Complex networks

Marija Vidmar

December 21, 2007

Mentor: prof. Rudolf Podgornik

Abstract

This work is a glimpse at science of networks. First, the empirical results: what is typical for networks, where do we find them, do different types of networks have anything in common. Second, the models: there are three basic models, random model, a mathematical concept, small-world model, an everyday experience, and scale-free model, an attempt to converge towards experiments' distributions. Third, the models are being upgraded all the time and some concepts, which are not primarily the interest of networks, like Bose-Einstein condensation, can be explained through them. And last but not least, the impact of this knowledge on our lives, because networks are not just an abstract idea that exists in articles; they are a part of our lives and even more, we are a part of their lives.

Contents

1 Introduction

Networks are everywhere. We are surrounded by complex weblike structures. Internet is a complex network of routers and computers linked by physical or wireless links, Wordwide web is an enourmous virtual network of webpages connected by hyperlinks. Human contacts form a network, where human beings are the nodes and relationships are the edges. Even a cell is a complex network of chemicals connected by chemical reactions.
In the past few years many scientists have tried to develop models of networks and to investigate the mechanisms that determine the topology of complex networks.

Traditionally the study of complex networks has been a territory of graph theory. Since 1950's largescale networs were described as random graphs. This was first studied by Paul Erdos and Alfred Renyi who intruduced the Erdos - Renyi model. But lately we witnessed computerization of data and increased computing power. The boundaries between disiplines are breaking down and there is a need to move beyond reductionism and understand the behaviour of a system as a whole. In physics we tend to explain the behaviour of a system as a whole from the properties of its constituents. Scientists dicovered that tools of statistical mechanics offer good description of complex networks.

There are three importants concepts that we have to define when thinking about complex networks:

Small world: despite the large size in many networks there is a relatively short path between any two nodes. The distance between two nodes is desribed as the number of edges along the shortest path connecting them. The popular manifestation of small world is "six degrees of separation" concept, that is: the path of acquaintances with typical length about six exist between most pairs in USA. The actors in Hollywood also form a small world network, where the edges are costarings.

Clustering: social networks form cliques, circles of friends or aquaintacies. The question, connected with this property, is: Are two my firends also friends to each other? The meausure of clustering is *clustering coefficient*, defined in article [4]: we have a single node i that is connected to k_i other nodes. There are E_i edges that actually exist between neighbours. There are $k_i(k_i-1)/2$ possible edges between neigbours. The *clustering coefficient* is a ratio between actual and all possible links:

$$C_i = \frac{2E_i}{k_i(k_i-1)}. \tag{1}$$

The C of a whole network is the average over all i-s.

Degree distribution: not all nodes have the same number of edges. The node degree is characterized by a distribution function $P(k)$, which tells the probabilty that a randomly selected node has exactly k edges. In random graphs edges are placed randomly and the majority of nodes have approximately the same degree, close to the average $\langle k \rangle$. The degree of a random graph is *Poisson distribution* with a peak at $P(\langle k \rangle)$. On the other hand real network show distribution significantly different from Poisson's. For instance,

the www has a distribution with a *power-law tail*:

$$P(k) \sim k^{-\gamma}. \tag{2}$$

Such networks are called *scale-free*. In the following chapters we will explore random graphs which are variants of Erdos - Renyi model, small- world networks that interpolate between random graphs and highly clustered graphs and scale-free models with power-law degree. After that we will focus on evolving of networks and some aplication. But first let's take a look at some empirical results from real networks.

2 Empirical results

The study of networks has begun because people wanted to understand various real systems, from communication networks to ecological webs. The databases for study span from several disciplines. There are three robust measures for network's topology, used in this comparison: *average path length*, *clustering coefficient* and *degree distribution*.

2.1 Word - Wide Web

The interest in WWW network has boomed after it has been discovered that the degree distribution of webpages follows a power - law order over a several orders of magnitude (article [3]). The nodes of the network are the documents (webpages) and the edges are hyperlinks (URLs) that point form one document to another - Figure **??** (i). The edges are directed, hence there are two degree disptributions: the distribution of outgoing edges, $P_{out}(k)$ - the probability that a document has k outgoing edges and the distribution of incoming edges, $P_{in}(k)$ - the probabilty that a document has k incoming edges. Both distirbution degrees have power - law tais:

$$P_{out}(k) \sim k^{-\gamma_{out}} \qquad \text{and} \qquad P_{in}(k) \sim k^{-\gamma_{in}}$$

There are several studies, containing different amounts of nodes and obtaining γ_{in} and γ_{out}, shown in Table 1. Apparantly γ_{in} is the same for all measurments but γ_{out} has an increasing tendency with the sample size or time. On figure 1 (ii) there is a comparison od degree distributions between Albert and Broder survey, where squares represent Albert and circles Broder. Figure 1 (ii, a) is for outgoing and 1 (ii, b) for incoming nodes.

study	number of nodes	γ_{in}	γ_{out}
Albert, Jeong, Barabasi (1999) - [3]	325 729	2.1	2.45
Kumar(1999)	40 million	2.1	2.38
Broder(2000)	200 million	2.1	2.72

Table 1: γ_{in} and γ_{out} on different subsets in different studies

It is difficult to measure the clustering coefficient using Eq. (1), because the links are directed in WWW. Making each edge bidirectional, as Adamic purposed in 1999, gives us $C = 0.1078$ on 153,127 sites. This differs a lot from a random graph model for WWW

