Chapter 2

Structural properties of networks

2.1 Introduction

Complex networks are usually the outcome of a stochastic process. Nevertheless complex network are not completely random, because they are self-organized to perform specific tasks. The structural properties of complex networks have very important effects on the dynamics that can be defined on them. Therefore, a general paradigm of complex network theory is that the function of a network is reflected and affected by its structural properties. For this reason one of the most fundamental roles of network theory is to define a series of properties of complex networks able to characterize their structure.

2.2 Network size and total number of links

The most fundamental structural properties of a network are the *network size* N indicating the total number of nodes in the network, and the total number of links L in the network.
The large majority of complex networks are formed by a sufficiently large number of nodes N linked by non regular interactions, that the characterization of these networks usually requires computational power.

2.2.1 The "minimal complex networks"

The number of genes of the "minimal cell" reconstructed in the laboratory of C. Venter includes $N = 256$ genes, and the smallest known neural network of the worm *c.elegans* includes $N = 302$ neurons. Already these " minimal networks" are sufficiently complex to perform incredible complex functions. The number

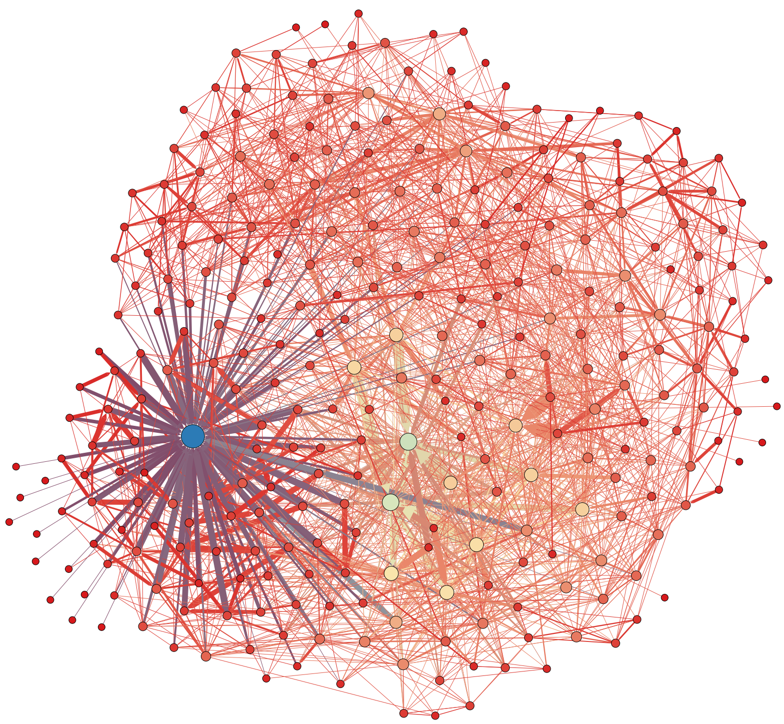

Figure 2.1: An example of “minimal complex network”: the *c.elegans* neural network with $N = 302$ nodes. The thickness of the links is proportional to the weights of the links, the size of the nodes is proportional to their degree. The figure does not show the direction of the links.

of genes in the human DNA is larger $N = 23,299$ but surprisingly small compared to expectations before the launch of the Genome Project.

2.2.2 Large Complex networks

Many other complex networks are significantly larger. For example the human brain is formed by $10^{10} - 10^{11}$ neurons or the online social networks have reached very large network sizes of $N \simeq 10^8$. Nevertheless these network sizes remain much smaller than the Avogadro Number $N_A \simeq 6 \times 10^{23}$ that indicates the total number of molecules in a mole of a substance.

In table 2.1 we indicate the order of magnitude of a series of complex networks.

Networks	**Network size N**
Brain	up to 10^{11}
Metabolic Networks	10^3
Social Networks	up to 10^9
Power-grids	up to 10^5
Internet	up to 10^5
WWW	10^9
Online social networks	10^8

Table 2.1: The network size of several complex networks

2.2.3 The total number of links L in the network

The total number of links in a network can be expressed in terms of the adjacency matrix $\mathbf{A}$. For a *undirected network* each link (i, j) with $i \neq j$ is represented by two matrix elements $A_{ij} = A_{ji} = 1$, while each tadpole incident to node i is represented by a single matrix element A_{ii}. Therefore we have

$$L = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} + \frac{1}{2} \sum_{i=1}^{N} A_{ii}, \tag{2.1}$$

where $\delta_{ij} = 1$ if $i = j$ and zero otherwise. This expression, in absence of tadpoles reduces to

$$L = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij}. \tag{2.2}$$

For a *directed network* each directed link between node j and node i is represented by a single matrix elements $A_{ij} = 1$, while each directed tadpole is represented by a single matrix element A_{ii}. Therefore we have

$$L = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij}. \tag{2.3}$$

2.3 Degrees, Degree Sequences and degree distributions

2.3.1 Degrees

The degrees are very fundamental local properties of the nodes.

Definition 23. *The* degree k_i of node i in an undirected network *is given by the total number of links incident to node i. In a directed network we distinguish between* in-degrees *and* out-degrees. *The* in-degree k_i^{in} of node i in a directed network *is given by the total number of nodes pointing to node i. The* out-degree k_i^{out} of node i in a directed network *is given by the total number of nodes to which node i points.*

For simplicity here we consider only unweighted networks. In this case the degree (or in-degree/out degree) of a node can be calculated directly from the adjacency matrix $\mathbf{A}$. Let us consider in the following the case of directed and undirected networks separately.

