this dynamics select succesfull elements and make them even more successful (hereafter called **preferential attachment**).

In detail, take a set of elements each of which is characterized by a certain number $X_i(t)$. As a possible example this could be the number of different genera that have i species per genera. The set can also be a set of vertices in a graph (i.e. the WWW) and the number X_i can represent the number of pages whose in-degree is i. Now let us introduce a rule that introduces new elements in the set, these elements will not be shared equally between the older ones, but rather will be assigned more to those that already have many.

Let us consider that $X_i(t)$ gives the number of pages with certain degree i (the total number of vertices is t). The probability p_+ that the number $X_i(t)$ increases by edge addition is given by two terms. The first one correspond to a random choice of the edges and a selection of a page with in-degree $i-1$. The second one correspond to a choice proportional to the degree. A similar form holds also for the probability p_- that this number $X_i(t)$ decreases (pages with in-degree i become pages with in-degree $i+1$).

$$\begin{aligned} p_+ &= \alpha X_{i-1}/t + (1-\alpha)(i-1)X_{i-1}/t \\ p_- &= \alpha X_i/t + (1-\alpha)(i)X_i/t \end{aligned} \tag{4.24}$$

So that the balance equation can be written as

$$\frac{dX}{dt} = p_+ - p_- = \frac{\alpha(X_{i-1} - X_i) + (1-\alpha)((i-1)X_{i-1} - iX_i)}{t} \tag{4.25}$$

At this point we make the crucial hypothesis that in the steady state the $X_i(t)$ are linear in time i.e. $X_i(t) = c_i t$.

Starting from the first one X_0 for which

$$\frac{dX_0}{dt} \equiv c_0 = 1 - \frac{\alpha X_0}{t} = 1 - \alpha c_0 \tag{4.26}$$

we obtain the first value $c_0 = 1/(1+\alpha)$. The successive values are obtained through a recurrence equation

$$c_j(1 + \alpha + j(1 - \alpha)) = c_{j-1}(\alpha + (j - 1)(1 - \alpha)). \tag{4.27}$$

For large values we have

$$\frac{c_j}{c_{j-1}} = 1 - \frac{2 - \alpha}{1 + \alpha + j(1 - \alpha)} \simeq 1 - \left(\frac{2 - \alpha}{1 - \alpha}\right)\left(\frac{1}{j}\right) \tag{4.28}$$

The result of the computation is that

$$c_j \propto j^{-\frac{2-\alpha}{1-\alpha}}. \tag{4.29}$$

Along this lines Simon (Simon, 1955; Simon, 1960) proposed this model not only for species abundance, but also for distribution of population in the cities, of words in documents, of papers published by scientists and incomes in a society. As Simon model the same idea becomes then familiar also in the field of economics after the biological sciences.

R.K. Merton (Merton, 1968; Merton, 1988) applied the same idea (even if not at that level of mathematical formulation) in the field of scientific production. The same author consider this principle, very reminiscent of one sentence in the Saint Matthew Gospel. This sentence of the Gospel states: *For to everyone who has, more will be given and he will grow rich [..] Matthew 25,29 (The New Testament, 2002).* The idea of this work can be summarized by a quotation by a Nobel laureate (from the first of the above cited papers): "The world is peculiar in this matter of how it gives credit. It tends to give the credit to the [already] famous people" (Zuckerman, 1977).

Interestingly, exactly this problem of the study of formation of scientific consesus is at the basis of the work of D. J. de Solla Price (de Solla Price, 1965) that few years before of Merton published another version of this "rich gets richer" model.

The latest and most interesting formulation of this concept is explained in great detail in the Chapter 5 where the Barabási-Albert model (Barabási and Albert, 1999) is presented and analysed. In this latter case the focus is on graphs. Elements to be rewarded are the vertices and the measure of success is their degree.

5

GRAPH GENERATING MODELS

Christmas won't be Christmas without any present and science will not be science without any model. The scientific approach is based on the idea that phenomena could be described quantitatively by means of a schematic, simplified representation usually called "model".

The very idea of graph is already a powerful schematisation for different phenomena. Graph models are then introduced in order to describe how these structure are originated and how they can evolve. Under this point of view, this part of the modelization is rather similar to the task of assigning the rule in a board game. The final goal is to find suitably rules such that the final outcome of the game could be compared with real data. If we fail in such a goal, therefore the basic hypotheses (rules) needs to be changed. By the use of a model we can then reduce the complexity of the real world to a stage that we can understand and get used to.

This line of proceeding has been introduced well before the beginning of modern science. Presocratic Greek phylosophers like Thales, Anaximander, Heraclitus in their quest for the first principle made the first example of reduction of the whole universe to a simple representation.

Sometime models work remarkably well against any expectation. One of the major success of Newtonian mechanics (that is the study of motion) starts from the very unrealistic approximation of no friction. Even so, we can predict with great accuracy the time evolution of a body in motion.

In the field of network there has been a continuous feedback between data properties and mathematical abstraction. This is at the basis of the activity in this field and produced a variety of different models that we are going to present. For all the models listed in this Chapter we give at least the information on degree distribution, diameter distribution and clustering. Some models are more likely to be treated analytically or computationally than others. Some others are simply more used and for that reason more studies are available. Therefore in particular cases we can present the description of more quantities.