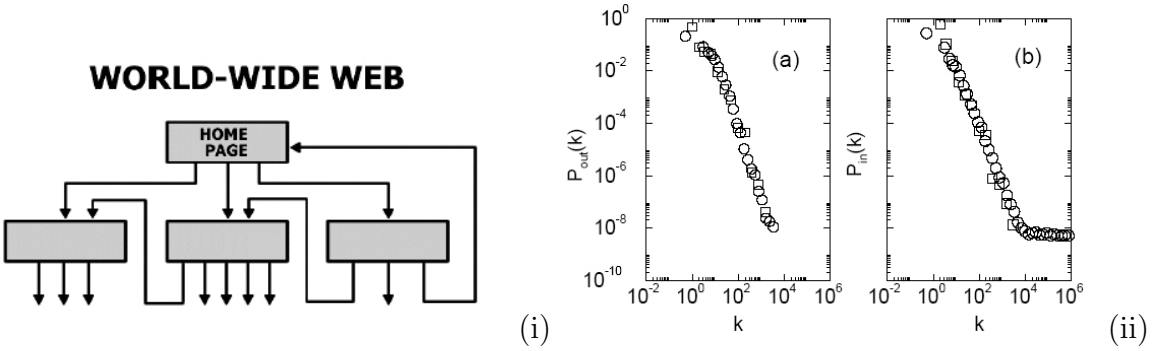


Figure 1: World-Wide Web as a network (i): the nodes are the web pages and the edges are the URLs; (ii) degree distribution of incoming (b) and outgoing (a) edges; squares - Albert, circles Broder study

which gives $C_{rand} = 0.00023$. Besides, WWW displays a small world property. The average path length for a sample 325,729 nodes (Barabasi, 1999) is 11,2 but in random model would be around 19.

2.2 Internet

Internet is the network of physical links between computers or other telecommunication devices. At router level the nodes are routers and the edges are physical links between them. At the domain level, each domain is represented by a single node and an edge is drawn between two domains if at least one route connects them. Again, there were several studies, both at router and domain level, obtaining results: on domain level $\gamma_i = 2.15$ (in 1997) and $\gamma_i = 2.2$ (in 1998) and on router level $\gamma_r = 2.48$ (degree distribution - Figure 2 a). The Internet also displays clustering and small path length. The study of internet at the domain level have found C ranged between 0.18 and 0.3, while random networks with similar parameters have $C_{rand} \sim 0.001$. The average path length is between 3.70 and 3.77.

2.3 Movie actor collaboration network

The movie collaboration network is based on the Internet Movie Database, that contains all movies and their casts since 1890's. The nodes are the actors and the two nodes have a common edge if the corresponding actors have acted in a movie together. The average path length is close to a random graph with the same size and average degree (3.65 compared to 2.9), but clustering coefficient is more than 100 higher than a random graph. The degree distribution has a power-law tail with $\gamma_{actor} = 2.3 \pm 0.1$ (Figure 2 b).

2.4 Science collaboration graph

In this case the nodes are the scientists and two notes are connected if the two scientists have written an article together. Newman (2001) have studied four databases spanning physics, biomedical research, high- energy physics and computer science over a 5 year window (1995 - 1999). All these networks show small average path length but high clustering coefficient. The degree distribution of high - energy physics is almost a perfect power- law

with an exponent of 1.2 (Figure 2 c). Barabasi studied neuroscientists publishing between 1991 and 1998 and got average path $\ell = 6$ and cluster coefficient $C = 0.76$. The exponent is $\gamma = 2.5$ (Figure 2 d).

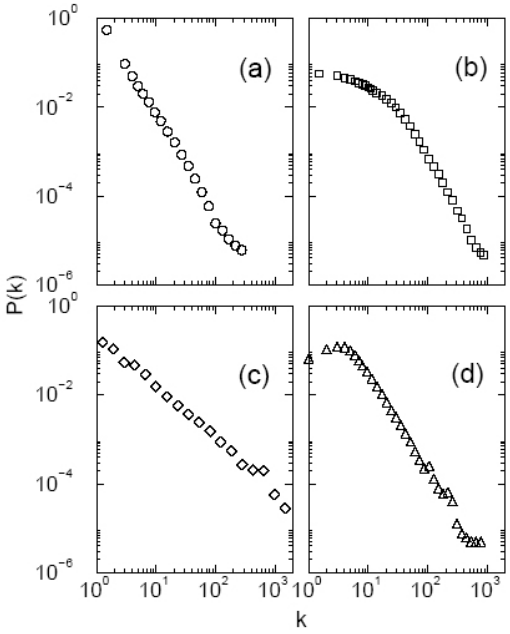


Figure 2: Degree distribution of several real networks: (a) internet at the router level; (b) movie actor collaboration network; coauthorship network of (c) high energy physicists; (d) neuroscientists

There are several other real complex networks that have a power-law tail or high clustering coefficient or average path length, for instance the web of human sex contacts, cellular networks, ecological networks, citation networks, networks in linguitics and many more. The existance of networks in many diefferent fields gives us further motivation to understand the topology of these beautiful and ubiquitous objects.

3 Models of networks

In the previous section we have overlooked some empirical results from real complex networks that surround us. In this section we will focus on different theoretical models of networks. There are three basic classes of models: the oldest is the *random graph model*, purposed by Erdos and Renyi in 1960's, the next is *small-world model* from Watts and Strogatz in 1998 and the last one is *scale-free model*, introduced by Albert and Barabasi in 1999.

3.1 Random graph theory

In mathematical terms a network is represented by a graph, that is as a pair of sets $G = \{P, E\}$, where P is a set of N nodes (vertices, points) and E is a set of edges (links, lines), that connect two elements of P.

3.1.1 Erdos - Renyi model

Erdos and Renyi define a random graph as N labeled nodes connected by n edges which are chosen randomly from the $N(N-1)/2$ possible edges. In total there are $C^n_{N(N-1)/2}$ probabal graphs with $\{N, n\}$. The binomial model definition is an alternative: here we start with N nodes, every node is connected with probabilty p. The total number of connections is a random variable with the expectation value $E(n) = p^{N(N-1)/2}$.If $G_0 = \{N, n\}$, the probability of obtaining it is $P(G_0) = p^n(1-p)^{N(N-1)/2-n}$.

