To summarize, we have seen that for most real networks it has been showed that the average path length scales as:

$$\langle l \rangle \approx \ln N$$

with the logarithm of the number of nodes in the network, not just with the number of nodes. Or in some cases as $\langle l \rangle \approx \ln(\ln(N))$. How is it possible? A paper which explain it is "Collective dynamics of small world networks" of Watts and Strogatz. Their idea is what is called the **Watss and Strogatz model**.
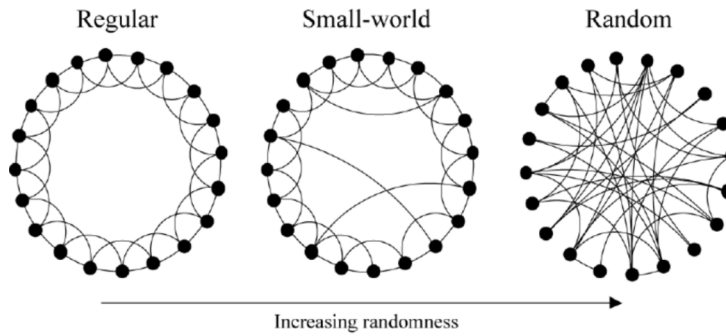
Figure 1: Idea of Watts and Strogatz model.

Let us focus with the first regular ring in Fig. 1 in which we have each node connected with its neighbour in its left and right. The structure is totally regular. If we want to measure the longest distance that we can find in the network we have that:

$$\left\langle l^{circle} \right\rangle \sim \frac{N}{4m}$$

What happens if we just rewire one of the connection? We connect it with a another random nodes in the network as in small-world circle in Fig. 1. What do happen for the distance? Just rewiring one connection reduce the size in an incredible way. If we can control the number of connection with a probability p, what happens is that every time we rewire the connection, the average steps is reduced by a factor 2. If we repeat this process several time we obtain a logarithmic scaling. At the end we obtain a random network which scales as:

$$\langle l \rangle \sim \log N$$

It is represented by the random circle in Fig. 1.

0.1 Degree distribution over networks

Now the question is how degrees are distributed over the newtwork. Let us consider a small network, the plot on the left of Fig. 2 represents the distribution of the degrees in this small network. How is this quantity distributed in real networks?

The first assumption that we can make is building the connections at random. Hence, there is no rule behind the degree distribution but a probability distribution as the one on the right of Fig. 2.

0.1.1 Erdös and Rényi Model: random graphs

Let us consider the Erdös and Rényi model which represents the evolution of a graph where links between nodes are drawn at random. In particular, the algorithm for creating such a network is:
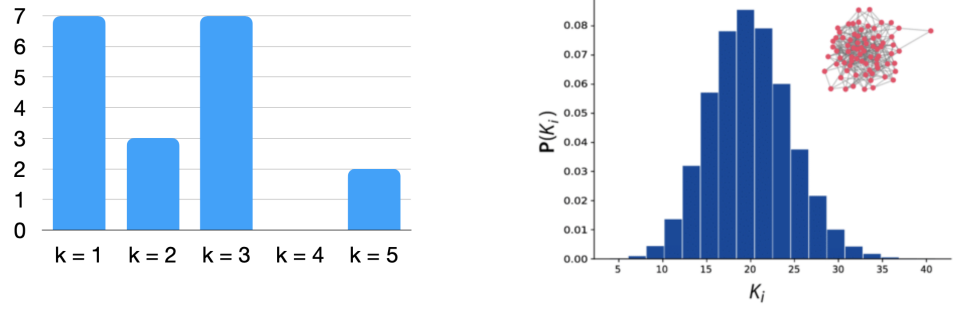
Figure 2: **Left:** degree distribution in a small network. **Right:** degree distribution in a network with random connections.

- create an empty graph with N nodes;

- connect each possible couple of nodes with probability p;

- avoid self-loops and multiple edges.

What are the properties of this graph? Let us consider a graph $G(N, p)$ where N are the number of nodes and p is the probability of connection. Before 1959 (the year of the publication of Erdös and Rényi's paper) people assumed that connection were regular, hence this is the first time in which random connections are been considered. If links are drawn at random with probability p, the probability that a node has k neighbors p_k is given by a binomial distribution:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{1}$$

The average and variance of such a distribution are:

$$\langle k \rangle = p(N-1), \qquad \sigma_k^2 = p(1-p)(N-1) \tag{2}$$

As we can see, the average and the variance scales in the same way with the size of the network. The problem of the last distribution is that it is difficult to be treated analitically when N increases, indeed:

$$\frac{\sigma_k}{\langle k \rangle} = \sqrt{\frac{1-p}{p(N-1)}} \overset{N \to \infty}{\longrightarrow} 0$$

which is very narrow as N increases, so we need a kind of approximation.