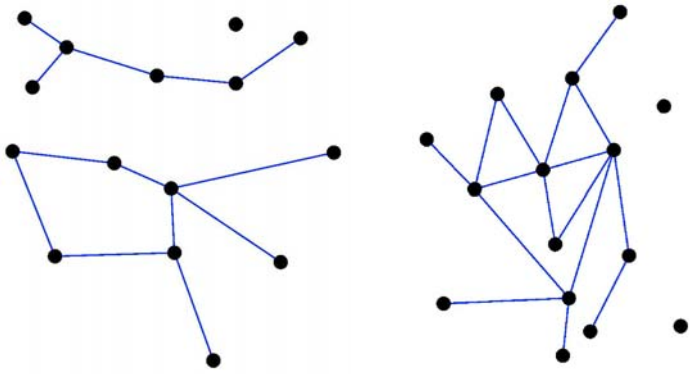

FIGURE 5.1. Two different realization of a Random Graph both with $n = 16$ and $p = 0.125$.

This chapter is devoted specifically to the model of networks. Some of them produce scale-free network some do not. In general through the modelisation we can understand the properties of real data. The first two models presented that is to say the Random Graph model and the Small World model do not form scale free networks. Starting from the Barabási-Albert model we present a series of other models as the fitness or the copying model that instead produce scale-free degree distributions.

5.1 Random Graph Model

The simplest model that has been introduced is due to the two mathematicians Paul Erdős[27] and Alfréd Renyi[28] (Erdős and Rényi, 1959; Erdős and Rényi, 1960; Erdős and Rényi, 1961).

One problem that can be addressed through this model is for example how to wire with telephone cable different cities in the most economic way.

As a first action we take a fixed amount n of vertices in the graph. After that we try to determine a rule in order to decide how many edges (cables) must be drawn. The simplest of the possible choices is to say that all the edges have the same probability to exist. This correspond to test every one of the possible $n(n-1)/2$ edges and drawing the edge with a certain fixed probability p.

A closely related model is obtained by assigning the number of vertices and consider all the graph of order n regardless the number m of the edges (size of the graph). Now every graph is a point in the probability space and we assign to it the same probability to be extracted. Whenever differently specified in the

[27] Paul Erdős was born in Budapest in 1913 and died in Warsaw in 1996. He has been one of the most productive mathematicians of his era, and was generally regarded by colleagues as one of the most brilliant minds in his field

[28] Hungarian mathematician, born in Budapest in 1921 and died in the same place in 1970. He has been the founder and director of the Hungarian Academy of Science

following we shall adopt under the name **Random Graph** a graph obtained according to the first recipe.

In Fig.5.1 we present two different realizations of a random graph. Every realization is different from the others. For that reason we instead focus on their statistical properties that remains valid for the whole class. Since the extraordinary amount of effort in this field, many quantities have been computed analytically. Here we present only the very simple ones without devoting too much space to formal derivation and proofs.

- A first trivial quantity that can be computed is the expected value of the size of the graph. This value gives the total number of edges present at the end of the graph construction. Since the probability is p and the total number of trials is given by the $n(n-1)/2$ possible edges, we have that the expectation value is given by

$$E(n) = p\frac{n(n-1)}{2}. \tag{5.1}$$

  Consequently, the probability to have at the end of the process a graph $G(n,m)$ is

$$P(G(n,m)) = p^m(1-p)^{\frac{n(n-1)}{2}-m}. \tag{5.2}$$

- Following a similar derivation the **average degree value** is given by

$$\langle k\rangle = 2m/n = p(n-1) \simeq pn. \tag{5.3}$$

- We can go a little bit further and computing the **degree probability distribution**. To obtain a vertex whose degree is k we must have have k times a succesful event whose probability is p and $(n-1-k)$ times an unsuccesful event whose probability is $(1-p)$. Since this can happen in $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ combinations we have

$$P_k = \binom{n}{k} p^k(1-p)^{n-1-k} = \frac{n!}{(n-k)!k!}p^k(1-p)^{n-1-k} \tag{5.4}$$

  The probability density distribution is very well known in Statistics and it is called Binomial Distribution. It accounts for processes where you have two distinct possible outcomes mutually exclusive (in this case either you draw a link or not). We remember that the k is an integer number, and therefore the degree distribution is a discrete one. For our purpose we note only that this distribution is bell-shaped and not self-similar.

  This distribution is usually approximated by means of the Poisson Distribution in the two limits $n \to \infty$ and $p \to 0$.

$$P_k = \frac{n!}{(n-k)!k!}p^k(1-p)^{n-1-k} \simeq \frac{(pn)^n e^{-pn}}{n!} \tag{5.5}$$

- No extimates have been given for the **clustering coefficient**. Essentially, if the graph is enough sparse we can say that the probability that two neighbours of a site are also neighbours each other (thereby forming a triangle) is still p apart some correcting terms (we are neglecting the fact that they have a common neighbour).
- Finally as regards the **diameter**, one can give a rough extimate of it. Consider the number of neighbours of a vertex i. This number $N^1(i)$(that is its degree) is given on average by

$$N^1(i) = p(n(n-1)/2) = \langle k \rangle. \tag{5.6}$$

Now consider the set $N^2(i)$ of the vertices that are at most two edges apart from i. We make an approximation, we say that this set is essentially those of the neighbours of the first neighbour. In general that is false, because in the latter one there will be counted twice all the first neighbours that are connected each other. It is important to note that this an "upper" extimate of the number of second neighbours. Through this extimate nevertheless we obtain that

$$N^2(i) \leq N^1(i)\langle k \rangle = (\langle k \rangle)^2. \tag{5.7}$$

In general through this approximation (that becomes more and more crude as the distance from i grows) we can say that for any distance d the number of the vertices that are $d-th$ neighbours grows as $(\langle k \rangle)^d$. This procedure must end at least when d is equal to the diameter D (it can stop well before). Since the sum of all the neighbours from $d = 1$ to $d = D$ must be equal to the total number n we have

$$N^D(i) = n \leq (\langle k \rangle)^D. \tag{5.8}$$

Taking the logarithm of the above formula we obtain that

$$D \leq \frac{ln(n)}{ln(\langle k \rangle)}. \tag{5.9}$$

That is the diameter D grows at most with the logarithm of the size of a random graph. Therefore if the number of vertices passes from 10000 to 100000, the diameter will simply pass from about 4 to 5. In the case of sparse Random Graph, under the hypothesis that $np \to \infty$ it has been shown (Chung and Lu, 2001) that almost surely the diameter D goes as $(1 + o(1))\frac{log(n)}{log(np)}$

