Introduction to Network Theory

Ginestra Bianconi

May 2, 2021

These are extensive lecture notes for the module MTH6142.

Although it is recommended that the student reads all the material, the examinable sections are denoted with a (E) the non-examinable sections are denoted with a (NE).

Please email me at g.bianconi@qmul.ac.uk with subject line "Lecture notes" if you notice any typo in these notes. Your input to improve these lecture notes is highly appreciated.

Chapter 1

Networks: A prelude

Graphs are mathematical entities formed by a set of *nodes* (vertices) connected by *links* (edges). Graph theory is a branch of mathematics that started at a precise date. It was the genius of Leonhard Euler (1707-1783) that first solved a combinatorial problem making use of structural properties of graphs. In particular Leonhard Euler in 1735 solved the problem of the Seven Bridges of Königsberg.

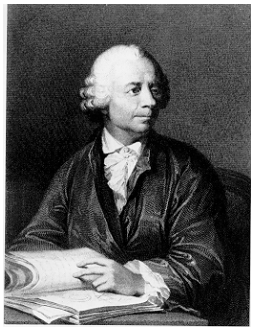

Figure 1.1: A portrait of Leonhard Euler (1707-1783), the founding father of graph theory.

At the time, the city of Königsberg, Prussia had seven bridges connecting the banks of the Pregel River to two large islands. The problem of the Seven Bridges of Königsberg is to decide whether it is possible to follow a path that crosses each bridge exactly once and returns to the starting point. It is not possible: there is no Eulerian cycle. Euler proved that such path does not exist by mapping the problem into a problem defined on a graph, in this way giving rise to the field of mathematics that goes under the name of *graph theory*.
Therefore, graph theory is almost 300 years old, but here in this module we will mainly discuss of the new field of *Complex Networks* which instead is only fifteen years old.
At the end of the '90s in fact two papers (Watts & Strogatz "small word" paper and Barabasi & Albert "Emergence of scale-free networks" paper) have proposed a new paradigm for studying complex systems. The fundamental idea behind these works was that complex systems as different as the Internet, complex infrastructures, social networks, cellular networks all have an underlying network structure describing the complex set of interactions between the elements forming these different systems. A network differs from a graph because it is a specific graph describing the interactions present in a specific complex system. In the

two cited seminal papers it was shown that, despite the diversity of the complex systems that can be described by complex networks, there are some properties of these networks that are *universal*: i.e. they are common to a large variety of complex systems. These universalities suggest that the organizational properties of *self-organized* complex systems that emerge as the outcome of biological evolution (like the cellular networks or the brain) or as the product of non-centrally-organized human activity (like the Internet or the social networks) present similarity beyond expectations. These properties of complex systems are responsible for their robustness and for their efficiency.

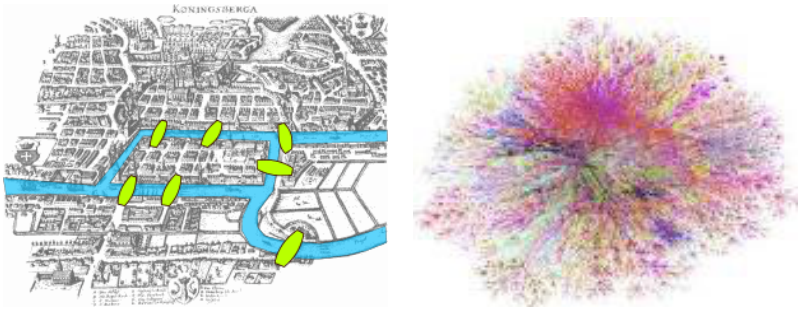

Figure 1.2: (left panel) The city of Königsberg. The problem of the seven bridges of Königsberg (the Eulerian cycle problem) was solved by L. Euler in 1735. (right panel) A visualization of the Internet as a *complex network*. The characterization of the navigability of this network and of the quantification of its robustness are new challenges for the scientific community of the XXI century working in complex network theory.

1.1 Graphs and Networks (E)

In graph theory a graph is defined as follows:

Definition 1. *A* graph *is an ordered pair* $G = (V, E)$ *comprising a set* V *of* vertices *connected by the set* E *of* edges.

In network theory, the graphs that describe a specific complex system are called *networks* . The vertices of these graphs are usually called *nodes* and the edges are usually called *links*. In the following we give the network definition.

Definition 2. *A* network *is the graph* $G = (V, E)$ *describing the set of interactions between the constituents of a complex system. The vertices of a network are called nodes and the edges links. The network size* N *is the total number of nodes in the network* $N = |V|$. *The total number of links* L *is given by* $L = |E|$.

Although, there is a clear difference in the definition of a network (describing a complex system) and a graph (that is the abstract mathematical object formed by vertices and edges), in the field of complex networks the term networks is often used as a synonym of a graph. In the table 1.1 you can find a "dictionary" between graph theory terms and complex network terms. In this course we will adopt the complex network terminology.

Graph theory term	**Complex network term**
Graphs	Network
Vertices	Nodes
Edges	Links
Cycles	Loops
Loops	Tadpoles

Table 1.1: Graph-theory/Network-theory Dictionary of common terms

1.2 Examples of Graphs and Networks (E)

Complex networks are found in a large variety of complex systems, it is sufficient that you start *"to think networks"* and you will discover that networks are ubiquitous. For example, our society is formed by a complex set of interactions between the individuals linked by friendship ties, family ties, collaborative interaction, etc... Also in animal societies networks are important. In ecology we can define food-webs where the interactions are of the type prey-predator but we can also define mutualistic networks where different species cooperate to increase reciprocally their fitness (e.g. the bee and the flower).
Networks are not only found in human and animal societies,also if we want to describe the complex organization of a cell or of the brain it is essential to consider networks. The cell is formed by several molecules, the DNA formed by genes, the proteins, and the metabolites. These constituents of the cell are interacting and we will not have living organisms without these interactions. It is nowadays usual to consider several cellular networks described briefly in the following.

- The *metabolic network* of a certain organism is formed by its metabolites that react chemically thanks to some proteins called enzymes. The metabolic network is responsible for providing the energy to the cell and producing the biomass necessary to allow its duplication.
- The *protein interaction network* of a certain organism describes the set of interactions between proteins. Proteins bind to each other and form protein complexes to perform complex cellular functions.
- The *transcription network* of a certain organism connects the genes of that organism and is responsible for what is called the "cellular regulation".

Complex networks	Nodes	Links
Actors network	Actors	Co-acting on a movie
Collaboration networks	Scientists	Co-authors in one paper
Citation networks	Scientific papers	Citation
Facebook network	Individuals	Facebook friends
Metabolic network	Metabolites	Common chemical reaction
Protein-Interaction networks	Proteins	Physical interaction
Transcription networks	Genes	Regulation
Brain network	Neurons	Synaptic connections
Internet	Routers	Physical lines
World-Wide-Web	Webpages	URL's addresses
Airport network	Airports	Flight connections
Power-grids	Power plants	Electric grid

Table 1.2: Examples of complex networks

Cellular regulation is something very crucial for the cell and allows for example the cells in the heart or in the brain of the same person to be different although they have an identical DNA. In transcription networks some genes when transcribed produce a special type of proteins called *transcription factors*. These proteins can bind to the DNA and might switch on or off the transcription of other genes.

Also man-made technological objects can be described as networks. Major examples include:

- The *Internet* formed by routers connected by physical lines;
- The *World-Wide-Web (WWW)* that is distinct from the Internet and is the virtual network of webpages and URL addresses between the webpages;
- The *Airport network* that is formed by airports connected by flight connections;
- The *Power grids* formed by power-plants connected by the electric grid.

1.3 Labelled, Simple, Undirected, Directed, Weighted and Signed Networks (E)

1.3.1 Labelled networks

Networks represent real complex systems, so usually the nodes of the network have a specific "name": Name of an individual in social networks, name of a protein in protein interaction networks, name of a species in food webs, etc. In this case we say that networks are labelled.

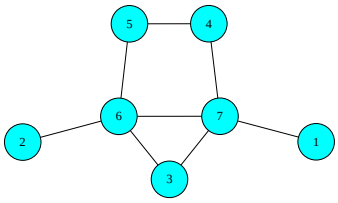

Figure 1.3: Labelled network of $N = 7$ nodes $i = 1, 2 \ldots, 7$ and $L = 8$ links

.

Definition 3. *A* labelled network*, of network size N, is formed by a set V of N distinguishable nodes indicated by a different and unique label $i = 1, 2, \ldots, N$ and by a set of links E characterizing the interactions between pairs of nodes.*

In figure 1.3 you can see an example of a labelled network of $N = 7$ nodes and $L = 8$ links.

A labeled network can be simple, undirected, directed, weighted and signed depending on the properties of their links.

1.3.2 Directed and Undirected networks

A link can be either *directed or undirected.*

Definition 4. *A* directed link *indicates an interaction between nodes that is not symmetrical. The graphical representation of directed links is an arrow. If node j points to node i the arrow starts from node j and points to node i (see Fig.* 1.4*).*
A directed network *is a network where all the links are directed.*

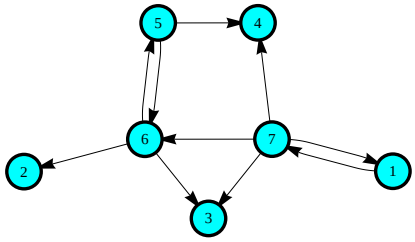

Figure 1.4: Labelled directed network of $N = 7$ nodes $i = 1, 2 \ldots, 7$ and $L = 10$ directed links

.

For example in a social network two individuals might be connected if one calls the other one by mobile phone. This is a case of directed interaction since it is not guaranteed that if j calls i then i calls j in the given time window over which the network is aggregated. Interestingly, in several sociological surveys where for example students in a school are asked to indicate their friends, it was found that a large fraction of friendships was not reciprocated

suggesting that social networks might be intrinsically directed. In biology, transcription networks are also directed because regulation is a non symmetric type of interaction: if a gene j regulates gene i it is not true in general that gene i regulates gene j. In technological networks the World-Wide-Web is a beautiful example of directed networks because if a webpage j points to a webpage i it is not guaranteed that the reciprocal is also true.

Definition 5. *An* undirected link *indicates a symmetric interaction. The graphical representation of undirected links is a line. If node j is linked to node i then also node i is linked to node j (see Fig.* 1.3*).*
An undirected network *is a network in which every link is undirected.*

Although the friendship might be considered as a directed interaction several social networks are undirected. Collaboration networks are undirected both in the case of co-acting actors or collaborating scientists, also usually friendship in online social networks indicates only undirected links. For example, if an individual i is a Facebook friend with node j then also j is a Facebook friend with i. In biology, protein interaction networks are also undirected because if protein i binds to protein j to form a protein complex, also the reciprocal is true. In technological networks the Internet is an example of an undirected network because if the router j is linked to the router i the reciprocal is also true.

1.3.3 Weighted and Unweighted networks

Links can be weighted or unweighted. The weight of the link is either an integer or a real number. Weighted networks describe the situation in which different interactions have different intensity. Therefore weights are fundamental if we want to characterize a variety of systems, because in many situations not all the links have the same relevance.

Definition 6. *A* weighted link *between node i and node j is a link to which we assign an integer or real number w_{ij} indicating the intensity of the interaction. When the weight is integer the weighted link is also called* multiple link*. Weighted networks can be either directed or undirected.*

The graphical representation of a multiple link between node i and node j is given by a number w_{ij} of lines between node i and node j (see Fig. 1.5). The graphical representation of a weighted link between node i and node j is a line associated with the weight w_{ij} (see Fig. 1.5).

Given the definition of weighted links we can define a weighted network as in the following.

Definition 7. *A* weighted network *is a network where all the links are weighted.*

Weights are very important in social networks where there is an important scientific debate regarding the role of weak and strong ties in society for the efficiency of the communication and spreading of opinions and behaviours. In

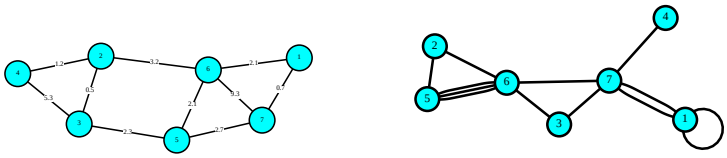

Figure 1.5: Two examples of weighted networks. In one case (left panel) the weights are real numbers. In the other case (right panel) the network contains multiple links such as the links connecting the pair of nodes $(6, 5)$ and $(7, 1)$. The link connecting node 1 to itself is a tadpole.

a collaboration network two people can just co-author a single paper of co-author multiple papers, also a pair of actors might act in few or several films together. Weights are also important for transportation networks. For example in the airport networks, we can associate to each flight connection the number of travellers flying each week in the given direction.
If all the links of a network have the same intensity, and a link is either present or absent, we say that the links of the network are unweighted, i.e. it is unnecessary to define a weight of the links.

Definition 8. *An* unweighted network *is a network in which all the links are unweighted. Unweighted networks can be either directed or undirected.*

1.3.4 Signed or unsigned networks (NE)

Links can also be signed or unsigned. Signed links are used when interactions of opposite type are both present in the same network. In a social network we can associate a sign ± 1 to links indicating whether an interaction is positive (such as friendship) or negative (such as enmity).

Definition 9. *A* signed link *is a link associated with a sign (either positive or negative). Signed networks can be also weighted and/or directed. A* signed network *is a networks where all the links are signed.*

As we mentioned already social networks might be signed. Another major example of signed network is the transcription network in which one gene can be either an *activator* or an *inhibitor* of any other regulated gene. An expressed activator gene activates the transcription of the regulated gene, while an expressed inhibitor gene inhibits the transcription of the regulated gene.

Definition 10. *An* unsigned network *is a network in which all the links have the same "sign". Therefore we can neglect the specification of the sign of the links.*

1.3.5 Tadpoles

A special type of links are *tadpoles*.

Definition 11. Tadpoles *are links that connect a node with itself. Tadpoles can be directed or undirected.*

Tadpoles are present for example in transcription networks where it can happen that one gene regulates its own transcription. In figure 1.5 (right panel) we show an example of weighted network with a tadpole.

1.3.6 Simple networks

In social networks, collaboration networks, and in transportation networks there are no tadpoles. For example the network in Figure 1.6 is a simple network. Simple networks are the most basic mathematical entities that can be called networks.

Definition 12. *A* simple network *is an undirected, unweighted, and unsigned network without tadpoles.*

1.4 Representation of a network: The Edge list and the Adjacency matrix (E)

A labelled network can be represented by means of an *edge list* or by means of and *adjacency matrix*. In the following we will consider first the simple networks, then directed, weighted, signed and finally bipartite networks.

1.4.1 Simple networks

Edge list of a simple network

Definition 13. *An* edge list of a simple network *is a list of L pairs of node labels (j, i) indicating that between node j and node i there is a link. Here L indicates the total number of links (edges) in the network.*

The edge list of a simple network does not allow for redundancies. If a link is listed as (j, i) it cannot be listed also as (i, j) and vice versa. In the following we give the edge list of the labelled network in Figure 1.6: i.e.

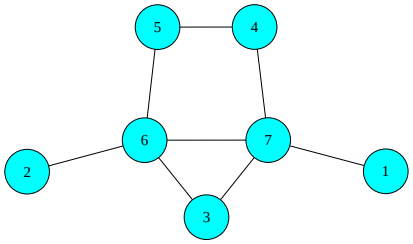

Figure 1.6: Labelled network of $N = 7$ nodes $i = 1, 2 \ldots, 7$ and $L = 8$ links.

.

Edge list

$$\begin{array}{r} (5,4) \\ (2,6) \\ (6,3) \\ (6,5) \\ (7,1) \\ (3,7) \\ (4,7) \\ (7,6). \end{array} \tag{1.1}$$

Adjacency matrix of a simple network

Definition 14. *The* adjacency matrix of a simple network *is an* $N \times N$ *matrix* $\mathbf{A}$ *of elements*

$$A_{ij} = \begin{cases} 1 & \text{if node } j \text{ is linked to node } i \\ 0 & \text{otherwise} \end{cases}$$

Since in a simple network, if node j is linked to node i also node i is linked to node j, the adjacency matrices of simple networks are always *symmetric*. Since in a simple network there are no tadpoles then the diagonal matrix elements are equal to zero. As an example in the following we provide the adjacency matrix of the network in Figure 1.6.

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

1.4.2 Directed networks

Edge List of directed networks

Definition 15. *An* edge list of a directed network *is a list of* L *pairs of ordered node labels* (j, i)*. Each ordered pair of labels* (j, i) *indicates that node* j *points to node* i*.* L *indicates the total number of directed links (edges) in the network.*

In the following we give the edge list of the labelled network in Figure 1.7: i.e.

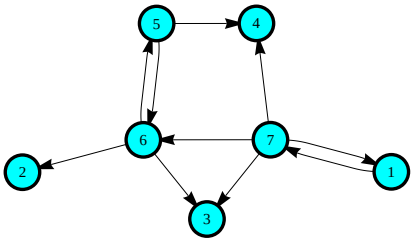

Figure 1.7: Labelled directed network of $N = 7$ nodes $i = 1, 2 \ldots, 7$ and $L = 10$ links.

Edge list

$(5, 4)$
$(6, 2)$
$(7, 6)$
$(6, 5)$
$(5, 6)$
$(7, 1)$
$(1, 7)$
$(7, 4)$
$(7, 3)$
$(6, 3)$.

Adjacency matrix of a directed network

Definition 16. *The* adjacency matrix of a directed network *is an* $N \times N$ *matrix* $\mathbf{A}$ *of elements*

$$A_{ij} = \begin{cases} 1 & \text{if node } j \text{ points to node } i \\ 0 & \text{otherwise} \end{cases}$$

The adjacency matrix of a directed network is *asymmetric*. In fact if node j points to node i it is not true in general that node i points to node j. As an example in the following we provide the adjacency matrix of the network in Figure 1.7.

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

1.4.3 Weighted networks

Edge List of a Weighted Network

Definition 17. *An* edge list of a weighted network *is a list of* L *triplets* (j, i, w_{ij}). *For an undirected network each triple* (j, i, w_{ij}) *indicates that between node* j *and node* i *there is a link of weight* w_{ij}. *Here* L *indicates the total number of links in the network. For a directed network each triple* (j, i, w_{ij}) *indicates that node* j *points to node* i *with a link of weight* w_{ij}, *where* L *indicates the total number of directed links (edges) in the network.*

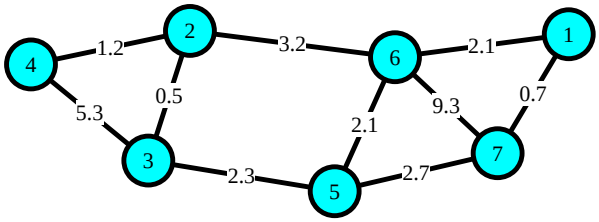

Figure 1.8: Example of weighted network

.

The edge list for the weighted network in figure 1.8 is given by

Edge list

$(2, 4, 1.2)$
$(4, 3, 5.3)$
$(3, 2, 0.5)$
$(6, 2, 3.2)$
$(5, 6, 2.1)$
$(5, 3, 2.3)$
$(1, 7, 0.7)$
$(7, 6, 9.3)$
$(7, 5, 2.7)$
$(6, 1, 2.1)$.

Adjacency matrix of a weighted network

Definition 18. *The* adjacency matrix of a weighted network *is a* $N \times N$ *matrix* **A** *of elements* A_{ij} *defined as follows:*

Undirected Weighted network

$$A_{ij} = \begin{cases} w_{ij} & \text{if node } j \text{ is linked to node } i \text{ with a link of weight } w_{ij}, \\ 0 & \text{otherwise} \end{cases}$$

Directed Weighted network

$$A_{ij} = \begin{cases} w_{ij} & \text{if node } j \text{ points to node } i \text{ with a link of weight } w_{ij}, \\ 0 & \text{otherwise} \end{cases}$$

The adjacency matrix of a weighted undirected network is *symmetric* the adjacency matrix of a weighted directed network is *asymmetric*. For example the adjacency matrix of the undirected unweighted network in Figure 1.8 is given by

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 2.1 & 0.7 \\ 0 & 0 & 0.5 & 1.2 & 0 & 3.2 & 0 \\ 0 & 0.5 & 0 & 5.3 & 2.3 & 0 & 0 \\ 0 & 1.2 & 5.3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2.3 & 0 & 0 & 2.1 & 2.7 \\ 2.1 & 3.2 & 0 & 0 & 2.1 & 0 & 9.3 \\ 0.7 & 0 & 0 & 0 & 2.7 & 9.3 & 0 \end{pmatrix}.$$

1.4.4 Signed Networks

Signed Networks can be treated as weighted networks with weights including the sign of the links.

1.4.5 Tadpoles

A tadpole linking node i to itself can be represented in a edge list as (i, i) and in a weighted edge list as (i, i, w_{ii}), and can be represented by a diagonal element of the adjacency matrix $A_{ii} = 1$ for an unweighted network and as $A_{ii} = w_{ii}$ for a weighted network.

1.5 Bipartite Networks (E)

In several cases a complex systems has an underlying bipartite network structure.

Definition 19. *A* bipartite network $G_B = (V, U, E)$ *is a network formed by two non overlapping sets of nodes* U *and* V *and by a set of links* E*, such that every link joins a node in* V *with a node in* U*.*

We call the number of nodes in V, $|V| = N_V$ and the number of nodes in U, $|U| = N_U$. We indicated the nodes in the set V by N_V integer numbers $1, 2, i, \ldots, N_V$ and the nodes in U by N_U latin letters $a, b, c \ldots$. Bipartite networks can be extremely useful in order to represent the membership of one node i to a group a. In this case N_U indicates the number of groups in the system. In Table 1.5 a series of complex systems that can be represented as a bipartite network is listed.

1.5.1 Incidence matrix

A bipartite network $G_B = (V, U, E)$ is described by an *incidence matrix*.

Definition 20. *The* incidence matrix *of a bipartite network* $G_B = (V, U, E)$ *is an* $N_V \times N_U$ *matrix of elements* $B_{i,a}$ *defined as follows:*

Bipartite network	**Nodes** $i \in V$	**Groups** $a \in U$
Bipartite Actors network	Actors	Films
Bipartite Collaboration networks	Scientists	Papers
Bipartite Board of Directors	Directors	Board of a company
Bipartite Metabolic network	Metabolites	Chemical reaction

Table 1.3: Examples of bipartite complex networks

Unweighted, Undirected Bipartite network

$$B_{i,a} = \begin{cases} 1 & \text{if node } i \text{ is linked to node } a \\ 0 & \text{otherwise.} \end{cases}$$

Weighted, Undirected Bipartite network

$$B_{i,a} = \begin{cases} w_{i,a} & \text{if node } i \text{ is linked to node } a \text{ with weight } w_{i,a} \\ 0 & \text{otherwise.} \end{cases}$$

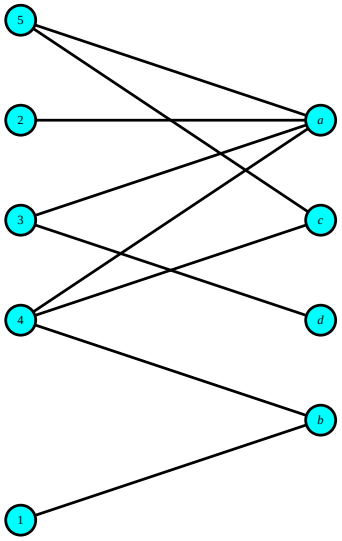

Figure 1.9: Example of a bipartite undirected and unweighted network of $N_V = 5$ nodes and $N_U = 4$ groups.

In Figure 1.9 we show an example of a bipartite undirected network with $N_V = 5$ and $N_U = 4$. The incidence matrix B of dimension $N_V \times N_U$ of the bipartite network in Figure 1.9 is

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

When using bipartite networks to describe the membership of some nodes in same groups, the unweighted and undirected bipartite networks are used.However, there are occasions in which also directed bipartite networks are useful. For example this is the case for the bipartite metabolic networks, where metabolites can be either the reactant or the product of a chemical reaction, and in this case we can adopt a convention according to which the reactants of a chemical reaction point to the reaction itself

while the products of the reaction are pointed at by the chemical reaction. The definition can be naturally extended also to directed bipartite networks by considering at the same time the incidence matrix B of elements $B_{i,a}$ indicating if node a points to the node i and the incidence matrix $\mathbf{B}'$ of elements $B'_{a,i}$ indicating if node i points to node a.

1.5.2 Projections of the unweighted and undirected bipartite network (NE)

From a single unweighted and undirected bipartite network $G_B = (V, U, E)$ we can extract two projection networks: $G_V = (V, E_V)$ and $G_U = (U, E_U)$. We will call the projection network G_V the *node projection network*, and the projection network G_U the *group projection network*.

The Node Projection Network (NE)

Definition 21. *The node projection network is a network between the nodes $i \in V$. Two nodes i and j are linked in the node projection network if they belong to at least one common group $a \in U$ in the bipartite network.*

If the original bipartite network is unweighted, the node projection network is weighted and contains tadpoles. Moreover, it is characterized by the weighted undirected adjacency matrix $\mathbf{P}$ of size $N_V \times N_V$ and matrix elements P_{ij} indicating the number of groups to which both node i and node j belong, i.e.

$$P_{ij} = \sum_{a \in U} B_{i,a} B_{j,a} = \sum_{a \in U} B_{i,a} B^{\top}_{a,j} = \left[\mathbf{B}\mathbf{B}^{\top}\right]_{ij} . \tag{1.2}$$

Therefore we have

$$\mathbf{P} = \mathbf{B}\mathbf{B}^{\top}. \tag{1.3}$$

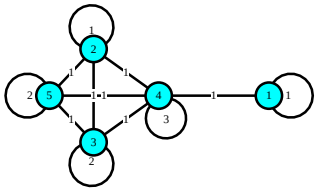

Figure 1.10: The node projection of the bipartite network in Figure 1.9.

From the Eq. (1.2), it follows that the matrix $\mathbf{P}$ has diagonal elements P_{ii} indicating the number of groups to which node i belongs. The 5×5 adjacency matrix $\mathbf{P}$ of the node projection network (shown in Figure 1.10) of the bipartite network in Figure 1.9 is given by

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 3 & 1 \\ 0 & 1 & 1 & 1 & 2 \end{pmatrix} .$$

The Group Projection Network (NE)

Definition 22. *The group projection network is a network between the groups $a \in U$. Two groups a and b are linked in the group projection network if at least one node belongs to both groups.*

If the original bipartite network is unweighted, the group projection network is weighted and contains tadpoles. Moreover, it is characterized by the weighted undirected adjacency matrix $\tilde{\mathbf{P}}$ of size $N_U \times N_U$ and matrix elements, $\tilde{P}_{ab}$ indicating the number of nodes belonging to both groups a and b

$$\tilde{P}_{a,b} = \sum_{i \in V} B_{i,a} B_{i,a} = \sum_{i \in V} B^{\top}_{a,i} B_{i,b} = \left[\mathbf{B}^{\top}\mathbf{B}\right]_{ab}. \tag{1.4}$$

Therefore we have

$$\tilde{\mathbf{P}} = \mathbf{B}^{\top}\mathbf{B}. \tag{1.5}$$

From the Eq. (1.4), it follows that the matrix $\tilde{\mathbf{P}}$ has diagonal elements $\tilde{P}_{aa}$ indicating the number of nodes belonging to group a. The 4×4 adjacency matrix $\tilde{\mathbf{P}}$ of the group projection network (shown in Figure 1.11) of the bipartite network in Figure 1.9 is given by

$$\tilde{\mathbf{P}} = \begin{pmatrix} 4 & 1 & 2 & 1 \\ 1 & 2 & 0 & 0 \\ 2 & 1 & 2 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}.$$

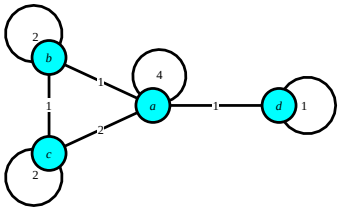

Figure 1.11: The group projection network of the bipartite network in Figure 1.9.