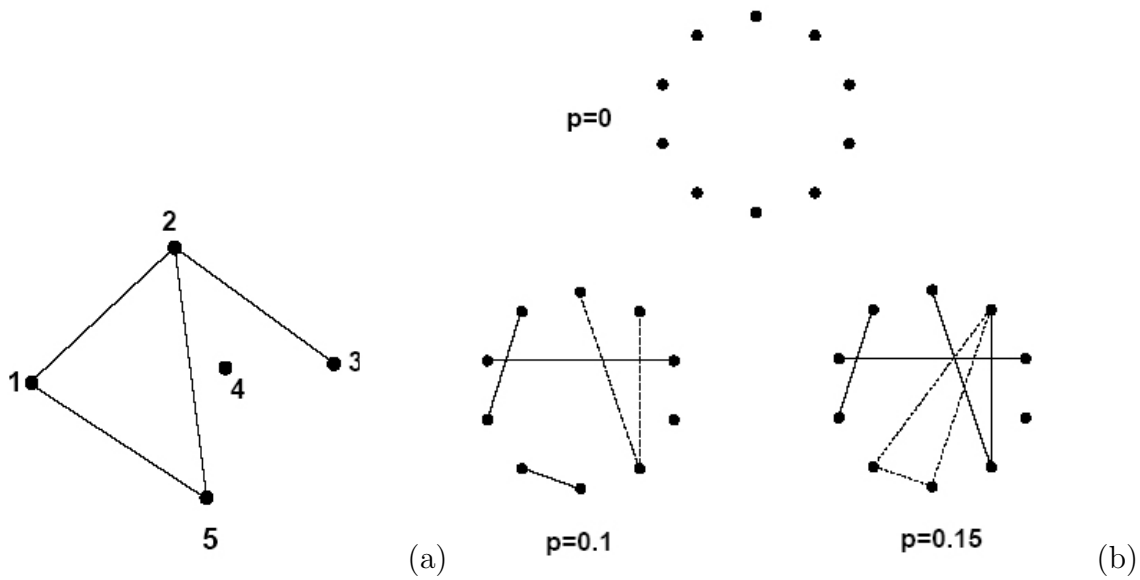


Figure 3: Random graphs: (a) illustration of a graph with $N = 5$ nodes and $n = 4$ edges. The set of nodes is $P = \{1, 2, 3, 4, 5\}$ and the set of edges is E = {{1,2}, {1,5},{2,3},{2,5}}; (b)the evolution of a graph. We start with N isolated nodes and then connect every pair with proability p. Dashed lines show the emergence of trees and cycles.

Random graph theory studies the properties of graphs with N nodes as $N \to \infty$. Almost every graph has a property Q if the probability of having Q approaches 1 as $N \to \infty$. The construction of a random graph is called *evolution*: starting with a set of N isolated vertices, the graph develops by adding random edges. During evolution the connection probability p becomes larger and larger. The crucial task is to determine at which p a particular property will most likely arise. Erdos and Renyi discovered, that most properties appear quite suddenly. Usually a critical probability $p_c(N)$ exists. If $p(N)$ grows slower than $p_c(N)$ as $N \to \infty$, then almost every graph with a connection $p(N)$ fails to have the property Q and vice versa. In mathematical terms:

$$\lim_{N\to\infty} P_{N,p}(Q) = \begin{cases} 0 & ; \quad \frac{p(N)}{p_c(N)} \to 0, \\ 1 & ; \quad \frac{p(N)}{p_c(N)} \to \infty \end{cases}$$

A similar model is percolation. The proper value of p_c is obtained by finite size scaling: the limit $p_c = p_c(N \to \infty)$. In physics most systems have a final dimension, but networks are, on the contrary, infinite dimensional and sometimes there is no unique, N

- independent threshold. But on the other hand, the average degree of a graph

$$\langle k \rangle = \frac{2n}{N} = p(N-1) \approx pN$$

does have a critical value, independent of a system size.

3.1.2 Degree distribution

In a random graph with connection probabilty p the degree k_i of a node i follows a binomial distribution with parameters $N-1$ and p:

$$P(k_i = k) = C_{N-1}^{k} p^k (1-p)^{N-1-k}. \tag{3}$$

This probability represents the number of ways in which k edges can be drawn from a certain node: the probability of k edges is p_k, the probability of absence of additional edges is $(1-p)^{N-1-k}$ and there are C_{N-1}^{k} equivalent ways of selecting k endpoints for these edges. To find the degree distribution of a graph we have to study the number of nodes with degree k, X_k. The expectation value of the number of nodes with degree k is

$$E(X_k) = NP(k_i = k) = \lambda_k.$$

From the conditions of subgraphs it can be derived that the distribution of the X_k values, $P(X_k = r)$ approaches a Poisson distribution

$$P(X_k = r) = \exp(-\lambda_k)\frac{\lambda_k^r}{r!}.$$

The mean value of a distribution is λ_k. The expectation value is a function and not a constant. A good approximation is also a binomial distribution

$$P(k) = C_{N-1}^{k} p^k (1-p)^{N-1-k} \tag{4}$$

3.1.3 Diameter and average path length

The diameter of a graph is the maximal distance between any pair of its nodes. Random graphs tend to have small diameters, provided p not too small. For most values of p, almost all graphs have precisely the same diameter. When we consider all graphs with N nodes and connection probability p, the range of values in which the diameters vary is very small, concentrated around

$$d = \frac{\log(N)}{\log(pN)} = \frac{\log(n)}{\log(\langle k \rangle)}.$$

Another characterisation of a random graph is the average path length, that is the average distance between any pair of nodes. The average path length scales with the number of nodes in the same way as diameter

$$\ell_{rand} \sim \frac{\log(N)}{\log(\langle k \rangle} \tag{5}$$

The average path length of real networks is close to the average path length of random graphs with the same size. On Figure 5(a) we see a comparison between real networks (some of them were mentoned in the previous section) and the prediction of random graph theory (dashed line). The trend of data is obviously similiar with the theoretical prediction, with several exceptions.