5.2 The Small-World model

The Small-World model (Watts and Strogatz, 1998; Watts, 1999) describes particular graphs whose diameter remains very small when the number of vertices increases (small-world effect). Note that in general also in the Random Graph

model the diameter remains very small. Actually in any graph model the diameter grows more or less logarithmically with the number of vertices. The point is that with the small-world model it is particularly simple to investigate how this happens and to understand the onset of this feature.

In the Small-World model, one starts with a portion of an ordered grid[29]. The connections in this grid are increased by adding new edges to second, third and in general j^{th} neighbours. Another way to see this process is to consider that a site is connected with a first (nearer because we have a distance defined in a grid) layer of neighbours. We can add connection also to neighbours of the second, third and in general j^{th} layer. After this procedure the site of the grid becomes directly connected with those sites as shown in Fig. 5.2

On top of that new random connections are also established between random sites either by rewiring existing links (original formulation) or by adding brand new ones (recent formulation) with probability p.

The structure obtained in this procedure is a **Small-World Graph**, where the edges are represented by the bonds and the vertices by the sites in the lattice. Hereafter we indicate as Small-World graph the one obtained with addition of shortcuts rather than rewiring existing links.

The parameters of the model are the *coordination number* z and the *shortcuts probability* p.

The *coordination number* z gives the number of vertices directly connected in the regular structure. In a one-dimensional ($d = 1$) system with $j = 3$ (connections arrive to the third layer) every vertex has connections with $z = 6$ other vertices (three from one side and three from the other). This number of connections grows also with the dimensionality. In general we can write

$$z = 2jd. \tag{5.10}$$

In the above example in Fig. 5.2 j is 2 and therefore $z = 4d$, that is to say $z = 4$ for the one-dimensional system and $z = 8$ for the two-dimensional system. If the initial number of vertices (order) is n, the size (i.e. the number of edges) is $m = nz/2$.

The *shortcuts probability* p, gives the probability per existing edge to draw a new edge (shortcut) between two random vertices. This means that the total number of shortcuts is given by

$$mp = nzp/2. \tag{5.11}$$

Since the 2 in the formula is rather boring we follow the convention of various papers (Newman, Moore and Watts, 2000) where the coordination number is defined as $z' = z/2$. In this way the total number of shortcuts is given by $nz'p$.

As regards the quantities of this model we assume to consider the situation where p is so small that essentially we are considering a regular grid. In the

[29]In the original formulation it is used a $d-$dimensional simple cubic lattice

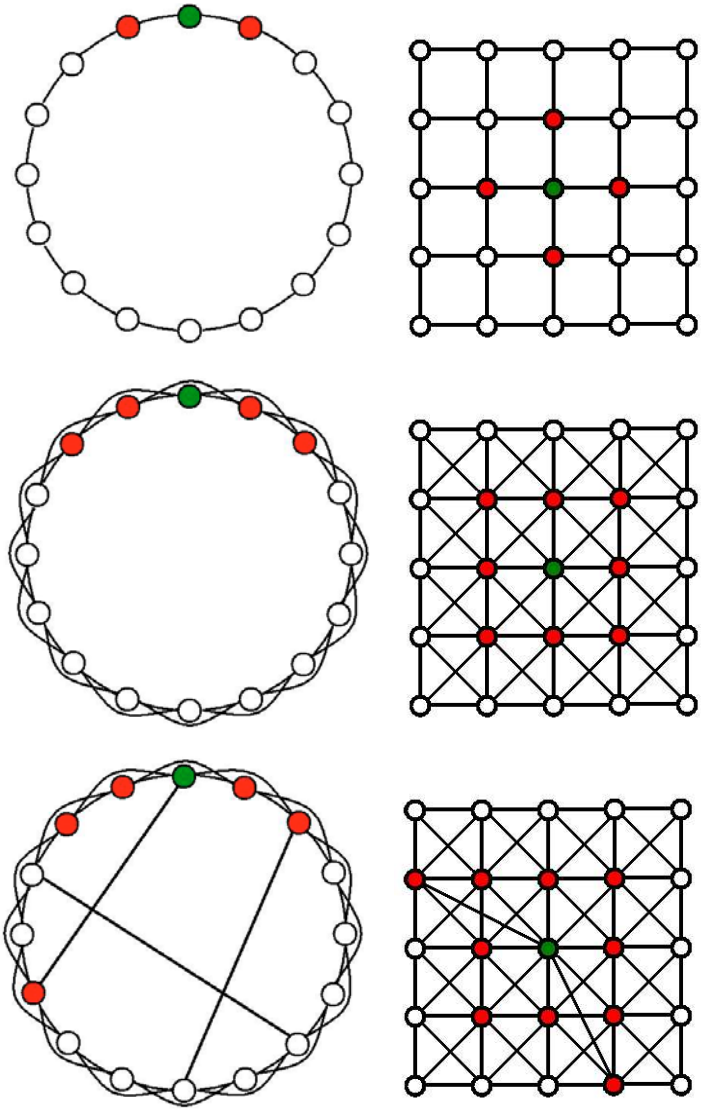

FIGURE 5.2. On the first row a regular one-dimensional lattice on the left and a regular two dimensional lattice on the right. In the second row the same lattices with extra links increasing the local connectivity. On the last line we have the small world lattices with shortcuts.

opposite limit of such many shortcuts to destroy the underlying structure we obtain a Random Graph.