Chapter 2

Structural properties of networks

2.1 Introduction (E)

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 (E)

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

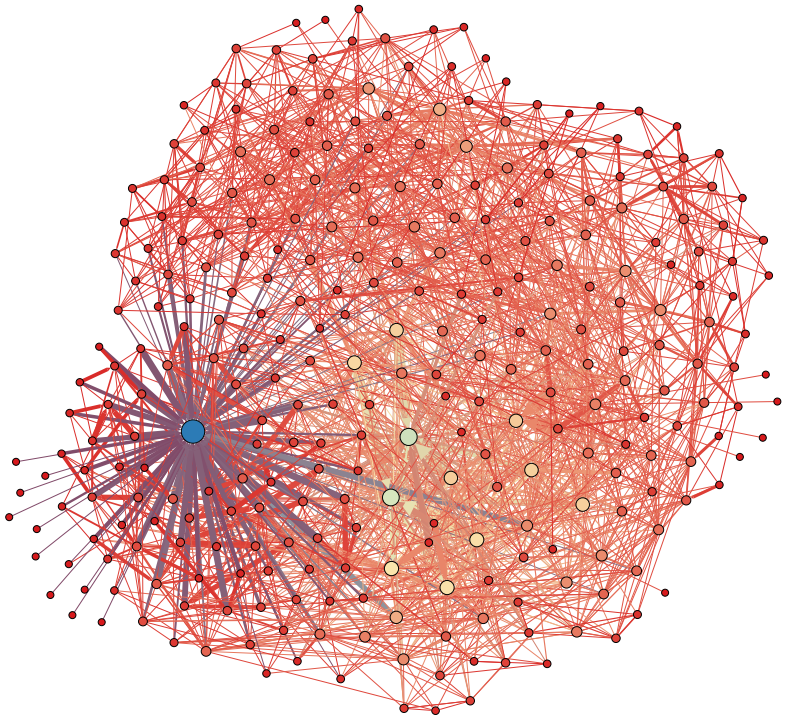

Figure 2.1: An example of "minimal complex network": the *c.elegans* neural network with $N = 302$ nodes. The thickness of the links is proportional to the weights of the links, the size of the nodes is proportional to their degree. The figure does not show the direction of the links.

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 (E)

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 **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

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

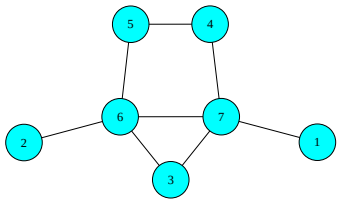

Figure 2.2: Undirected network of $N = 7$ nodes and $L = 8$ links

.

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,

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

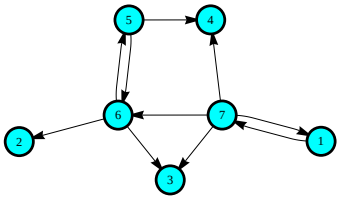

Figure 2.3: Directed network of $N = 7$ nodes and $L = 10$ links

.

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

$$K^{in} = \max_i k_i^{in},$$
$$K^{out} = \max_i k_i^{out}. \tag{2.11}$$

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 (E)

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

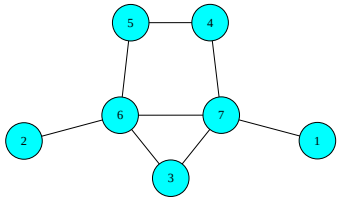

Figure 2.4: Undirected network of $N = 7$ nodes and $L = 8$ links

.

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

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

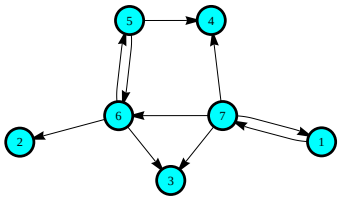

Figure 2.5: Undirected network of $N = 7$ nodes and $L = 10$ links

.

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

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

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

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.

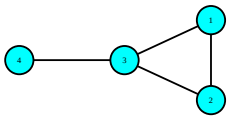

Figure 2.6: An undirected network of $N = 4$ nodes containing cyclic paths.

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

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

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

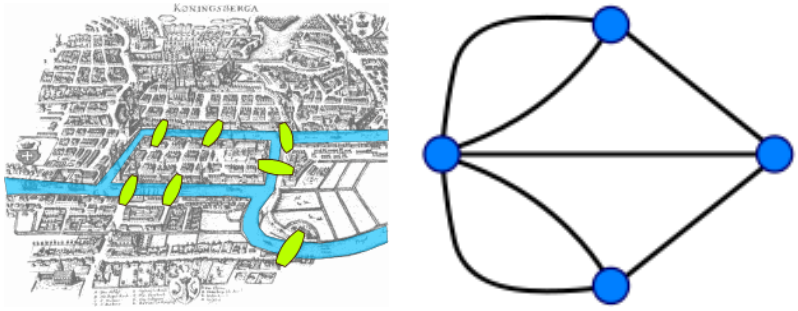

Figure 2.7: The city of Königsberg is shown on the left panel . The problem of the seven bridges of Könisberg (the Eulerian path problem) was solved by L. Euler in 1736 by mapping the mapping the problem in to the problem of finding an Eulerian path in the network shown in the right panel.

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 (E)

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.

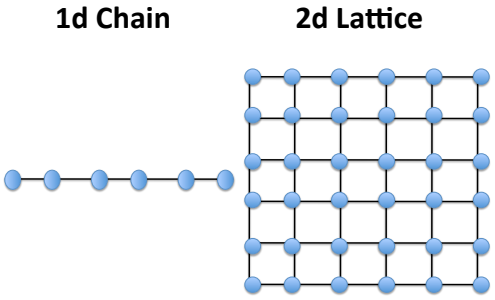

Figure 2.8: A 1d lattice (a chain) and a 2d lattice.

2.6 Network subgraphs, Loops, Cliques

2.6.1 Network subgraphs (E)

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.

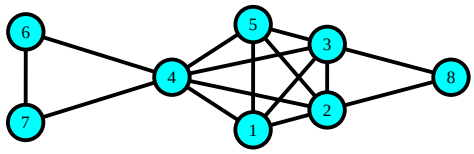

Figure 2.9: A network of $N = 8$ nodes including loops and cliques of various sizes.

2.6.2 Loops (E)

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.

2.6.3 Cliques (E)

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.

2.6.4 k-Core (NE)

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.

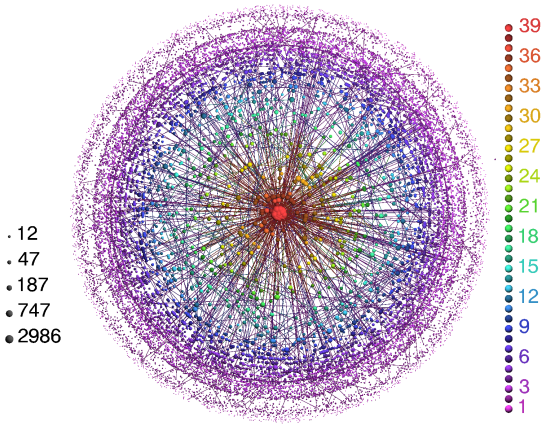

Figure 2.10: The k-core structure of the Internet at the Autonomous System Level. Data from DIMES. Figure produced with the LaNetVi visualization tool of networks, the k-cores are visualized with different color code and the node sizes indicates the degrees of the nodes.

2.7 Connected Components (E)

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-

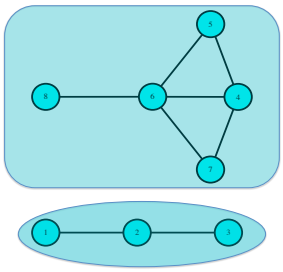

Figure 2.11: A disconnected network of $N = 8$ nodes and two connected components.

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. *A* strongly connected component *of a directed network is the subgraph induced by a set of nodes V' with cardinality $|V'| \geq 2$ such that every pair of nodes in the component is connected by at least one path going in each direction and such that no other node of the network can be added to V' preserving this property.*

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.*

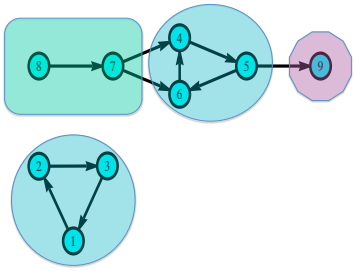

Figure 2.12: A disconnected network of $N = 9$. This network has two *weakly connected components*,including respectively the nodes $1, 2, 3$ and $4, 5, 6, 7, 8, 9$.Moreover this network contains two *strongly connected components* highlighted in cyan. The first strongly connected components includes 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 $203,549,046$.

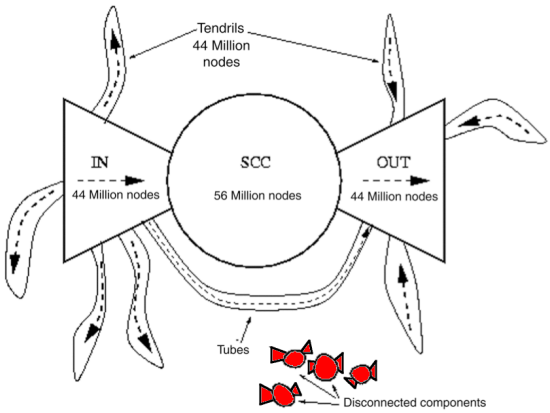

Figure 2.13: The bow-tie structure of the World-Wide-Web as described in the paper A. Broder et al. *Graph structure in the web*,Proceeding Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications networking, 309-320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

2.8 Special types of networks (E)

2.8.1 Trees and Forests

Definition 44. *A* tree *is a connected 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.*

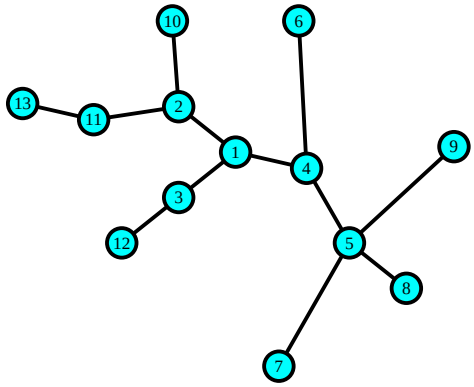

Figure 2.14: A tree of $N = 13$ nodes.

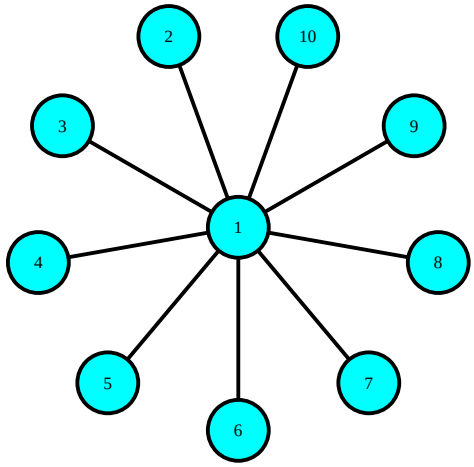

Figure 2.15: A star network of $N = 10$ nodes. Node 1 is the central node of the network.

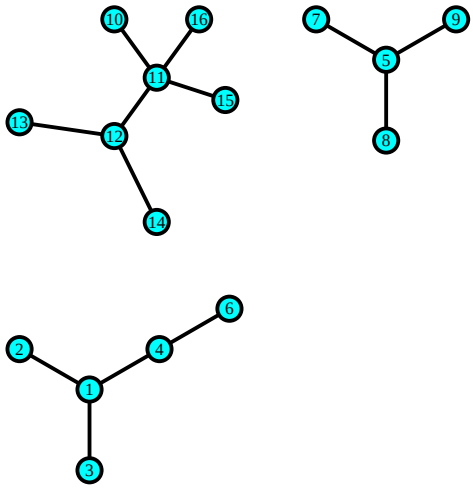

Figure 2.16: A forest of $N = 16$ nodes and 3 connected components.

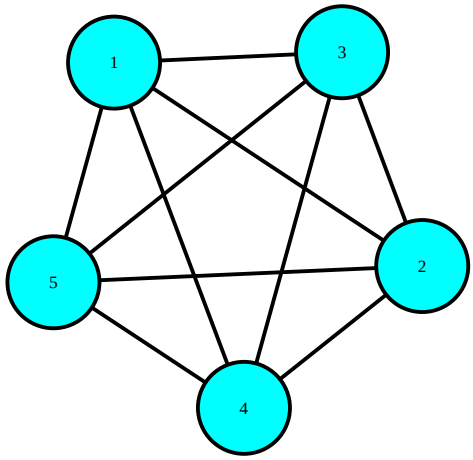

Figure 2.17: A complete network of $N = 5$ nodes.

Chapter 3

Centrality measures

3.1 Introduction (E)

Ranking algorithms are becoming increasingly important in our society. The traditional alphabetic order of the Encyclopaedia is nowadays substituted in everyday life by websearch algorithm that rank webpages (i.e. the nodes of the World-Wide-Web) in order of decreasing relevance to the query. In the context of network theory, in may cases we are interested to find the most *central* nodes of a given network. Centrality of a node might reflect the importance of the node in keeping a network connected, in decreasing the shortest distance between the nodes, or it can just be a property of the node like its degree. In the nowadays most successful centrality algorithm, the PageRank algorithm, used by Google for ranking the results of a query, a node is more central if many central nodes point to it. The works related to centrality of the nodes in a network are very often related to the sociological literature where the problem of find "important" and influential nodes was first introduced. Nevertheless the centrality of nodes is very important also in biology where for example a gene like the p53 gene is an essential node for most of cellular functions and its mutations are related with the onset of cancer.

3.2 Degree centrality (E)

Nodes with high degree, called *hubs nodes* usually play an important role in the network. Therefore in undirected network the degree $k_i = \sum_j A_{ij}$ of node i is considered often as a good proxy of its centrality. In directed networks, it is possible to consider both the in-degree $k_i^{in} = \sum_j A_{ij}$ of a node or the out-degree $k_i^{out} = \sum_j A_{ji}$ of a node as centrality measures.

3.3 Eigenvector centrality (E)

Consider a undirected network or a strongly connected network, then it is possible to refine the measure of centrality by considering the *eigenvector centrality*. In the eigenvector centrality a node is more important is already important nodes point to it.

Definition 46. *The* eigenvector centrality $\mathbf{x}$ *can be obtained starting from an initial guess of the centrality of the nodes* $x_i^{(0)}$ *and by considering the the following recursive process:*

$$x_i^{(n)} = \sum_j A_{ij} x_j^{(n-1)}. \tag{3.1}$$

Whereas the limit $\lim_{n\to\infty} \sum_j x_j^{(n)}$ *exits the following procedure defines the eigenvector centrality. If* $\lim_{n\to\infty} \sum_j x_j^{(n)} > 0$ *the eigenvector centrality* x_i *of a node* i *is found by performing the limit*

$$x_i = \lim_{n\to\infty} \frac{x_i^{(n)}}{\sum_j x_j^{(n)}}, \tag{3.2}$$

if instead $\lim_{n\to\infty} \sum_j x_j^{(n)} = 0$ *then all the nodes have zero eigenvector centrality, i.e.*

$$\mathbf{x} = \mathbf{0}.$$

A typical initial guess is

$$x_i^{(0)} = \frac{1}{N}$$

for all $i \in \{1, 2, \ldots, N\}$*. However if the choice of this initial guess does not lead to a well defined* $\lim_{n\to\infty} \sum_j x_j^{(n)}$ *another choice of the initial guess should be made.*

Proposition 2. *In an undirected network or in a network with at least one strongly connected component, the eigenvector centrality* $\mathbf{x}$ *of a network is proportional (*$\mathbf{x} \propto \mathbf{v}^1$*)to the leading eigenvector* $\mathbf{v}^1$ *satisfying,*

$$\lambda_1 \mathbf{v}^1 = \mathbf{A}\mathbf{v}^1 \tag{3.3}$$

where the real eigenvalue λ_1 *is the Perron-Frobenious eigenvalue of the adjacency matrix* $\mathbf{A}$*. This implies that* λ_1 *is the only real eigenvalue of* $\mathbf{A}$ *with* $\lambda_1 \geq |\lambda_i|$ *where* λ_i *with* $i = 1, 2, \ldots, N$ *are all the eigenvalues of the adjacency matrix of the network. For the Perron-Frobenious theorem the eigenvector* $\mathbf{x}$ *has all non-negative elements* $x_i \geq 0 \forall i = 1, 2 \ldots, N$*.*

Proof. The recursive process defined in (3.1) implies that $x_i^{(n)}$ is given by

$$\mathbf{x}^{(n)} = \mathbf{A}^n \mathbf{x}^{(0)} \tag{3.4}$$

where $\mathbf{x}^{(0)} = \mathbf{1}/N$ and $\mathbf{1}$ is the column vector of elements $\mathbf{1}_j = 1$. Expressing the vector $\mathbf{1}/N$ into the base of right eigenvectors of the matrix $\mathbf{A}$, $\{\mathbf{v}^\mu\}_{\mu=1,2...,N}$ where each right eigenvector is associated with the eigenvalues λ_μ, we obtain:

$$\frac{1}{N}\mathbf{1} = \sum_{\mu=1}^{N} c^\mu \mathbf{v}^\mu. \tag{3.5}$$

Therefore $\mathbf{x}^{(n)}$ if given by

$$\mathbf{x}^{(n)} = \sum_{\mu=1}^{N} c^\mu \mathbf{A}^n \mathbf{v}^\mu = \sum_{\mu=1}^{N} c^\mu (\lambda_\mu)^n \mathbf{v}^\mu. \tag{3.6}$$

Finally since all the eigenvalues λ_μ with $\mu \neq 1$ satisfy $|\lambda_\mu| \leq |\lambda_1|$ and since the eigenvector associated with the eigenvalue λ_1 is the only one with non-negative elements,it follows that,as long as $\lambda_1 > 0$

$$\lim_{n\to\infty} \frac{\mathbf{x}^{(n)}}{\sum_i x_i^{(n)}} = \frac{\mathbf{v}^1}{\sum_i (\mathbf{v}^\mathbf{1})_i}. \tag{3.7}$$

Therefore the eigenvector centrality $\mathbf{x} = c'\mathbf{v}^1$ where c' is a normalization constant. □

Any connected undirected network has an irreducible adjacency matrix. Therefore for the Ferron-Frobenious theorem in a connected undirected network every node i has a a positive eigenvector centrality $x_i > 0$. A directed network which is strongly connected has also an irreducible matrix, therefore also in this case every node i of the network has a positive eigenvector centrality $x_i > 0$. Nevertheless, if the directed network when there is no strong component in the network then $\lambda_1 = 0$ and therefore $\mathbf{x} = \mathbf{0}$. When, instead there is at least one non vanishing strongly connected component in the network $\mathbf{x} \neq \mathbf{0}$, but not all the nodes have positive centrality.

Proposition 3. *Given a directed network all the nodes of the in-component of the network have centrality $x_i = 0$.*

Proof. Consider the iterative procedure for constructing the eigenvector centrality starting form the initial guess $x_i^{(0)} = 1/N$ by calculating the vector $\mathbf{x}^{(n)}$ given by

$$x_i^{(n)} = \sum_j A_{ij} x_j^{(n-1)} \tag{3.8}$$

at every step $n \geq 1$ of the iteration. If we start from the leaves nodes i of the in-component that have in-degree zero, all these nodes have clearly zero centrality $x_i^{(n)} = 0$, at any iteration $n \geq 1$. Moreover all the nodes i of the in-components that are pointed only by the leaves nodes, must also necessarily

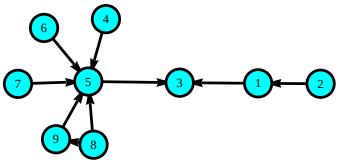

Figure 3.1: The pathology of the eigenvector centrality. In a directed network in which there is no strongly connected component, as in the example provided in this figure, all the nodes have zero centrality.

have zero centrality $x_i^{(n)} = 0$ for all $n \geq 2$. By assuming that all the nodes i in the in-component that are distant at maximum d from the leaves nodes have zero centrality $x_i^{(n)} = 0$ at any iteration $n \geq d$, it follows that all the nodes i in the in components that are distant at maximum $d+1$ from the leaves nodes have zero centrality $x_i^{(n)} = 0$ as well for any $n \geq d+1$, in fact at any iteration $n \geq d+1$ they are only pointed by nodes with zero eigenvector centrality. □

This property of the eigenvector centrality defined on directed network is usually an undesired feature of this centrality definition. Therefore while in social network studies the eigenvector centrality is quite popular to analyse undirected network datasets, it use is limited to analyse directed network datasets.

3.4 Katz Centrality (E)

In order to solve the undesired property of the eigenvector centrality , i.e. the vanishing of the eigenvector centrality for the nodes in the in-component of directed network, the Katz centrality has been proposed. The Katz centrality assign to each node a small centrality value β just for being a node of the network, then the centrality of the node increases if many important nodes point to it.

Definition 47. *The* Katz centrality $\mathbf{x}$ *satisfies the following equation*

$$x_i = \alpha \sum_{j=1}^{N} A_{ij} x_j + \beta \tag{3.9}$$

where $\beta > 0$, $\alpha \in (0, 1/\lambda_1)$ *where* λ_1 *is the Perron-Frobenious eigenvalue of the adjacency matrix* $\mathbf{A}$.

Using the matrix formalism we can write Eq. (3.9) as

$$\mathbf{x} = \alpha \mathbf{A}\mathbf{x} + \beta \mathbf{1}, \tag{3.10}$$

where **1** indicates the column vector such that $1_i = 1\ \forall i = 1, 2, \ldots, N$. By performing simple algebraic calculation it is possible to express the Katz centrality $\mathbf{x}$ as

$$\mathbf{x} = \left(\mathbb{I} - \alpha\mathbf{A}\right)^{-1} \beta\mathbf{1} = \beta \sum_{n=0}^{\infty} \left(\alpha\mathbf{A}\right)^n \mathbf{1}. \tag{3.11}$$

The matrix $\left(\mathbb{I} - \alpha\mathbf{A}\right)^{-1}$ diverges for $\det(\mathbb{I} - \alpha\mathbf{A}) = 0$. Since we have $\mathbb{I} - \alpha\mathbf{A} = \alpha(\alpha^{-1}\mathbb{I} - \mathbf{A})$ we have that $\left(\mathbb{I} - \alpha\mathbf{A}\right)^{-1}$ diverges when α is the inverse of an eigenvalue of the adjacency matrix $\mathbf{A}$. Therefore, in order to ensure the convergence of $\left(\mathbb{I} - \alpha\mathbf{A}^{-1}\right)$ and a well defined Katz centrality we must consider for the parameter α the following range: $\alpha \in (0, 1/\lambda_1)$.

3.5 PageRank Centrality (E)

The PageRank centrality is the main algorithm beyond the Google search engine, and has played a key role in determining the success of Google. The PageRank centrality is based on the same idea as the Katz centrality, i.e. if many important nodes j point to node i, node i increases its centrality. Nevertheless in the World-Wide-Web we have sometime very important webpages with many links, and in general the downstream nodes of these links are not more important than the important nodes from which the link is coming. In order to account for this fact, in the PageRank the downstream nodes only acquire a fraction of the centrality of the very central node that is pointing to them. Therefore we have the following definition

Definition 48. *The* PageRank centrality $\mathbf{x}$ *satisfies the following equation*

$$x_i = \alpha \sum_{j=1}^{N} A_{ij} \frac{1}{\kappa_j^{out}} x_j + \beta \tag{3.12}$$

where $\kappa_j^{out} = \max(k_j^{out}, 1)$, $\beta > 0$ *and* $\alpha \in (1, 1/\lambda_1')$ *where* λ_1' *is the Perron-Frobenious eigenvalue of the matrix* $\mathbf{A}\mathbf{D}^{-1}$ *with* $\mathbf{D}$ *given by the diagonal matrix of elements* $D_{ii} = \kappa_i^{out} = \max(k_i^{out}, 1)$.

In the matrix formalism we obtain the expression

$$\mathbf{x} = \alpha\mathbf{A}\mathbf{D}^{-1}\mathbf{x} + \beta\mathbf{1} \tag{3.13}$$

where **1** is the column vector with elements $1_i = 1\ \forall i = 1, 2 \ldots, N$. By performing simple algebraic calculation it is possible to express the Katz centrality $\mathbf{x}$ as

$$\mathbf{x} = \left(\mathbb{I} - \alpha\mathbf{A}\mathbf{D}^{-1}\right)^{-1} \beta\mathbf{1} = \beta \sum_{n=0}^{\infty} \left(\alpha\mathbf{A}\mathbf{D}^{-1}\right)^n \mathbf{1}. \tag{3.14}$$

The condition $\alpha < 1/\lambda_1'$ guarantees that the PageRank centrality is well defined for every node of the network. In undirected network $\lambda_1' = 1$. Therefore $\alpha \in (0,1)$, in directed network λ_1' is order one. In the original PageRank algorithm of Google $\alpha \simeq 0.85$. Let us now show that indeed $\lambda_1' = 1$ for a undirected network.

Theorem 3.5.1. *In an undirected network the maximal eigenvalue λ_1' of the matrix $\mathbf{A}\mathbf{D}^{-1}$ is equal to one, i.e. $\lambda_1' = 1$.*

Proof. The matrix $\mathbf{A}\mathbf{D}^{-1}$ is irreducible and with non negative elements, therefore the Perron-Frobenious theorem applies. In this case the leading eigenvector is the only eigenvector with non vanishing elements. The us show that the vector $\mathbf{x}$ with elements $x_i = k_i \geq 0$ is the eigenvector of $\mathbf{A}\mathbf{D}^{-1}$ associated with the eigenvalue $\lambda_1' = 1$. In fact if $\kappa_j = \max(1, k_j)$ we have, for the definition of the degree of a node

$$\sum_j A_{ij} \frac{1}{\kappa_j} k_j = k_i = \lambda_1' k_i, \tag{3.15}$$

with $\lambda_1' = 1$. It follows then that $\mathbf{x}$ is the leading eigenvector and that $\lambda_1' = 1$ is the leading eigenvalue. □

3.6 Example of calculation of the centrality of the nodes

Numerically the calculation of the centrality measures of the nodes are evaluated within a specific error by considering the following approximations

$$\begin{aligned}
\mathbf{x}_{eig} &= C\mathbf{A}^n\mathbf{1} \\
\mathbf{x}_{Katz} &= \beta \sum_{n'=0}^{n} (\alpha\mathbf{A})^{n'}\mathbf{1} \\
&= \beta[\mathbb{I} + \alpha\mathbf{A} + \alpha^2\mathbf{A}^2 + \ldots \alpha^n\mathbf{A}^n]\mathbf{1} \\
\mathbf{x}_{PageRank} &= \beta \sum_{n'=0}^{n} \left(\alpha\mathbf{A}\mathbf{D}^{-1}\right)^{n'}\mathbf{1} \\
&= \beta\left[\mathbb{I} + \alpha\mathbf{A}\mathbf{D}^{-1} + \alpha^2(\mathbf{A}\mathbf{D}^{-1})^2 + \ldots \alpha^n(\mathbf{A}\mathbf{D}^{-1})^n\right]\mathbf{1}.
\end{aligned} \tag{3.16}$$

In specific exercises we might aim at calculating exactly the centralities of the nodes. Two examples are given below.