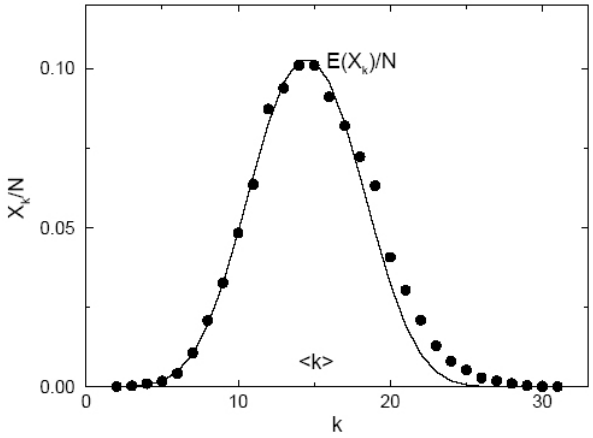
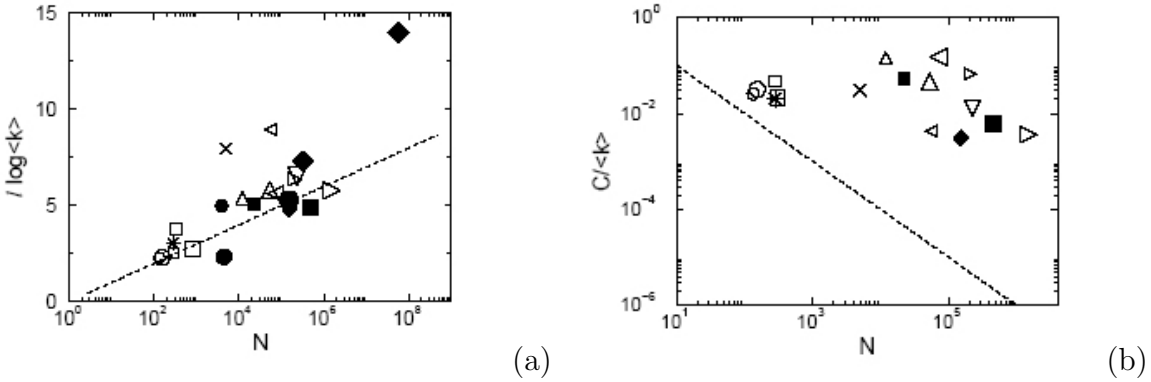


Figure 4: The degree distribution for a random graph with $N = 10000$ nodes; the points are numerical simulation and the line is Poisson distribution. The deviation is small.



(a) (b)

Figure 5: Comparison between random graphs (dashed line) and real networks (symbols): (a) ℓ_{rand} is close to ℓ for real networks; (b) comparing clsuering coefficient, real networks do not follow the predicition of random graphs.

3.1.4 Clustering coefficient

The clustering coefficient of a random graph is

$$C_{rand} = p = \frac{\langle k \rangle}{N}. \tag{6}$$

The explanation of the above formula is simple: the probability that a node in a random graph is connedted to its first neighbours is equal to the probability that two randomly selected nodes are connected. Again we can compare the C of random graph with real networks (Figure 5 b). Unlike the average path length the plot inticateds that real networks are not close to random graphs concerning clustering coefficient. For random graphs the fraction $C/\langle k \rangle$ decreases as N^{-1}, but for real networks it seems to be independent of N. This property is characteristic to large ordered lattices.

3.2 Small world model

On Figure 5(b) we have seen that real-world networks have unusually large clustering coefficients in comparison to random graphs and even more, the clustering coefficient seems to be independent of the network size. For example, a one-dimensional lattice with periodic boundary conditions, in which every node is connected to the K nodes closes to it, most immediate neighbours are also neighbours to each other. The cluster coefficient is

$$C = \frac{3(K-1)}{4(K-1}),$$

which converges to 3/4 for large K. Such low-dimensional regular lattices do not have short path lengts. The first atempt to describe graphs with high clustering coefficients but small average length was proposed by Watts and Strogatz.

3.2.1 The Watts - Strogatz (WS) model

is a one-parameter model which interpolates between an ordered finite-dimensional lattice and a random graph. The algorithm behind the graph:

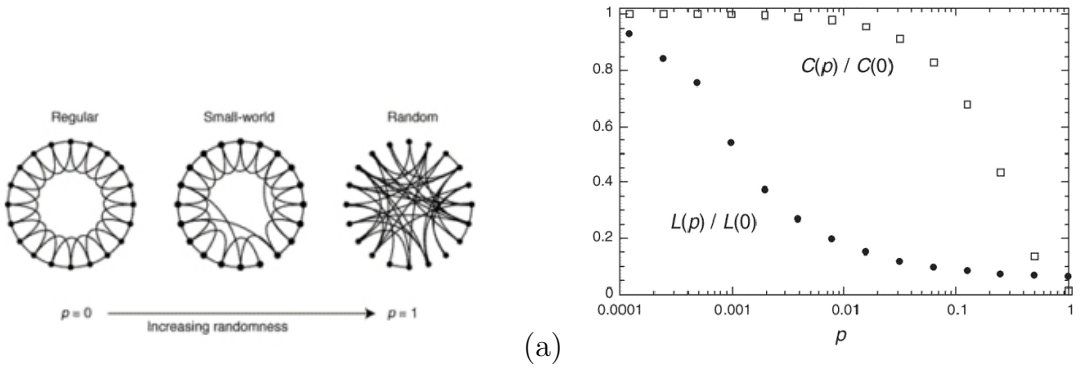
(a) *start with order:* we start with a ring lattice with N nodes in which every node is connected to its first K neighbours ($K/2$ on each side). For sparse but connected network it is $N \gg K \gg \log(N) \gg 1$.

(b) *Randomize:* we randomly rewire each edge of the lattice with probability p such that self-connections and duplicate edges are excluded. For $p = 0$ we obtain order and for $p = 1$ we obtain randomness.

The construction is shown on Figure 6 (a). To understand to coexistance of small path length and clustering, we study the behaviour of the clusttering coefficient $C(p)$ and the average path length $\ell(p)$ as a function of rewiring probability p. Let's look at two limit cases of the ring lattice (Figure 6(a)). First, when $p = 0$, ℓ scales linearly with the system size, $\ell(0) \sim N/2K \gg 1$, and clustering coefficient is large, $C(0) = 3/4$. On the other hand, when $p \to 1$, the model converges to a random graph, for which $\ell(1) \sim \log(N)/\log(K)$ and $C(1) \sim K/N$. From these results we may conclude that large (small) C is asociated with large (small) ℓ. But Watts and Strogatz found out that there is a borad interval of p over which $\ell(p)$ is close to $\ell(1)$ yet $C(p) \gg C(1)$ (Figure 6(b)). This is a good agreement with the caracteristics of real networks. In the following, we will summarize the main results, regarding the prosperties of small world models.