- We have that the **degree distribution** is a function peaked around the fixed value z characteristic of the regular grid. With no shortcuts, this function is extremely peaked around z, being a delta function. When shortcuts completely destroys the structure of underlying grid we must expect a behaviour similar to that of Random Graph. That is to say in this limit we have a Poissonian distribution. It is easy to realize that also for intermediate values of p the Small-World model does not produce scale-free

networks.

- Exactly the shortcut presence is the "active ingredient" at the basis of the small-world effect. Even if these shortcuts are very few, their effect is dramatic. Using numerical simulation we can compute the variation on the **diameter**. It has been shown that a system with $N = 1000$ vertices ($d = 1$) a coordination number $z = 10$ and a rewiring probability $p = 1/4 = 0.25$ has a diameter as small as $d = 3.6$. with no rewiring at all the diameter of the same system is $d = 50$. Even with p as small as $p = 1/64 = 0.015625$ one still finds a small diameter $d = 7.6$.

  In ref. (Newman, Moore and Watts, 2000) it is proposed an analytical expression for the mean distance l

  $$l = \frac{n}{z'} f(npz') \tag{5.12}$$

  where $z' = z/2$ and the function $f(x)$ is

  $$f(x) = \frac{1}{2\sqrt{x^2 + 2x}} tanh^{-1} \frac{x}{\sqrt{x^2 + 2x}} \tag{5.13}$$

- The *clustering coefficient* is usually very high and it is reminiscent of the regular connection of the underlying grid. As long as z stays reasonably small and in particular $z < \frac{2}{3}n$ (as it is the case when $n \to \infty$ we have

  $$C = \frac{3(z-2)}{4(z-1)} \tag{5.14}$$

  To understand why, let us start with the one-dimensional grid by counting the triangles which a vertex belongs to. It is simpler to do such a computation if we can exclude that vertices on our left are connected to vertices on our right. As shown in Fig. **??** this can be done if the network is such large that we cannot close a triangle between a vertex on the left with a vertex on the right. That is to say $3jd < n \to 2jd < 2/3n \to z < 2/3n$. Note that in the large z limit C tends to $3/4$.

5.3 The Barabási-Albert model

The Barabási-Albert model (Barabási, Albert and Jeong, 1999; Albert, Jeong and Barabási, 2000) is specifically suited to reproduce one striking evidence of some real networks, that is to say their growth in time. the final version of the graph is built after successive time-steps. During any time-step one or more vertices are added to the system. As new vertices are added they establish edges with the old ones. The latter ones are chosen with a probability that is proportional to their existing degree.

In Fig. 5.4 we present two steps of this construction. The two ingredients of the model are the **growth** and **preferential attachment**. Growth implies that new vertices enter the network at some rate. Preferential attachment tells that

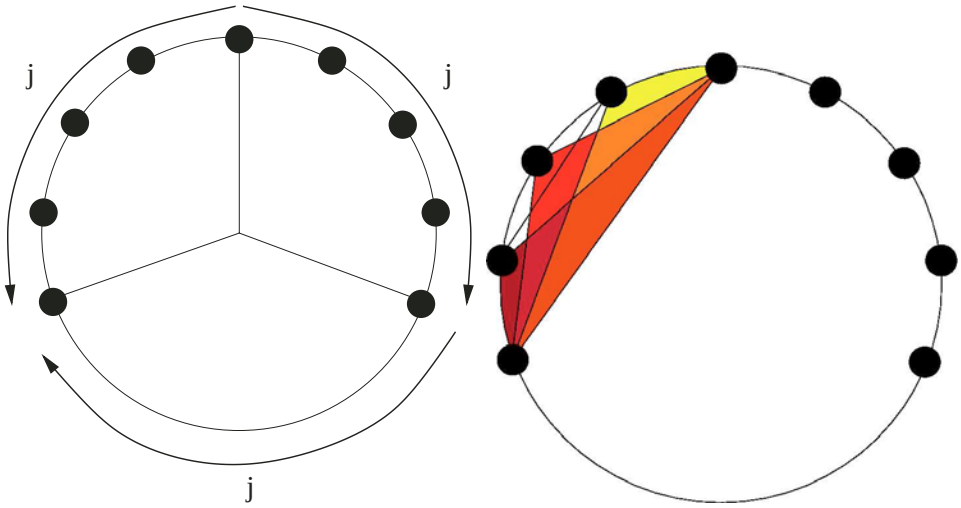

FIGURE 5.3. The computation of the clustering coefficient. On the left the condition to have separate triangles between left and right sides.

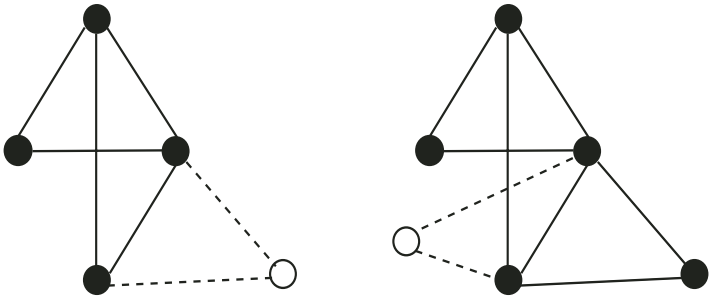

FIGURE 5.4. Two steps in the construction of a Growing network with preferential attachment.

these new-comers establish their connectionse preferentially with vertices that already have a large degree (*rich-get-richer*). Of course this latter rule is in the spirit of the St. Matthew effect described in Section 4.5. Actually not only in the spirit, but also in the formulation and in the onset of the scale-invariance this model is in the spirit of the Yule process (Yule, 1925) described in the same section.