3.6.1 First example

Let us consider the directed network with adjacency matrix

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

Eigenvector centrality (E)

The network described by the adjacency matrix $\mathbf{A}$ given by Eq. (3.17), is a directed network without any strongly connected components, therefore $x_i = 0$ $\forall i = 1, 2, 3$. To see how the iterative procedure for calculation $x_i^{(n)}$ work in this case we start with the "democratic ansatz" $x_i^{(0)} = 1/3$ $\forall i = 1, 2, 3$. Using $\mathbf{x}^{(n)} = \mathbf{A}^n \mathbf{x}^{(0)}$, we obtain

$$\mathbf{x}^{(1)} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ \frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix}$$
$$\mathbf{x}^{(2)} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{2}{3} \\ \frac{1}{3} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} \\ 0 \\ 0 \end{pmatrix}$$
$$\mathbf{x}^{(3)} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Therefore $\mathbf{x}^{(n)} = \mathbf{0}$ for $n \geq 3$.

Katz centrality

The Katz centrality $\mathbf{x}$ can be calculated by

$$\mathbf{x} = \beta \left(\mathbb{I} - \alpha \mathbf{A} \right)^{-1} \mathbf{1} = \beta \sum_{n=0}^{\infty} \alpha^n \mathbf{A}^n \mathbf{1}. \tag{3.17}$$

Now we have by definition $\mathbf{A}^0 = \mathbb{I}$, moreover $\mathbf{A}^2$ and $\mathbf{A}^3$ are given by

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

and therefore $\mathbf{A}^n = 0$ for $n \geq 3$. Using the Eq. (3.17), we have therefore

$$\begin{aligned} \mathbf{x} &= \beta \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \alpha \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \alpha^2 \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \right] \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \\ &= \beta \begin{pmatrix} 1 + 2\alpha + \alpha^2 \\ 1 + \alpha \\ 1 \end{pmatrix}. \end{aligned}$$

PageRank Centrality (E)

The PageRank centrality $\mathbf{x}$ can be calculated by

$$\mathbf{x} = \beta \left(\mathbb{I} - \alpha \mathbf{A} \mathbf{D}^{-1} \right)^{-1} \mathbf{1} = \beta \sum_{n=0}^{\infty} \alpha^n \left(\mathbf{A} \mathbf{D}^{-1} \right)^n \mathbf{1}. \tag{3.18}$$

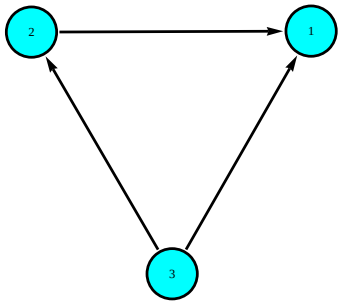

Figure 3.2: The graphical representation of the network with adjacency matrix **A** given by Eq. (3.17).

with $D_{ii} = \max(1, k_i^{out})$. Therefore we have

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \quad \mathbf{D}^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix}.$$

Therefore the matrices $\left(\mathbf{AD}^{-1}\right)^n$ are given by

$$\mathbf{AD}^{-1} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix},$$
$$(\mathbf{AD}^{-1})^2 = \begin{pmatrix} 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

and finally $\left(\mathbf{AD}^{-1}\right)^n = 0$ for $n \geq 3$. Therefore the PageRank centrality is given by

$$\begin{aligned} \mathbf{x} &= \beta \left[\mathbb{I} + \alpha \mathbf{AD}^{-1} + \alpha^2 \left(\mathbf{AD}^{-1}\right)^2\right] \mathbf{1} \\ &= \beta \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \alpha \begin{pmatrix} 0 & 1 & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix} + \alpha^2 \begin{pmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right] \\ &= \beta \begin{pmatrix} 1 + \frac{3}{2}\alpha + \frac{1}{2}\alpha^2 \\ 1 + \frac{1}{2}\alpha \\ 1 \end{pmatrix}. \end{aligned}$$

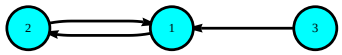

Figure 3.3: The graphical representation of the network with adjacency matrix **A** given by Eq. (3.17).

3.6.2 Second example (NE)

Let us consider the directed network with adjacency matrix

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

This network is shown in Figure 3.3. It has a strongly connected component induced by the nodes $\{1, 2\}$. In similar cases, when the network includes a very large number of nodes N, we might still use the expansion given by Eq. (3.16) to calculate numerically the eigenvector, Katz and the PageRank centralities, but for small networks it is possible to work directly with the eigenvectors and the inverse of the matrices finding exact results as shown in this case.

Eigenvector centrality (NE)

The network described by the adjacency matrix **A** given by Eq. (3.19), is a directed network has a strongly connected component, induced by nodes $1, 2$. Instead node 3 is in the in-component , therefore we expect $x_3 = 0$. To see this let us find the eigenvector $\mathbf{v}_1$ corresponding to the largest eigenvalue of the matrix.

Let us first evaluate the largest eigenvalue λ_1. To this end, let us find the spectrum of the matrix by solving the eigenvalue problem

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0. \tag{3.19}$$

which reads

$$\det(\mathbf{A} - \lambda\mathbf{I}) = \begin{vmatrix} -\lambda & 1 & 0 \\ 1 & -\lambda & 1 \\ 0 & 0 & -\lambda \end{vmatrix}. = -\lambda(\lambda^2 - 1) = 0.$$

This equation can be solved giving the eigenvalues $\lambda_1 = 1, \lambda_2 = 0, \lambda_3 = -1$ with $\lambda_1 > \lambda_2 > \lambda_3$. Therefore the leading eigenvalue is $\lambda_1 = 1$. The corresponding eigenvector $\mathbf{v}_1$ can be found by solving the equation

$$\mathbf{A}\mathbf{v}_1 = \lambda_1\mathbf{v}_1 \tag{3.20}$$

which reads

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 1 \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

where

$$\mathbf{v}_1 = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}. \tag{3.21}$$

Eq. (3.21) has solution

$$\mathbf{v}_1 = C \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \tag{3.22}$$

where C is a normalization constant fixed by the condition $\sum_i x_i = 1$. We have therefore $C = 1/2$ and

$$\mathbf{v}_1 = \frac{1}{2} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}. \tag{3.23}$$

As expected we have $x_3 = 0$.

Katz centrality

The Katz centrality $\mathbf{x}$ can be calculated by

$$\mathbf{x} = \beta \left(\mathbb{I} - \alpha \mathbf{A}\right)^{-1} \mathbf{1} = \beta \sum_{n=0}^{\infty} \alpha^n \mathbf{A}^n \mathbf{1}. \tag{3.24}$$

Now, it is convenient for small networks with a strongly connected component to calculate directly $\beta \left(\mathbb{I} - \alpha \mathbf{A}\right)^{-1} \mathbf{1}$ with our using the power expansion. Let us first calculate $\mathbb{I} - \alpha \mathbf{A}$. This matrix is given by

$$\mathbb{I} - \alpha \mathbf{A} = \begin{pmatrix} 1 & -\alpha & 0 \\ -\alpha & 1 & -\alpha \\ 0 & 0 & 1 \end{pmatrix}$$

The determinant of the matrix is given by

$$\det(\mathbb{I} - \alpha \mathbf{A}) = 1 - \alpha^2. \tag{3.25}$$

The the matrix of cofactors is given by

$$\mathbf{C} = \begin{pmatrix} 1 & \alpha & 0 \\ \alpha & 1 & 0 \\ \alpha^2 & \alpha & 1 - \alpha^2 \end{pmatrix}.$$

The inverse is given by

$$\det(\mathbb{I}-\alpha\mathbf{A})^{-1}=\frac{1}{1-\alpha^2}\begin{pmatrix}1 & \alpha & \alpha^2\\ \alpha & 1 & \alpha\\ 0 & 0 & 1-\alpha^2\end{pmatrix}.$$

Therefore the Katz centrality $\mathbf{x}$ is given by

$$\mathbf{x}=\beta\left(\mathbb{I}-\alpha\mathbf{A}\right)^{-1}\mathbf{1} \tag{3.26}$$

i.e.

$$\begin{pmatrix}x_1\\ x_2\\ x_3\end{pmatrix}=\frac{1}{1-\alpha^2}\begin{pmatrix}1 & \alpha & \alpha^2\\ \alpha & 1 & \alpha\\ 0 & 0 & 1-\alpha^2\end{pmatrix}\begin{pmatrix}1\\ 1\\ 1\end{pmatrix}=\frac{\beta}{1-\alpha^2}\begin{pmatrix}1+\alpha+\alpha^2\\ 1+2\alpha\\ 1-\alpha^2\end{pmatrix}.$$

PageRank Centrality (NE)

The PageRank centrality $\mathbf{x}$ can be calculated by

$$\mathbf{x}=\beta\left(\mathbb{I}-\alpha\mathbf{A}\mathbf{D}^{-1}\right)^{-1}\mathbf{1}=\beta\sum_{n=0}^{\infty}\alpha^n\left(\mathbf{A}\mathbf{D}^{-1}\right)^n\mathbf{1}. \tag{3.27}$$

with $D_{ii}=\max(1,k_i^{out})$. Again here for a small network with a strongly connected component we can solve exactly for the PageRank centrality by inverting the matrix $\mathbb{I}-\alpha\mathbf{A}\mathbf{D}^{-1}$. Let us calculate the matrix $\mathbf{D}$. This is given by

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

Therefore in this case the PageRank centrality is the same as the Katz centrality and is given by

$$\mathbf{x}=\frac{\beta}{1-\alpha^2}\begin{pmatrix}1+\alpha+\alpha^2\\ 1+2\alpha\\ 1-\alpha^2\end{pmatrix}. \tag{3.28}$$

3.7 Closeness Centrality and Efficiency (E)

The main idea beyond the definition of closeness centrality is to rank the importance of a node depending on its distance to the other nodes of the network. Therefore the closeness centrality is smaller for nodes that have larger average distance to the other nodes of the network.

Definition 49. *The* Closeness Centrality Cl_i *of a node i in a undirected network is given by*

$$Cl_i=\frac{N-1}{\sum_j d_{ij}}=\frac{1}{\frac{1}{N-1}\sum_j d_{ij}}=\frac{1}{\ell_i} \tag{3.29}$$

where d_{ij} is the shortest distance between node i and node j, and ℓ_i is the averaged shortest distance between node i and the other nodes of the network. The normalization to $N-1$ takes into account that we have always $d_{ii}=0$

The closeness centrality has the following properties:

- *It has small dynamic range.*
  Since most networks are characterized by the " small world" properties and have typically small average distances, the values of the closeness centrality spans a relatively small dynamic range in any small world network.

- *It is zeros if the network is disconnected.*
  If the network is disconnected for every node i there will be a node j such that $d_{ij}=\infty$. In this case the closeness centrality is zero. To avoid this problem, the closeness centrality of disconnected network is calculated for every node i only considering the nodes j belonging to the connected component $\mathcal{C}(i)$ of the network including node i, i.e.

$$Cl_i = \frac{|\mathcal{C}(i)-1|}{\sum_{j\in\mathcal{C}(i)} d_{ij}}. \tag{3.30}$$

An alternative definition of the closeness centrality is also called the *Efficiency* defined as in the following

Definition 50. *The* efficiency E_i *of a node i in a undirected network is given by*

$$E_i = \frac{1}{N-1}\sum_{j\neq i}\frac{1}{d_{ij}}. \tag{3.31}$$

The efficiency, is non vanishing also for disconnected networks, but has always numerical values spanning a small dynamic range in small world networks. Using this definition one can define the global efficiency E of a network defined as in the following

Definition 51. *The* global efficiency *E of a network is the averaged of the efficiencies E_i of its nodes, i.e.*

$$E = \frac{1}{N}\sum_{i=1}^{N} E_i = \frac{1}{N(N-1)}\sum_{i,j|i\neq j}\frac{1}{d_{ij}}. \tag{3.32}$$

3.8 Betweenness Centrality (E)

The *betweenness centrality* of a node i is very large if node i lies on many shortest paths between the other nodes of the network.

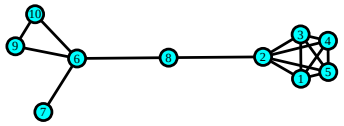

Figure 3.4: Node 8 is a node with small degree $k_8 = 2$ but large betweenness centrality $b_i = 10^2 - 5^2 - 4^2 = 59$because it connects two groups of nodes formed respectively by 5 and 4 nodes that are otherwise disconnected.

Definition 52. *The* betweenness centrality b_i *of node* i *in a network is given by*

$$b_i = \sum_{r,s} \frac{n_{rs}^i}{g_{rs}} \tag{3.33}$$

where n_{rs}^i *are the number of shortest paths between node* r *and node* s *that pass through node* i*, and* g_{rs} *are the total number of shortest paths between node* r *and node* s*.*

The betweenness centrality is not in general related with the degree of a node, in fact a low degree node connecting two otherwise disconnected regions of the network has high betweenness centrality (See Figure 3.4). Moreover, the betweenness centrality can acquire a wide range of values also if we consider only connected and undirected networks.
Given a network of N nodes, the maximal value of the betweenness is obtained for the central node of a star network with one central node and $N-1$ leaves. For any pair of nodes r, s since the star network is a tree we have a single shortest path linking the two nodes, i.e. $g_{rs} = 1$. Moreover all the paths joining nodes $r \neq s$ pass through the central node. Therefore only the paths starting from nodes r different from the central node and arriving at r do not pass through the central node. Therefore the betweenness centrality of the central node i of such star network is given by

$$b_i = N^2 - (N-1) = N^2 - N + 1. \tag{3.34}$$

The minimal betweenness of a node in a connected and undirected network of N nodes is given by the betweenness centrality of a leaf node. The leaf node i lies only on shortest paths that start or end with itself. These paths are given by $(N-1) + N = 2N - 1$ because there are $N-1$ shortest paths starting from other nodes and ending in i and there are N shortest paths starting from node i. Therefore the ratio between the largest and lowest betweenness centrality in an undirected and connected network is given by

$$\frac{N^2 - N + 1}{2N - 1} \simeq \frac{1}{2} N \tag{3.35}$$

where we have approximated the above ration in the limit $N \gg 1$ of a large network. It follows that the betweenness centrality of a undirected connected network can spam a wide range of values.

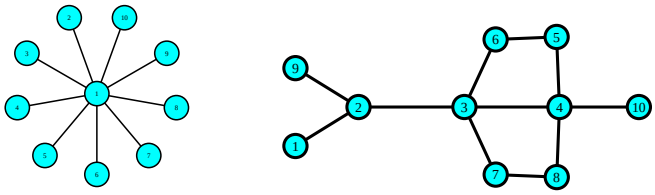

Figure 3.5: (left panel) A star network of $N = 10$ nodes, the node 1 is linked to all the others $N-1 = 9$ nodes and has betweenness centrality $b_1 = N^2-N+1 = 91$. (right panel) Node 1 in the network on the right is a leaf node, i.e. a node of degree 1, therefore $b_1 = (2N-1) = 19$.

3.9 Review of Algebra

3.9.1 Perron-Frobenious Theorem (E)

Here we will state the Perron-Frobenious Theorem, which is very important to guarantee the desired properties for the eigenvector centrality. Before stating the Perron-Frobenious Theorem valid for a square matrix $\mathbf{A}$, with non-negative elements $A_{ij} \geq 0$, we define the irreducible matrices.

Definition 53. *An square matrix with non negative matrix elements $A_{ij} > 0$ is called* irreducible *if, for every pair of indices (i, j), there is a value of $n = n(i, j)$ such that $[\mathbf{A}^n]_{ij} > 0$.*

If $\mathbf{A}$ is an adjacency matrix and is irreducible, the underlying network must be formed by a unique strongly connected component. In fact the condition $\left[\mathbf{A}^N\right]_{ij} > 0$, implies that there is at least one path of length n going from node j to node i. It follows that if the adjacency matrix $\mathbf{A}$ of a network is irreducible, it must be possible from each node j of the network to reach any node i of the network in a finite number of steps, i.e. the network is formed by a unique strongly connected component.

Theorem 3.9.1. *Let $\mathbf{A}$ be a $N \times N$ irreducible matrix.*
Then the Perron-Frobenious theorem *states that:*

- $\mathbf{A}$ *has a real positive eigenvalue λ_1 such that all other eigenvalues λ_i with $i = 2, \ldots, N$ satisfy*

$$\lambda_1 \geq |\lambda_i|. \tag{3.36}$$

- *The eigenvalue λ_1 has algebraic and geometric multiplicity equal to one, and has an left eigenvector $\mathbf{x}$ with all positive elements, i.e. $x_i > 0\ \forall i = 1, 2 \ldots, N$ and a left eigenvector $\mathbf{v}$ with all positive elements, i.e. $v_i = 0\ \forall i = 1, 2 \ldots, N$.*

- *Any non-negative right eigenvector is a multiplex of $\mathbf{x}$, any non-negative left eigenvector is a multiple of $\mathbf{v}$.*

Moreover, if $\mathbf{A}$ is the adjacency matrix of a directed network with a single strongly connected component, we have that

- $\mathbf{A}$ has a real positive eigenvalue λ_1 such that all other eigenvalues λ_i with $i = 2, 3, \ldots, N$ satisfy

$$\lambda_1 \geq |\lambda_i|. \tag{3.37}$$

- The eigenvalue λ_1 has algebraic and geometric multiplicity equal to one, and has an left eigenvector $\mathbf{x}$ with non-negative elements, with $x_i > 0$ is node i belongs to the strongly connected component of the network, or the out-components of the network, and $x_i = 0$ if node i belongs to the in-component of the strongly connected component of the network.

3.9.2 Introduction to Matrix Functions (E)

Let $f(z)$ be a complex valued function of $z \in \mathbb{C}$.
Let $\mathbf{A}$ be a $N \times N$ complex valued matrix.
If $f(z)$ is analytic in the disk $|z| < R$, it can be represented as a convergent Taylor series

$$f(z) = \sum_{n=0}^{\infty} c_n z^n \quad \text{for} \quad |z| < R. \tag{3.38}$$

A formal substitution of the $N \times N$ matrix into this series yields the $N \times N$ matrix $\mathbf{f}(\mathbf{A})$ given by

$$\mathbf{f}(\mathbf{A}) = \sum_{n=0}^{\infty} c_n \mathbf{A}^n \tag{3.39}$$

where $\mathbf{A}^0 = \mathbb{I}$. We say that the series in Eq. (3.39) is convergent is all N^2 scalar elements that make the matrix $\mathbf{f}(\mathbf{A})$ are convergent. For example we can consider the following matrix functions

$$\begin{aligned} (\mathbb{I} - \alpha\mathbf{A})^{-1} &= \sum_{n=0}^{\infty} \alpha^n \mathbf{A}^n \\ e^{\alpha\mathbf{A}} &= \sum_{n=0}^{\infty} \frac{1}{n!} \alpha^n \mathbf{A}^n. \end{aligned} \tag{3.40}$$

3.9.3 Inverse of a Matrix (NE)

The inverse of a square $N \times N$ invertible matrix $\mathbf{M}$, is a square $N \times N$ matrix indicated with $\mathbf{M}^{-1}$ that satisfies the equation

$$\mathbf{M}\mathbf{M}^{-1} = \mathbf{M}^{-1}\mathbf{M} = \mathbf{I}. \tag{3.41}$$

It can be found by applying the formula

$$\mathbf{M}^{-1} = \frac{1}{\det \mathbf{M}} \mathbf{C}^{\mathbf{T}} \tag{3.42}$$

where $\mathbf{C}$ is the matrix of cofactors and $\mathbf{C}^T$ represents the transpose of the matrix of cofactors. The matrix elements C_{ij} of the cofactor matrix $\mathbf{C}$ are given by the product of the sign $(-1)^{i+j}$ and the minor of the entry of the i-th row and j-th column. Here the minor indicates the determinant of the square matrix obtained from the matrix $\mathbf{M}$ by removing the i-th row and the j-th column.

Chapter 4

Random graphs

4.1 Introduction (E)

Many complex networks from the World-Wide-Web to the brain are intrinsically stochastic. In fact we put a new link on our webpage with a certain probability, using only partial information of the structure of the full network, and we do not follow a global design principle. Similarly, the structure of brain networks is not fully encoded in the genome, therefore stochastic effects play an essential role in brain development.

But are complex networks completely random?
The answer to this question is no, most complex networks are not completely random: they encode information in their structure and they follow complex organization principles related to their robustness and their efficiency. Network theory is essentially needed for discovering these non-random properties of complex networks.

In the their 1959 Erdös and Reńyi for the first time introduced probability arguments in graph theory. It took therefore several hundred years for graph theory (from Euler work in 1735 to Erdös and Rényi work of 1959) to take this important step.

The importance of random graph theory for network theory is capital and relies on the fact that random graph theory

- *a)* makes use of probabilistic arguments;
- *b)* characterizes the properties of networks in the limit of large network sizes $N \gg 1$.

These two aspects of random graph theory are fundamental in network theory, and are essential to characterize the evolution and the structure of not only random networks but also more complex and realistic network models.

Random graphs are models of maximally random graphs that have several important properties. In this chapter we will define random graphs, determine the distribution of the number of links and the degree distribution, characterize

the emergence of the giant component of the network as a function of the average degree of the nodes. Finally we will characterize the local structure of random graphs by finding the average number of important subgraphs such as loops or cliques.

4.2 Random Graph Ensembles (E)

In random graph theory two different ensembles of random graphs are considered. Here to be consistent with our use of terminology we will call these graphs random networks interchangeably.
A random graph ensemble is given when for every simple network $G = (V, E)$ of $N = |V|$ nodes it is assigned a probability $P(G)$.

The $\mathbb{G}(N, L)$ ensemble is formed by all the simple networks $G(V, E)$ with $N = |V|$ labelled nodes and $L = |E|$ links. In other words, all the networks with N nodes and L links are taken with equal probability while all the other networks have zero probability.
The total number of simple networks with N nodes and L links is given by $Z = \begin{pmatrix} N(N-1)/2 \\ L \end{pmatrix}$, therefore we have the following definition.

Definition 54. *In the $\mathbb{G}(N, L)$ ensemble the probability of a simple network $G = (V, E)$ is given by*

$$P(G) = \begin{cases} \frac{1}{Z} & \text{for } |V| = N \text{ and } |E| = L \\ 0 & \text{otherwise} \end{cases},$$

where $Z = \begin{pmatrix} N(N-1)/2 \\ L \end{pmatrix}$ is the total number of networks of the ensemble with non-zero probability.

The $\mathbb{G}(N, p)$ ensemble is formed by all the simple networks $G(V, E)$ with $N = |V|$ labelled nodes where each pair of nodes is linked with probability p.

Definition 55. *In the $\mathbb{G}(N, p)$ ensemble the probability of a simple network $G = (V, E)$ with total number of nodes $N = |V|$ is given by*

$$P(G) = p^L (1-p)^{N(N-1)/2-L} \quad \text{with } |E| = L. \tag{4.1}$$

Any network in this ensemble can be seen as a result of $N(N-1)/2$ independent coin tossings, one for each link, with a probability of success, (i.e. drawing a link) equal to p.

In the limit of large network limit $N \gg 1$, these the $\mathbb{G}(N, L)$ ensemble and the $\mathbb{G}(N, p)$ ensemble where we take $p\frac{N(N-1)}{2} = L$, are asymptotically equivalent, and share most of their statistical properties. For this reason, here we will restrict our attention to the $\mathbb{G}(N, p)$ ensemble,while leaving to the next chapters a more detailed treatment of these and other generalized random ensembles.

4.3 Distribution p_L of the number of links L and average number of links $\langle L \rangle$ of the $\mathbb{G}(N,p)$ ensemble, (E)

While in the $\mathbb{G}(N,L)$ ensemble all the networks with non-zero probability have the same number of links, in the $\mathbb{G}(N,p)$ ensemble different networks have different number of links. It is therefore important to calculate the distribution p_L of the number of links L and average number of links $\langle L \rangle$ of the random networks in the $\mathbb{G}(N,p)$ ensemble.

Proposition 4. *The probability p_L that a network in the $\mathbb{G}(N,p)$ ensemble has L links is a binomial distribution given by*

$$p_L = \left(\begin{array}{c} \frac{N(N-1)}{2} \\ L \end{array} \right) p^L (1-p)^{N(N-1)/2-L}, \tag{4.2}$$

i.e. $L \sim B\left(\frac{N(N-1)}{2}, p\right)$.

Proof. The total number of links in the ensemble can take any value L between zero and $N(N-1)/2$. Moreover, since each link of a network in the $\mathbb{G}(N,p)$ can be seen as a result of an independent coin flip,with probability of success given by p, the total number of links in one network realization of the $\mathbb{G}(N,p)$ ensemble can be seen as the result of $N(N-1)/2$ of such coin flips. Given a certain network in the $\mathbb{G}(N,p)$ ensemble the probability that it has L links to a given set of nodes is $p^L(1-p)^{N(N-1)/2-L}$. The number of possibilities of choosing L links out of $N(N-1)/2$ is given by $\left(\begin{array}{c} N(N-1)/2 \\ L \end{array} \right)$. Therefore it follows that the probability p_L is the binomial distribution given by Eq. (4.2), i.e. $L \sim \mathrm{B}\left(\frac{N(N-1)}{2}, p\right)$. □

Proposition 5. *The average number of links $\langle L \rangle$ of a network in the $\mathbb{G}(N,p)$ ensemble is given by*

$$\langle L \rangle = \sum_{L=0}^{N(N-1)/2} L p_L = p \frac{N(N-1)}{2}. \tag{4.3}$$

Proof. We can calculate the average $\langle L \rangle$ over the distribution p_L by using the generating function of the P_L distribution. The generating function $G_0(x)$ of the distribution p_L is given by

$$\begin{aligned} G_0(x) &= \sum_{L=0}^{N(N-1)/2} p_L x^L \\ &= \sum_{L=0}^{N(N-1)/2} \left(\begin{array}{c} N(N-1)/2 \\ L \end{array} \right) p^L (1-p)^{N(N-1)/2-L} x^L \\ &= (1-p+px)^{N(N-1)/2}, \end{aligned} \tag{4.4}$$

where we used the Newton binomial. Using the properties of the generating functions we have

$$\begin{aligned} \langle L \rangle &= G_0'(x)|_{x=1} = p\frac{N(N-1)}{2}\,(1-p+px)^{N(N-1)/2-1}\Big|_{x=1} \\ &= p\frac{N(N-1)}{2}. \end{aligned} \tag{4.5}$$

□

4.4 Degree distribution of the $\mathbb{G}(N,p)$ ensemble (E)

In the $\mathbb{G}(N,p)$ ensemble, the nodes have in general different degree. Therefore it is important to characterize the degree distribution of the network.