Undirected networks

The degree k_i of a generic node i in an undirected network is given by

$$k_i = \sum_{j=1}^{N} A_{ij} = \sum_{j=1}^{N} A_{ji}, \tag{2.4}$$

where we have use the fact that in this case the adjacency matrix A is symmetric, and the fact that every tadpole incident to node i increases its degree by 1. In a simple network of N nodes the maximal possible degree is $k = N - 1$, the minimal degree is $k = 0$.

Directed network

The in-degree k_i^{in} and the out-degree k_i^{out} of node i in a directed network is given by

$$\begin{aligned} k_i^{in} &= \sum_{j=1}^{N} A_{ij}, \\ k_i^{out} &= \sum_{j=1}^{N} A_{ji}, \end{aligned} \tag{2.5}$$

where here, the in-degree and the out-degree are in general different because the adjacency matrix is asymmetric. In a directed network of N nodes the maximal possible in-degree is $k^{in} = N - 1$, the minimal degree is $k^{in} = 0$,the maximal possible out-degree is $k^{in} = N - 1$, the minimal degree is $k^{in} = 0$.

2.3.2 Degree sequence, Average Degree, Maximum Degree

Definition 24. *The* degree sequence of an undirected network *is the ordered sequence* $\{k_i\} = \{k_1, k_2, \ldots, k_i, \ldots, k_N\}$ *of the degrees* k_i *of all the nodes of the network* ($i = 1, 2, \ldots N$)*.*
The in degree sequence of an directed network *is the ordered sequence* $\{k_i^{in}\} = \{k_1^{in}, k_2^{in}, \ldots, k_i^{in}, \ldots, k_N^{in}\}$ *of the in-degrees* k_i^{in} *of all the nodes of the network* ($i = 1, 2, \ldots, N$)*. The* out degree sequence of an directed network *is the ordered sequence* $\{k_i^{out}\} = (k_1^{out}, k_2^{out}, \ldots, k_i^{out}, \ldots, k_N^{out})$ *of the out-degrees* k_i^{out} *of all nodes of the network* ($i = 1, 2, \ldots, N$)*.*

Undirected network

Given the degree sequence of an undirected network, we can definite the average degree $\langle k \rangle$ of the network defined as

$$\langle k \rangle N = \sum_{i=1}^{N} k_i = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{N} A_{ji}. \tag{2.6}$$

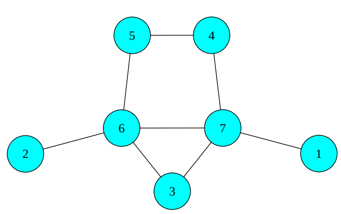

Figure 2.2: Undirected network of $N = 7$ nodes and $L = 8$ links

.

The average degree of a simple network is related to the total number of links in the network by the expression

$$L = \frac{1}{2} \langle k \rangle N. \tag{2.7}$$

The maximum degree of the networks will be indicated by K i.e.

$$K = \max_i k_i. \tag{2.8}$$

The degree sequence of the undirected network in Figure 2.2 is given by $\{1, 1, 2, 2, 2, 4, 4\}$. The average degree of this network is given by $\langle k \rangle = 16/7 \simeq 2.86$ and the total number of links L is given by $L = 8$.

Directed network

For a directed network the average in-degree $\langle k^{in}\rangle$ is equal to the average out-degree $\langle k^{out}\rangle$. In fact we have,

$$\langle k^{in}\rangle N = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N}\sum_{j=1}^{N} A_{ij} = \sum_{j=1}^{N} k_j^{out} = \langle k^{out}\rangle N. \tag{2.9}$$

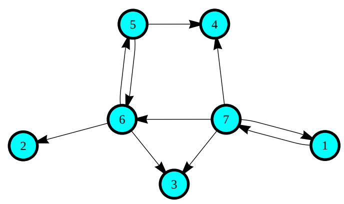

Figure 2.3: Directed network of $N = 7$ nodes and $L = 10$ links

.

The averaged in-degree and the average out-degree of the network are related to the total number of links in the network by the relation

$$L = \langle k^{in}\rangle N = \langle k^{out}\rangle N. \tag{2.10}$$

The maximum in-degree and maximum out-degree of the networks will be indicated by K^{in}, K^{out} respectively with i

$$\begin{aligned} K^{in} &= \max_i k_i^{in}, \\ K^{out} &= \max_i k_i^{out}. \end{aligned} \tag{2.11}$$

The in-degree sequence of the directed network in Figure 2.3 is given by $\{1, 1, 2, 2, 1, 2, 1\}$, while the out-degree sequence is given by $\{1, 0, 0, 0, 1, 2, 3\}$. The average in-degree and the average out-degree of this network is given by $\langle k^{in}\rangle = 10/7 \simeq 1.43$ and the total number of links L is given by $L = 10$.

2.3.3 The degree distribution of the network

The degree of a node is a local property of the network but by considering the degree sequence of the network we can characterize some important global property of the network. The global organizational structure induced by the degree sequence, is characterized by the *degree distribution* of the network.