Growth and Preferential Attachment are specifically suited to model the Internet and the World Wide Web, two networks that in a relativley short timespan (roughly fifteen years) have seen a huge growth of their elements. Furthermore new routers (for the Internet) and new webpages (for the WWW) tend to connect with authoritative pre-existing routers and webpages, where authoritativness is based on the consensus and can be weighted by the number of connections.

- The **degree distribution** is scale invariant for a precise choice of the preferential attachment rule, otherwise the degree is distributed according to a stretched exponential function.

- As regards the **diameter** D of Barabási-Albert networks an analytical computation (Bollobás and Riordan, 2004) shows that $D \propto log(n)/log(log(n))$.
- There is not extimate for the **clustering coefficient**

5.3.1 *Equations for the continuous degree*

The mathematical derivation of the degree distribution is easily obtained describing the degree as a continous variable. If new vertices enter the network at a constant rate, and each makes m connections to pre-existing nodes, the corresponding equation for the degree of vertex i is

$$\frac{dk_i}{dt} = m\frac{f(k_i)}{\sum_j f(k_j)} \tag{5.15}$$

where $\frac{f(k_i)}{\sum_j f(k_j)}$ is the probability that a new node entering the network at time t choses vertex i as a connection partner for one of its m edges. The $f(k_i)$ is a growing function of k_i in order to reproduce the preferential attachment rule. eqn 5.15 can be solved by giving an explicit expression for $f(k_i)$. The simplest form is $f(k_i) = k_i^\alpha$ with $\alpha \geq 0$. eqn 5.15 then becomes

$$\frac{dk_i}{dt} = m\frac{k_i^\alpha}{\sum_j k_j^\alpha} \tag{5.16}$$

The above equation can be simplified by observing that the denominator on its *right hand side (rhs)* grows linearly with t for any $\alpha \in [0, 1]$.

- If $\alpha = 0$ we have $\sum_j k_j^0 = \sum_j 1 = m_0 + t$ where m_0 is the number of vertices of the network at time $t = 0$. The coefficient of time is unitary because time can always be rescaled so that one new node enters the network per unit time.
- If $\alpha = 1$ we have $\sum_j k_j(t) = A + 2mt$ where $A = \sum_j^{m_0} k_j(0)$ is the sum of the degrees at time $t = 0$ of the vertices of present in the network at time $t = 0$; the $2mt$ counts the degrees of all new vertices because every new vertex adds m edges to the network, each of which accounts for an increase of two of the total degree.
- Given that $\sum_j k_j^\alpha$ is a growing function of both t and α, and that it is linearly bounded in t for both $\alpha = 0$ and $\alpha = 1$, it follows that $\sum_j k_j^\alpha = A(\alpha) + \mu(\alpha)t$ for $0 \leq \alpha \leq 1$.

Equation 5.16 can then be written

$$\frac{dk_i}{dt} = m\frac{k_i^\alpha}{A(\alpha) + \mu(\alpha)t} \tag{5.17}$$

whose solution is

$$k_i(t) = \left[m^{1-\alpha} + \frac{(1-\alpha)m}{\mu(\alpha)} \ln\left(\frac{A(\alpha) + \mu(\alpha)t}{A(\alpha) + \mu(\alpha)t_i}\right)\right]^{\frac{1}{1-\alpha}} \tag{5.18}$$

which has been obtained with the initial condition $k_i(t_i) = m$ where t_i is the time at which vertex i has entered the network. Degree k_i can be therefore labelled by means of its entry time t_i rather than by the vertex label i. We can then write $k_i(t) = k(t, t_i)$. Importantly, eqn 5.18 holds only for $0 \leq \alpha < 1$. The distribution of degrees, $P(k)$ can be obtained from eqn 5.18 by means of

$$P(k) = \frac{1}{m_0 + t} \int_0^t \delta(k - k(t, t_i)) dt_i = \frac{1}{t + m_0} \left(\frac{\partial k(t, t_i)}{\partial t_i} \right)^{-1} \Bigg|_{t_i = t_i(k,t)} \tag{5.19}$$

where $t_i(k, t)$ is the solution of the implicit equation $k = k(t, t_i)$. Using eqn 5.18 and eqn 5.19 we have

$$P(k, t) = \frac{A(\alpha) + \mu(\alpha)}{m_0 + t} \frac{1}{m} k^{-\alpha} exp \left\{ - \frac{\mu(\alpha)}{(1 - \alpha)m} \left[k^{-(1-\alpha)} - m^{-(1-\alpha)} \right] \right\} \tag{5.20}$$

which in the $t \to \infty$ limit reaches a stationary state

$$P(k) = \frac{\mu(\alpha)}{m} e^{\frac{\mu(\alpha)}{(1-\alpha)m^{\alpha}}} k^{-\alpha} e^{-\frac{\mu(\alpha)}{(1-\alpha)m} k^{1-\alpha}} \tag{5.21}$$

which in turn reduces to a simple exponential for $\alpha = 0$. Solution in eqn 5.21 is completely determined by

$$\mu(\alpha) = \frac{1}{t} \sum_i k_i^{\alpha}(t) = \frac{1}{t} \int_0^t k^{\alpha}(t, t_i) dt_i \tag{5.22}$$

which becomes, in the infinite time limit, $\mu(\alpha) = \int_m^{\infty} k^{\alpha} P(k) dk$, that is a self-consistent equation for $\mu(\alpha)$, by using eqn 5.21.