3.2.2 Average path length

As dicsussed above, there is a change in the scaling of the characteristic path length as p increases. This is due to the appearance of shortcuts between nodes. Every randomly created shortcut is likely to connect widely separated parts of the graph. But the onset of the small world behaviour is also dependent on the system size. There exists a p-dependent crossover length N^* such that if $N < N^*$, $\ell \sim N$ but if $N < N^*$, $\ell \sim \log(N)$. Numerical simulations and analytical arguments concluded that the crossover length N^* scales with p as $N^* \sim p^{-\tau}$ where $\tau = 1/d$ and d is a dimension of the original lattice. For



(a) (b)

Figure 6: Small world: (a) the random rewiring process in WS model interpolates between an ordered and a random graph; (b) characteristic path length $\ell(p)$ and clustering coefficient $C(p)$ for the WS model. There is a rapid drop in $\ell(p)$ while $C(p)$ remains almost constant.

the original WS model, defined on a circle ($d = 1$) we have $\tau \sim 1$, the onset of small-world behaviour taking place at $p^* \sim 1/N$. It is now widely accepted that the characteristic path length obeys the general scaling form

$$\ell(N, p) \sim \frac{N}{K} f(pKN^d), \tag{7}$$

where $f(u)$ is a scaling function that obeys

$$f(u) = \begin{cases} \text{constant} & ; \quad u \ll 1, \\ \log(u)/u; \quad u \gg q. \end{cases}$$

Equation (7) tells us that ℓ is not dependent on theree parameters (p, K and N), but is determined by a single scalar function $f(u)$ of a single scalar variable. They both have simple physical interpretations: u is two times the average number of random links (shortcuts) on the graph for a given p, and $f(u)$ is the average of the fraction by which the distance between two notes is reduced for a given u.

3.2.3 Clustering coefficient

Small-world networks have a relatively high clustering coefficient. In a regular lattice ($p = 0$) C does not depend on the size of the network but only on its topology. As the edges are randomized, C remains close to $C(0)$ up to relatively large values of p. To see the dependance of $C(p)$ on p we will use a slightly different definition of C, introduced by Barrat and Weigt in article [5]: $C'(p)$ is the fraction between mean number of edges between the neighbours of the node and the mean number of possible edges between those neighbours, that is

$$C'(p) = \frac{3(k-1)}{2(2k-1)}(1 - p^3).$$

In a more graphic formulation that is

$$C' = \frac{3 \times \text{number of triagnles}}{\text{number of connected triplets}} \tag{8}$$

Here triangles are trios of nodes in which each node is connected to both of others and connected triplets are trios in which at least one is connected to both others and we have the factor 3 because each triangle contributes to 3 connected triples.

3.2.4 Degree distribution

In the WS model for $p = 0$ each node has the same degree K, thus the degree distribution is a delta function centered at K. A nonzero p brings disorder in the network, but the average degree still equals K. For $K > 2$ there are no isolated nodes and the network is usually connected. There are no isolated clusters like in random graphs. For $p > 0$ the degree k_i can be written as $k_i = K/2 + c_i$ where c_i can be divided in two parts: $c_i^1 \leq K/2$ edges have been left in place with probability $1 - p$, while $c_i^2 = c_i - c_i^1$ edges have been rewired towards i with probability $1/N$. The probability distributions of c_i^1 and c_i^2 are

$$P_1(c_i^1) = C_{K/2}^{c_i^1}(1-p)^{c_i^1} p^{K/2-c_i^1}$$

and

$$P_2(c_i^2) = C_{pNK/2}^{c_i^2} \left(\frac{1}{N}\right)^{c_i^2} \left(1 - \frac{1}{N}\right)^{pNK/2-c_i^2}.$$

If we combine these two factors together we obtain a shape of a distribution that is similar to the random graph. It has a pronounced peak at $\langle k \rangle = K$ and decays exponentially for large k. We can see the results of numerical simulation on Figure 7.

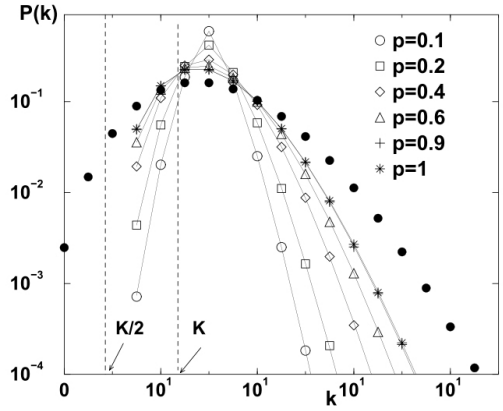


Figure 7: Degree distribution of the WS model for $K = 3$ and various p for $N = 1000$. Degree distribution of a random graph with the same parameteres is plotted with filled symbols.

3.3 The scale-free model

As we have seen in section 2 many large real networks are scale-free, that is, their degree distribution follows a power-law for large k. Even those networks for which $P(k)$ has an exponential tail, the degree distribution significantly deviates from a Poisson. Random graph theory and WS model cannot reproduce this feature. The mechanism that is responsible for the emergence of scale-free networks is modeling the network asembly and evolution. In the previous models we were modeling topology but now we put emphasis

on capturing the network dynamics. Dynamics takes the driving role and the topology is only a byproduct.

3.3.1 Definition of the scale-free (SF) model

Barabasi and Albert in 1999 [6] argued that the scale-free nature of real networks is rooted in two generic mechanisms, that is *growth* and *preferential attachment*. So far we have studied networks with fixed number N of vertices that are then randomly rewired, without modifying N. But most real world networks are open systems which grow by continuous addition of new nodes. For example, the WWW grows exponentially in time by the addition of new web pages. On the other hand, models discussed so far assumed that the probability that two nodes are connected is independent of the node's degree. But most real networks exhibit preferential attachment, the likelihood of connecting to a node depends on a node degree. If we look at the WWW exaple again, a web page will more likely include hyperlinks to popular documents with already high degree.

The algorithm of the SF model is the following:

(a)*Growth*: starting with a small number (m_0) of nodes, at every timestep we add a new node with $m(\leq m_0)$ edges that link the new node to m different nodes already present in the system.