Definition 25. *The* degree distribution $P(k)$ *of a undirected network is the fraction of nodes of degree* k*. It also indicates the probability that a randomly chosen node of the network has degree* k*.*
The in-degree distribution $P^{in}(k)$ *of a directed network if the fraction of nodes of in-degree* k*. It also indicates the probability that a randomly chosen node of the network has in-degree* k*.*
The out-degree distribution $P^{out}(k)$ *of a directed network is the fraction of nodes of out-degree* k*. It also indicates the probability that a randomly chosen node of the network has out-degree* k*.*

Undirected networks

Let us indicate with $N(k)$ is the total number of nodes of the network with degree k, i.e.

$$N(k)=\sum_{i=1}^{N}\delta(k,k_i), \tag{2.12}$$

where $\delta(k,k_i)$ indicates the Kronecker delta, i.e. $\delta(k,k_i)=1$ if $k=k_i$ and $\delta(k,k_i)=0$ otherwise. The degree distribution of an undirected network is given by $P(k)$ given by

$$P(k)=\frac{1}{N}N(k)=\frac{1}{N}\sum_{i=1}^{N}\delta(k,k_i). \tag{2.13}$$

The degree distribution non-negative $P(k)\geq 0\ \forall k$, and normalized

$$\sum_{k=0}^{K}P(k)=1. \tag{2.14}$$

Starting from a given degree sequence the calculation of the degree distribution is therefore very simple. For example, starting from the degree sequence of the undirected network in Figure 2.2, i.e. $\{1,1,2,2,2,4,4\}$ we can evaluate the degree distribution $P(0)=0, P(1)=2/7, P(2)=3/7, P(3)=0, P(4)=2/7$ and $P(k)=0$ for $k>4$.

Directed networks

Let us indicate with $N^{in/out}(k)$ is the total number of nodes of the network with in/out-degree k, i.e.

$$\begin{aligned} N^{in}(k) &= \sum_{i=1}^{N}\delta(k,k_i^{in}),\\ N^{out}(k) &= \sum_{i=1}^{N}\delta(k,k_i^{out}), \end{aligned} \tag{2.15}$$

where $\delta(k,k_i)$ indicates the Kronecker delta. The in/out-degree distribution of an directed network is given by $P^{in/out}(k)$

$$\begin{aligned} P^{in}(k) &= \frac{1}{N}N^{in}(k)=\frac{1}{N}\sum_{i=1}^{N}\delta(k,k_i^{in})\\ P^{out}(k) &= \frac{1}{N}N^{out}(k)=\frac{1}{N}\sum_{i=1}^{N}\delta(k,k_i^{out}). \end{aligned} \tag{2.16}$$

The in/out-degree distributions are non negative $P^{in/out}(k) \geq 0\ \forall k$ and normalized, i.e.

$$\sum_{k=0}^{K^{in}} P^{in}(k) = 1,$$
$$\sum_{k=0}^{K^{out}} P^{out}(k) = 1. \tag{2.17}$$

Starting from a given in/out degree sequence the calculation of the in/out degree distribution is therefore very simple. For example the in-distribution of the directed network in Figure 2.3 with in-degree sequence $\{1, 1, 2, 2, 1, 2, 1\}$ is given by $P^{in}(0) = 0, P^{in}(1) = 4/7, P^{in}(2) = 3/7$ and $P^{in}(k) = 0$ for $k > 2$. The out-degree distribution of the same network can be calculated starting from the out-degree sequence $\{1, 0, 0, 0, 1, 2, 3\}$ and is given by $P^{out}(0) = 3/7, P^{out}(1) = 2/7, P^{out}(2) = 1/7, P^{out}(3) = 1/7$ and $P^{out}(k) = 0$ for $k > 3$. The degree distribution of complex networks have large impact on their robustness properties under random failure or targeted attacks and on the behaviour of dynamical processes defined on them. Moreover statistical properties of the degree distribution can change also the local properties of the networks such as the number of subgraphs such as loops of cliques find in the networks. The different classes of degree distributions will be discussed in Chapter 4.

2.4 Paths

Networks can used to search and navigate complex systems and in general to transmit information. For example, when we "browse the Internet" we follow paths on the World-Wide-Web, when we take a connecting flight we explore paths in the airport network, when we discover that two of our friends are already friends essentially we discover a path in our social network.

Definition 26. *A* path *of a network, is a sequence of nodes, such that every consecutive pair of nodes is connected by a link. A* directed path *of a directed network, is a path, with the links being directed from each node to the following one.*

Each path, either directed or undirected has its *path length.*

Definition 27. *The* path length *is equal to the number of links traversed along the path, including eventual repetitions in the case of paths that intersect themselves.*

Finite paths have an initial node and a final node. Eventually paths can come back to the starting node. In this case we say that the path is a cyclic path.

Definition 28. *A path that starts from a node and finish on the same node is called a* cyclic path*, paths that start and finish on different nodes are called* acyclic*.*

Acyclic paths that do not visit any node more than once are called *self-avoiding paths*. Cyclic paths that do not visit any node different from the starting node more than once are called *self-avoiding cyclic paths*.

Undirected networks

In Figure 2.4 we show an undirected network. In the following we describe different paths on this network

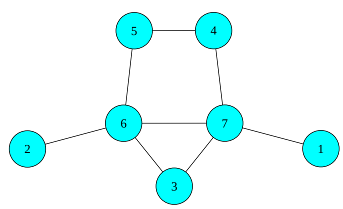

Figure 2.4: Undirected network of $N = 7$ nodes and $L = 8$ links
.

$$
\begin{aligned}
\textbf{Path } \mathcal{P}_1 &= (2,6,5,4) \\
\textbf{Path } \mathcal{P}_2 &= (2,6,7,4) \\
\textbf{Path } \mathcal{P}_3 &= (2,6,3,7,1) \\
\textbf{Path } \mathcal{P}_4 &= (2,6,7,1) \\
\textbf{Path } \mathcal{P}_5 &= (2,6,7,3,6,7,1) \\
\textbf{Path } \mathcal{P}_6 &= (6,7,3,6)
\end{aligned}
$$

Both paths $\mathcal{P}_1$ and $\mathcal{P}_2$ have initial node $i = 2$ and final node $j = 4$. Moreover both paths have the same length $\ell = 3$. The three paths $\mathcal{P}_3, \mathcal{P}_4, \mathcal{P}_5$ have initial node $i = 2$ and final node $j = 1$ but they have different lengths given by $4, 3, 6$ respectively. Finally the path $\mathcal{P}_6$ is a cyclic path of length 3. All the listed paths except the path $\mathcal{P}_5$ are self-avoiding paths.