Proposition 6. *The degree distribution $P(k)$ of the $\mathbb{G}(N,p)$ ensemble is a binomial distribution given by*

$$P(k) = \left(\begin{array}{c} N-1 \\ k \end{array} \right) p^k (1-p)^{N-1-k}, \tag{4.6}$$

i.e. $k \sim B(N-1,p)$.

Proof. The degree of a node take any number k between zero and $N-1$. Moreover, since each link of a network in the $\mathbb{G}(N,p)$ can be seen as a result of an independent coin flip probability of success given by p, the degree of each node can be seen as the result of $N-1$ of such coin flips. Given a certain node the probability that it is linked to other k given nodes is $p^k(1-p)^{N-1-k}$. The number of possibilities of choosing k nodes out of $N-1$ is given by $\left(\begin{array}{c} N-1 \\ k \end{array} \right)$. Therefore it follows that the degree distribution is the binomial distribution given by Eq. (4.2), i.e. $k \sim \mathrm{B}(N-1,p)$. □

Proposition 7. *The average degree $\langle k \rangle$ and the second moment of the degree distribution $\langle k(k-1) \rangle$ of a network in the $\mathbb{G}(N,p)$ ensemble are given by*

$$\begin{aligned} \langle k \rangle &= \sum_{k=0}^{N-1} kP(k) = p(N-1), \\ \langle k(k-1) \rangle &= \sum_{k=0}^{N-1} k(k-1)P(k) = p^2(N-1)(N-2). \end{aligned} \tag{4.7}$$

Proof. We can calculate $\langle k \rangle$ and $\langle k(k-1) \rangle$ over the distribution $P(k)$ given by Eq. (4.6) using the generating function of the degree distribution $P(k)$. The

generating function $G_0(x)$ of the distribution $P(k)$ is given by

$$\begin{aligned} G_0(x) &= \sum_{k=0}^{N-1} P(k)x^k \\ &= \sum_{k=0}^{N-1} \left(\begin{array}{c} N-1 \\ k \end{array} \right) p^k (1-p)^{N-1-k} x^k \\ &= (1-p+px)^{N-1}, \end{aligned} \tag{4.8}$$

where we used the Newton binomial. Using the properties of the generating functions we have

$$\begin{aligned} \langle k \rangle &= G_0'(x)|_{x=1} = p(N-1)\ (1-p+px)^{N-1-1}\Big|_{x=1} \\ &= p(N-1). \end{aligned} \tag{4.9}$$

Moreover we have

$$\begin{aligned} \langle k(k-1) \rangle &= G_0''(x)|_{x=1} = p^2(N-1)(N-2)\ (1-p+px)^{N-3}\Big|_{x=1} \\ &= p^2(N-1)(N-2). \end{aligned} \tag{4.10}$$

□

4.5 Poisson networks (E except proof of Prop 8)

In many cases we are interested in characterizing large networks, with $N \gg 1$. Moreover in many cases it is important to compare networks of different size N but with the same average degree $\langle k \rangle = p(N-1)$. In the framework of random graph ensemble it is possible to consider the implications of the complete randomness of the interactions as $N \to \infty$ while $p(N-1) = c$.

Proposition 8. *The degree distribution of a network in the $\mathbb{G}(N,p)$ ensemble, with $p = \frac{c}{N-1}$, and c is independent on N, can be approximated, in the large network limit (i.e. $N \to \infty$) by a Poison distribution with average c, notably*

$$P(k) = \frac{1}{k!} c^k e^{-c}, \tag{4.11}$$

i.e. $k \sim Poi(c)$. These networks with Poisson degree distribution are also called Poisson networks.

Proof. The degree of a network in $\mathbb{G}(N,p)$ ensemble, with $p = \frac{c}{N-1}$ and c independent on N follows the binomial distribution

$$\begin{aligned} P(k) &= \left(\begin{array}{c} N-1 \\ k \end{array} \right) p^k (1-p)^{N-1-k} \\ & \frac{1}{k!} \frac{(N-1)!}{(N-1-k)!} p^k (1-p)^{N-1-k}. \end{aligned} \tag{4.12}$$

i.e. $k \sim \mathrm{B}(N-1,p)$. Let us now show that in the limit $N \to \infty$, we have

$$\ln\left[\frac{(N-1)!}{(N-1-k)!}p^k(1-p)^{N-1-k}\right] \to \ln[e^{-c}c^k]. \tag{4.13}$$

In fact using the Stirling approximation for the factorials $\ln n! \simeq n\log n - n$, the fact that $N-1-k \simeq N-1$ for $(N-1) \gg k$ and the expansion $\ln(1+x) \simeq x$ for $x \ll 1$ we get

$$\begin{aligned}
\ln\left[\frac{(N-1)!}{(N-1-k)!}p^k(1-p)^{N-1-k}\right] &\simeq (N-1)\ln(N-1) - (N-1) - (N-1-k)\ln(N-1-k) \\
&\quad +(N-1-k) + k\ln\left[\frac{c}{N-1}\right] \\
&\quad +(N-1-k)\ln\left[1-\frac{c}{N-1}\right] \\
&\simeq (N-1)\ln(N-1) - (N-1)\log(N-1-k) \\
&\quad +k\ln(N-1-k) - k + k\log c - k\ln(N-1) - N\frac{c}{N} \\
&\simeq -(N-1)\ln\left(1-\frac{k}{N-1}\right) - k + k\ln c - c \\
&\simeq k\log c - c = \log\left[c^k e^{-c}\right]. \qquad (4.14)
\end{aligned}$$

It follows that in the limit $N \to \infty$ the degree distribution is a Poisson distribution with average degree $\langle k\rangle = c$.

$$P(k) = \frac{1}{k!}c^k e^{-c} \tag{4.15}$$

□

In figure 4.1 we plot the degree distribution $P(k)$ with average degree $\langle k\rangle = c$.

Proposition 9. *The average degree $\langle k\rangle$ and the second moment of the degree distribution $\langle k(k-1)\rangle$ of a network in the $\mathbb{G}(N,p)$ ensemble with $p = \frac{c}{N-1}$ are given, in the large network limit $N \gg 1$ by*

$$\begin{aligned}
\langle k\rangle &= \sum_{k=0}^{N-1} kP(k) = c, \\
\langle k(k-1)\rangle &= \sum_{k=0}^{N-1} k(k-1)P(k) = c^2. \qquad (4.16)
\end{aligned}$$

Therefore the standard deviation of the degree distribution $\sigma = \sqrt{\langle k^2\rangle - \langle k\rangle^2}$ is given by

$$\sigma = \sqrt{c}. \tag{4.17}$$

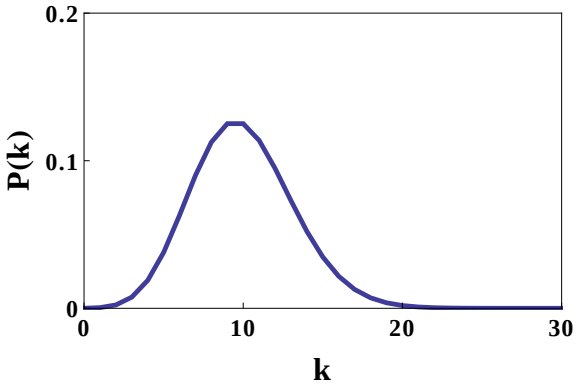

Figure 4.1: A Poisson degree distribution $P(k)$ with average degree $\langle k \rangle = c$.

Proof. In the hypothesis that $p = \frac{c}{N-1}$ and that $N \gg 1$ the degree distribution of a network in the $\mathbb{G}(N,p)$ ensemble, can be approximated by a Poisson distribution, i.e. $k \sim \text{Poi}(c)$. Therefore can calculate the $\langle k \rangle$ and $\langle k(k-1) \rangle$ over the Poisson degree distribution $P(k)$ given by Eq. (4.11) using the generating function approach. The generating function $G_0(x)$ of the Poisson degree distribution $P(k)$ is given by

$$\begin{aligned} G_0(x) &= \sum_{k=0}^{\infty} P(k) x^k \\ &= \sum_{k=0}^{\infty} \frac{1}{k!} c^k e^{-c} x^k \\ &= e^{-c+cx} = e^{-c(1-x)}, \end{aligned} \tag{4.18}$$

where we used the Taylor expansion of the exponential.
Using the properties of the generating function we find

$$\begin{aligned} \langle k \rangle &= G_0'(x)|_{x=1} = c e^{-c(1-x)}\Big|_{x=1} \\ &= c \\ \langle k(k-1) \rangle &= G_0''(x)|_{x=1} = c^2 e^{-c(1-x)}\Big|_{x=1} \\ &= c^2. \end{aligned} \tag{4.19}$$

Therefore we have that the variance of the degree distribution σ^2 is given by

$$\begin{aligned} \sigma^2 &= \langle k^2 \rangle - \langle k \rangle^2 = \langle k(k-1) \rangle + \langle k \rangle - \langle k \rangle^2 \\ &= c^2 + c - c^2 = c, \end{aligned} \tag{4.20}$$

and the standard deviation is given by $\sigma = \sqrt{c}$. □

It is easy to see that the expressions found for $\langle k \rangle$ and $\langle k(k-1) \rangle$ in Eqs. (4.16) coincide with the expression found previously Eqs. (4.7) for $p = \frac{c}{N-1}$ and $N \gg 1$. In order to show that in this network large fluctuation in the degree are not allowed, let us assume that a social network is described by a Poisson network with average degree $\langle k \rangle = 100$. Then the standard deviation σ of the degree distribution is given by $\sigma = \sqrt{c} = \sqrt{100} = 10$. Therefore observing people with $k = 1000$ social contacts would correspond to an event distant 90σ from the average and it would be very unlikely in the network!
Since such large fluctuations in the degree of individulas in social networks are observed, the Poisson network cannot be a very good model for social networks. Moreover this observation can be extended to any complex network characterized by large fluctuations in the degrees of the nodes.

4.6 Emergence of the giant component in random graphs (E)

If we consider the connected components of a Poisson random graph as a function of the average degree $\langle k \rangle = c$ of the network we observe a dramatic structural change at $c = 1$. In fact for $c < 1$ the network is formed by a large number of small connected components, formed by trees. Instead for $c > 1$ the largest connected component acquires and extensive number of nodes, i.e. a number of nodes of the same order of magnitude as N. This largest connected component is called the giant component of the network. Moreover the giant component is not any more tree-like and contains large loops. This transition occurs exactly at $c = 1$ in the limit $N \to \infty$, and this dramatic change of structure in the network is an example of "phase transition" in the structure of the network. (In physics phase transitions correspond to different phases of matter and describe phenomena like the gas-liquid phase transition, or the magnetism of certain materials).

Given a complex system, it usually an important requirement that the network contains a giant component because it is important to have paths joining a large fraction of nodes in the network.

4.6.1 Giant component

A central role in network theory is played by the *giant component* of a network.

Definition 56. *A network has a* giant component *when its largest connected component $H = (V', E')$ is induced by a set of nodes V' formed by an extensive number of nodes, i.e.*

$$|V'| \sim \mathcal{O}(N). \tag{4.21}$$

In this case the giant component of the network coincides with the largest connected component.

4.6.2 Emergence of the giant component in Poisson random networks

In the structure of random networks with average degree $\langle k\rangle = c$ we observe a "phase transition" as a function of c. In fact when $c = 0$, every node has degree zero and is in a different disconnected component. In this case we do not have a giant component in the network. As the value c of the average degree of the network increases, we observe the emergence of the giant component.

Proposition 10. *A random network in the $\mathbb{G}(N,p)$ with average degree $\langle k\rangle \to c$ for $N \to \infty$ contains n the limit $N \to \infty$ a fraction of nodes S in the giant component determined by the equation*

$$S = 1 - e^{-cS}. \tag{4.22}$$

Proof. Let us consider a random network in the $\mathbb{G}(N,p)$ ensemble with $p = \frac{c}{N-1}$ and with a number of nodes $N \gg 1$. In the limit $N \to \infty$ this network has an average degree $\langle k\rangle = p(N-1)$ tends to $\lim_{N\to\infty} p(N-1) = \lim_{N\to\infty} c$. Therefore if c is a constant $\langle k\rangle \to c$ as $N \to \infty$. If a fraction of nodes S is in the giant component, S can be also interpreted as the probability that a random node in the network belongs to the giant component. A node i of a $\mathbb{G}(N,p)$ is *not in the giant component* if none of its links is connected to nodes that are part of the giant component.
This means that for every other node $j \neq i$ of the network one of the two following conditions must be verified :

- *(a)* there is no link joining node i to node j;
- *(b)* there is a link between node i and node j but node j *is not* in the giant component.

If $p = \frac{c}{N-1}$ is the probability that any two nodes of the network are linked, and S is the probability that a random node in the network is in the giant component, then we have that the two events *(a)* and *(b)* occur respectively with probability $1-p$ and $p(1-S)$. We call $p^j_{ab} = 1-p+p(1-S = 1-pS$ the probability that either condition *(a)* or condition *(b)* are satisfied for node j. It follows that the probability $1-S$ that a node is not in the giant component of the network is given by

$$1 - S = \prod_{j\neq i} p^j_{ab} = [1 - pS]^{N-1} \tag{4.23}$$

$$1 - S = \left[1 - \frac{c}{N-1}S\right]^{N-1} \tag{4.24}$$

$$1 - S = \left[1 - \frac{c}{N-1}S\right]^{N-1}, \tag{4.25}$$

where we have used $p = \frac{c}{N-1}$. Finally, taking Eq. (4.25) and performing the limit $N \to \infty$ we obtain

$$S = 1 - e^{-cS}. \tag{4.26}$$

□

Starting from Eq. (4.22) we can find that the critical average degree for having a giant component in a Poisson network is given by $c = 1$.

Proposition 11. *A Poisson random network with average degree $\langle k \rangle = c$ has a giant component if and only if $c > 1$.*

Proof. As we have just observed the fraction of nodes S in the giant component of a Poisson network with average degree $\langle k \rangle = c$ satisfies Eq. (4.22), i.e.

$$S = 1 - e^{-cS}. \tag{4.27}$$

This equation is always satisfied for $S = 0$, but, depending on the value of the average degree c it can have another non-trivial solution $S > 0$. Unfortunately this equation cannot be solved analytically for arbitrary value of S. For this reason we will make use of some graphical argument. The solution of Eq. (4.27) can be seen as the value of S where the two functions $y = f(S)$ with $f(S) = S$ and $y = g(S)$ with $g(S) = 1 - e^{-cS}$ cross. In Figure 4.2 we plot the function $y = S$ and $y = 1 - e^{-cS}$ for different values of c. We see indeed that $S = 0$ is always a solution of $S = 1 - e^{-cS}$ and that for high enough value of the average degree c another solution $S > 0$ emerges continuously from the value $S = 0$. In order to detect when this new solution emerges, we impose that the two functions $y = f(S) = S$ and $y = g(S) = 1 - e^{-cS}$ are tangent to each other at $S = 0$, i.e. we impose

$$\begin{aligned} \left.\frac{dS}{dS}\right|_{S=0} &= \left.\frac{d(1-e^{-cS})}{dS}\right|_{S=0}, \\ 1 &= \left. ce^{-cS}\right|_{S=0}, \\ 1 &= c. \end{aligned} \tag{4.28}$$

Therefore in a Poisson network with average degree c we observe a giant component if and only if $c > 1$. □

In Figure 4.3 we plot the fraction of nodes S in the giant component of a Poisson network as a function of c. At $c = 1$ there is a continuous phase transition in the network. In fact the value $c = 1$ separates the values of the average degree c for which there is no giant component in the network ($c < 1$) from the values of the average degree c for which there is a giant component in the network ($c > 1$).

Proposition 12. *For $c = 1 + \epsilon$ and $\epsilon \ll 1$, the fraction of nodes in the giant component increases with the average degree c as*

$$S \propto (c-1)^{\beta}, \tag{4.29}$$

with $\beta = 1$.

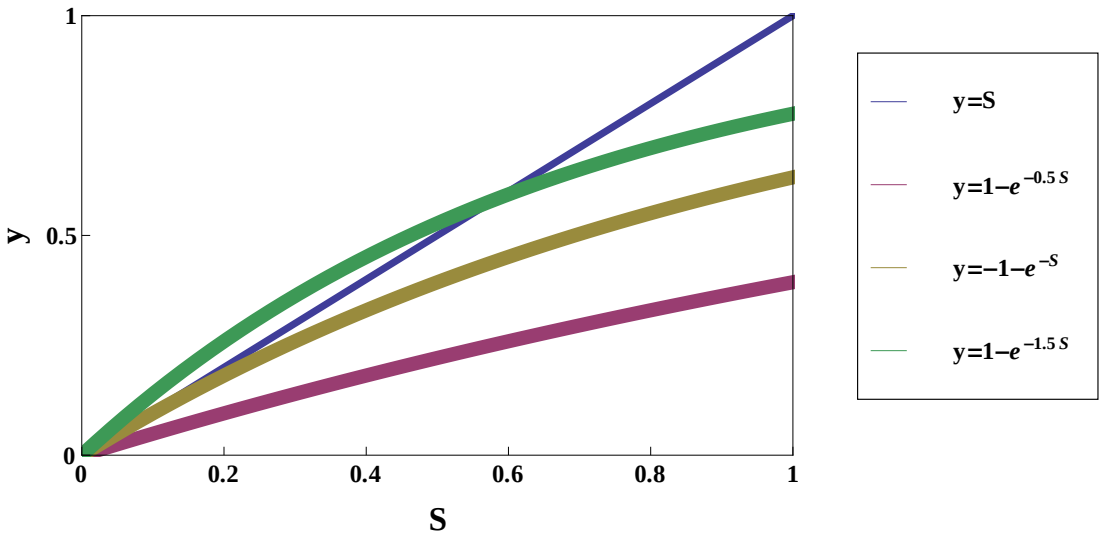

Figure 4.2: Graphical solution of Eq. (4.22). The point where the functions $y = f(S) = S$ and $y = g(S) = 1 - e^{-cS}$ cross is the solution of the Eq. (4.22). For $c \leq 1$ there is only the solution $S = 0$ and there is no giant component in the network, whereas for $c > 1$ another non-trivial solution $S > 0$ emerge and in the network there is a non-vanishing giant component.

Proof. Developing Eq. (4.22) for $c - 1 \ll 1$, and $S \ll 1$, we get

$$\begin{aligned} S &= 1 - \left(1 - cS + \frac{1}{2}c^2S^2 + \dots\right), \\ (c-1)S &= \frac{1}{2}c^2S^2, \\ S &= \frac{2}{c^2}(c-1) \\ S &\propto (c-1). \end{aligned} \tag{4.30}$$

□

Random network can be distinguished into subcritical, supercritical and critical

- In a random graph, if $\lim_{N\to\infty}\langle k\rangle < 1$ there is no giant component in the network, i.e. the largest connected component in the network contains a vanishing fraction of all the nodes. These random networks called *subcritical*. For example Poisson networks with $\langle k\rangle = c < 1$ are subcritical.

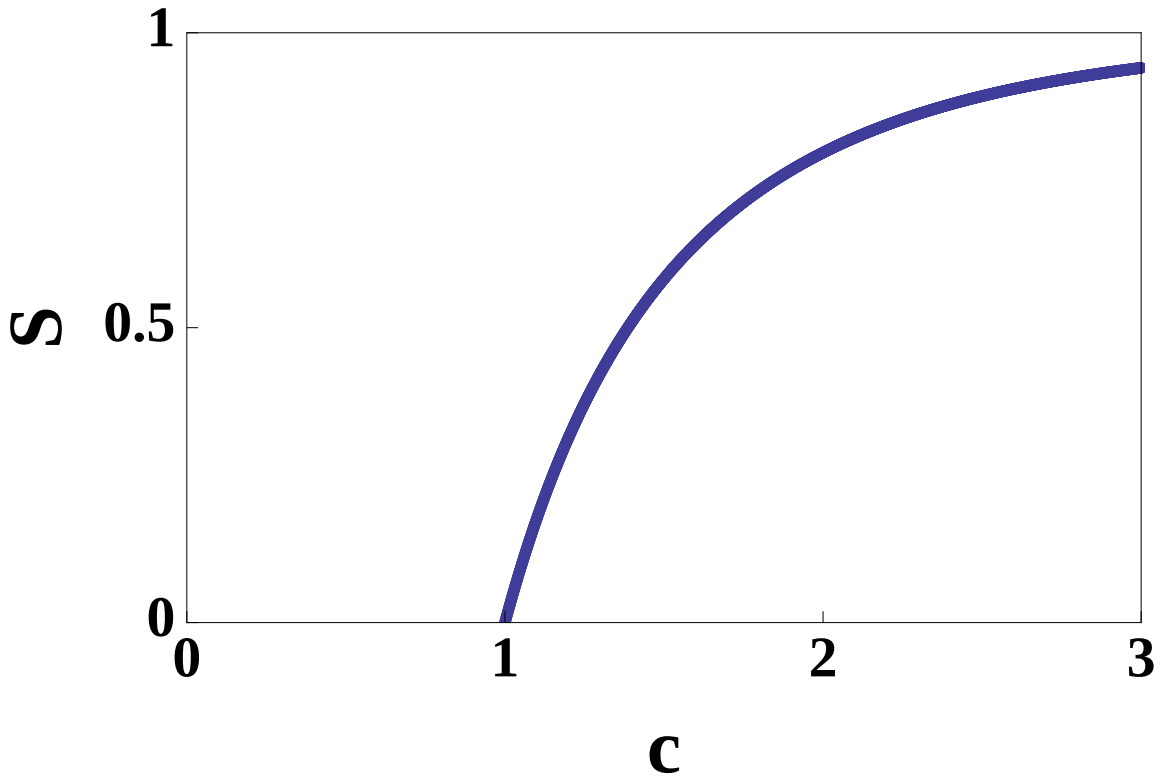

Figure 4.3: Fraction of nodes S in the giant component of Poisson networks as a function of their average degree $\langle k \rangle = c$.

- On the contrary, if $\lim_{N\to\infty}\langle k\rangle > 1$ we observe a giant component in the random networks, i.e. the largest connected component of the network contains a finite fraction of the total number of nodes in the network. These random networks are called *supercritical*. For example Poisson networks with $\langle k\rangle = c > 1$ are supercritical.

- Finally, if $\lim_{N\to\infty}\langle k\rangle = 1$ we observe the emergence of the giant component as a continuous phase transition. These random networks are called *critical*. Poisson networks with $\langle k\rangle = c = 1$ are critical.

In figure 4.4 we draw example of Poisson subcritical, critical and supercritical network, showing the emergence of the giant component as a function of the average degree in the network c.

Given a finite network, for a sufficiently high average degree, the network contains a single component, we say in this case that the network is fully connected.

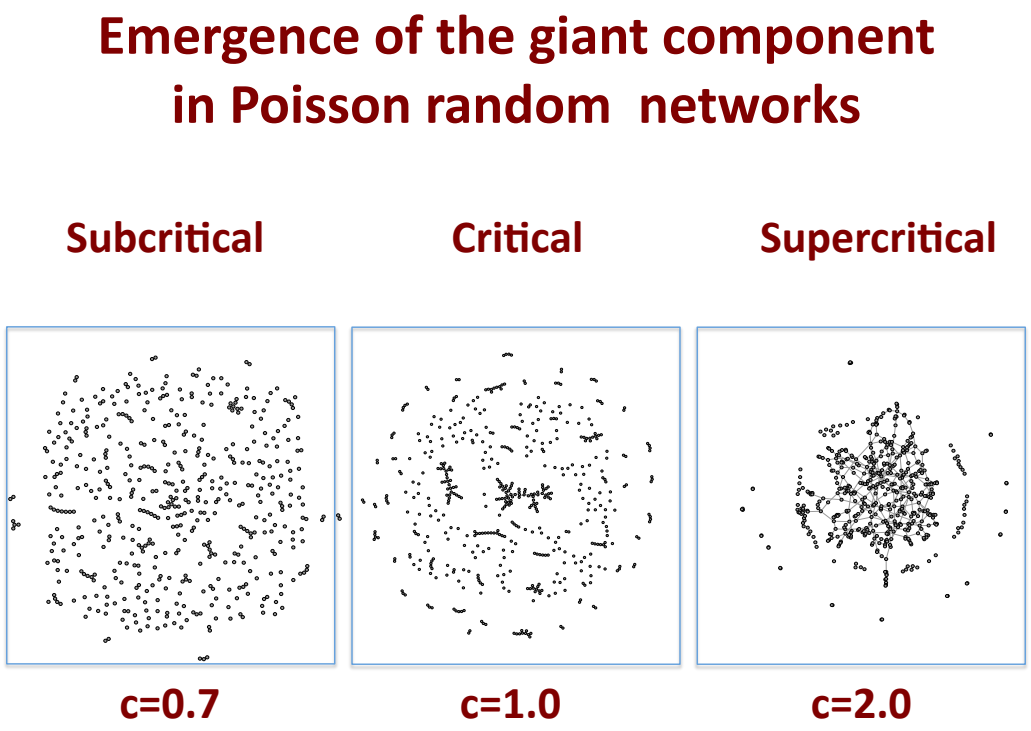

Figure 4.4: Three Poisson networks of $N = 500$ nodes and average degree respectively $c = 0.7$ (subcritical), $c = 1$ (critical) and $c = 2$ (supercritical).

This happens at $c > \ln(N)$ where N is the network size.

Proposition 13. *For $\langle k \rangle \simeq \ln N$ a network of size N belonging to the $\mathbb{G}(N,p)$ ensemble contains a single connected component.*

Proof. In a finite random network of size N we expect to have a single connected component when the fraction of nodes in that is not in giant component $1 - S < \frac{1}{N}$, meaning that we expect less than one node not to be in the giant component of the network. Considering the fact that S must satisfy Eq. (4.23), we have

$$1 - S = (1 - pS)^{N-1}, \tag{4.31}$$

Putting $S = 1 - \frac{1}{N}$ we find the minimal average degree $\langle k \rangle$ that is necessary to

have less than one node outside the giant component, i.e. the minimal degree that is required to have a fully connected network. In fact starting from Eq. (**??**) we have

$$\frac{1}{N} \simeq (1-p)^{N-1} = e^{(N-1)\ln(1-p)}. \tag{4.32}$$

Using $p \ll 1$ and $\ln(1-x) \simeq -x$ for x $ll1$ wee get

$$\begin{aligned} \frac{1}{N} &= e^{-\langle k \rangle} \\ \langle k \rangle &= \ln(N) \end{aligned} \tag{4.33}$$

where the first equation is derived in the limit $N \gg 1$. Finally if the average degree $\langle k \rangle > \ln(N)$ a random network in the $\mathbb{G}(N,p)$ contains just a single connected component. □

4.7 Expected number of cliques in random graphs (E)

Given a random network in the $\mathbb{G}(N,p)$ ensemble, it is sometimes important to characterized also its local properties. In particular it is interesting to consider the average number of some special subgraph as loops or cliques. Let us consider here the cases of cliques of size n.