(b)*Preferential attachment*: we assume the probability Π that a new node will be connected to the node i depends on the degree k_i such that

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}. \tag{9}$$

After t timesteps the algorithm resultes in a network with $N = t+m_0$ nodes and mt edges. Numerical simulations give a power-law degree distribution with an exponent $\gamma_{SF} = 3$. The scaling parameter is independent of m, the only parameter in the model. We see numerical simulation on Figures 8(i).

3.3.2 Theoretical approaches

The dynamical properties of the scale-free model can be addressed using various analytic approaches. First is the *continuum theory* proposed by Barabasi and Albert that focuses on the dynamics of node degrees, followed by *master equation approach* and *rate equation approach*. Since the last ones are completely equivalent and offer the same asymptotic results as the continuum theory, we will take a better look of the first one.

The continuum approach calculates the time dependance of the degree k_i of a given node i. This degree increases every time a new node enters the system and links to node i, the probability of this process being $\Pi(k_i)$. Assuming that k_i is a continuous real variable, the rate at which k_i changes is expected to be proportional to $\Pi(k_i)$. The dynamical equation is

$$\frac{\partial k_i}{\partial t} = m\Pi(k_i) = m\frac{k_i}{\sum_{j=1}^{N-1} k_j}.$$

The sum in the denomirator goes over all nodes in the system except the newly introduced one, thus $\sum_j k_j = 2mt - m$, leading to

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t}.$$

The solution of the above equation with the initial condition that every node i at its introduction has $k_i(t_i) = m$, is

$$k_i(t) = m\left(\frac{t}{t_i}\right)^{\beta}, \qquad \text{with} \qquad \beta = \frac{1}{2}. \tag{10}$$

The above equation shows that the degree of all nodes evolves the same way. Using this equation, the probability that a node has a degree $k_i(t)$ smaller than k, $P(k_i(t) < k)$, can be written as

$$P(k_i(t) < k) = P\left(t_i > \frac{m^{1/\beta}}{k^{1/\beta}}\right). \tag{11}$$

We add nodes at equal time intervals to the network, the t_i values have a constant probability density

$$P(t_i) = \frac{1}{m_0 + t}. \tag{12}$$

Substituting this into Eq. (11) we obtain

$$P\left(t_i > \frac{m^{1/\beta}}{k^{1/\beta}}\right) = 1 - \frac{m^{1/\beta}}{k^{1/beta}(t + m_0)}. \tag{13}$$

The degree distribution $P(k)$ can be obtained using

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^{1/\beta t}}{m_0 + t}\frac{1}{k^{1/\beta+1}},$$

predicting that asymptotically ($t \to \infty$)

$$P(k) \sim 2m^{1/\beta}k^{-\gamma}, \qquad \text{with} \qquad \gamma = \frac{1}{\beta} + 1 = 3$$

being independent of m, in agreement with numerical results. The degree distribution is asymptotically also independent of time (and system size, $N = t + m_0$), indicating that despite continuous growth the network reaches a stationary scale-free state.

3.3.3 Limiting cases of the SF model

We used two generic mechanism in SF model, growth and preferential attachment. In this section we will investigate two limiting cases, which contain only one of these two mechanisms.

Model A keeps the growing character without preferential attachment. That means the new node connects with equal probability to the nodes already present in the system, i.e. $\Pi(k_i) = 1/(m_0 + t - 1)$, independent of k_i. Using continuum theory with this probability we obtain

$$P(k) = \frac{e}{m}\exp\left(-\frac{k}{m}\right). \tag{14}$$

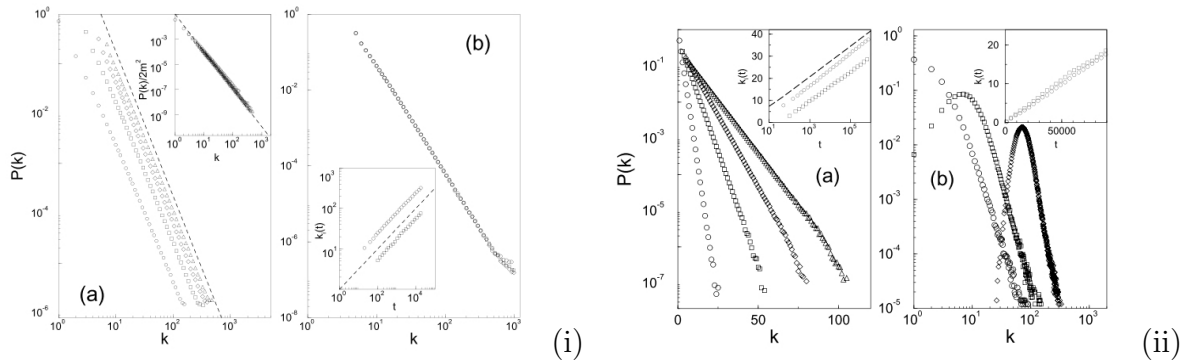


Figure 8: Degree distribution of SF model: (i, a) N = 300,000 and $m_0 = m = 1$ (circles), 3 (squares), 5 (diamonds), 7 (triangles). The slope of the dashed line is $\gamma = 2.9$; (i, b) $m_0 = m = 5$, $N = 100 \cdot 10^3$ (circles), $150 \cdot 10^3$ (squares), $200 \cdot 10^3$ (diamonds). (ii,a) Limiting model A, for different m, $N = 8 \cdot 10^5$; (ii, b) limiting model B for $t = N$ (circles), 5 N (squares), 40 N (diamonds). For later times $P(k)$ becomes Gaussian.

The absence of preferential attachment also eliminates the scale-free character. Figure 8 (ii, a).

Model B starts with N nodes and no edges. At each time step a node is selected randomly and connected with probability to a node i in the system. This model eliminates growth process. N is constant and the number of edges increases with time, after $T \sim N^2$ timesteps system reaches a state in which all nodes are connected. The time-evolution of the individual degrees can be calculated analytically, using the continuum theory, resulting in

$$k_i(t) \sim \frac{2}{N} t. \tag{15}$$

The continuum theory predicts that after a transistent period all nodes should have the same value, given by Eq. (15). Therefore we expect that the degree distribution becomes a Gaussian around its mean value. The Figure 8 (ii,b) shows that the shape changes from initial power-law to Gaussian.