Directed Networks

In Figure 2.5 we show an directed network. In the following we describe different directed paths on this network

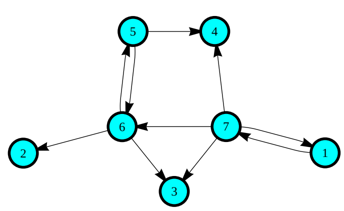

Figure 2.5: Undirected network of $N = 7$ nodes and $L = 10$ links
.

$$
\begin{aligned}
\textbf{Path } \mathcal{P}_a &= (6,5,6) \\
\textbf{Path } \mathcal{P}_b &= (6,5,6,5) \\
\textbf{Path } \mathcal{P}_c &= (1,7,6,2) \\
\textbf{Path } \mathcal{P}_d &= (7,4) \\
\textbf{Path } \mathcal{P}_e &= (7,1,7) \\
\textbf{Path } \mathcal{P}_f &= (7,1,7,3).
\end{aligned}
$$

The directed paths $\mathcal{P}_a$ and $\mathcal{P}_e$ are self-avoiding cyclic paths, while the directed path $\mathcal{P}_b$ is a non-self-avoiding cyclic path.

2.4.1 Number of paths between two nodes

The number of paths (directed path in a directed network) of length n joining two nodes j and i in a given network, can be expressed in terms of the adjacency matrix $\mathbf{A}$ of the network.

In particular here we want to prove the following theorem:

Proposition 1. *In a unweighted network, the number of paths of length n joining node j to node i is given by*

$$\mathcal{N}_{ij}^{n} = [\mathbf{A}^n]_{ij} \tag{2.18}$$

where $[\mathbf{A}^n]_{ij}$ indicates the matrix element i, j of the matrix $\mathbf{A}^n$.

Proof. The theorem is true for path of length $n = 1$. In fact the number of paths of length $n = 1$ between two given nodes can be either 1 or 0. Moreover the matrix element $[\mathbf{A}]_{ij} = 1$ if there is a path between node j and node i and zero otherwise by definition. Therefore the theorem is true for $n = 1$.
Let us now show that the theorem is also true for $n = 2$. The product $A_{i,r}A_{r,j} = 1$ if and only if both $A_{ir} = 1$ and $A_{rj} = 1$, i.e. if an only if there is a path (j, r, i) of length $n = 2$ joining node j to node i. If the there is no path j, r, i, then $A_{ir}A_{rj} = 0$. Calculating the number of paths of length $n = 2$ in the network means performing the sum of $A_{ir}A_{rj}$ over all possible intermediate nodes r. Therefore we have

$$\mathcal{N}_{ij}^{2} = \sum_{r=1}^{N} A_{ir}A_{rj} = [\mathbf{A}^2]_{ij}. \tag{2.19}$$

Therefore the theorem is true also for $n = 2$. We can generalize this argument to path of a generic length n between node j and node i. Such paths are of the form $(j, r_1, r_2, \ldots, r_{n-1}, i)$. The product $A_{ir_1}A_{r_1r_2}\ldots A_{r_{n-2}r_{n-1}}A_{r_{n-1},i} = 1$ if and only if the path $(j, r_1, r_2, \ldots, r_{n-1}, i)$ exist, otherwise the product is zero. Calculating the number of paths of length n in the network means performing the sum of $A_{ir_1}A_{r_1r_2}\ldots A_{r_{n-2}r_{n-1}}A_{r_{n-1},i}$ over all possible intermediate nodes $r_1, r_2, \ldots, r_{n-1}$. Therefore we have

$$\mathcal{N}_{ij}^{n} = \sum_{r_1=1}^{N}\sum_{r_2=1}^{N}\ldots\sum_{r_{n-1}=1}^{N} A_{ir_1}A_{r_1r_2}\ldots A_{r_{n-2}r_{n-1}}A_{r_{n-1},i} = [\mathbf{A}^n]_{ij}. \tag{2.20}$$

Therefore the theorem is valid for paths of any length n. □

From this theorem, used in the case in which the starting node and the ending node of the path is the same, we get that the number $\mathcal{N}_{ii}^{n}$ of *cyclic paths* of length n starting from node i and coming back to i, are given by

$$\mathcal{N}_{ii}^{n} = [\mathbf{A}^n]_{ii} \tag{2.21}$$

where $[\mathbf{A}^n]_{ii}$ indicates the matrix element i, i of the matrix $\mathbf{A}^n$. Finally, the total number of cyclic paths of length n in a network of adjacency matrix $\mathbf{A}$ is given by

$$\sum_{i=1}^{N} \mathcal{N}_{ii}^{n} = \mathrm{Tr}\mathbf{A}^n. \tag{2.22}$$

In figure 2.6 we show an undirected network of $N = 4$ containing cyclic paths.

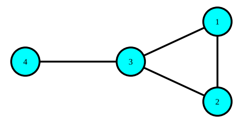

Figure 2.6: An undirected network of $N = 4$ nodes containing cyclic paths.