Proposition 14. *In a random network of the $\mathbb{G}(N,p)$ ensemble the average number $\langle \mathcal{N}_n^{cliques} \rangle$ of cliques of size n is given by*

$$\langle \mathcal{N}_n^{cliques} \rangle = \left(\begin{array}{c} N \\ n \end{array} \right) p^{n(n-1)/2}.$$

Proof. In fact there are $\left(\begin{array}{c} N \\ n \end{array} \right)$ ways to choose n nodes out of the N nodes of the network, and the probability that any set of n nodes is fully connected is given by $p^{n(n-1)/2}$ because we have to draw $n(n-1)/2$ links. □

Proposition 15. *The average number of triangles a Poisson network with average degree c, is given by*

$$\langle \mathcal{N}_3^{triangle} \rangle = \frac{1}{6} c^3, \tag{4.34}$$

i.e. the average number of triangles in a Poisson network does not depend on the network size. In other words the average number of triangles is finite also in an infinite network. This implies that triangles are negligible in the network.

Proof. A triangle is a clique of size $n = 3$. A Poisson network with average degree c is a network of the $\mathbb{G}(N,p)$ ensemble with $p = \frac{c}{N-1}$. Therefore, using

Eq. (4.34) we find that the average number of triangles a Poisson network with average degree c is given by

$$\begin{aligned}\langle \mathcal{N}_3^{\text{triangle}} \rangle &= \frac{N!}{3!(N-3)!}\left(\frac{c}{N-1}\right)^3 \\ &= \frac{1}{6}c^3. \end{aligned} \tag{4.35}$$

□

This result can be extended to any loop of finite size n, finding that the average number of these loops remain finite. Therefore the only relevant loops in Poisson network, are loops of length $n \geq \log N$. For this reason, we say that Poisson networks are "locally tree-like".

Proposition 16. *In a random network of the $\mathbb{G}(M,p)$ ensemble, with $p = \frac{a}{N^z}$, we have in the limit $N \to \infty$,*

$$\langle \mathcal{N}_n^{cliques} \rangle = 0 \tag{4.36}$$

for every $n > \frac{2+z}{z}$.

Proof. Starting from Eq (4.34), putting $p = \frac{a}{N^z}$ and going in the limit $N \to \infty$ we get

$$\begin{aligned}\langle \mathcal{N}_n^{\text{cliques}} \rangle &= \binom{N}{n}\left(\frac{a}{N^z}\right)^{n(n-1)/2}, \\ &= \frac{N^n}{n!}\left(\frac{a}{N^z}\right)^{n(n-1)/2}, \\ &= \frac{a^{n(n-1)/2}}{n!}N^{n[1-z(n-1)/2]}. \end{aligned} \tag{4.37}$$

Therefore we have that $\langle \mathcal{N}_n^{\text{cliques}} \rangle \to 0$ as long as

$$1 - z(n-1)/2 < 0, \tag{4.38}$$

i.e. $n > \frac{z+2}{z}$. □

From this result it follows that we need at least to consider random networks with $p = \frac{a}{N^z}$ and $z > \frac{2}{n-1}$ to observe a clique of size n with non zero probability in the infinite network limit. It turns out that this is a sharp threshold, i.e. that as soon as $z = \frac{2}{n-1}$ we have some non-zero probability to observe a clique of size n in the network. In Figure 4.5 we show the value of z for which different network subgraph start to have non zero probability in a random graph $\mathbb{G}(N,p)$ with $p = \frac{a}{N^z}$.

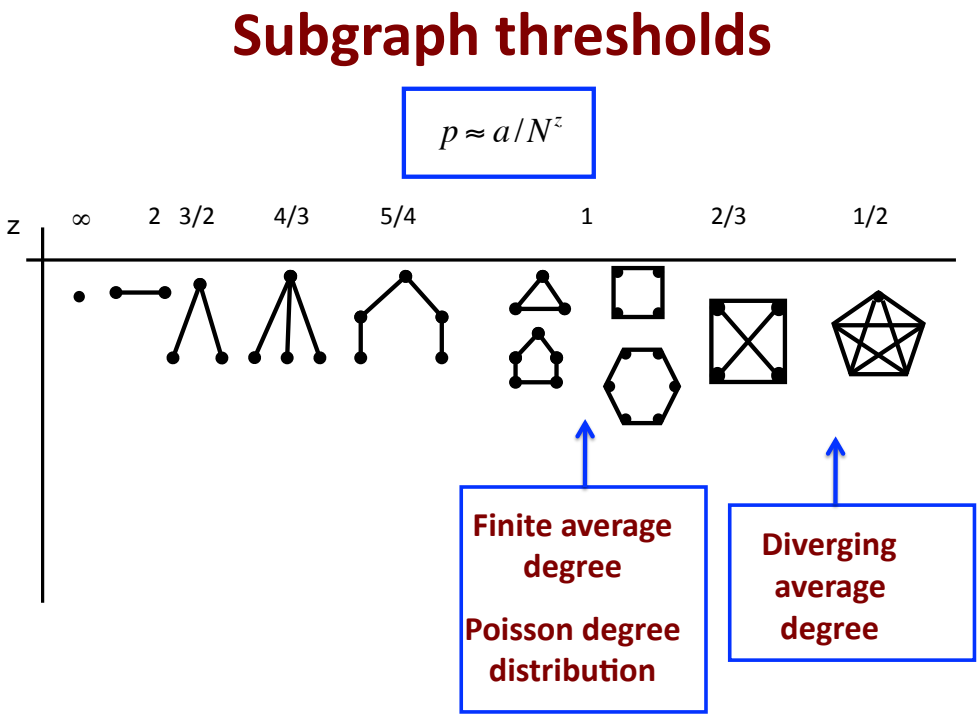

Figure 4.5: Threshold values of z needed in order to observe the corresponding network subgraph in a random network of the $\mathbb{G}(N,p)$ ensemble with $p = \frac{a}{N^z}$.

4.8 Generating functions (E)

Definition 57. *Given a distribution p_n its* generating function $G_0(x)$ *is given by*

$$G_0(x) = \sum_{n=0}^{\infty} p_n x^n. \tag{4.39}$$

Proposition 17. *The generating function of an arbitrary degree distribution has the following properties*

- (i) *The generating function calculated for $x = 1$ is equal to one,i.e.*

$$G_0(1) = 1. \tag{4.40}$$

- (ii) *The $m-$moments of the distribution can be extracted by differentiating the generating function n times, i.e.*

$$\left.\frac{d^m G_0(x)}{dx^m}\right|_{x=1} = \langle n(n-1)(n-2)\dots(n-m+1)\rangle. \tag{4.41}$$

Proof. The relation *(i)* follows immediately from the normalization of the distribution p_n. In fact

$$G_0(1) = \sum_{n=0}^{\infty} p_n = 1. \tag{4.42}$$

The relation (ii) can be proved easily by observing that

$$\begin{aligned}\left.\frac{d^m G_0(x)}{dx^m}\right|_{x=1} &= \sum_{n=0}^{\infty} p_n \left.\frac{d^m x^n}{dx^n}\right|_{x=1} \\ &= \sum_{n=0}^{\infty} p_n n(n-1)(n-2)\dots(n-m+1).\end{aligned} \tag{4.43}$$

□

As a consequence of this proposition we get that if we consider the generating function $G_0(x)$ of a degree distribution $P(k)$ the average degree $\langle k\rangle$ in the network and the second moment of the degree distribution $\langle k(k-1)\rangle$ are given respectively by

$$\begin{aligned}\langle k\rangle &= \left.\frac{dG_0(x)}{dx}\right|_{x=1} \\ \langle k(k-1)\rangle &= \left.\frac{d^2 G_0(x)}{dx^2}\right|_{x=1}.\end{aligned} \tag{4.44}$$

Chapter 5

Scale-free networks and the Barabasi-Albert model

Despite the elegance and the importance of random graph theory, many real networks cannot be described by random graphs. In particular most complex networks are sparse, i.e. they have a finite average degree $\langle k \rangle$ but they cannot be described by Poisson networks. This observation has been made in the late nineties, by two very important papers:

- *Collective dynamics of "small world" networks* paper by D. J. Watts and S. Strogatz (Nature 1998) in which it has been shown that many networks have at the same time small average distance and large clustering coefficient, while Poisson networks have vanishing clustering coefficient in the limit of large network sizes;

- *Emergence of scaling in random networks* paper by A.-L. Barabasi and R. Albert (Science 1999) in which it was shown that many complex networks have a broad degree distribution with diverging $\langle k^2 \rangle$ that differers substantially from the Poisson degree distribution of random graphs;

5.1 Introduction (E)

In this chapter we will introduce scale-free networks. In the paper *Emergence of scaling in random networks* by A.-L. Barabasi and R. Albert (Science 1999) it has been shown that many complex networks have a broad degree distribution with a power law tail $P(k) \simeq k^{-\gamma}$ for $k \gg 1$. Moreover it was shown that most of these networks have a finite average degree $\langle k \rangle$ but a diverging second moment $\langle k^2 \rangle$ in the large network limit $N \to \infty$. Therefore the power-law exponent γ is typically in the range $\gamma \in (2, 3]$. For these networks the average degree $\langle k \rangle$ is finite but the standard deviation of the degree $\sigma = \sqrt{\langle k^2 \rangle - \langle k \rangle^2}$ diverge in the large network limit, implying that the there is no characteristic scale for the

degree in the network. Therefore these networks are called *scale-free networks*. Scale-free networks are characterized by the presence of nodes with high-degree.

5.2 Power-law networks (E)

Definition 58. Power-law networks *have power-law degree distribution* $P(k)$ *given by*

$$P(k) = Ck^{-\gamma}, \tag{5.1}$$

with $k_{min} \leq k \leq K$ *and* $\gamma > 1$. *The quantity* k_{min} *is the* minimal degree *of the network, while* K *is the* maximal degree *of the network, also called the* cutoff. *The normalization constant* C *is fixed by the condition* $\sum_k P(k) = 1$. *Therefore we have*

$$C = \left(\sum_{k=k_{min}}^{K} k^{-\gamma} \right)^{-1}. \tag{5.2}$$

In the limit of large network sizes in which the cutoff $K \to \infty$ *then we have*

$$C \quad \to \quad \frac{1}{\zeta(\gamma, k_{min})}, \tag{5.3}$$

where ζ *is the incomplete Riemann zeta function.*
The condition $\gamma > 1$ *is necessary for guarantee the normalization of the distribution.*

Proposition 18. *The degree distribution of power-law networks can be plotted as a straight line in a log-log plot.*

Proof. In fact we have that if $P(k)$ is given by Eq. (5.1) then

$$\ln P(k) = \ln C - \gamma \log k. \tag{5.4}$$

Therefore if $y = \ln P(k)$ and $x = \ln k$ the degree distribution is described by the equation

$$y = A - \gamma x \tag{5.5}$$

where $A = \ln C$, i.e. the degree distribution can be plotted as a straight line in a log-log plot. The slope of the line is given by the power-law exponent γ. □

In order to characterize several quantities (normalization constants, moment of the distribution, natural cutoff) of power-law networks is it sometimes useful to perform estimates assuming that the degree is a continuous variable.

Definition 59. *In the* continuous approximation, *the degree k of a node in the network is assumed to be a continuous variable taking real positive values. In this approximation the degree distribution of a node is given by*

$$P(k) = Ck^{-\gamma}, \tag{5.6}$$

with $\gamma > 1$ and $k \in (k_{min}, K)$, and the normalization constant C is fixed by the condition

$$1 = \int_{k_{min}}^{K} dk P(k) \tag{5.7}$$

or equivalently

$$1 = C \int_{k_{min}}^{K} dk k^{-\gamma}. \tag{5.8}$$

Definition 60. *The* natural cutoff *of a power-law network with power-law exponent $\gamma > 2$ is the maximal degree expected in the network of N nodes if we assume that the degree of each node is drawn randomly from a power-law degree distribution $P(k) = Ck^{-\gamma}$.*

Proposition 19. *The natural cutoff K of a power-law network with power-law exponent $\gamma > 1$ diverges in the large network limit, and can be estimated to be*

$$K \simeq \min(N, k_{min} N^{1/(\gamma-1)}). \tag{5.9}$$

Therefore,

$$K \simeq \begin{cases} k_{min} N^{1/(\gamma-1)} & for \quad \gamma > 2 \\ N & for \quad \gamma \in (1, 2]. \end{cases}$$

Proof. Given a network of N nodes the maximum degree K should be lower than N. Nevertheless it might be typically smaller than N. In order to characterize the typical scale of the maximum degree K we can impose that in such network the number of expected nodes with degree $k > K$ is just 1. This implies that we can estimate the value of the cutoff by imposing

$$N \sum_{k=K}^{\infty} P(k) = NC \sum_{k=K}^{\infty} k^{-\gamma} = 1. \tag{5.10}$$

In the continuous approximation we can substitute the sum over the degree with an integral over the continuous variable k, getting

$$\begin{aligned} C \int_{K}^{\infty} dk k^{-\gamma} &= \frac{1}{N} \\ C \frac{1}{\gamma - 1} K^{1-\gamma} &= N^{-1} \\ K &\simeq k_{min} N^{1/(\gamma-1)}. \end{aligned} \tag{5.11}$$

Now this quantity will indicate the typical scale of the maximum degree K of the network as long as this expression is smaller than the total number of nodes in the network N. □

5.3 The average degree $\langle k \rangle$ and the second moment of the degree distribution $\langle k^2 \rangle$ of the power-law networks (E)

Proposition 20. *The moments $\langle k^n \rangle$ of the degree distribution of power-law networks, with degree distribution $P(k) = Ck^{-\gamma}$, with $k \in [k_{min}, K]$ are given in the continuous approximation by*

$$\langle k^n \rangle = \begin{cases} C\frac{1}{n+1-\gamma}\left(K^{n+1-\gamma} - k_{min}^{n+1-\gamma}\right) & for \quad n \neq \gamma - 1, \\ C\ln\left(\frac{K}{k_{min}}\right) & for \quad n = \gamma - 1, \end{cases} \tag{5.12}$$

with the normalization constant C given by

$$C = (\gamma - 1)\frac{1}{k_{min}^{1-\gamma} - K^{1-\gamma}}. \tag{5.13}$$

Therefore in the limit of large network sizes $N, K \to \infty$, then $\langle k^n \rangle$ diverges for $n \geq \gamma - 1$ and converges to a constant for $n < \gamma - 1$.

Proof. Let us evaluate first the normalization constant in the continuous approximation. The normalization constant is determined by

$$\begin{aligned} 1 &= \int_{k_{min}}^{K} dk P(k) \\ 1 &= C\int_{k_{min}}^{K} dk k^{-\gamma}, \\ 1 &= C\frac{1}{1-\gamma}\left[K^{1-\gamma} - (k_{min})^{1-\gamma}\right]. \end{aligned} \tag{5.14}$$

Therefore

$$C = (\gamma - 1)\frac{1}{(k_{min})^{1-\gamma} - K^{1-\gamma}}. \tag{5.15}$$

Since $\gamma > 1$ for $N \to \infty$, i.e. $K \to \infty$ we have

$$C \to (\gamma - 1)(k_{min})^{\gamma - 1}, \tag{5.16}$$

therefor C is converging to a finite value for every $\gamma > 1$.

Let us evaluate $\langle k^n \rangle$ in the continuous approximation. We have

$$\begin{aligned}
\langle k^n \rangle &= \int_{k_{min}}^{K} dk k^n P(k) \\
&= C \int_{k_{min}}^{K} dk k^{n-\gamma} \\
&\begin{cases} C \frac{1}{n+1-\gamma} \, k^{n+1-\gamma} \big|_{k_{min}}^{K} & \text{for} \quad n \neq \gamma - 1 \\ \\ C \ln k \big|_{k_{min}}^{K} & \text{for} \quad n = \gamma - 1, \end{cases} \\
&= \begin{cases} C \frac{1}{n+1-\gamma} \left(K^{n+1-\gamma} - k_{min}^{n+1-\gamma} \right) & \text{for} \quad n \neq \gamma - 1, \\ C \ln \left(\frac{K}{k_{min}} \right) & \text{for} \quad n = \gamma - 1. \end{cases}
\end{aligned} \tag{5.17}$$

For $N \to \infty$ we have that $K \to \infty$ and the $\langle k^n \rangle$ converges to a finite value is $n < \gamma - 1$ otherwise $\langle k^n \rangle$ diverges. □

Proposition 21. *The* average degree $\langle k \rangle$ *and the second moment* $\langle k^2 \rangle$ *of a power-law network, with degree distribution* $P(k) = Ck^{-\gamma}$*, with* $k \in [k_{min}, K]$ *and cutoff* $K \to \infty$ *and* $N \to \infty$ *are either finite or diverging as* $N, K \to \infty$ *depending on the value of the power-law exponent* γ*. In fact we have*

- For $\gamma > 3$
  The average degree $\langle k \rangle$ is finite *in the limit* $N, K \to \infty$.
  The second moment $\langle k^2 \rangle$ *of the degree distribution* is finite *in the limit* $N, K \to \infty$.

- For $\gamma \in (2, 3]$
  The average degree $\langle k \rangle$ is finite *in the limit* $N, K \to \infty$.
  The second moment $\langle k^2 \rangle$ *of the degree distribution* is diverging *in the limit* $N, K \to \infty$.

- For $\gamma \in (1, 2]$
  The average degree $\langle k \rangle$ is diverging *in the limit* $N, K \to \infty$.
  The second moment $\langle k^2 \rangle$ *of the degree distribution* is diverging *in the limit* $N, K \to \infty$.

Proof. In fact we have to calculate the moment $\langle k^n \rangle$ with $n = 1, 2$.

- For $\gamma > 3$ we have for $n = 1$, $1 < \gamma - 1$. In fact $\gamma - 1 > 1$. Therefore $\langle k \rangle$ is finite in the limit $N, K \to \infty$. Moreover we have also for $n = 2$, $2 < \gamma - 1$. In fact $\gamma - 1 > 2$. Therefore $\langle k^2 \rangle$ is finite in the limit $N, K \to \infty$.

- For $\gamma \in (2, 3]$ we have for $n = 1$, $1 < \gamma - 1$. In fact $\gamma - 1 > 1$. Therefore $\langle k \rangle$ is finite in the limit $N, K \to \infty$.

Nevertheless for $n = 2$, we have $2 \geq \gamma - 1$. In fact $\gamma - 1 \leq 2$. Therefore $\langle k^2 \rangle$ is diverging in the limit $N, K \to \infty$.

- For $\gamma \leq 2$ we have for $n = 1$ $1 \geq \gamma - 1$, i.e. $\gamma - 1 \leq 1$, and therefore $\langle k \rangle$ is diverging in the $N, K \to \infty$ limit.
  Moreover we have for $n = 2$, $2 \geq \gamma - 1$. In fact $\gamma - 1 \geq 2$. Therefore $\langle k^2 \rangle$ is diverging in the limit $N, K \to \infty$.

□

5.3.1 Other types of degree distributions

Examples of relevant degree distributions with finite average degree $\langle k \rangle$ are *Poisson, Exponential distributions* $P(k)$ given by

$$\begin{array}{lll} \textbf{Poisson Distribution} & P(k) = \frac{1}{k!} c^k e^{-c}, & c > 0 \\ \textbf{Exponential Distribution} & P(k) = (1-b) b^k, & b \in (0,1) \end{array} \tag{5.18}$$

(Note these distributions are normalized assuming that the maximal degree $K = \infty$). These degree distributions have both a finite average degree $\langle k \rangle$ and finite second moment $\langle k^2 \rangle$.

- *Poisson networks*
  For networks with Poisson degree distribution $P(k) = \frac{1}{k!} c^k e^{-c}$ we have that $\langle k \rangle = c$ and $\langle k(k-1) \rangle = c^2$, therefore $\langle k^2 \rangle = c(c+1)$ and $\sigma = c$. This means that if we want to model a social network with $\langle k \rangle = 100$ the standard deviation of the degree distribution will be $\sigma = 10$ and observing a person with $1,000$ friends would be a event 90 standard deviation from the mean, i.e. very unlikely

- *Exponential networks*
  The average degree $\langle k \rangle$ of a network with exponential degree distribution $P(k) = (1-b)b^k$ is given by $\langle k \rangle = \frac{b}{1-b}$, while the $\langle k^2 \rangle = \frac{b(b+1)}{(1-b)^2}$. Both moments are finite and $\sigma = \frac{\sqrt{b}}{1-b}$ For example consider the a network as the Internet with average degree $\langle k \rangle = 4$. Then we would have $b = \frac{5}{4}$ and the standard deviation of the degree distribution would be $\sigma = \sqrt{5}/2 \simeq 1.12$. Therefore it will be very unlikely to observe nodes of degree $k = 100$.

5.4 Scale-free networks (E)

Poisson and exponential networks cannot account for the large fluctuation in the degree of the nodes observed in a large variety of complex networks, from the Internet, to citation networks, to movie actor networks, to collaboration networks, to the World-Wide-Web, to airport networks ect. In fact many networks have at the same time a finite average degree $\langle k \rangle$ but a very large second

moment of the degree distribution, i.e. $\langle k^2 \rangle$. These networks have the degree distribution that for large degrees can be approximated by a power-law with power-law exponent $\gamma \in (2, 3]$. These networks are called scale-free networks.

Definition 61. Scale-free networks *have a degree distribution $P(k)$ that for large values of the degree can be approximated by a power-law*

$$P(k) \simeq ck^{-\gamma} \tag{5.19}$$

with $\gamma \in (2, 3]$.

5.5 The Barabasi-Albert model (E)

5.5.1 Motivation and Definition

Albert-Laszlo Barabasi and Reka Albert in their paper *Emergence of scaling in random networks* (Science 199) have shown that scale-free network constitute a complex networks universality class. In fact they show that scale-free networks are ubiquitous and are found in systems as different as the Internet, the World-Wide-Web, the movie actor networks and the scientific collaboration networks, just to cite a few. Moreover they identify the following two main characteristics of a large number of scale-free networks.

- **GROWTH:**
  A large variety of scale-free networks are growing, i.e. the number of nodes N in these networks is increasing with time.
  Example of growing scale-free networks are the World-Wide-Web, the Internet, Wikipedia, the citation networks, the movie actor networks, online social networks, ect..

- **PREFERENTIAL ATTACHMENT**
  In many of these networks the“ popularity is attractive”, meaning that the new links are not attached randomly but they follow the so called *preferential attachment*, i.e. node of high degree are more likely to acquire new nodes.
  For example, a new webpage is more likely ot connected to a well known website (e.g. BBC, New york Times ect.) than to an rather unknown one. Similarly, highly cited papers are more likely to be cited again.

These two main ingredients are responsible for the emergence of scale-free networks. The Barabasi-Albert model (BA model) is the basic growing network model that capture these two main characteristic and display a scale-free degree distribution.

Definition 62. *In the* Barabasi-Albert model *is the most fundamental growing network model including the preferential attachment mechanism. At time $t = 1$ the network is formed by n_0 nodes connected by $m_0 > m$ links.*
The model evolves in time by

- **GROWTH:**
  At each time $t > 1$ a new node is added to the network and connected to the existing network with exactly m links to distinct nodes of the network;

- **PREFERENTIAL ATTACHMENT:**
  Every new link of the new node is attached to an existing node i of the network not already linked to the new node with probability

$$\Pi_i = \frac{k_i}{\sum_j k_j} \tag{5.20}$$

  where k_i is the degree of node i and where the sum over the nodes j is extended over all the nodes of the network not already linked to the new node.

5.5.2 Mean-field approximation for growing network models

Definition 63. *The* mean-field approximation for growing network models *consist in taking two approximations.*

- *The* continuous approximation *for the degrees of the nodes k that can take now any positive real value $k > k_{min}$ and for the time of arrival of the nodes in the network that can take any real positive value.*

- *The* strictly speaking mean-field approximation *for which the degree $k_i(t)$ of node i at time t acquires a deterministic value equal to the average degree of node i at time t in the original stochastic growing network model.*

Proposition 22. *In the mean-field approximation, the degree $k_i(t)$ of node i at time t satisfies the following differential equation*

$$\frac{dk_i(t)}{dt} = m\frac{k_i}{\sum_j k_j} \tag{5.21}$$

Proof. In fact in the mean-field approximation the degree $k_i(t)$ of node i at time t is the average degree of the original BA model. Now the average number of links that a node i acquires at time t is given by $m\Pi_i$ with $\Pi_i = \frac{k_i}{\sum_j k_j}$. It follows that $k_i(t)$ satisfies Eq. (5.21). □

Proposition 23. *In the mean-field approximation the degree k_i of node i arrived in the network at time t_i increases with time as a power-law. In particular*

$$k_i = m\left(\frac{t}{t_i}\right)^{\beta} \tag{5.22}$$

with $\beta = 1/2$ for $t \geq t_i$. This implies that older nodes have higher degree.

Proof. The degree $k_i(t)$ of node i at time t satisfies the mean-field Eq. (5.21). In the limit of large times, $t \gg 1$ we can always neglect the probability that the new links end on the same nodes. Since at each time we add a node with m new links we have that

$$\sum_j k_j = 2(m_0 + mt) \simeq 2mt \tag{5.23}$$

where the last expression is derived for $t \gg 1$. Therefore Eq. (5.21) can be written as

$$\frac{dk_i}{dt} = m\frac{k_i}{2mt} = \frac{1}{2}\frac{k_i}{t}, \tag{5.24}$$

with initial condition $k_i(t_i) = m$ because the degree of node i at the time t_i when it arrives in the network is given by m. Integrating by parts Eq. (5.24) we have

$$\begin{aligned} \int_m^{k_i(t)} \frac{d\tilde{k}_i(t)}{\tilde{k}_i} &= \frac{1}{2}\int_{t_i}^{t} d\tilde{t}\frac{1}{\tilde{t}} \\ \ln k_i(t) - \ln m &= \frac{1}{2}\left[\ln t - \ln t_i\right] \\ k_i(t) &= m\sqrt{\frac{t}{t_i}} \end{aligned} \tag{5.25}$$

Therefore we have that the degree $k_i(t)$ of node i in the mean-field approximation follows the evolution dictated by the following equation

$$k_i(t) = m\sqrt{\frac{t}{t_i}}. \tag{5.26}$$

for $t \geq t_i$. The node arrived in the network at time $t_i = 1$ is the most connected, so the cutoff of this network, can be estimated to be $K = m\sqrt{t}$ and is diverging as $t \to \infty$. □

Proposition 24. *In the mean-field approximation the degree distribution of the BA model is given by*

$$P(k) = 2m^2k^{-3} = Ck^{-\gamma}, \tag{5.27}$$

with $\gamma = 3$.