Equation 5.21 predicts a stretched exponential distribution for any $\alpha < 1$. In the limit $\alpha \to 1$ eqn 5.21) becomes $P(k) \sim k^{-3}$. Indeed a full solution of eqn 5.17 for $\alpha = 1$ gives

$$k(t, t_i) = m \left(\frac{A(1) + 2mt}{A(1) + 2mt_i} \right)^{\frac{1}{2}} \tag{5.23}$$

from which it can be found, by means of eqn 5.19, that $P(k) = 2m^2 k^{-3}$.

A linear preferential attachment rule ($\alpha = 1$) is therefore the necessary ingredient to find scale-free networks in the Barabasi-Albert model. At first sight the exponent $\gamma = -3$ is not satisfactory, since most networks exhibit exponents $2 \leq \gamma < 3$ and in some cases even $\gamma > 3$. Yet the good news are that the γ exponent of the Barabasi-Albert model is actually sensitive to the details of the preferential attachment rule. If for example $f(k) = a + k$, with a a constant, the resulting exponent is $\gamma = 3 + a/m$, showing that any exponent $\gamma > 2$ can actually be obtained. The Barabasi-Albert model endowed with a linear preferential attachment rule defines therefore a class of scale-free networks of *tunable* degree distribution exponent.

One last comment is necessary for the case $\alpha > 1$. In this case the solution strategy used above is not amenable and it is necessary to use the complete master equation for the probability distribution $P(k,t)$. It is just important to remark here that for $\alpha > 2$ a *winner-take-all* situation emerges, where almost all the vertices of the network are connected to single vertex.

The preferential attachment rule, although simple and nice, poses some problems: new nodes entering the network need to have a complete knowledge of the network in order to make their connection choices. This is unlikely to be realistic, casting a shadow on a direct interpretation of the PA rule. In the next section a different model will be presented, likely to capture some real process behind the formation of the WWW and of protein interaction networks, in which preferential attachment emerges as an effective law from microscopic rules where only a local knowledge of the network is needed.

5.4 Copying or Duplication-divergence Models

Consider as a particular case of study that of the World Wide Web. If you want to add your webpage to the graph (i.e. add a vertex and some links to the graph) you are very likely to take as template one existing page therefore keeping all the list of existing hyperlinks (Aiello, Chung and Lu, 2000). This copying procedure is then followed by some specific change of the page to meet the interests of the owner of the new page. Consequently some old hyperlinks can be lost and some new ones acquired.

In the completely different context of protein interaction networks (Vazquez, Flammini, Maritan and Vespignani, 2003), the same mechanism is in agreement with the current view of genome evolution. When organisms reproduce, the duplication of their DNA is accompanied by mutations. Those mutations can sometimes entail a complete duplication of a gene. Since in this case the corresponding protein can be produced by two different copies of the same gene, point-like mutations on one of them can accumulate at a rate faster than normal since a weaker selection pressure is applied. Consequently, proteins with new, properties can arise by this process. The new proteins arising by this mechanism share many *physico-chemical* properties with their ancestors. Many interactions remain unchanged, some are lost and some are acquired.

This growth process works by replicating (with some tolerance) nodes and relative edges already present in the graph. The duplication-divergence model is governed by a local rule, where no global knowledge of the network is needed.

A possible formulation of the model is the following. At every time-step a randomly chosen vertex is duplicated at random. Each of its m out-going connections is either kept with probability $1-\alpha$ or it is rewired with probability α. The rate of change of the in-degree of a node is then given by

$$\frac{\partial k_{in,i}(t)}{\partial t} = (1-\alpha)\frac{k_{in,i}(t)}{N} + m\frac{\alpha}{N} \tag{5.24}$$

where the first term on the *r.h.s* of (5.24) is the probability that a vertex pointing to vertex i is duplicated and its link toward i retained. The second term on the *r.h.s* represents the probability that the duplicated vertex points toward i by one of its rewired out-going edges. For linearly growing networks we have that $N \simeq t$. The solution of Eq.5.24 is

$$k_{in,i}(t) = \frac{m\alpha}{1-\alpha}\left[\left(\frac{t}{t_i}\right)^{1-\alpha} - 1\right] \tag{5.25}$$

where t_i is the time when vertex i has entered the network and $k_{in,i}(t_i) = 0$ is the initial condition used to solve Eq.5.24. From Eq.5.25 it is possible to finally show that $P(k_{in}) \sim [k_i n + m\alpha/(1-\alpha)]^{-(2-\alpha)/(1-\alpha)}$.

- Extensive numerical simulations of such models have shown that the resulting **degree distribution** is scale-free. The onset of such scale-invariance is related to the preferential attachment mechanism of the Barabasi-Albert model. This can be seen by considering that every time-step a vertex chosen at random and its connections are duplicated. Since any vertex can be chosen, the probability to be a neighbor of a vertex of degree k is k/N, where N is the number of nodes in the network. Therefore the probability that a vertex increases its degree (by a unit in a time-step) is proportional to the degree itself. The preferential attachment rule emerges at an effective level from local principles.
  The case of undirected networks is more complex than the one of directed ones. Indeed, in Eq.5.24 the first term on the *r.h.s.* does not change, whereas the second term depends on the degree of the duplicated vertex. The resulting networks are always characterized by heavy-tail degree distributions, although not necessarily strictly scale-free.

5.5 Fitness Model

Although in some contexts preferential attachment can be a very reasonable assumption, in many others it is certainly not. In particular, in some situations, the information about the degree of each and every single vertex is not available to newly added sites, neither in a direct nor in an effective way. Furthermore in the case of many social interactions this information whenever available is likely not to play an active role in the link connection.