The adjacency matrix $\mathbf{A}$ of the network is

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

The first powers of this matrix are

$$\mathbf{A}^2 = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 1 & 1 & 0 & 1 \end{pmatrix}, \ \mathbf{A}^3 = \begin{pmatrix} 2 & 3 & 4 & 1 \\ 3 & 2 & 4 & 1 \\ 4 & 4 & 2 & 3 \\ 1 & 1 & 3 & 0 \end{pmatrix}.$$

Therefore the number of cyclic paths of length $n = 3$ starting and ending on node $i = 1, 2, 3, 4$, is given by $\mathcal{N}_{11}^3 = \mathcal{N}_{22}^3 = \mathcal{N}_{33}^3 = 2$ and $\mathcal{N}_{44}^3 = 0$.

2.4.2 Euleurian and Hamiltonian Cycles

In networks there are some special types of cyclic paths, the *Eulerian and Hamiltonian cycle* of the network. As we mentioned in chapter 1 the existence of an Eulerian cycle in the network formed by the seven bridges of Könisberg, the mainland and the two island on the Pregel River, was the original problem solved by Euler and signing the start date of graph theory.

Definition 29. *An* Eulerian cycle *of a network is a cyclic path that traverse each link of the network exactly once.*

The following theorem was first proven by Euler (in particularly he stated the theorem and he proven the necessary condition).

Theorem 2.4.1. *An undirected network has an* Eulerian cycle *if and only if all its nodes have even degrees and each pair of its non-zero-degree nodes can be connected by at least one path (i.e. they belong to a single connected component).*

Proof. Here we will prove only the necessary condition that is very easy to prove. In fact, if there is a Eulerian cycle in the network, the Eulerian cycle will visit every non-zero degree node of the network at least one time. If the Eulerian path visit a node reaching it from a link, it should be able to leave the node following another link not yet traversed by the cyclic path. Since the Eulerian cycle must visit all the links, it follows that if a network has an Eulerian cycle, the degree of every node must be necessarily even.

□

The seven bridges of Könisberg cannot be traversed exactly once in a single path. In fact the problem can be mapped to the problem of finding an Eulerian path in a networks with nodes of odd degrees, as the Figure 2.15 shows. Another fascinating combinatorial problem on network is relating to finding the

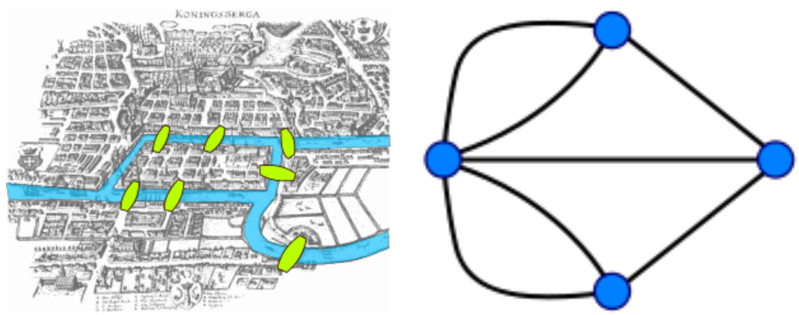

Figure 2.7: The city of Königsberg is shown on the left panel . The problem of the seven bridges of Könisberg (the Eulerian path problem) was solved by L. Euler in 1736 by mapping the mapping the problem in to the problem of finding an Eulerian path in the network shown in the right panel.

Hamiltonian cycle in a network. Assume that you have to organise a diplomatic dinner around a round table. Your goal is to make the dinner a success, so you want to place close to each other only diplomats of countries with friendly and peaceful relations. If you consider the network of friendly and peaceful relation, this problem reduces to the problem of finding a Hamiltonian cycle in this network. In fact the Hamiltonian cycle of a network is defined as follows.

Definition 30. *A* Hamiltonian cycle *is a cyclic path that visit each node of a network exactly once.*

Determining weather such paths exist in a given network is the Hamiltonian path problem, which is hard combinatorial problem (NP-complete).

2.5 Distances, Mean Average distance and Diameter of a network

The concept of distance in a network does not depend on an embedding space, but only on the shortest length of the paths connecting them. For example, the shortest distance between the two small cities of Trieste in the North-Est of Italy and of Baden-Baden in Germany in the airport network is larger than the distance between the city of Trieste with the city of London, although Trieste and Baden-Baden are closer in space. Many complex networks are characterized by small shortest distance between the nodes. For example in social networks any two people in the Earth are separated by only few shaken hands, or in the World-Wide-Web any pair of webpages are only few clicks apart despite these networks contain more than 10^8 nodes. Here we introduce the terms necessary to quantify these important properties of complex networks.

2.5.1 Shortest distance between two points

Given two nodes i and j of the network first we define their shortest path and their shortest distance.

Definition 31. *A* shortest path *between node j and node i is a path (directed path in the case of a directed network) of minimum length. The* shortest distance *d_{ij} between node j and node i is the length of any shortest path between node j and node i.*

If node there is no path between node j and node i we set $d_{ij} = \infty$.

2.5.2 Average distance and Diameter of a Network

The average distance of a network and its diameter are global quantities that characterize important properties of the distances in the network. Let us limit our discussion to connected networks, i.e. network for which there is a path from every node of the network to any other node.