Proof. In the mean-field approximation each node i have degree given by $k_i(t) = m\sqrt{\frac{t}{t_i}}$. In the mean field approximation also the time at which new nodes arrive in the network is continuous. In this approximation the probability that a random node i of the network is arrived in the network at time $t_i < \tau$ is given by

$$P(t_i < \tau) = \frac{\tau}{t} \tag{5.28}$$

since the new nodes arrive in the network at a uniform rate of one. Let us evaluate the probability $P(k_i(t) > k)$ that if we take a random node in the network it will have degree $k_i(t) > k$. We have

$$\begin{aligned} P(k_i(t) > k) &= P\left(m\sqrt{\frac{t}{t_i}} > k\right) \\ &= P\left(t_i < t\frac{m^2}{k^2}\right) \\ &= \frac{1}{t}\left(t\frac{m^2}{k^2}\right) = \frac{m^2}{k^2}, \end{aligned} \tag{5.29}$$

where in the last expression we have been using Eq. (5.28). This implies that the probability $P(k_i(t) \leq k)$ that if we take a random node in the network it has degree $k_i(t) \leq k$ is given by

$$P(k_i(t) \leq k) = 1 - \frac{m^2}{k^2}. \tag{5.30}$$

Therefore we have that the degree distribution of the network $P(k)$ is given, in the mean-field approximation by

$$\begin{aligned} P(k) &= \frac{dP(k_i(t) \leq k)}{dk} \\ &= \frac{d}{dk}\left(1 - \frac{m^2}{k^2}\right) \\ &= \frac{2m^2}{k^3} \end{aligned} \tag{5.31}$$

Therefore $P(k) = Ck^{-\gamma}$ with $\gamma = 3$. This implies that the generated network is *scale-free* with a finite average degree $\langle k \rangle$ and a diverging second moment $\langle k^2 \rangle$ as the network size $t \to \infty$. □

5.5.3 The master equation approach

The mean-field approximation is a drastic approximation, that is principle is not very well controlled. Nevertheless for growing network models gives very good qualitative insights on the dynamics of the degree is the nodes, and the power-law exponent of the degree distribution. Here we present a more rigorous approach base on the master equation of the model that will predict exactly the degree distribution of the BA model in the limit $t \to \infty$. This approaches is called the master equation approach.

Proposition 25. *The master equation is the equation describes the evolution of the average number $N_k(t)$ of nodes that at time t have degree k . This equations reads for the BA model,*

$$\begin{aligned} N_k(t+1) &= N_k(t) + m\Pi(k-1)N_{k-1}(t) - m\Pi(k)N_k(t) \quad \textit{for} \quad k > m \\ N_k(t+1) &= N_k(t) - m\Pi(k)N_k(t) + 1 \qquad\qquad\qquad\qquad\quad \textit{for} \quad k = m. \end{aligned}$$

where

$$\Pi(k) = \frac{k}{\sum_j k_j}. \tag{5.32}$$

Proof. With $N_k(t)$ we indicate the average number of nodes that at time t have degree k. At time $t+1$ the average number of nodes have degree k is equal to $N_k(t)$ plus the average number of nodes that acquire degree k at time $t+1$ minus the average number of nodes that had degree k at time t and they acquire a degree $k+1$ at time $t+1$.

- The average number of nodes that had degree $k-1$ at time t and acquire degree k at time $t+1$ is given by

$$m\Pi(k-1)N_{k-1}(t) \tag{5.33}$$

  where $\Pi(k-1) = \frac{k-1}{\sum_j k_j}$. In fact at time t we add m new links the probability that we attach the each new link to a node of degree $k-1$ is $\Pi(k-1)$ and the average number of nodes that at time t have degree $k-1$ is $N_{k-1}(t)$.

- The average number of nodes that had degree k at time t and acquire degree $k+1$ at time $t+1$ is given by

$$m\Pi(k)N_k(t) \tag{5.34}$$

  where $\Pi(k) = \frac{k}{\sum_j k_j}$. In fact at time t we add m new links the probability that we attach the each new link to a node of degree k is $\Pi(k)$ and the average number of nodes that at time t have degree k is $N_k(t)$.

- Finally, if $k = m$, the average number of nodes $N_k(t)$ of degree $k = m$ is increasing by one because of the arrival of the new node in the network of degree $k = m$.

Therefore we get

$$\begin{array}{lcll} N_k(t+1) & = & N_k(t) + m\Pi(k-1)N_{k-1}(t) - m\Pi(k)N_k(t) & \text{for} \quad k > m \\ N_k(t+1) & = & N_k(t) - m\Pi(k)N_k(t) + 1 & \text{for} \quad k = m. \end{array}$$

where

$$\Pi(k) = \frac{k}{\sum_j k_j}. \tag{5.35}$$

□

Proposition 26. *The degree distribution $P(k)$ of the BA model in the limit $t \to \infty$ is given by*

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}, \tag{5.36}$$

for $k \geq m$. Therefore for large values of k we have $P(k) \simeq Ck^{-\gamma}$ with $\gamma = 3$.

Proof. Considering the master equation we observe that for $t \gg 1$ we can approximate

$$\Pi(k) = \frac{k}{\sum_j k_j} \simeq \frac{k}{2mt} \tag{5.37}$$

because

$$\sum_j k_j = 2(mt + m_0) \simeq 2mt. \tag{5.38}$$

Using this approximation valid for $t \gg 1$ we can write the master equation as

$$\begin{array}{rcll} N_k(t+1) & = & N_k(t) + \frac{k-1}{2t} N_{k-1}(t) - \frac{k}{t} N_k(t) & \text{for} \quad k > m \\ N_k(t+1) & = & N_k(t) - \frac{k}{2t} N_k(t) + 1 & \text{for} \quad k = m. \end{array}$$

Now we observe that for sufficiently large values of $t \gg 1$ we have

$$N_k(t) \simeq tP(k) \tag{5.39}$$

where $P(k)$ is the degree distribution of the network and the total number of nodes in the network is given by $N \simeq t$. Substituting Eq. (5.58) into the master equation (5.39), we obtain

$$\begin{array}{rcll} (t+1)P(k) & = & tP(k) + \frac{(k-1)}{2} P(k-1) - \frac{k}{2} P(k) & \text{for} \quad k > m \\ (t+1)P(k) & = & tP(k) - \frac{k}{2} P(k) + 1 & \text{for} \quad k = m. \end{array}$$

Let us write this last equation for $k > m$. We obtain

$$P(k) = \frac{k-1}{2+k} P(k-1) \tag{5.40}$$

for $k > m$. This recursive equation for $k > m$ can be solved in terms of $P(1)$ giving

$$\begin{aligned} P(k) &= \prod_{j=1+m}^{k} \left[\frac{j-1}{2+j} \right] P(m) \\ &= \frac{\Gamma(k)\Gamma(m+3)}{\Gamma(k+3)\Gamma(m)} P(m) \\ &= \frac{(k-1)(k-2)(k-3)\ldots(m+3) \times (m+2) \times (m+1) \times m}{(2+k)(1+k)k(k-1)(k-2)\ldots(m+3)} P(m) \\ &= \frac{m(m+1)(m+2)}{k(k+1)(k+2)} P(m). \end{aligned} \tag{5.41}$$

Taking the last equation of (5.40) for $k = m$ we obtain

$$\begin{aligned} \left(1 + \frac{m}{2}\right) P(1) &= 1 \\ P(m) &= \frac{2}{2+m}. \end{aligned} \tag{5.42}$$

Therefore it follows that the degree distribution of the BA model in the limit $t \gg 1$ is given by

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}. \tag{5.43}$$

□

5.6 Growing network model without preferential attachment (E)

5.6.1 Definition of the model

The preferential attachment in growing network model is necessary to have a scale-free degree distribution. To show this let us consider the case of a growing network without preferential attachment, where the new links of the new nodes are attached to a random node of the network.

Definition 64. *In the* growing network without preferential attachment *we have that the new links are attached with uniform probability to the existing nodes of the network. At time $t = 0$ the network is formed by n_0 nodes connected by $n_0 > m$ links.*
The model evolves in time by

- **GROWTH:**
  At each time $t > 0$ a new node is added to the network and connected to the existing network with exactly m links to distinct nodes of the network;
- **UNIFORM ATTACHMENT:**
  Every new link of the new node is attached to an existing node i of the network not already linked to the new node with uniform probability. For $t \gg 1$ we have that the number of nodes in the network are $N(t) = t + n_0$ and we can assume that the probability that the new link is attached to a node i is given by

$$\Pi_i = \frac{1}{N(t)} = \frac{1}{t + n_0} \tag{5.44}$$

5.6.2 Mean-field approach

Proposition 27. *In the mean-field approximation, the degree $k_i(t)$ of node i at time t satisfies the following differential equation*

$$\frac{dk_i(t)}{dt} = m\frac{1}{N(t)}, \tag{5.45}$$

with initial condition $k_i(t_i) = m$.
In the mean-field approximation, the degree $k_i(t)$ of node i arrived in the network

at time t_i evolves in time according to the equation

$$k_i(t) = m + m \ln \left(\frac{t}{t_i} \right). \tag{5.46}$$

In the mean-field approximation the degree distribution $P(k)$ of the growing network model with uniform attachment is exponential and is given by

$$P(k) = \frac{e}{m} e^{-k/m}, \tag{5.47}$$

for $k \geq m$.

Proof. In the mean-field approximation, the degree $k_i(t)$ is given by the average degree of node i at time t in the stochastic model. Since at each time the degree of node i increases in average of a quantity $m\frac{1}{t+n_0} \simeq \frac{m}{t}$ for $t \gg 1$, we have that $k_i(t)$ satisfies the following mean-field equation

$$\frac{dk_i(t)}{dt} = \frac{m}{t} \tag{5.48}$$

for $t \gg 1$ with initial condition $k_i(t_i) = m$ where t_i is the time at which the node i has arrived in the network. Integrating this equation we get the time evolution of the degree of node i in the mean-field approximation, i.e.

$$\begin{aligned} k_i(t) - m &= \int_m^{k_i(t)} d\tilde{t} \frac{m}{\tilde{t}} \\ k_i(t) &= m + m \ln \left(\frac{t}{t_i} \right). \end{aligned} \tag{5.49}$$

In this growing network model with uniform attachment, the probability $P(t_i < \tau)$ that a random node of the network observed at time t has arrived in the network at time $t_i < \tau$ is given by

$$P(t_i < \tau) = \frac{\tau}{t}. \tag{5.50}$$

Therefore in order to find the mean-field results for the degree distribution in this model we can following similar steps performed for the BA model. Defining $P(k_i(t) > k)$ as the probability that a random node of the network has degree $k_i(t) > k$ we obtain

$$\begin{aligned} P(k_i(t) > k) &= P\left(m \ln \left(e\frac{t}{t_i} \right) > k \right) \\ &= P\left(t_i < te^{-k/m+1} \right) \\ &= ee^{-k/m} \end{aligned} \tag{5.51}$$

The degree distribution of the model in the mean-field approximation is then given by

$$\begin{aligned} P(k) &= \frac{dP(k_i(t) < k)}{dk} \\ &= \frac{e}{m} e^{-k/m} \end{aligned} \tag{5.52}$$

for $k \geq m$. Therefore this model displays an *exponential degree distribution* which displays finite average degree $\langle k \rangle$ and finite $\langle k^2 \rangle$. □

5.6.3 Master equation approach

Proposition 28. *The master equation is the equation describes the evolution of the average number $N_k(t)$ of nodes that at time t have degree k . This equations reads for the uniformly growing network model,*

$$\begin{array}{rcll} N_k(t+1) & = & N_k(t) + m\Pi(k-1)N_{k-1}(t) - m\Pi(k)N_k(t) & for \quad k > m \\ N_k(t+1) & = & N_k(t) - m\Pi(k)N_k(t) + 1 & for \quad k = m. \end{array}$$

where

$$\Pi(k) = \frac{1}{N(t)}. \tag{5.53}$$

where $N(t) = t + n_0$ is the total number of nodes in the network at time t.

Proof. The derivation of this result follows the same steps used already for deriving the master equation for the BA model. The difference is all encoded in the different functional dependence of $\Pi(k)$. □

Proposition 29. *The degree distribution $P(k)$ of the growing network model with uniform attachment in the limit $t \to \infty$ is given by*

$$P(k) = \left(\frac{m}{1+m}\right)^{k-m} \frac{1}{1+m} \tag{5.54}$$

for $k \geq m$. Therefore for large values of k we have $P(k)$ is decaying exponentially.

Proof. For $t \gg 1$ we can approximate $\Pi(k)$ as

$$\Pi(k) = \frac{1}{N(t)} = \frac{1}{t+n_0} \simeq \frac{1}{t}. \tag{5.55}$$

Therefore the master equation for the average number of nodes $N_k(t)$ that have degree k at time t reads for $t \gg 1$

$$N_k(t+1) = N_k(t) + \frac{m}{t}N_{k-1}(t) - \frac{m}{t}N_k(t), \tag{5.56}$$

for $k > m$ and

$$N_k(t+1) = N_k(t) - \frac{m}{t}N_k(t) + 1, \tag{5.57}$$

for $k = m$. Now we observe that for sufficiently large values of $t \gg 1$ we have

$$N_k(t) \simeq tP(k) \tag{5.58}$$

Therefore the master equation becomes

$$(1+m)P(k) = mP(k-1) \tag{5.59}$$

for $k > m$. This recursive equation has solution

$$P(k) = \left(\frac{m}{1+m}\right)^{k-m} P(1). \tag{5.60}$$

Moreover the master equation (5.53) for $k = m$ can be written has solution

$$P(1) = \frac{1}{1+m}. \tag{5.61}$$

$$\tag{5.62}$$

Therefore the degree distribution of this model is given by

$$P(k) = \left(\frac{m}{1+m}\right)^{k-m} \frac{1}{1+m} \tag{5.63}$$

for $k \geq m$. □

Chapter 6

Evolving networks

6.1 Introduction (E)

Some people have a special talent to turn a random meeting into a long lasting social interaction, acquiring new friends at a faster rate than other people. Similarly, some scientific papers are highly innovative and they attract citations at a faster rate than other ones. In very competitive environments, like in the World-Wide-Web, some nodes like Google or Facebook have acquired links at an incredible fast rate, becoming the leading websites of the entire network. We tend to associate these differences with some perceived quality of the nodes, such as the social skills of an individual, the content of a web page, or the content of a scientific article. We will call this the node's *fitness*, describing its ability to compete for links at the expense of other nodes.

In this chapter we will provide describe the Bianconi-Baraási model that is the most basic model for network evolution that allow very fit nodes, even latecomers, to become the hub of the network. This model can explain the fundamental mechanism by which the nodes of a network can acquire links at different rates. Interestingly enough this model, as a function of the distribution of the fitness of the nodes can display a structural phase transition, called a condensation transition. When the condensation phenomena occurs the fittest node of the network becomes a super hub, i.e. it acquires a finite fraction of links of the network. Despite the clear differences between this model and the physics of a quantum Bose gas, the model can be mathematically mapped to a Bose gas. In this mapping the condensation phase transition observed in the structure of the network can be mapped to a well known phase transition in quantum physics, the Bose-Einstein condensation of a quantum Bose gas.

The remaining of this chapter will be devoted on one side in describing additional mechanisms for complex network evolution describing different phenomena as the addition of internal links or removal of nodes from the network, on the other side, we will explore the evolution of Cayley trees, described by a Fermi-Dirac distribution, emphasizing important mechanism by which quantum

statistics can emerge in evolving networks.

6.2 The Bianconi-Barabási model (E)

The fitness is the ability of a person to turn a random encounter in a lasting friendship, or the content of a scientific article or the innovation of a webpage or of a company. In the Barabási-Albert model, it is assumed that the rate at which a node acquires new links only depend on its degree (*preferential attachment*). Therefore in the Barabási-Albert model the nodes have a 'first mover advantage', i.e. the nodes arrived at the beginning of the network evolution are also the nodes with higher degree. This is not the case for complex networks in general, for example in the World-Wide-Web, Google was certainly a latecomer arriving in the network after very successful search engines like Altavista an Inktomi. Nevertheless, Google, thanks to the PageRank algorithm, established itself and overcome previous search engines like Altavista and Inktomi. Similarly, Facebook had an even later start and become the Web's biggest hub only in 2011. The Bianconi-Barabási model describes the success of latecomers in complex networks through the introduction of a parameter, the fitness η_i of each node i characterizing the ability of a node to acquire new links. Therefore the Bianconi-Barabási model is a growing network model evolving according to a fitness based preferential attachment,describing the fact that new links are more likely to be attached to high degree and high fitness nodes.

Definition 65. *The* Bianconi-Barabási model *is the most fundamental growing network model including the effect of a preferential attachment mechanism biasing the choice of the target node of the new links toward nodes having high degree and high fitness value. At time $t = 1$ the network is formed by $n_0 > m$ nodes connected by m_0 links. Each node i is assigned a fitness value η_i drawn from the distribution $\rho(\eta)$ and kept fixed over time.*
The model evolves in time by

- **GROWTH:**
  At each time $t > 1$ a new node is added to the network and connected to the existing network with exactly m links to distinct nodes of the network. The new node has a fitness η drawn from the distribution $\rho(\eta)$. ;

- **PREFERENTIAL ATTACHMENT TO NODES OF HIGH DEGREE AND HIGH FITNESS VALUE:**
  Every new link of the new node is attached to an existing node i of the network not already linked to the new node with probability

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j} \tag{6.1}$$

  where k_i is the degree of node i and where the sum over the nodes j is extended over all the nodes of the network not already linked to the new node.

6.2.1 Mean-field solution of the model

Let us solve the model in the mean-field approximation. We will first define the differential equation that the degree $k_i(t)$ of node i needs to satisfy, and then solve this equation using a self-consistent approach.Finally we find the degree distribution of the model.

Proposition 30. *In the mean-field approximation, the degree $k_i(t)$ of node i at time t satisfies the following differential equation*

$$\frac{dk_i(t)}{dt} = m\frac{\eta_i k_i}{\sum_j \eta_j k_j} \tag{6.2}$$

Proof. In fact in the mean-field approximation the degree $k_i(t)$ of node i at time t is the average degree of the original stochastic model. Now the average number of links that a node i acquires at time t is given by $m\Pi_i$ with $\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}$. It follows that $k_i(t)$ satisfies Eq. (6.2). □

Proposition 31. *In the mean-field approximation the degree k_i of node i arrived in the network at time t_i increases with time as a power-law with exponent $f(\eta) = \eta/C$. In particular*

$$k_i = m\left(\frac{t}{t_i}\right)^{\eta_i/C} \tag{6.3}$$

for $t \geq t_i$. In Eq. (6.3) C is a constant depending on the fitness distribution $\rho(\eta)$ satisfying the following equation

$$1 = \int d\eta \rho(\eta) \frac{1}{C/\eta - 1}. \tag{6.4}$$

The dynamical solution of the model implies that nodes of higher fitness value acquire links at a faster rate than nodes with low fitness. Therefore latecomers, with high fitness will become the hubs of the network giving rise to the so called *fit-get-rich* phenomena.

Nevertheless if one compares nodes with the same fitness older nodes have more links than younger ones.

Proof. We assume self-consistently that the normalization sum $\sum_j \eta_j k_j$ has the limiting behaviour $\sum_j \eta_j k_j \simeq mCt$ for $t \gg 1$, therefore

$$\lim_{t\to\infty} \frac{\sum_j \eta_j k_j}{mt} = C \tag{6.5}$$

with $C > 0$. (Self-consistently means that we will check at the end of the calculation if this is indeed consistent with the solution of the model.) In this hypothesis, and in the limit $t \gg 1$ we can write the dynamical man-field equation for the degree $k_i(t)$ of node i, getting

$$\frac{dk_i}{dt} = \frac{\eta_i k_i}{Ct}, \tag{6.6}$$

with initial condition $k_i(t_i) = m$. This equation has solution

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\eta/C}. \tag{6.7}$$

The us note that, since the total number of links is increasing linearly with time, the degree of the nodes in the network cannot grow faster than linearly in time. Therefore let us assume $\eta/C < 1$. In order to check if our assumption in Eq. (6.5) is consistent with the solution (6.7) we note that Eq. (6.5) implies

$$\lim_{t\to\infty} \frac{\left\langle \sum_j \eta_j k_j \right\rangle}{mt} = C \tag{6.8}$$

Now the quantity $\left\langle \sum_j \eta_j k_j \right\rangle$ can be calculated using the solution Eq. (6.7) and the continuous approximation , getting

$$\begin{aligned} \left\langle \sum_j \eta_j k_j \right\rangle &\simeq \int d\eta\rho(\eta) \int_1^t dt_j \eta m \left(\frac{t}{t_j}\right)^{\eta/C} \\ & m \int d\eta\rho(\eta) \frac{\eta}{1-\eta/C}\left[t - t^{\eta/C}\right] \end{aligned} \tag{6.9}$$

Since $\eta/C < 1$, if we perform the limit in Eq. (6.8), we get a self-consistent equation for the constant C given by

$$C = \int d\eta\rho(\eta) \frac{\eta}{1-\eta/C} \tag{6.10}$$

this equation can be also written as

$$1 = \int d\eta\rho(\eta) \frac{1}{C/\eta - 1}. \tag{6.11}$$

□

Proposition 32. *In the mean-field approximation the degree distribution of the Bianconi-Barabási model is given by*

$$P(k) = \int d\eta\rho(\eta) \frac{C}{\eta} m^{C/\eta} \frac{1}{k^{C/\eta+1}}. \tag{6.12}$$

Moreover it can be shown that this networks always generate a scale-free network with finite $\langle k\rangle$ and diverging $\langle k^2\rangle$, i.e. the degree distribution is described by a power-law $P(k) \propto k^{-\gamma}$ with $\gamma \in (2,3]$ including sometime some logarithm corrections to this scaling.

Proof. The probability $P(k_i(t) > k|\eta_i = \eta)$ that a random node with fitness value $\eta_i = \eta$ has degree $k_i(t) > k$, in the mean-field approximation can be calculated as follow

$$P(k_i(t) > k|\eta_i = \eta) = P\left(m\left(\frac{t}{t_i}\right)^{\eta/C} > k\right) = P\left(t_i < t\left(\frac{m}{k}\right)^{C/\eta}\right) = \left(\frac{m}{k}\right)^{C/\eta} \tag{6.13}$$

Therefore, the degree distribution $P(k|\eta)$ of the nodes with fitness value η is given by

$$P(k|\eta) = \frac{dP(k_i(t) < k|\eta_i = \eta)}{dk} = \frac{d}{dk}\left(\frac{m}{k}\right)^{C/\eta} = \frac{C}{\eta} m^{C/\eta} \frac{1}{k^{C/\eta+1}}. \quad (6.14)$$

Finally the degree distribution of the entire network is given by

$$P(k) = \int d\eta \rho(\eta) P(k|\eta) = \int d\eta \rho(\eta) \frac{C}{\eta} m^{C/\eta} \frac{1}{k^{C/\eta+1}}. \quad (6.15)$$

□

Let us consider the Bianconi-Baraási model for simple example of fitness distribution.

- *1) Case in which all the fitness are the same*
  In this case $\rho(\eta) = \delta(\eta, 1)$, i.e. $\eta_i = 1\ \forall i$. In this limit the Bianconi-Barabási model becomes the Baraási-Albert model with $C = 2$ solution of the equation

  $$1 = \frac{1}{1/C - 1}. \quad (6.16)$$

  The degree of the nodes increase in time with the same rate

  $$k_i(t) = m\left(\frac{t}{t_i}\right)^{1/2}, \quad (6.17)$$

  i.e. older node are the nodes have higher degree than young nodes. The degree distribution is power-law with exponent

  $$\gamma = 1 + C = 3. \quad (6.18)$$

- *2) Case of uniform distribution of fitness*

  In this case $\rho(\eta) = 1$, and $\eta \in [0, 1]$. The with higher fitness increases their degree faster than nodes with lower fitness (see figure 6.1). The mean-field evolution is described by

  $$k_i(t) = m\left(\frac{t}{t_i}\right)^{\eta_i/C}. \quad (6.19)$$

  The constant C satisfies the equation

  $$\begin{aligned} 1 &= \int_0^1 \frac{\eta/C}{1 - \eta/C} d\eta \\ 1 &= C\int_0^{1/C} \left(-1 + \frac{1}{1-x}\right) dx \\ \frac{2}{C} &= -\ln\left(1 - \frac{1}{C}\right) \\ 1 - \frac{1}{C} &= e^{-2/C}, \end{aligned} \quad (6.20)$$

whose solution is $C^\star = 1.255\ldots$. The degree distribution is given by

$$P(k) = \int_0^1 d\eta m^{-1} \frac{C^\star}{\eta} \left(\frac{m}{k}\right)^{C^\star/\eta+1} \simeq \frac{k^{-1-C^\star}}{\ln k}, \tag{6.21}$$

i.e. the power-law exponent $\gamma = 1 + C^\star = 2.255\ldots$ but the degree distribution has additional logarithm corrections to a pure power-law scaling. In fact the integral in Eq. (6.21) can be evaluated by saddle point by writing it has

$$P(k) = \int_0^1 d\eta m^{-1} e^{F_k(\eta)} \tag{6.22}$$

where

$$F_k(\eta) = (C^\star/\eta + 1) \ln\left(\frac{m}{k}\right) + \ln\left(\frac{C^\star}{\eta}\right). \tag{6.23}$$

Expanding $F_k(\eta)$ around its maximum in the interval $\eta \in [0,1]$ achieved for $\eta = 1$ we get

$$\begin{aligned} F_k(\eta) &\simeq F_k(1) + F_k'(1)(\eta - 1) \\ &= \left[(C^\star + 1)\ln\left(\frac{m}{k}\right) + \ln(C^\star)\right] \\ &\quad + \left[C^\star \ln(k/m) - 1\right](\eta - 1) \end{aligned} \tag{6.24}$$

where we stop at the first order of the expansion. Therefore

$$\begin{aligned} P(k) &= \int_0^1 d\eta m^{-1} e^{F_k(\eta)} \\ &\simeq \int_0^1 d\eta m^{-1} e^{F_k(1) + F_k'(1)(\eta-1)} \\ &= m^{-1} e^{F_k(1)} \frac{1}{F_k'(1)} (1 - e^{-F_k'(1)}) \\ &= m^{C^\star} C^\star k^{-1-C^\star} \frac{1}{C^\star \ln(k/m) - 1} \left[1 - e\left(\frac{m}{k}\right)^{C^\star}\right] \end{aligned} \tag{6.25}$$

and for $k \gg 1$ we have

$$P(k) \propto k^{-1-C^\star} \frac{1}{\ln(k)} \tag{6.26}$$

6.3 Evolving networks

6.3.1 Model with initial attractiveness of the nodes (E)

On of the most simple variations of the Barabási-Albert model is the model with initial attractiveness of the nodes $A > -m$.