The failure of the limiting cases to lead to scale-free distribution indicates growth and preferential attachment are needed simultaneously.

3.3.4 Average path length

The average path length is smaller in the SF network than in a random graph for any N. That incicates that the heterogeneous SF topology is more efficient in bringing the nodes close than the homogenous topology of random graphs. The average path length increases logarithmically with N, the best fit following logarithmic form

$$\ell = A \log(N - B) + C \tag{16}$$

On Figure 9(a) we see ℓ of a network with average degree $\langle k \rangle = 4$ generated by the SF model as a function of a network size, compared with ℓ of a random graph with the same size and average degree. Apart from the empirical fit (16) there is no theoretical expression that would give a good approximation for the path length.

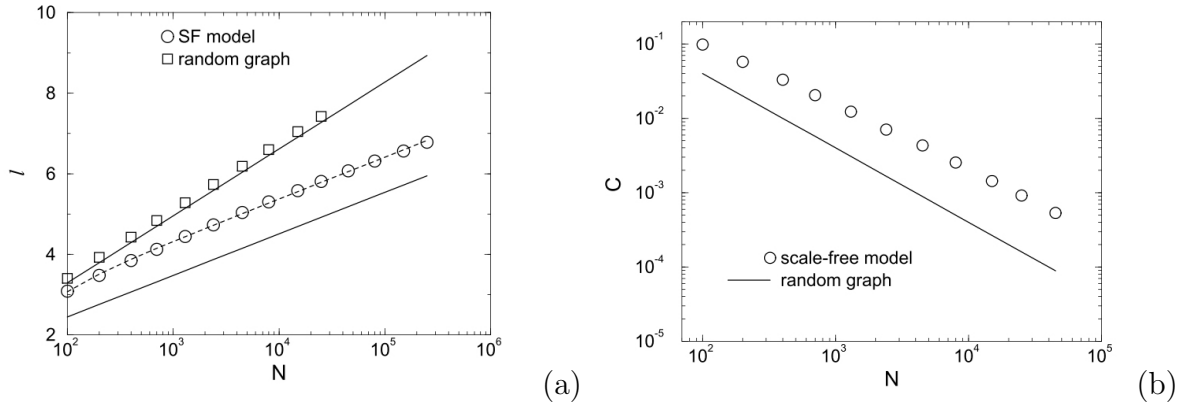


Figure 9: SF compared to random graph: (a) average path length ℓ for $\langle k \rangle = 4$; (b) clustering coefficient C for $\langle k \rangle = 4$

3.3.5 Cluster coefficient

has been much investigated for the WS model but for SF model there is no analytical prediction. Figure 9 (b) shows the clustering coefficient of the SF network with average degree $\langle k \rangle = 4$ and different sizes, compared with the clustering coefficient for a random graph (eq.(6)). C of the SF network is about 5 times higher than that of random graph and this factor slowly increases with number of nodes. But C of SF network decreases with the network size, following approximately a power-law $C \sim N^{-0.75}$. This behaviour is still different from C for small-world models and real networks (Figure 5 (b)).

4 Evolving networks

The SF model from the previous section is a minimal model that captures the mechanism responsible for the power-law degree distribution. But it has limitations: it predicts a power-law distribution with a fixed exponent, while the exponents measured for real networks vary between 1 and 3. There have been made several robust corrections that will be briefly described in this section.

Preferential attachment: the SF model assumes that the probability $\Pi(k)$ that a node attaches to node i is linear in k - Eq.(9) - which gives us the coefficient $\gamma = 3$. Non-linear models sugest $\Pi(k) \sim k^\alpha$ or $\Pi(k) = A + k^\alpha$. The last equation allows the new node to attach to an isolated node.

Growth: the SF model increases number of nodes and edges linearly in time, consequently the average degree of the network is constant. Non-linear or accelerated growth can also have an impact on the exponent of degree distribution.

Local events: in real networks nodes can be rewired and nodes and edges can be removed. Such microscopic events shape the network evolution but do not take place in SF model. The models which include this events, may obtain degree distribution, which has both scale-free and exponential regime, phase diagram being separeted by a single parameter. This is interesting because some real network do show an exponential character (for ex-

ample, Figure 2 (b), (d)).

Competition: the SF model assumes that all nodes increase their degree following a power-law time dependance with the same dynamic exponent $\beta = 1/2$ - Eq. (10). Therefore, the oldest nodes have the highest number od edges. But in real networks a node's degree does not depend on age alone. Web pages with good content and marketing acquire a large number of edges in short time. We will take a better look of one models by which a Bose-Einstein condesation can be explained.

4.1 Fitness model

has been introduced by Bianconi and Barabasi. They argued that real networks have a competitive aspect, as each node has an intrinsic ability to compete for edges at the expense of other nodes. Each mode is assigned a *fitness parameter* η_i which does not change in time. At every timestep a new node j is added to the system with the fitness parameter η_j that is chosen from a distribution $\rho(\eta)$. The probability to connect to a node i thus becomes

$$\Pi_i = \frac{\eta_i k_i}{\sum_j \eta_j k_j}. \tag{17}$$

The rate of change of the degree of node i is (using the continuum theory)

$$\frac{\partial k_i}{\partial t} = m \frac{\eta_i k_i}{\sum_j \eta_j k_j}, \tag{18}$$

assuming that the time-evolution of k_i follows (10) with a fitness dependent $\beta(\eta)$:

$$k_{\eta_i}(t, t_i) = m \left(\frac{t}{t_i}\right)^{\beta(\eta_i)}, \qquad \beta(\eta) = \frac{\eta}{C} \qquad \text{with} \qquad C = \int \rho(\eta) \frac{\eta}{1-\beta(\eta)} \mathrm{d}\eta. \tag{19}$$

The nodes with larger η increase their degree faster than those with smaller fitness. The degree distribution of the model depends on the choice of the fitness distribution. For a uniform fitness distribution the Eq. (19) gives $C = 1.255$ and $\beta(\eta) = \eta/1.255$ and the degree distribution is

$$P(k) \sim \frac{k^{-(C+1)}}{\log k}, \tag{20}$$

a power-law with a logarithmic correction.