Definition 32. *The* average shortest distance *ℓ of a connected network is the average of the shortest distances between any two distinct nodes of the network. Therefore, in a connected network we have*

$$\ell = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1,N|j\neq i} d_{ij}. \tag{2.23}$$

Definition 33. *The* diameter *D of a connected network is the maximum of the shortest distances between any two nodes of the network. Therefore we have*

$$D = \max_{i,j\neq i} d_{ij}. \tag{2.24}$$

From the definitions of the average shortest distance and of the diameter of a connected network, it clearly follows that

$$D \geq \ell, \tag{2.25}$$

i.e. the diameter of a connected network is never smaller than the average distance of the network.
As a useful reference point we can consider lattices, that are regular symmetric structures widely studied in physics or whenever it is necessary to approximate a continuous Euclidean space with a network. In Figure 2.8 we show two examples of lattices of dimension respectively one and two. For the 1d chain, the diameter $D = N-1$, for the 2d finite lattices of $N = l \times l$ nodes (in the figure we have the example of $l = 6$ nodes) the diameter is given by $D = 2(l-1) = 2(\sqrt{N}-1) \simeq 2N^{1/2}$ where the last relation is valid for $N \gg 1$. This result can be easily generalized for large lattices $N \gg 1$ of dimension d giving

$$D \simeq N^{1/d}. \tag{2.26}$$

Instead, as we will see in the following chapters ,many complex networks are small world, i.e. they are characterized by a diameter D scaling with the number of nodes as

$$D \simeq \mathcal{O}(\ln N), \tag{2.27}$$

or

$$D \simeq o(\ln N) \tag{2.28}$$

i.e.

$$\lim_{N\to\infty} \frac{D}{\ln N} = const. \tag{2.29}$$

This property of complex networks is called the *small world distance property*. Example of networks that have this property are ubiquitous, from the Internet and the World-Wide-Web to the social networks or the neural network of c.elegans.

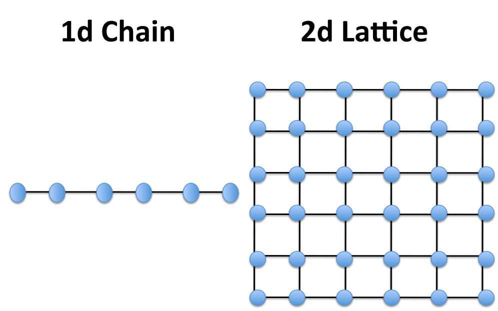

Figure 2.8: A 1d lattice (a chain) and a 2d lattice.

2.6 Network subgraphs, Loops, Cliques

Given a network $G = (V, E)$ formed by the set of nodes V different from the null set and by the set of edges E, it is always possible to define a subgraph.

Definition 34. *A subgraph $H = (V', E')$ of a network $G = (V, E)$ is formed by a set of node $V' \in V$ and by a set of links E' such that $E' \in E$ and that all the link in E' are incident only to nodes included in V'.*

Sometimes in useful to consider the subgraph composed by all the links incident to a subset V' of the set of nodes V of the original network. In this case we say that the subgraph is induced by the subset of vertices V'. Therefore we have the following definition.

Definition 35. *A subgraph $G' = (V', E')$ of the network $G = (V, E)$ is induced by the nodes in the set $V' \in V$ if and only if the set E' of its links includes all the links of G incident to the nodes in V'.*

In many situations it is interesting to consider special types of subgraphs such as loops, cliques, and k-cores. In Figure 2.9 we present a networks including cliques and loops of various size.

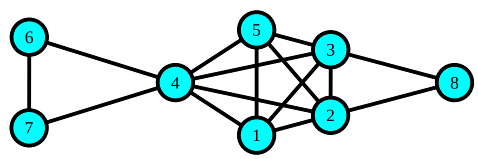

Figure 2.9: A network of $N = 8$ nodes including loops and cliques of various sizes.

2.6.1 Loops

Definition 36. *A* undirected loop *is a subgraph $H = (V', E')$ of an undirected network such that every node $i \in V'$ has degree 2 in the subgraph and such that every node can be reached by all the other nodes. A* directed loop *is a subgraph $H = (V', E')$ of a directed network such that every node $i \in V'$ has a in-degree 1 and a out-degree 1 and such that every node can be reached by all the other nodes.*

In a loop the number of nodes $|V'|$ is equal to the number of links $|E'|$, we call this number the length of the loop n.

Theorem 2.6.1. *The number of undirected loops of length 3 in an undirected network is given by*

$$\mathcal{L}_n = \frac{1}{6} Tr\mathbf{A}^3. \tag{2.30}$$

The number of directed loops of length 3 in a directed network is given by

$$\mathcal{L}_n = \frac{1}{3} Tr\mathbf{A}^3. \tag{2.31}$$

Proof. For every undirected loop of length 3 there are 6 distinct undirected cyclic path of length 3 in the network. In fact we can consider the cyclic paths departing from each of the 3 nodes of the loop and going either clockwise or counter-clockwise. Therefore the number of undirected loops of length 3 is given by the number of cyclic paths of length 3 divided by $6 = 3 \times 2$. The number of undirected cyclic paths is given by Eq.(2.22), i.e. $\mathrm{Tr}\mathbf{A}^n$. Therefore, it follows (2.30).
For every directed loop of length 3 there are three distinct cyclic paths of length 3 in the network. In fact we can consider all the cyclic paths departing from each of the 3 nodes of the loop and going in the direction of the directed loop. Therefore the number of loops of length 3 is given by the number of cyclic paths of length 3 divided by 3. The number of directed cyclic paths is given by Eq.(2.22), i.e. $\mathrm{Tr}\mathbf{A}^n$. Therefore, it follows (2.31). □

In Figure 2.9 there are 12 loops of size 3, 15 loops of size 4 and 12 loops of size 5.