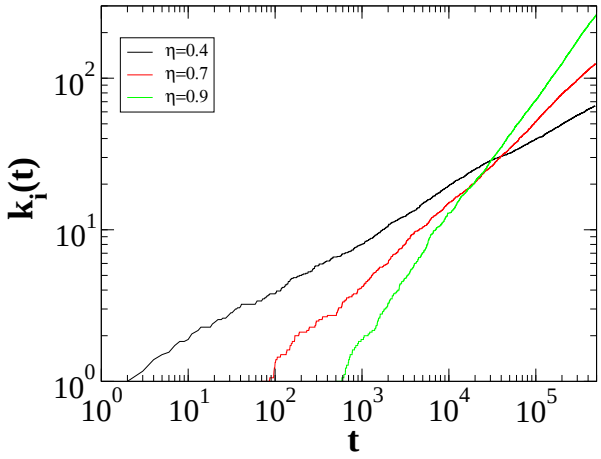

Figure 6.1: The effect of fitness in the dynamics of the degree of the nodes, node with higher fitness also if arrived later in the network, can take over nodes with lower value of the fitness, because their degree increases at a faster rate than the degree of less fit nodes.

Definition 66. *In the model with initial attractiveness of the nodes A the preferential attachment of the BA model is slightly modified At time $t = 0$ the network is formed by n_0 nodes connected by $m_0 > m$ links.*
The model evolves in time by

- **GROWTH:**
  At each time $t > 0$ a new node is added to the network and connected to the existing network with exactly m links to distinct nodes of the network;
- **PREFERENTIAL ATTACHMENT WITH INITIAL ATTRACTIVENESS:**
  Every new link of the new node is attached to an existing node i of the network not already linked to the new node with probability

$$\Pi_i = \frac{k_i + A}{\sum_j (k_j + A)} \tag{6.27}$$

  where k_i is the degree of node i and where the sum over the nodes j is extended over all the nodes of the network not already linked to the new node.

This model can be easily solved using the same methods adopted for the BA model (either mean-field of master equation approach). We leave as an exec rice for the student that the following conclusion can be drawn form the mean-field solution of the model.

- *Evolution of the degrees $k_i(t)$ in the mean-field approximation*
  The degree $k_i(t)$ of node i evolves, in the mean-field approximation according to the following dynamic behaviour

$$k_i(t) = (m + A)\left(\frac{t}{t_i}\right)^{\frac{1}{2+A/m}} - A \tag{6.28}$$

  for $t \gg 1$.

- *The degree distribution $P(k)$ of the network the mean-field approximation*
  The degree distribution $P(k)$, is given in the mean field approximation is scale-free

$$P(k) \ \propto k^{-\gamma} \tag{6.29}$$

  with power-law exponent $\gamma = 3 + \frac{A}{m}$.

6.3.2 Krapivsky-Redner model with non-linear preferential attachment (NE)

In the Krapivsky-Redner mode a non-linear preferential attachment with probability $\Pi_i \propto k_i^\alpha$ and $\alpha > 0$ has been considered. The results is that only in the case $\alpha = 1$ a scale-free network can be generated by the model. Therefore the model is defined as follows.

Definition 67. *In the Krapivsky-Redner model includes a non linear preferential attachment. At time $t = 0$ the network is formed by n_0 nodes connected by $m_0 > m$ links.*
The model evolves in time by

- **GROWTH:**
  At each time $t > 0$ a new node is added to the network and connected to the existing network with exactly m links to distinct nodes of the network;
- **NON-LINEAR PREFERENTIAL ATTACHMENT :**
  Every new link of the new node is attached to an existing node i of the network not already linked to the new node with probability

$$\Pi_i = \frac{k_i^\alpha}{\sum_j k_j^\alpha} \tag{6.30}$$

  where k_i is the degree of node i and where the sum over the nodes j is extended over all the nodes of the network not already linked to the new node.

The degree distribution of the model depends on the value of the exponent α modulating the non-linear preferential attachment.

- *Linear preferential attachment:* $\alpha = 1$
  The model reduces to the BA model.
  The network is scale-free with power-law exponent $\gamma = 3$.
- *Sublinear preferential attachment* $\alpha < 1$
  The degree distribution is described by a stretched exponential function with finite average degree $\langle k \rangle$ and finite $\langle k^2 \rangle$. The degree distribution is homogeneous, the network is not scale-free.
- *Superlinear preferential attachment* $\alpha > 1$
  There is a gelation phenomena in the network, in which one node, the oldest node of the network acquires a number of links $k_1(t) \simeq t$. The degree distribution is highly inhomogeneous but not described by a scale-free degree distribution. The gelation s a phenomena close the the condensation of the links but the fraction of links on the gelated node is always almost equal to 1. In the model there is not fitness, therefore only the oldest node can acquire a finite fraction of the links, while in a network displaying the Bose-Einstein condensation a node with very high fitness, also if it is not the first can acquire a finite fraction of the links of the network.

6.3.3 Effect of node deletion (E)

In the growing network model, new nodes are continuously added to the network but sometimes in real network nodes are also leaving the network. For example in scientific collaboration networks, some node might stop the activity due to change of job or retirement. In the contest of growing network model where the

rate at which new nodes are added in the network is 1, we have the following scenario, depending on the rate r at which the nodes are removed from the network.

- *Case* $r < 1$.
  The network is growing in size, despite the removal of some nodes at rate r. In presence of preferential attachment the network remains scale-free.

- *Case* $r = 1$.
  In this case the model is dominated by fluctuations. In some realization the networks can be reduced to zero. In general we do not expect scale-free degree distribution since the network is not growing typically.

- *Case* $r > 1$.
  The network is shrinking in size, describing a network that is disappearing.

Chapter 7

Small world properties of complex networks

7.1 Introduction (E)

In this chapter we will discuss the small-world properties of a number of complex networks, and discuss the Small World Network Model proposed by Watts and Strogatz for modelling complex networks. In order to introduce the small-world properties of complex networks and their implication we will first define the clustering coefficient, then state the small-world network properties which define the wide universality class of *small-world networks* with small average distance and large clustering coefficient. Subsequently we calculate the clustering coefficient the diameter and the average distance of several reference networks. Finally we will discuss the small-world network model first proposed by Watts and Strogatz to model complex networks.

7.2 Clustering coefficient of a network (E)

Definition 68. *The* local clustering coefficient C_i *of node* i *of degree* k_i *is given by*

$$C_i = \begin{cases} \frac{\#\ \textit{of triangles passing though node } i}{\frac{1}{2}k_i(k_i-1)} & \textit{for} \quad k_i > 1, \\ 0 & \textit{for} \quad k_i = 0, 1. \end{cases} \tag{7.1}$$

where $\frac{1}{2}k_i(k_i-1)$ *enumerates the number pairs of distinct nodes which are neighbours of node* i *if* $k_i > 1$.

The clustering coefficient measure "partial transitivity" in networks, e.g. in social networks measures the fraction of the total pairs of friends of a node (neighbours in the social network) that are each other friends. In the figure

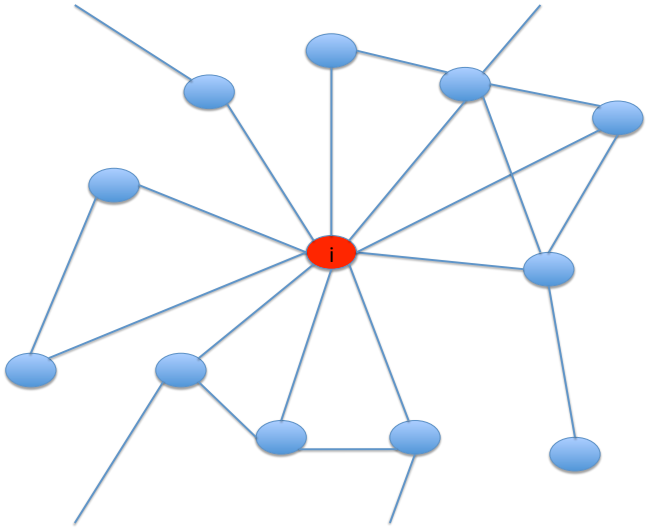

Figure 7.1: The local clustering coefficient of node i is given by $C_i = 7/45$.

7.1 we show how to calculate the clustering coefficient of a node in a concrete example.

The local clustering coefficient C_i of node i is given by $C_i = 1$ if all the pairs of neighbours nodes of node i are connected, i.e. if the set of nodes formed by node i and its neighbours are in a clique. If, instead, $C_i = 0$ none of the neighbours of node i are linked together. For example in a tree network all the nodes i of the network have local clustering coefficient $C_i = 0$.

Definition 69. *The* Watts-Strogatz clustering coefficient C_{WS} *of a network of size* N *is the average of the local clustering over all the nodes of the network and is given by*

$$C_{WS} = \frac{1}{N}\sum_{i=1}^{N} C_i. \tag{7.2}$$

This is the definition of the clustering coefficient used by Watts and Strogatz. Sometimes in the literature and alternative global clustering coefficient is introduced.

Definition 70. *The* global clustering coefficient C *of a network of size* N *is given by*

$$C = \frac{3\#\textit{of triangles in the network}}{\#\textit{of distinct paths of length 2}}. \tag{7.3}$$

This last definition of the clustering coefficient is used sometimes to avoid the fact that the Watts and Strogatz clustering coefficient is dominated by the contribution of nodes with small degree. For regular lattices where the local structure of the network around node i is independent on the choice of the node i, then the Watts-Strogatz clustering coefficient is equal to the global clustering coefficient and the local clustering coefficient, i.e. $C_i = C_{WS} = C$.

7.3 The Small World Network Properties

In their paper Watts and Strogatz show that many complex networks with finite average degree $\langle k \rangle$ from the neural networks of C.elegans to the actor collaboration network have two coexisting important properties:

Definition 71. *The networks that have the diameter D scaling with the number of nodes as*

$$D = \mathcal{O}(\ln N), \tag{7.4}$$

or

$$D = o(\ln N) \tag{7.5}$$

i.e. they display the small world distance property*.*

Proposition 33. *If a network displays the small-world distance property, then, since we have always $\langle d \rangle = \ell \leq D$, then either*

$$\ell = \mathcal{O}(\ln N) \tag{7.6}$$

or

$$\ell = o(\ln N). \tag{7.7}$$

Since, as we will see in the following paragraphs, the average distance of a random Poisson network with average degree $\langle k \rangle$ is given by

$$\ell_{rand} = \frac{\ln N}{\ln \langle k \rangle} \tag{7.8}$$

a network has a small world distance property if its diameter and its average distance are of the same order of magnitude of the average distance of a random Poisson network with the same average degree $\langle k \rangle$.

Definition 72. *The network with Watts and Strogatz clustering coefficient C_{WS}*

$$C_{WS} \gg \frac{\langle k \rangle}{N} \tag{7.9}$$

have high clustering coefficient

C.elegans neural network	282	14	2.65	2.25	0.28	0.05
Power-grids	4941	2.67	18.7	12.4	0.08	0.005
Internet (snapshots)	3015-6209	3.52-4.11	3.7-3.75	6.36-6.18	0.18-0.3	0.001
WWW (snapshot)	153127	35.21	3.1	3.35	0.1078	0.00023
World, synonyms	22311	13.48	4.5	3.84	0.7	0.006

Table 7.1: Small-world universality: Many complex networks have the two small-world properties, as first observed by Watts and Strogatz in 1998.

Since, as we will see in the following paragraphs, the clustering coefficient of a random Poisson network with average degree $\langle k \rangle$ is given by

$$C_{rand} = \frac{\langle k \rangle}{N} \tag{7.10}$$

a network has a high clustering coefficient if its clustering coefficient is much larger than the clustering coefficient of a random Poisson network with the same average degree $\langle k \rangle$. The following is the strict definition of small-world networks

Definition 73. *Stricktly speaking* small-world networks *have at the same time the small-world distance property and high clustering coefficient.*

The small-world networks are ubiquitous, and for this reason the small-world networks are said to constitute a complex networks universality. In Table **??** we provide a table of a large set of real complex networks that display the small world properties. Example include the neural network of the worm c.elegans, the movie actor network, the collaboration networks of scientists, food webs, the power-grid, and words synonyms.

7.4 Regular one dimensional lattice with nodes of degree k (E)

Here we will consider a specific example of regular lattice: a regular one-dimensional lattice with nodes of degree k.

Definition 74. *A one dimensional regular lattice with nodes of degree k (with k even) is a chain of N nodes labelled $i = 1, 2 \ldots, N$ with N even, and such that*

$$A_{ij} = \begin{cases} 1 & \text{if} \quad i \neq j \ \& \ \left| \frac{N}{2} - \left| \frac{N}{2} - |i-j| \right| \right| \leq k/2 \\ 0 & \text{otherwise} \end{cases}$$

An example of the regular one dimensional lattice with nodes of degree $k = 4$ is shown in Figure 7.2.

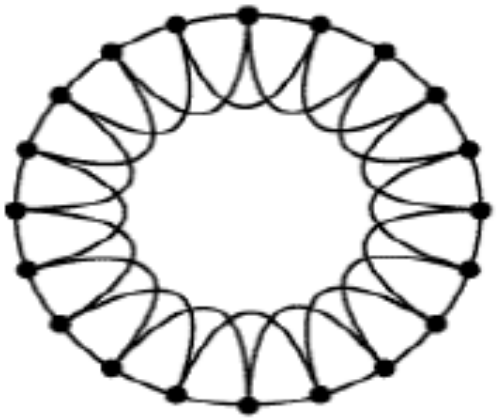

Figure 7.2: One dimensional regular lattice with nodes of degree $k = 4$

7.4.1 Clustering coefficient

Proposition 34. *The Watts-Strogatz clustering coefficient C_{WS} in a regular one dimensional lattice with nodes of degree k is equal is given by*

$$C_{WS} = \frac{3}{4}\frac{k-2}{k-1}. \tag{7.11}$$

Proof. The number of triangles passing through a given node in the network is given by

$$\left(\frac{k}{2}-1\right)\frac{3}{4}k \tag{7.12}$$

In fact there $(k/2-1)+(k/2-n)$ triangles passing through the node i and a node j distant $n \in [1, k/2]$ steps around the ring from the original node i of the network as shown in Figure 7.3. If we sum over the nodes j distant $n \in [1, k/2]$ either on the right or on the left of the original node i we obtain the total number of triangle passing through the node i multiplied by two, because every triangle is counted twice.

Therefore the number of triangles passing through a node is given by

$$\sum_{n=1}^{k/2} k-1-n = (k-1)\frac{k}{2} - \frac{k}{4}\left(\frac{k}{2}+1\right) = \tag{7.13}$$

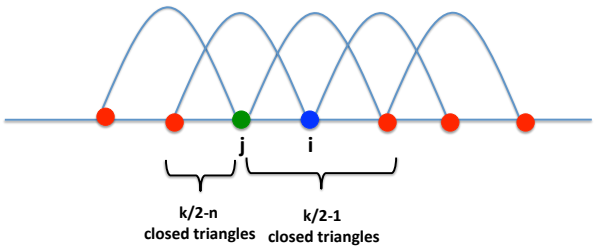

Figure 7.3: Number of triangles passing through two linked nodes i and j distant $n = mod(i-j, N/2)$ on the ring. The number of closed triangles are $(k/2-1)+(k/2-n)$ for a regular one-dimensional lattice of nodes of degree k. The figure show a region of a one dimensional lattice with nodes of degree $k = 4$. For our concrete example $n = 1$.

Therefore the clustering coefficient is given by

$$C_i = \frac{\frac{3}{4}k\left(\frac{k}{2}-1\right)}{\frac{1}{2}k(k-1)} = \frac{3}{4}\frac{k-2}{k-1}. \tag{7.14}$$

for every i. It follows that

$$C_{WS} = \frac{3}{4}\frac{k-2}{k-1}. \tag{7.15}$$

□

Therefore the clustering coefficient is a constant independent on the network size, and therefore C_{WS} if finite in the limit $N \to \infty$.

7.4.2 Diameter and Average Distance

Proposition 35. *The diameter of the regular one dimensional lattice with nodes of degree k has a diameter that scales in the large network limit $N \gg 1$, as*

$$D \simeq \frac{N}{k}, \tag{7.16}$$

therefore the network does not have the small-world distance property.

Proof. The pair of nodes that are further apart are the ones at distance $N/2$. Since nodes at distance $k/2$ are directly linked we can extimate the diameter of the network as

$$D \simeq \frac{N/2}{k/2} = \frac{N}{k}. \tag{7.17}$$

□

Similarly it can be shown that the the following statement for the average distance of the regular one dimensional lattice with nodes of degree k.

Proposition 36. *The average distance ℓ of the regular one dimensional lattice with nodes of degree k increases linearly with the network size, i.e.*

$$\ell \simeq \frac{N}{k}. \tag{7.18}$$

7.5 Cayley tree (E)

A Cayley tree is a simple example of symmetric regular tree in which the nodes have either degree k or degree 1 (see figure 7.4 for and example of a Cayley tree with $k = 3$).

Definition 75. *A Cayley tree is a symmetric regular tree constructed starting from a central node of degree k.*
In a Cayley tree network every node at distance d from the central node has degree k until we reach the nodes at distance $\mathcal{P}$ that have degree one and are called the leaves of the network. The quantity $b = k - 1$ is called the branching ratio *of the tree.*

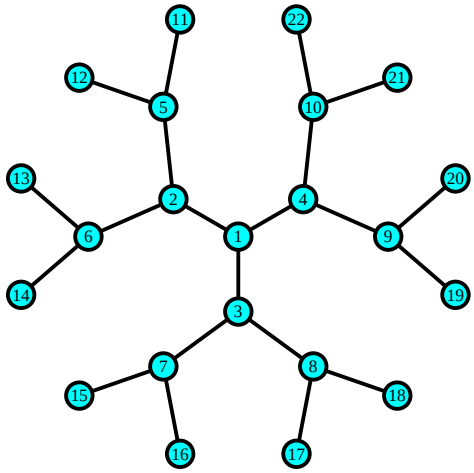

Figure 7.4: A Cayley tree network with $k = 3$ and $\mathcal{P} = 3$.

7.5.1 Diameter of a Cayley tree

The Cayley tree has a small-world distance property, in fact we will prove the following proposition.

Proposition 37. *The diameter D of a Cayley tree can be expressed in terms of the total number of nodes N in the network giving in the large network limit $N \gg 1$*

$$D \simeq \frac{\ln N}{\frac{1}{2}\ln(k-1)}, \tag{7.19}$$

i.e. $D = \mathcal{O}(\ln N)$, and therefore the Cayley tree has the small-world distance property.

Proof. The proof of the proposition as several steps.

- a) Let us prove by iteration that the number of nodes N_d at distance d from the central node is given by

$$N_d = \begin{cases} kb^{d-1} & \text{for} d \geq 1, \\ 1 \quad \text{for} \quad d = 0, \end{cases}$$

where the quantity $b = k-1$ is called the " branching ratio" of the Cayley tree. The above relation is valid for the nodes at distance $d = 0$. In fact the number of nodes at distance zero from the central node is given by one, i.e. it is only the central node itself. Moreover the relation given by Eq. (7.20) is also valid for distances $d = 1$ from the central node, since the degree of the central node is k.
Let us now show that if the relation (7.20) is valid for the nodes at distance $d \geq 1$, then it must be valid for the nodes at distance $d+1$.
In fact we have that every node i at distance $d < \mathcal{P}$ from the central node has degree k, i.e. is linked to other k nodes.
Since the Cayley tree is connected and does not contain loops, only one of these k links is attached to a node at distance $d-1$ from the central node while the other $b = k-1$ links are attached to nodes at distance $d+1$ from the central node. It follows that each node at distance d will branch into $b = k-1$ nodes at distance $d+1$.
Moreover since the Cayley tree network does not contain loops any node at distance $d+1$ from the central node can be reached only by one node at distance d.
Therefore we will have

$$N_{d+1} = N_d(k-1) = kb^{d-1}b = kb^d \tag{7.20}$$

- b) Using the formula for the sum of the first terms of a geometric series, we now show that the total number of nodes in the network is given by

$$N = 1 + k\left[\frac{b^{\mathcal{P}} - 1}{b-1}\right]. \tag{7.21}$$

The total number of nodes in the Cayley tree is given by the sum of 1 (indicating that there is only one central node) at the sum of all the number of nodes N_d with distances $d \in [1, \mathcal{P}]$ from the central node. Therefore we have

$$N = 1 + \sum_{d=1}^{\mathcal{P}} kb^{d-1} = 1 + k\frac{1-b^{\mathcal{P}}}{1-b} = 1 + k\left[\frac{(k-1)^{\mathcal{P}} - 1}{k-2}\right]. \tag{7.22}$$

- c) We then observe that the diameter D of the Cayley tree is given by $D = 2\mathcal{P}$.

In fact, the maximum distance in the network is the distance between any two leaves nodes connected to the central node by non-overlapping

paths. Since the distance of any leaf node from the central node is $\mathcal{P}$, the diameter of the Cayley tree is given by $D = 2\mathcal{P}$.

- d) Finally we find an expression for the diameter D of the network in terms of the total number of nodes N.
  Using $D = 2\mathcal{P}$ and using Eq. (7.21) we can derive the expression of D as a function of N, i.e.

$$\begin{aligned} N &= 1 + \frac{k}{k-2}\left[(k-1)^{D/2} - 1\right] \\ \frac{(N-1)(k-2)}{k} &= (k-1)^{D/2} - 1 \\ \frac{(N-1)(k-2)}{k} + 1 &= (k-1)^{D/2} \\ \frac{D}{2}\ln(k-1) &= \ln\left[1 + \frac{(N-1)(k-2)}{k}\right] \\ D &= \frac{2}{\ln(k-1)}\ln\left[1 + (N-1)\frac{k-2}{k}\right]. \end{aligned} \quad (7.23)$$

  This final expression in the limit $N \gg 1$ is given by

$$D \simeq \frac{\ln N}{\frac{1}{2}\ln(k-1)} \quad (7.24)$$

□

7.5.2 Clustering coefficient

Nevertheless the Cayley tree is not a strickly speaking small-world network because it has zero clustering coefficient.

Proposition 38. *The local clustering coefficient C_i of any node i of a Cayley tree is zero, i.e. $C_i = 0$. Moreover also the Watts and Strogatz clustering coefficient and the global clustering coefficient of the Cayley tree are zero, i.e. $C_{WS} = C = 0$.*

Proof. The Cayley network is a tree, i.e. does not contains any loop and therefore does not contain any triangle. It follows that in a Cayley tree is $C_i = C_{WS} = C = 0$ for every node i of the network. □

7.6 Random Poisson network (E)

The Random Poisson network is a network of the $\mathbb{G}(N,p)$ ensemble with $p = \frac{\langle k \rangle}{N}$.

7.6.1 Average distance and Diameter of a Random Graph

A random Poisson network is locally tree-like (the number of finite loops is finite in the limit of large network sizes $N \to \infty$.) Therefore it is possible to extend and generalize the argument used for calculating the diameter of a Cayley tree to evaluate the average distance and the diameter of random graphs. Here we give the following proposition, whose proof will be given in chapter 7.

Proposition 39. *In a random Poisson network with average degree $\langle k \rangle = c$ the number N_d of nodes at distance d from a given node i of degree k_i can be approximated to be*

$$N_d \sim \begin{cases} k_i c^{d-1} & \text{for} \quad d > 0 \\ 1 & \text{for} \quad d = 0 \end{cases}$$

The Eq. (7.25) is very close to the expression given by Eq. (7.20) where we have substituted the branching ratio $b = k - 1$ with the average branching ratio of the Poisson network $\bar{b} = \frac{\langle k(k-1) \rangle}{\langle k \rangle} = c$.

Proposition 40. *The* average distance *$\langle d \rangle = \ell$ of a Poisson network with average degree $\langle k \rangle = c$ can be approximated in the limit of large network sizes $N \gg 1$ to*

$$\ell \sim \frac{\ln N}{\ln c} \tag{7.25}$$

Proof. We start from the expression given by Eq. (7.25) for the node at distance d from a given node i of degree i. Assuming that the node i is choose at random in the network we can evaluate the number of nodes at distance $d > 0$ from a random node as

$$N_d \sim k_i c^{d-1} \sim c^d. \tag{7.26}$$

Therefore we can evaluate the number of nodes at distance $d \leq d'$ from a random node as

$$N_{d \leq d'} = 1 + \sum_{d=1}^{d'} N_c = \sum_{d=0}^{d'} c^d = \frac{c^{d'+1} - 1}{c - 1}. \tag{7.27}$$

The average distance between the nodes of the network can be estimated by imposing that the number of nodes at distance $d \leq \ell$ must be equal to the total number of nodes N. Therefore we have

$$\begin{aligned} N &= \frac{c^{\ell+1} - 1}{c - 1} \\ c^{\ell+1} &= 1 + N(c-1) \\ \ell + 1 &= \frac{\ln[1 + N(c-1)]}{\ln c}. \end{aligned} \tag{7.28}$$

It follows that in the limit of large network sizes $N \gg 1$

$$\ell \sim \frac{\ln N}{\ln c}. \tag{7.29}$$

□

Here in the following we give the following proposition regarding the diameter of random networks without proof.

Proposition 41. *The* diameter *D of a Poisson network with average degree $\langle k \rangle = c$ can be show to have the same scaling behaviour of the average distance in the large network limit $N \gg 1$,i.e.*

$$D \sim \frac{\ln N}{\ln c} \tag{7.30}$$

7.6.2 Clustering coefficient of a Poisson network

The Poisson network has the small-world distance property but is not a a strickly speaking small-world network. In fact it has a small clustering coefficient.

Proposition 42. *The Watts and Strogatz clustering coefficient C_{WS} of a random network in the $\mathbb{G}(N,p)$ ensemble is in the large network limit is given by*

$$C_{WS} = p \tag{7.31}$$

Proof. Given a node i of a random network in the $\mathbb{G}(N,p)$ ensemble, let k_i indicates its degree. The number of distinct pair of neighbours nodes is given by $\frac{1}{2}k_i(k_i-1)$. Each of these pair of nodes is connected with probability p. Therefore the average number of triangles passing through node i is given by $p\frac{1}{2}k_i(k_i-1)$. Therefore we have that in average the local clustering coefficient of a node in the random network is given by

$$\langle C_i \rangle = \frac{p\frac{1}{2}k_i(k_i-1)}{\frac{1}{2}k_i(k_i-1)} = p. \tag{7.32}$$

In the large network limit we have therefore

$$C_{WS} = p. \tag{7.33}$$

□

Proposition 43. *In a Poisson random network with average degree $\langle k \rangle = c$ the Watts and Strogatz clustering coefficient in the large network limit is given by*

$$C_{WS} = p = \frac{c}{N} \tag{7.34}$$

Therefore on Poisson networks the Watts and Strogatz goes not have a high clustering coefficient and therefore is not a small world network. Also we observe that the clustering coefficient is vanishing in the large network limit $N \to \infty$.

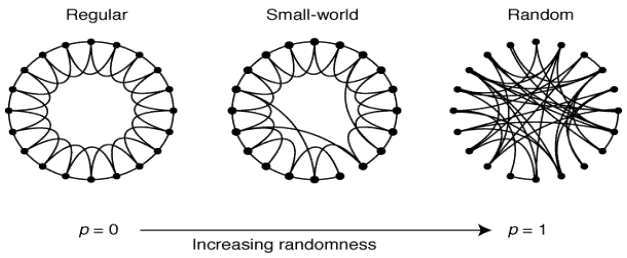

Figure 7.5: The small-world network model. Starting from a regular one dimensional lattice of degree k, links are rewired with probability p. For $p = 0$ the network is a regular one dimensional lattice of degree k. For $p = 1$ the network is a random network in the $\mathbb{G}(N, k/2N)$ ensemble. For a large value of intermediate values of p the network is small world.