4.2 Bose-Einstein condensation

The existence of close link between evolving networks and an equilibirum Bose gas was shown in by Bianconi and Barabasi in 2001. They start with the fitness model.The mapping to a Bose gas can be done by assigning an energy ε_i to each node, determined to its fitness through the relation

$$\varepsilon_i = -\frac{1}{\beta} \log \eta_i, \tag{21}$$

where $\beta = 1/T$ plays the role of inverse temperature. An edge between two nodes i and j, having energies ε_i and ε_j, correspondes to two non-interacting particles, one on each

energy level. Adding a new node, l to the system correspondes to adding a new energy level ε_l and $2m$ new particles. Half of them are deposited on the level ε_l, while the other half are distributed between the energy levels of the endpoints of the new edges, the probability that a particle lands on level i being given by

$$\Pi_i = \frac{\exp(-\beta\varepsilon_i)k_i}{\sum \exp(-\beta\varepsilon_i)k_i}. \tag{22}$$

The continuum theory predicts that the rate at which particles accumulate on energy level ε_i is given by

$$\frac{\partial k_i(\varepsilon_i, t, t_i)}{\partial t} = m\frac{\exp(-\beta\varepsilon_i)k_i(\varepsilon_i, t, t_i)}{Z_t}, \tag{23}$$

where $k_i(\varepsilon_i, t, t_i)$ is the occupation number of level i and Z_t is the partition function, defined as $Z_t = \sum_{j=1} t \exp(-\beta\varepsilon_j)k_j(\varepsilon_j, t, t_j)$. The solution of Eq.(23) is

$$k_i(\varepsilon_i, t, t_i) = m\left(\frac{t}{t_i}\right)^{f(\varepsilon_i)}, \tag{24}$$

where the dynamic exponent $f(\varepsilon)$ stisfies $f(\varepsilon) = \exp(-\beta(\varepsilon - \mu))$, μ plays the role of the chemical potential, satisfying the equation

$$\int \deg(\varepsilon)\frac{1}{\exp(\beta(\varepsilon - \mu)) - 1} = 1, \tag{25}$$

and $\deg(\varepsilon)$ is the degenaracy of the energy level ε. The above equation suggests that in the $t \to \infty$ limit the occupation number follows the familiar Bose statistics

$$n(\varepsilon) = \frac{1}{\exp(\beta(\varepsilon - \mu)) - 1}. \tag{26}$$

The existance of the solution (23) depends on a distribution of energy levels, determined by the fitness distribution $\rho(\eta)$. If Eq.(25) has no nonnegative solution for given parametrs we can observe a Bose Einstein condensation, indicating that a finite fraction of the particles condensate on the lowest energy level.

5 Conclusion

In this seminar I have presented the origin of the rising science of networks. This is a very young area of research, mainly because in the nineties people started observing web structures like WWW and internet, which grew very fast. Some networks have already existed before, like Science collaboration network or Movie actor collaboration network, but there was no proper tool for manipulating a huge set of data. Two mathematicians, Erdos and Renyi, proposed a mathematical model, based on graphs, in the sixties. It was a start, but results of numerical simulation showed that real networks aren't random. And the search for new models began. I've presented the small-world model which very nicely satisfies the large cluster coefficient, measured in real networks, and scale-free model, which on the other hand satisfies the degree distribution. But that is only the beginning. Models become better with every simulation and new relations are discovered all the time.

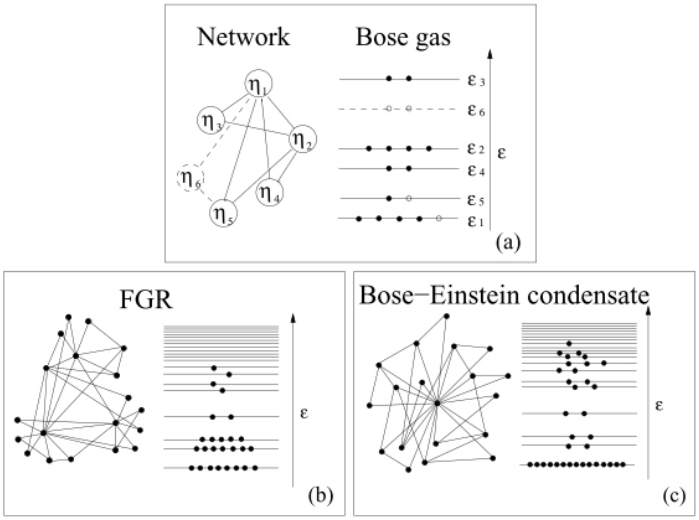


Figure 10: Mapping between the network model and the Bose gas; (a) on the left we have network of 5 nodes, each characterized by fitness η_i and energy ε_i. The network evolves by adding the sixth node which connects to $m = 2$ other nodes chosen following (9). In the gas this results in th addition of a new energy level ε_6 populated by $m = 2$ and the deposition of $m = 2$ other particles to energy level to which the new node is connected to. (b) The FGR= fit-get-rich phase, several high degree nodes linking the low degree nodes together. Occupation number decreases with increasing energy. (c) In the Bose- Einstein condesate the fittest node attracts a finite fraction of all edges. The ground level is highly populated, higher energies are sparsely populated.

There are also some analytical approaches, for instance the continuum theory for scale-free model. On the end, I've mentioned an explanation of Bose-Einstein condensation, based on fitness model, an example of an evolving network. I believe it is early days for the science of networks and the future will bring many interesting relations or, in the language of networks, links. By then, knowing that the whole is more than a sum of its parts, can be a good starting point.

References

[1] Reka Albert, Albert-Laszlo Barabasi: *Statistical mechanics of complex networks*, arXiv:cond-mat/0106096v1

[2] Newman, Barabasi, Watts : *The structure and dynamics of networks*

[3] Albert, Jeong, Barabasi: *Diameter of the World - Wide Web*, Nature 401 (1999)

[4] Watts, Strogatz: *Collective dynamics of 'small - world' networks*, Nature 393, 440 - 442 (1998).

[5] : *On the properties of small-world network models*, Eur. Phys. J. B 13, 547-560 (2000)

[6] : Albert, Barabasi: *Emergence of scaling in random networks*, Science Vol 286 (1999)

[7] Albert-Lazslo Barabasi : *Linked*