2.6.2 Cliques

Definition 37. *A* clique *is a subgraph $G' = (V', E')$ of an undirected network such that every node $i \in V'$ with cardinality $|V'| = n$ has degree $n - 1$, i.e. such that every node is connected to every other node. The number of nodes in the clique n is also called the* clique size*.*

A clique of size n is also called $\mathcal{K}_n$. An undirected loop of length 3 (a triangle) is a clique of size 3. In the Figure 2.9 there are 12 cliques of size 3, 5 cliques of size 4 and 1 clique of size 5.

2.6.3 k-Core

Some networks have regions more dense than others. For example this is the case of the Internet described at the Autonomous System Level where few Autonomous Systems are linked by a relative large number of links. In order to characterize these dense regions of the network, it is useful to define the k-cores.

Definition 38. *A* k-core *of an undirected network is the subgraph induced by a set nodes whose degree within the subgraph is at least k and such that from each node it is possible to reach any other node of the subgraph by following a path (i.e. the subgraph is connected).A k-core has also the property that no additional node can be added to it whose degree is at least k within the subgraph.*

The k-cores of a network can be obtained by iteratively removing all the nodes of the network of degree less than k.
Every finite network has a maximal k for the $k-$cores with at least one element.

In Figure 2.10 the k-core structure of the Internet at the Autonomous System Level is shown. The average degree of this network is small, but the maximal

k of the k-cores reaches value 39, indicating that in the network there are very densely connected regions.

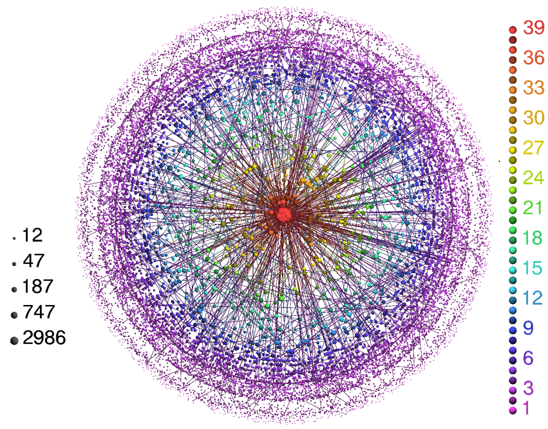

Figure 2.10: The k-core structure of the Internet at the Autonomous System Level. Data from DIMES. Figure produced with the LaNetVi visualization tool of networks, the k-cores are visualized with different color code and the node sizes indicates the degrees of the nodes.

2.7 Connected Components

2.7.1 Connected components in undirected networks

Definition 39. *An undirected network is* connected *if there is a path from every node of the network to any other node. A undirected network is* disconnected *if it is not connected.*

A network that is not connected contains several connected components.

Definition 40. *A* connected component of a unidirectednetwork *is a subgraph of the network induced by a set of nodes connected by each other by undirected path. Additionally, a connected component has maximum size, i.e. there is no node in the network that is connected to it by undirected paths but does not belong to it.*

For example the network in Figure 2.11 contains two connected components induced by the nodes $\{1,2,3\}$ and the nodes $\{4,5,6,7,8\}$.

2.7.2 Connected components in directed networks

Given a directed network we can either neglect the direction of the links or we can take into account the direction of the links. Therefore we can consider differ-

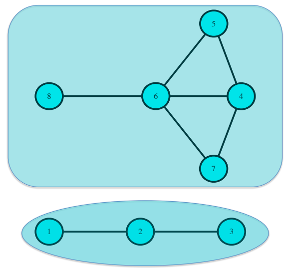

Figure 2.11: A disconnected network of $N = 8$ nodes and two connected components.

ent definitions of connected components called the *weakly connected components* and the *strongly connected components* of the directed network.

Definition 41. *The* weakly connected components *of a directed network are the connected components of the undirected network that can be constructed from the directed network by neglecting the direction of the links.*

Therefore two nodes are in the same weakly connected component if there is at least one path connecting them, where paths are allowed to go either way along the link.

Definition 42. *The* strongly connected components *of a directed network are the subgraphs induced by the maximal set of nodes that are connected to each other by paths going in both directions.*

Not all the nodes of a directed network are in a strongly connected component in general. In fact there are nodes in the weakly component of a directed network that can be reached from the other nodes following directed links but from which it is impossible to reach the other nodes or vice versa. For this reason is useful also to define the in-component (out-component) of a directed network relative to a given strongly connected component of the network.

Definition 43. *The* in-component *relative to a given strongly connected component is the set of nodes that are not reachable from the nodes of the strongly connected component by directed path, but from which there is a direct path to the nodes in the strongly connected component. The* out-component *relative to a given strongly connected component is the set of nodes that can be reached from the nodes of the strongly connected component by directed paths but from which there is no directed path to the nodes in the strongly connected components.*

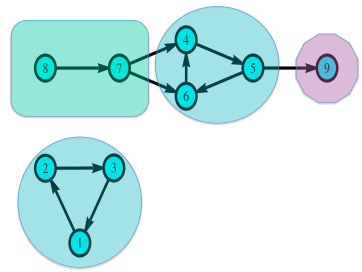

Figure 2.12: A disconnected network of $N = 9$. This network has two *weakly connected components*,including respectively the nodes $1, 2, 3$ and $4, 5, 6, 7, 8, 9$.Moreover this network contains two *strongly connected components* highlighted in cyan. The first strongly connected components includes the nodes $1, 2, 3$; the second connected component includes the nodes $4, 5, 6$. The in-components of the second strongly connected component is highlighted in green and includes nodes $7, 8$ while the out-component is highlighted in pink and contain only node 9.

In Figure 2.12 we highlighted in green the in-component relative to the strongly connected component formed by nodes $\{4, 5, 6\}$ and by pink its out-component.
In the case in which there is only one strongly connected component in the network we will refer to the in-component (out-component) relative to the strongly connected component as the *in-component (out-component) of the directed network*.