7.7 The Small-World Network Model (E)

In regular lattices, random short-cuts can reduce significantly the average distance in the network keeping the high the clustering coefficient. This has been the simple, clear intuition beyond the Small-World model proposed by D.J. Watts and S. Strogatz in 1998.

Definition 76. *A* Small-World model *is obtained starting from a regular one dimensional lattice with nodes of degree k. In other words we start from a lattice of N nodes, place on a ring and such that every node is linked to the k nearest neighbour on the ring. Each link of the network is removed from the lattice with probability p and its two ends are attached to randomly chosen distinct nodes of the network.*

In figure 7.5 we plot networks generated by the small world network model for different value of the rewiring probability p. The model has different regimes

- *Case* $p = 0$
  The network is a regular one dimensional lattice with nodes of degree k. In this case the clustering coefficient C_{WS} is finite in the large N limit,

and is given by Eq. (7.11) that we rewrite here for convenience

$$C_{WS} = \frac{3}{4}\frac{k-2}{k-1}, \tag{7.35}$$

i.e. the network has high clustering coefficient. This network has average distance ℓ given by Eq. (7.18)

$$\ell \sim \frac{N}{k}. \tag{7.36}$$

- *Case* $p = 1$
  All the links are randomly attached to the nodes of the network: the network is a random network in the $\mathbb{G}(N, L)$ with $L = k/2N$ links. The clustering coefficient C_{WS} and the average distance ℓ of the random network in the $\mathbb{G}(N, kN/2)$ network are the same of the ones of a Poisson network of average degree $\langle k \rangle = k$. Therefore we have a small clustering coefficient

$$C_{rand} = \frac{\langle k \rangle}{N}, \tag{7.37}$$

  and a small average distance ℓ given by

$$\ell_{rand} \sim \frac{\ln N}{\ln \langle k \rangle}. \tag{7.38}$$

- *In a wide range of values of* p The network is a small-world network , i.e. it has at the same time high clustering coefficient and small average distance ℓ. In order to show this, in Figure 7.6 the clustering coefficient C and the average shortest distance ℓ are shown as a function of p for a small world network model with $\langle k \rangle = 4, N = 10^3$. The data are averaged over 100 realizations.

The Small-World network capture a basic mechanism for shortening considerably the average distances starting from spatial, local interactions, and can be used to model a large variety of networks. Nevertheless, many small world networks display large fluctuations in the degrees of their nodes, while the degree distribution of the Small-World model does not display such large fluctuations. In Figure 7.7 we plot $\sigma = \sqrt{\langle k^2 \rangle - \langle k \rangle^2}$ for the small-world network as a function of p showing that σ is finite for every value of p. In fact the maximal value of σ is reached for $p = 1$ where $\sigma = \sqrt{k}$ is the standard deviation of a random Poisson network with average degree $\langle k \rangle = k$.

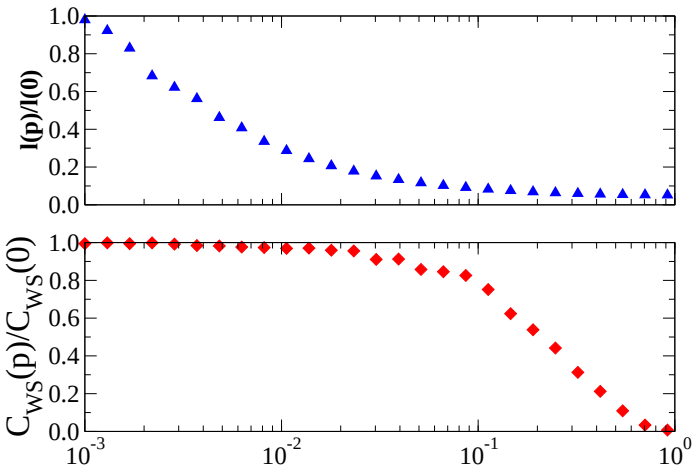

Figure 7.6: The average distance $\ell(p)$ normalized by the value $\ell(p=0)$ and the clustering coefficient $C_{WS}(p)$ normalized by the value $C_{WS}(p=0)$ for the small world network model with $\langle k\rangle = 4, N = 10^3$. The data are averaged over 100 realizations. There is a wide range of values of p in which the network is small-world displaying at the same time small average distance and high clustering coefficient.

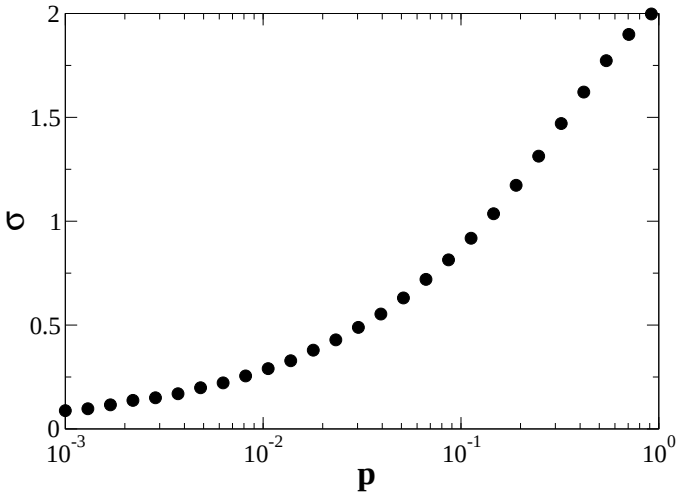

Figure 7.7: The standard deviation of the degree distribution $\sigma = \sqrt{\langle k^2 \rangle - \langle k \rangle^2}$ as a function of the rewiring probability p for a small world network model with $\langle k \rangle = 4, N = 10^3$. The data are averaged over 100 realizations. The standard deviation σ has a maximum for $p = 1$ where it reaches the Poisson random value $\sigma = \sqrt{k}$.

Chapter 8

The configuration model and the Molloy-Reed condition

8.1 The configuration model (E)

Definition 77. *The* configuration model *is the set (or ensemble) of networks with N nodes and with given degree sequence $\{k_i\} = (k_1, k_2, k_3, \dots k_N)$ where k_i is the degree of node $i = 1, 2 \dots, N$.*

A network in the configuration model can be generated by the following recursive procedure: Given a degree sequence $\{k_i\} = (k_1, k_2, \dots, k_N)$ with an even $\sum_j k_j$

- *Step a)*
  We place k_i stubs on each node i of the network.
- *Step b)*
  We match each stub of the network with another stub of the network.
- *Step c)*
  We repeat step b) until all the stubs of the network are matched. *Step d)* If the network constructed in this way contains multiedges and tadpoles repeat step b and step c.

8.2 Uncorrelated networks (E)

Definition 78. *A network with degree distribution $\{k_i\}$ is* uncorrelated *if the probability p_{ij} that a node i is connected to a node j is given by*

$$p_{ij} = \frac{k_i k_j}{\langle k \rangle N}, \tag{8.1}$$

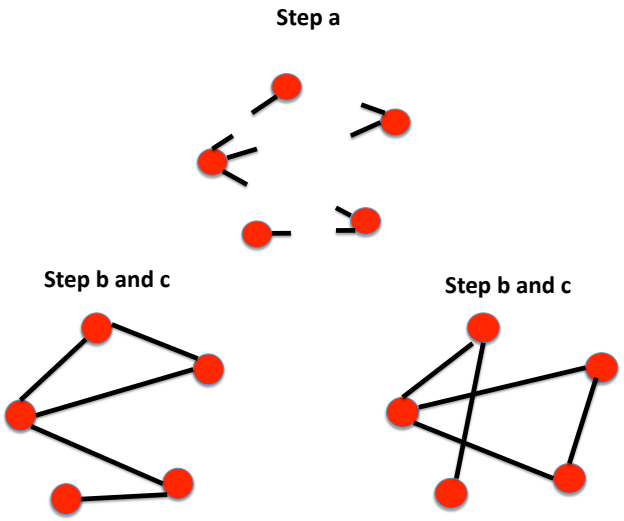

Figure 8.1: Construction of a network in the configuration model. In step a) the k_i stubs are placed on each node i of the network. In step b) and c) all the stubs are repeatedly matched until the full network if formed. In the figure we show two possible networks generated by the configuration model starting from the same degree distribution. Starting from the degree distribution in this figure, one can construct 6 different simple networks, here we just show two network realizations.

for any pair of nodes (i, j) of the network. Moreover, the probability q_{ij} that one link of node i is linked to a node j is given by

$$q_{ij} = \frac{k_j}{\langle k \rangle N} \tag{8.2}$$

Definition 79. *The* structural cutoff *of a network is a maximal allowed degree given by*

$$K = \sqrt{\langle k \rangle N}. \tag{8.3}$$

Therefore if a network has a structural cutoff we have

$$k_i \leq K = \sqrt{\langle k \rangle N} \tag{8.4}$$

for every node $i = 1, 2 \ldots, N$.

Proposition 44. *The networks generated with the configuration model are uncorrelated if and only if they have a structural cutoff.*

Proof. Here we will prove only the necessary condition, i.e. a necessary condition for generating uncorrelated networks with the configuration model is that the degree sequence has a structural cutoff. In fact if the networks are uncorrelated then the probability that a node i is connected to a node j is given by

$$p_{ij} = \frac{k_i k_j}{\langle k \rangle N}. \tag{8.5}$$

by putting $k_i = K$ and $k_j = K$ where K is the maximal degree of the network and imposing that p_{ij} is a probability, i.e. $p_{ij} \leq 1$ we have

$$p_{ij} = \frac{K^2}{\langle k \rangle N} \leq 1, \tag{8.6}$$

which imply that the maximal degree must be the structural cutoff

$$K \leq \sqrt{\langle k \rangle N}. \tag{8.7}$$

One can also prove that the presence of a structural cutoff in the networks is a sufficient condition for generating uncorrelated networks with the configuration model. □

Proposition 45. *In an uncorrelated network the probability $q(k)$ that by following a link in the network we reach a node of degree k is given by*

$$q(k) = \frac{k}{\langle k \rangle} P(k). \tag{8.8}$$

Proof. In fact, assume that we follow a link of a generic node i of the network, the probability q_{ij} that we reach a node j with degree $k_j = k$ is given by

$$q_{ij} = \frac{k}{\langle k \rangle N}, \tag{8.9}$$

and only depend on the degree $k_j = k$ of the target node. The number of nodes with the degree k is given by $NP(k)$.
Therefore the probability $q(k)$ that by following a link of the network we reach a node of degree k is given by

$$q(k) = \frac{k}{\langle k \rangle N} NP(k) = \frac{k}{\langle k \rangle} P(k). \tag{8.10}$$

□

Proposition 46. *The average degree of the neighbours of a random node in the configuration model with structural cut-off is given by*

$$\sum_k kq(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}. \tag{8.11}$$

Proof. In fact, $q(k)$ is the probability that a neighbour of a node has degree k, therefore the average degree of a neighbour of a node is

$$\sum_k kq(k) = \sum_k k \frac{k}{\langle k \rangle} P(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}. \tag{8.12}$$

□

Note that in such networks the neighbours of a random node have in average more links that the starting node. This phenomenon can be expressed by the sentence that applies to social networks: *Your friends have more friends than you do!* Or, more properly speaking, given a random person in a social network *His/her friends have more friends than him/her!* In fact let us assume that the social network is uncorrelated (note that this assumption does not hold since social networks tend to have hubs more likely connected to hubs than to nodes of small degree). In this network the average number of friends of a node is given by $\frac{\langle k^2 \rangle}{\langle k \rangle}$, while the average degree of a random node is given by $\langle k \rangle$. But we always have

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > \langle k \rangle, \tag{8.13}$$

as soon as the network is not a regular network with all the nodes having the same degree. In fact we have always

$$\langle k^2 \rangle - \langle k \rangle^2 = \left\langle (k - \langle k \rangle)^2 \right\rangle \geq 0, \tag{8.14}$$

where the equality holds only if all the nodes has the same degree.

8.3 Birth of the giant component and Molloy-Reed criterion (NE)

We have already seen that the giant component emerges in Poisson networks at a critical value of the average degree $\langle k\rangle = c = 1$. Poisson networks with average degree $c < 1$ do not have a giant component, i.e. the fraction of nodes in the giant component is zero in the limit $N \to \infty$ ($S = 0$) while Poisson networks with average degree $c > 1$ has a giant component and in these networks the fraction of nodes S in the giant component is positive ($S > 0$). This drastic change in the structure of the network, is characterized with the same tools used to study phase transitions in condensed matter, (ex. the transition between a ferromagnetic and a paramagnetic material as a function of the temperature). In this chapter we will study how to characterize the emergence of the giant component in sparse networks with generic degree distribution $P(k)$ and finite average degree $\langle k\rangle$. Interestingly we will show that the main parameter that determines whether or not there is a giant component in the network is not given in general by the average degree but it is given by $\frac{\langle k(k-1)\rangle}{\langle k\rangle}$. We start by defining a recursive criterion for determining is a node of the network is in the giant component. Applying this definition we will first find the equation for the fraction S of nodes in the giant component of a network with degree distribution $P(k)$, and secondly we will show that a network has a giant component $S > 0$ if and only if $\frac{\langle k(k-1)\rangle}{\langle k\rangle} > 1$ or, equivalently if and only if $\frac{\langle k^2\rangle}{\langle k\rangle} > 2$ which constituted the so called Molloy-Reed criterion.

Definition 80. *A node is in the giant component of the network if, at least one of the nodes reached by following one of its links is also in the giant component of the network. A node reached by following a link is in the giant component if at least one of its remaining links reaches a node in the the giant component.*

Proposition 47. *The probability S' that by following a link, in a locally tree-like network with degree distribution $P(k)$ we reach a node in the giant component, needs to satisfy the following equation:*

$$S' = 1 - \sum_k \frac{k}{\langle k\rangle} P(k)(1 - S')^{k-1}. \tag{8.15}$$

The fraction of nodes S that are in the giant component of the same network is given by

$$S = 1 - \sum_k P(k)(1 - S')^k, \tag{8.16}$$

where S' is the solution of Eq. (**??**).

Proof. To find the equation Eq. (8.15) for S' we use the recursive rule for determining is a node reached by following a link in the network is in the giant component. By following a link we reach a node of degree k with probability

$q_k = kP(k)/\langle k\rangle$, the probability that at least one of the remaining $k-1$ links of this node reach a node in the giant component is

$$1-(1-S')^{k-1}, \tag{8.17}$$

where we have assumed that the network is locally tree-like and neglected any possible correlations between the fact that two or more neighbours of the same node are/(are not) in the giant component. Therefore summing over all the possible degrees k of the node reached by following a link, we have

$$S' = \sum_k \frac{k}{\langle k\rangle} P(k)\left[1-(1-S')^{k-1}\right]$$

$$S' = 1-\sum_k \frac{k}{\langle k\rangle} P(k)(1-S')^{k-1}. \tag{8.18}$$

To find the expression for S, the fraction of nodes in the giant component of the network, we first notice that S indicates also the probability that a random node is in the giant component, when we consider the limit $N\to\infty$. A random node of the network has degree k with probability $P(k)$. The probability that a node of degree k is not in the giant component is given by the probability that all this links reach nodes that are not in the giant component, therefore we have

$$1-S = \sum_k P(k)(1-S')^k. \tag{8.19}$$

Finally the fraction S of nodes in the giant component can be written as

$$S = 1-\sum_k P(k)(1-S')^k. \tag{8.20}$$

□

Proposition 48. *The Molloy-Reed criterion for having a giant component is the following: a sparse random network with degree distribution $P(k)$ has a giant component if and only if*

$$\frac{\langle k^2\rangle}{\langle k\rangle} > 2. \tag{8.21}$$

Proof. The fraction of nodes S in the giant component satisfies Eq. (8.16), i.e.

$$S = 1-\sum_k P(k)(1-S')^k, \tag{8.22}$$

therefore there is a giant component in the network ($S>0$) if and only if $S'>0$. The probability S' satisfies Eq. (8.15) given by

$$S' = 1-\sum_k \frac{k}{\langle k\rangle} P(k)(1-S')^{k-1}. \tag{8.23}$$

This equation is always satisfied for $S' = 0$, but, depending on the properties of the degree distribution $p(k)$ it can have another non-trivial solution $S' > 0$. Unfortunately this equation cannot be solved analytically for arbitrary value of S'. For this reason we will make use of some graphical argument. The solution of Eq. (8.15) can be seen as the value of S' where the two functions $y = f(S')$ with $f(S') = S'$ and $y = g(S')$ with $g(S') = 1 - \sum_k \frac{k}{\langle k \rangle} P(k)(1 - S')^{k-1}$ cross.

Since the function $g(S')$ is an increasing function of S', with maximum slope at $S' = 0$, the non trivial solution $S' > 0$ emerges when the functions $y = f(S')$ and $y = g(S')$ are tangent to each other at $S' = 0$.

In order to detect when this new solution emerges, we impose therefore

$$\begin{aligned} \left.\frac{dS'}{dS'}\right|_{S'=0} &= \left.\frac{d(1 - \sum_k \frac{k}{\langle k \rangle} P(k)(1 - S')^{k-1})}{dS'}\right|_{S'=0}, \\ 1 &= \left.\sum_k \frac{k(k-1)}{\langle k \rangle} P(k)\right|_{S'=0}, \\ 1 &= \frac{\langle k(k-1) \rangle}{\langle k \rangle} \end{aligned} \tag{8.24}$$

Therefore a random network generated with the configuration model will have a giant component if an only if

$$\frac{\langle k(k-1) \rangle}{\langle k \rangle} > 1, \tag{8.25}$$

or

$$\frac{\langle k^2 \rangle}{\langle k \rangle} > 2. \tag{8.26}$$

□

8.3.1 Giant component in Poisson and scale-free networks

Proposition 49. *The Molloy-Reed condition for having a giant component in a Poisson network reduced to the already obtained necessary and sufficient condition*

$$c = \langle k \rangle > 1. \tag{8.27}$$

Proof. In fact for a Poisson degree distribution $P(k) = c^k e^{-c}/k!$ we have $\langle k(k-1) \rangle = c^2$ and $\langle k \rangle = c$. Therefore the Molloy-Reed condition can be written as

$$\begin{aligned} \frac{\langle k(k-1) \rangle}{\langle k \rangle} &> 1 \\ \frac{c^2}{c} = c = \langle k \rangle &> 1. \end{aligned} \tag{8.28}$$

□

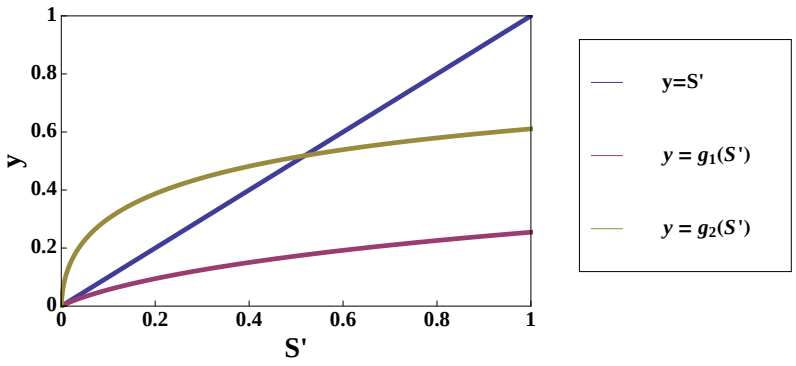

Figure 8.2: The graphical solution of Eq. (8.15).

In a Poisson network we found a critical value of the average degree $\langle k\rangle = c = 1$ necessary for having a giant component in the network. In scale-free network the situation is significantly different. In fact it is not the average degree that is determining if the network has a giant component, but the ratio $\langle k(k-1)\rangle/\langle k\rangle$. As we will see then scale-free networks with $\gamma \in (2,3]$ have always a non vanishing giant component independently on their average degree. This is one of the most important signals that these structures are also more robust to random damage.

Proposition 50. *Uncorrelated sparse scale-free networks with degree distribution $P(k) = Ck^{-\gamma}$ and $\gamma \in (2,3]$ have always a giant component in the limit $N \to \infty$.*

Proof. Let us consider uncorrelated power-law networks with degree distribution $P(k) = Ck^{-\gamma}$ with $\gamma > 2$ and $k \in [k_{min}, \sqrt{\langle k\rangle N}]$. For these network the average degree $\langle k\rangle$ is finite in the limit $N \to \infty$. Let us evaluate $\langle k^2\rangle$ in the continuous limit approximation. We have

$$\begin{aligned}\langle k^2\rangle &= \int_{k_{min}}^{K} dk k^2 P(k) = C\int_{k_{min}}^{K} dk k^{2-\gamma} \\ &= \begin{cases} C\frac{1}{3-\gamma}\left[K^{3-\gamma} - k_{min}^{3-\gamma}\right] & \text{for} \quad \gamma \neq 3 \\ C\left[\ln K - \ln k_{min}\right] & \text{for} \quad \gamma = 3 \end{cases}\end{aligned}$$

Using the expression for the structural cutoff $K = \sqrt{\langle k\rangle N}$ and evaluating $\langle k^2\rangle$ at the leading term for $N \gg 1$ we get

$$\langle k^2\rangle = \begin{cases} C\frac{1}{3-\gamma}\left[(\langle k\rangle N)^{(3-\gamma)/2}\right] & \text{for} \quad \gamma < 3 \\ \frac{C}{2}\left[\ln\langle k\rangle N\right] & \text{for} \quad \gamma = 3 \\ C\frac{1}{\gamma-3}\left[k_{min}^{(3-\gamma)}\right] & \text{for} \quad \gamma > 3 \end{cases}$$

Therefore for $\gamma \in (2,3]$, the ratio $\frac{\langle k^2 \rangle}{\langle k \rangle}$ diverges for $N \to \infty$ and the Molloy-Reed condition $\frac{\langle k^2 \rangle}{\langle k \rangle} > 2$ is always satisfied. This means that a scale-free network has always a giant component, independently on the value of the average degree $\langle k \rangle$ (fixed by the power-law exponent γ and the minimal degree of the network k_{min}). □

8.4 Local clustering coefficient of the uncorrelated configuration model (E)

The local clustering coefficient of a node of a network is the probability that two nearest neighbours of a node are connected together. In an uncorrelated network the average local clustering coefficient is independent on the degree of the starting node. In fact the average local clustering coefficient can be easily calculated.

Proposition 51. *The average local clustering coefficient of a node in a uncorrelated network is equal to the Watts-and Strogatz clustering coefficient and is given by*

$$C_{WS} = \frac{1}{\langle k \rangle N} \left(\frac{\langle k(k-1) \rangle}{\langle k \rangle} \right)^2 \tag{8.29}$$

Proof. Suppose that two nearest neighbours of a given node i are called node r and node m. Node r has degree k_ℓ with probability $q(k_r)$, node m has degree k_m with probability $q(k_m)$. These two nodes have each one link linked to the node i. Therefore node r has $k_r - 1$ remaining links and node m has $k_m - 1$ remaining links. In the configuration model stubs are randomly matched, therefore the probability that node r and node m are connected, given that they are both connected to node i is given by

$$\frac{(k_r - 1)(k_m - 1)}{\langle k \rangle N}. \tag{8.30}$$

If we want to evaluate the average clustering coefficient of a node in this network (equal to the Watts and Strogatz clustering coefficient C_{WS} of the network) we have to evaluate the probability that any two nearest neighbours of a generic

node i are connected therefore we have

$$\begin{aligned} C_{WS} &= \sum_{k_r,k_m} q(k_r)q(k_m)\frac{(k_r-1)(k_m-1)}{\langle k\rangle N} \\ &= \sum_{k_r,k_m} \frac{k_r}{\langle k\rangle}P(k_r)\frac{k_m)}{\langle k\rangle}P(k_m)\frac{(k_r-1)(k_m-1)}{\langle k\rangle N} = \\ &= \frac{1}{\langle k\rangle N}\left(\sum_k \frac{k(k-1)}{\langle k\rangle}P(k)\right)^2 \\ &= \frac{1}{\langle k\rangle N}\left(\frac{\langle k(k-1)\rangle}{\langle k\rangle}\right)^2 \end{aligned} \tag{8.31}$$

□

8.5 Average distance of an uncorrelated network (E)

Let us consider uncorrelated networks that are locally tree-like. In this networks the average number of short loops is finite in the limit $N \to \infty$. Therefore we must have $\langle k(k-1)\rangle/\langle k\rangle$ finite in the limit $N \to \infty$. In this case we can evaluate the average distance of the network following the same procedure that we have used for random networks and Cayley trees in chapter 5. In particular we will evaluate the average branching ratio of a node reached by following a random link of the network, and we will express the typical distance in the network as a function of this average branching ratio.

Definition 81. *The branching ratio b_k of a node of degree k is given by*

$$b_k = (k-1), \tag{8.32}$$

expressing the number of remaining links of the node if we reach the node by following a link.

Let us consider the average branching ratio of nodes reached by following a link satisfying the following definition

Definition 82. *The average branching ratio $\overline{b}$ of a node reached by following a links is given by*

$$\overline{b} = \sum_k q(k)b_k = \sum_k \frac{k(k-1)}{\langle k\rangle}P(k) = \frac{\langle k(k-1)\rangle}{\langle k\rangle}, \tag{8.33}$$

expressing the typical number of remaining links of a random node reached by following a link.

When $\langle k(k-1)\rangle$ is constant in the large network limit the networks generated with the configuration model have a negligible clustering coeffiencient and are locally tree-like. In this case it is possible to evaluate the average distance between the nodes of the network by approximating the number of nodes N_d at distance d from a given node i as

$$N_d = \begin{cases} k_i \left(\frac{\langle k(k-1)\rangle}{\langle k\rangle}\right)^{d-1} & \text{for} \quad d \geq 1 \\ 0 & \text{for} \quad d = 0. \end{cases}$$

Assuming that node i is taken randomly we can approximate the number of nodes at distance d from a random node as

$$\hat{N}_d = \begin{cases} \langle k\rangle \left(\frac{\langle k(k-1)\rangle}{\langle k\rangle}\right)^{d-1} & \text{for} \quad d \geq 1 \\ 0 & \text{for} \quad d = 0. \end{cases}$$

Notice that for a Poisson network with average degree $\langle k\rangle = c$ we have

$$\frac{\langle k(k-1)\rangle}{\langle k\rangle} = c. \tag{8.34}$$

Therefore

$$\hat{N}_d = c^d. \tag{8.35}$$

In general, as long as $\langle k(k-1)\rangle/\langle k\rangle$ is finite in the large network limit we have that the average distance in the network scales like

$$\ell = \frac{\ln(N)}{\ln \bar{b}}. \tag{8.36}$$

It has therefore the small-world distance property. Actually it can also be shown that scale-free networks with $\langle k(k-1)\rangle$ diverging in the large network limit have a typical distance that scales like

$$\ell \simeq \mathcal{O}(\ln \ln N) \tag{8.37}$$

for $\gamma \in (2, 3)$.