Instead, it is reasonable that two vertices are connected when the link creates a mutual benefit depending on some of their intrinsic properties (authoritativeness, friendship, social success, scientific relevance, interaction strength, etc). Therefore, it is reasonable to expect that for some of these systems the $P(k)$ scale free behavior (when existing) could have an origin unrelated to preferential attachment.

In order to explore this simple idea, Caldarelli et al. (Caldarelli, Capocci, De Los Rios and Muñoz, 2002) proposed the following network-building algorithm:

- Start by creating a total (large) number N of vertices. At every vertex i a fitness x_i, which is a real number measuring its importance or rank, is assigned. Fitnesses are random numbers taken from a given probability distribution $\rho(x)$.
- For every couple of vertices, i, j, a link is drawn with a probability $f(x_i, x_j)$ (f a symmetric function of its arguments) depending on the "importance" of both vertices, *i.e.* on x_i, x_j.

It is clear from that definition that a trivial realization of the above rules is the standard Erdős Rényi model (Erdős and Rényi, 1961). In this case the $f(x_i, x_j)$ is constant and equal to p for all vertex couples. While this particular choice does not produce SF networks, as soon as random fitness are introduced this triggers the onset of scale invariance.

Another feature of the model is that it can be defined as static as well as dynamic. INdeed either the size of the graph is fixed, or by adding new vertices at every time step we can link them to the old ones according to the above attaching rule.

A general expression for $P(k)$ can be easily derived. Indeed, the mean degree of a vertex of fitness x is simply

$$k(x) = N \int_0^\infty f(x, y)\rho(y)dy = NF(x) \tag{5.26}$$

(with $x_i \in (0, \infty)$). Assuming $F(x)$ to be a monotonous function of x, and for large enough N, we have the simple relation

$$P(k) = \rho \left[F^{-1}\left(\frac{k}{N}\right) \right] \frac{d}{dk} F^{-1}\left(\frac{k}{N}\right). \tag{5.27}$$

For finite values of N corrections to this equation emerge (**?**). As a particular example, consider $f(x_i, x_j) = (x_i x_j)/x_M^2$ where x_M is the largest value of x in the network. Then

$$k(x) = \frac{Nx}{x_M^2} \int_0^\infty y\rho(y)dy = N\frac{< x > x}{x_M^2} \tag{5.28}$$

and we have the simple relation

$$P(k) = \frac{x_M^2}{N < x >} \rho\left(\frac{x_M^2}{N < x >} k\right). \tag{5.29}$$

A particularly simple realization of the model emerges if we consider power-law distributed fitnesses. This choice can be naturally justified by arguing that power-laws appear rather generically in many contexts when one ranks, for example, people according to their incomes or cities according to their population, etc. This is the so-called Zipf law which establishes that the rank $R(x)$ behaves as $R(x) \propto x^{-\alpha}$ in a quite universal fashion (Zipf, 1949). The reason for the

ubiquitous presence of the Zipf law yields on the multiplicative nature of the intrinsic fluctuations which generically leads to flat distributions in logarithmic space and, consequently, to power-laws (Zipf, 1949).

Clearly, if $\rho(x) \sim x^{-\beta}$ (Zipf's behavior, with $\beta = 1 + 1/\alpha$ (Zipf, 1949)) then, using eq.(5.29), also the degree distribution $P(k)$ is a power-law and the network shows SF behavior. In Fig.**??** we show the degree distributions from simulations with $\beta = 2.5, 3, 4$ (corresponding to Zipf exponents $\alpha = 2/3, 1/2, 1/3$); the asymptotic behavior is, in all cases, well described by eq.(5.29). This result is hardly surprising: from SF fitnesses we generate SF networks, but still it provides a new generic path to SF networks and takes into account the widespread occurrence of the Zipf's behavior in nature. In order to extend this result and check whether SF networks can be generated even when $\rho(x)$ is not SF itself, we consider an exponential distribution of fitnesses, $\rho(x) = e^{-x}$ (representing a random, Poisson distribution) and $f(x_i, x_j) = \theta(x_i + x_j - z)$, where $\theta(x)$ is the usual Heaviside step function. This represents processes where two vertices are linked only if the sum of their fitnesses is larger than a given *threshold* z. Using these rules we obtain analytically (and confirm in computer simulations) that $P(k) \sim k^{-2}$ (**?**). This leads to the non-trivial result that *even non scale-free fitness distributions can generate scale-free networks* (see Fig.**??**). Also different implementations of the threshold rule, such as $f(x_i, x_j) = \theta(x_i^n + x_j^n - z^n)$ (where n is an integer number) give rise to the same inverse square behavior (although, in some cases, with logarithmic corrections).

In a future publication we will explore, in a more systematic way, the necessary and sufficient conditions for the fitness distribution and attaching rule under which well-behaved SF networks are generated.

Let us stress that the model, as defined, has a diverging average connectivity in the large N limit, as can be easily inferred from Eq.(5.26); *i.e.* it is severely *accelerated* (**?**). Nevertheless we can introduce in a rather natural way an upper cut-off accounting for the fact that every site has a limited information on the rest of the world and, therefore, connection is attempted with a finite number, m, of different sites. Alternatively, vertices can be linked with the above rule and, after that, links are kept with probability p (so that, for example, $pN = m$). By including this modification, the N factor in Eq.(5.26), is substituted by m, and the connectivity is finite in the thermodynamic limit. In order to generate different accelerated networks (with the averaged connectivity not reaching a stationary value but growing with N in different possible ways (**?**)) other selection rules can be easily implemented.