2.7.3 The Bow-Tie structure of the World-Wide-Web

The World-Wide-Web is a wonderful example of self-organized network, that over few decades has become central in today communication and diffusion of ideas and knowledge. The first maps of the World-Wide-Web (WWW) structure appear in the literature only around the year 2000, despite the fact that at that time the network was already extensively developed. In particular in the paper of A. Broder et al. *Graph structure in the web*,published in the Proceeding Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications networking, 309-320 (2000), for the first time the structure of the components of the WWW have been investigated. It was found that the WWW contains one major strongly connected component (SCC) that has one big in-component (IN) and one big out-component (OUT). Then there are TENDRILS departing from the in or the out components and tubes connecting directly the in-component with the out-component of the SCC. Finally there are small disconnected components (DISC). A schematic view of this "bow-tie" structure is represented in Figure 2.13. The sizes of the different regions of the WWW as reported in the cited paper are: SCC $56,463,993$, IN $43,343,168$, OUT $43,166,185$; DISC $16,777,756$ Total $203,549,046$.

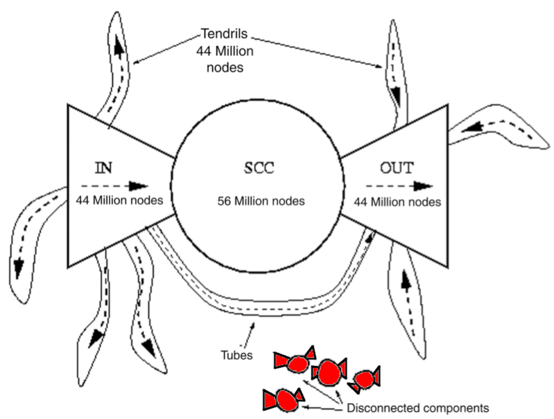

Figure 2.13: The bow-tie structure of the World-Wide-Web as described in the paper A. Broder et al. *Graph structure in the web*,Proceeding Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications networking, 309-320 (2000).

2.8 Special types of networks

2.8.1 Trees and Forests

Definition 44. *A* tree *is an network without loops.*
A tree in which a single node is connected to all remaining nodes is called a star network*. A* forest *is network formed by several trees forming the different connected components of the forest.*

2.8.2 Complete network

Definition 45. *A* complete network *of N nodes is a network in which every pair of nodes are connected by an undirected link.*

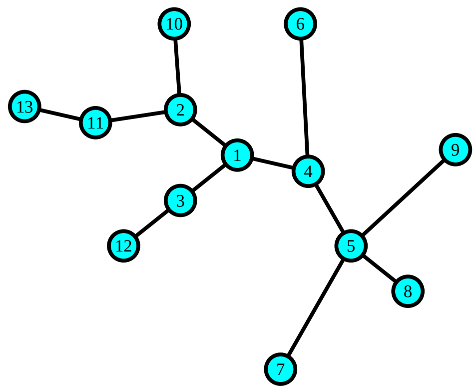

Figure 2.14: A tree of $N = 13$ nodes.

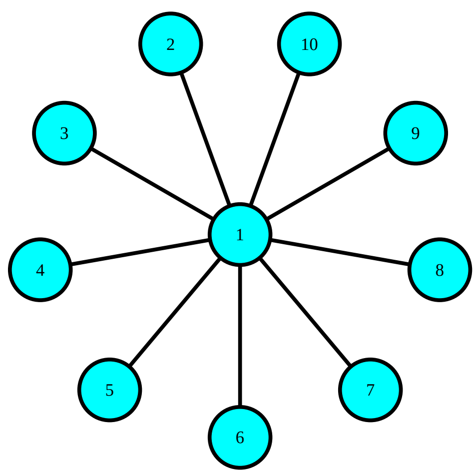

Figure 2.15: A star network of $N = 10$ nodes. Node 1 is the central node of the network.

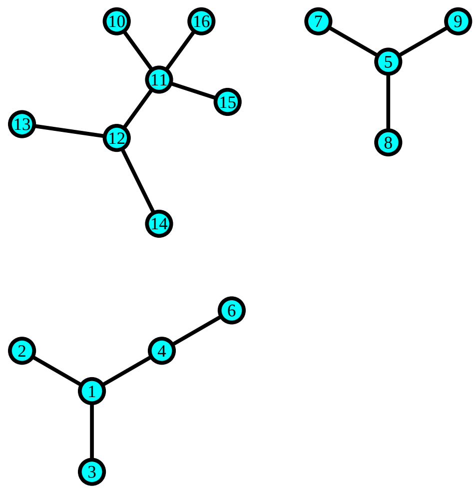

Figure 2.16: A forest of $N = 16$ nodes and 3 connected components.

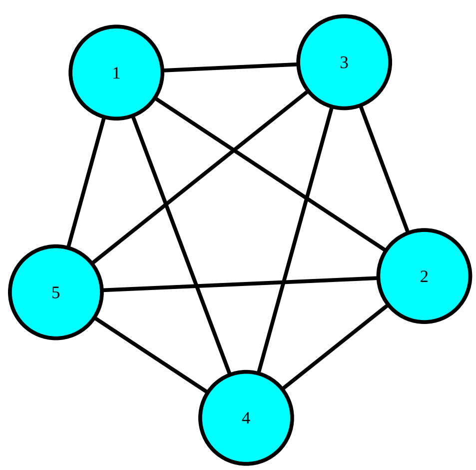

Figure 2.17: A complete network of $N = 5$ nodes.