To have a more extensive picture of the nature of the networks under consideration, we have studied the following topological properties (**?**), interest in which has been triggered by recent studies on the Internet structure (**?**; **?**):

- The average distance $< d >$, measuring the average minimum number of arcs needed to connect two given sites.
- The *average neighbor connectivity* $k_{nn}(k)$, measuring the average degree of vertices neighbor of a k-degree vertex.

- The *clustering coefficient* $c(k)$ that measures the degree of interconnectivity of nearest neighbors of k-degree vertices. More specifically the clustering coefficient c_i of a vertex i, whose degree is k_i, is the ratio between the number of edges e_i in the subgraph identified by its neighbors and the maximum possible number of edges in the subgraph. That is $c_i = 2e_i/k_i(k_i - 1)$ (**?**). $c(k)$ is obtained by averaging c_i for all vertices with fixed degree k.
- The probability distribution of the *betweenness*, b_i, defined as the total number of minimum paths between any couple of vertices in the network passing through vertex i (**?**). This quantity gives a measure of the amount of traffic passing through a vertex. We studied, as in the aforementioned papers, both the probability distribution $P(b)$ and its first moment $< b > /N$.

Computer simulations of our model show that networks with power-law distributed fitnesses, and different values of β, show nearly constant $k_{nn}(k)$'s and $c(k)$'s, just as occurs for the original BA model (**?**). The distribution of betweenness decays as a power law with an exponent $\gamma_b \approx 2.2$ for $\gamma = 2.5$ and $\gamma = 3$, and $\gamma_b \approx 2.6$ for $\gamma = 4$. This is in good agreement with what conjectured in Ref. (**?**): all networks with $3 \geq \gamma > 2$ can be classified in only two groups according to the value of γ_b ($\gamma_b = 2$ and $\gamma_b = 2.2$, respectively), while for larger values of γ, larger non-universal values of γ_b are reported.

The exponential case behaves in a different way: for a network of size $N = 10^4$, $z = 10$, and $m = N$ we find $< d >= 2$, $< c > \simeq 0.1$ and $< b > /N \simeq 0.1$, but a power-law behavior is found for the clustering magnitudes, *i.e.* $< k_{nn} > \propto k^{-0.85}$ and $c(k) \propto k^{-1.6}$. The betweenness distribution instead, shows an unexpected behavior, giving a power-law tail with an exponent $\gamma_b \approx 1.45$ (see Fig.4). It is worth remarking that our model having $\gamma = 2$ is not included in the previously discussed classification of betweenness exponents (**?**).

Having explored the most basic properties of the model and some particular realizations, let us comment now on possible applications.

E-mail networks (**?**) are a good candidate to be represented by our model. In this case growth may occur, but agents (e-mail senders) do not have any access or knowledge of the degree of the receivers. Rather than preferential attachment there should be some intrinsic feature of the receiver playing a role in the phenomenon.

To further emphasize the utility of this new mechanism let us mention the following possibility: one can imagine situations where a Poisson network is seen as SF just because the exploration method implicitly implements a probabilistic rule depending on the fitnesses (this applies for example when links are detected by "picking" them one by one, but not if the network is explored by crawling on it). Let us think, for example, of the case with threshold type of attaching rule. If only links with corresponding fitnesses over threshold are "seen" by the exploration method then, for example, an Erdős Rényi network with exponentially distributed fitnesses can be seen as SF (with, obviously, a connectivity upper cutt-off related to the maximum connectivity of the underlying network; in cases

in which this connectivity is high, one can generate hubs in the "apparent" SF network). In particular this scenario could be of relevance to protein networks. Let us argue why.

The way comprehensive protein networks have been obtained to date is through a bait-prey method, named "two-hybrid" method: two proteins are hybridized with two fragments of a transcription factor (a protein that binding to a gene promotes its transcription into the corresponding RNA). The spliced promoter does not bind to the gene, transcription is inhibited and the corresponding RNA is absent. Yet, if the two proteins interact they bring together the two promoter fragments allowing it to bind to the gene and transcription to start. The presence of the corresponding RNA signals the interaction between the two proteins. We can imagine that the interaction strength between the two proteins has to be above a given threshold, else the typical promoter binding time will be too short for the RNA polymerase to bind to the gene and initiate transcription. In turn it is reasonable to assume that the interaction strength is a function of some properties of the two proteins (such as, for example, their hydrophobicity, or their Accessible Surface Area). This possibility has still to be checked through an analysis of the detailed physics behind the two-hybrid method.

In summary, we have presented an alternative model to justify the ubiquity of SF networks in nature.

5.6 Networks from degree sequence

A definite degree distribution can be the result of some construction rules, as showed above. A possible different approach is to use the degree distribution as a starting ingredient, and build a network according to it. Once fixed the number n nodes, we can draw a degree sequence from a given distribution $P(k)$. The edges are then added by joining pairs of nodes at random until all degrees are satisfied. Of course there are very strict conditions to be considered in the construction procedure. Firstly, the sum of the degree extracted must be an even number, since it must give twice the number of edges. Secondly, if self and double edges are forbidden[30], not all random sequences produce acceptable networks. Therefore the procedure must be repeated until all the constraints are fulfilled. We have to note that for scale free networks with exponents $\gamma < 3$ is almost impossible to produce a network avoiding self and double edges.

The procedure outlined above, although very effective, has the drawback that the it does not allow to change the size of the network during its construction. Any new vertex at the end finds no free vertices to share edges. Therefore the final state of the network is locked. Note also that in principle also in a random graph it is possible to add new nodes at any time, connecting them to already existing ones with probability p.

[30]That is to say we are not considering multigraphs