Fortunately, for sparse networks we have $k \ll N$, hence the binomial (N, p) distribution can be approximated by a Poisson distribution with $\lambda = pN$. Indeed, since $\langle k \rangle = p(N-1)$, having $k \ll N$ implies $p \ll N$ and thus:

$$(1-p)^{N-1-k} \approx e^{(N-1-k)\log(1-\langle k \rangle/(N-1))} \overset{N \to \infty}{\longrightarrow} e^{-\langle k \rangle}$$

and

$$\binom{N-1}{k} \approx \frac{(N-1)^k}{k!}$$

Obtaining:

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle}{k!} \tag{3}$$

As before, the average and the variance scales in the same way with the size of the network. This is telling us that all the nodes more or less are the same. If we have a bounded variance means that all the nodes have more or less the same degree. That is the point. In particular, as p is increased the graph undergoes a transition from disconnected to fully connected:

- if $Np < 1$, the graph will almost surely have no connected compontens of size larger than $O(\log(N))$;

- if $Np = 1$, the graph will almost surely have a giant component of size $O(N^{2/3})$;

- if $Np \to c > 1$, the graph will almost surely have a giant component comprising a large fraction of the nodes;

- if $p < \frac{(1-\varepsilon)\ln N}{N}$, the graph will almost surely contain isolated vertices;

- if $p > \frac{(1-\varepsilon)\ln N}{N}$, the graph will almost surely be connected.
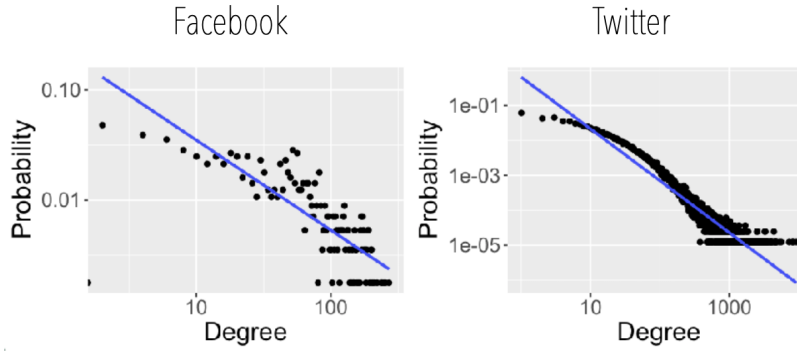
Figure 3:   Real network of Facebook and Twitter.
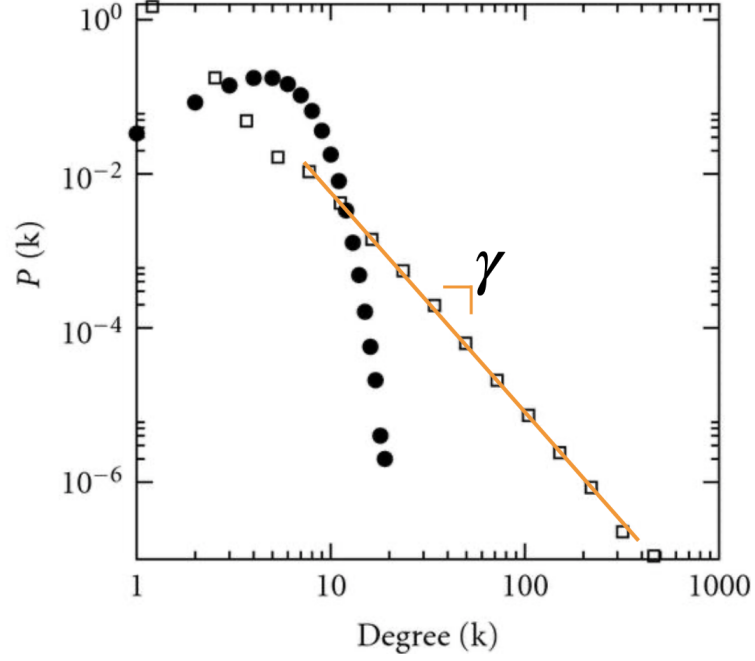
0.1.2   Scale-free networks

The point now is: what is the degree distribution of real networks? In the last decades we start to have very very complex and large networks. Their structure is nothing like the structure that you see for a random network. In Fig. 4 are shown real network of social networks as Facebook and Twitter. All this plots are in log-log scale. We can conclude that most of the networks scales in the same way. But, what is the form of this connection? Let us consider Fig. 4, the black curve represents the Poissonian distribution that we saw before, and then we plot the power-law (heavy tailed distribution) $P(k) \sim k^{-\gamma}$. We see that the Poissonian distribution is not able to reproduce the hetherogeneity you see in the data, while the power-law it is. Hence, in most contexts real networks are highly heterogenous and degrees can vary several orders of magnitude. In particular, the γ coefficient of the power-law has an important role, since it represents the slope of the curve in log-log scale. Thus, we have similar structures at different scales, this leads to the concept of scale-free networks. Most real networks have small values of γ, i.e. $\gamma \leq 3$.

Heterogeneous means that most of the nodes have a very low connectivity, less than a random net. However, the probability of having very large degrees is not zero (**hubs**). Even for small networks you can have large Hubs and this is something we have to take into account for the spreading of diseases since we can have shortcuts for spreading or super-spreaders.

We can study the limiting cases of this scale-free networks. For instance, we can study some of the properties we have seen before, as the average degrees or we can prove how the largest degree scales as the size of the network. Let us consider the power-law:

$$P(k) = C_0 k^{-\gamma} \qquad \text{with} \quad C_0 = (\gamma - 1)k_{min}^{\gamma-1} \tag{4}$$

Figure 4: Difference between random networks and scale-free networks.

To know how k_{max} scales with N, we have to study when:

$$\int_{k_{max}}^{\infty} P(k)\,\mathrm{d}k = \frac{1}{N} \quad \rightarrow \quad \left(\frac{k_{min}}{k_{max}}\right)^{\gamma-1} = N$$

Thus, when:

$$k_{max} = k_{min}N^{\frac{1}{\gamma-1}} \tag{5}$$

The most important thing is that more of the network we have $\gamma = 2, 3$ as we can see in the previous plot of Fig. **??**.

The general n^{th} moment of the distribution is:

$$\langle k^n \rangle = \int_{k_{min}}^{\infty} k^n P(k)\,\mathrm{d}k = \int_{k_{min}}^{\infty} C_0 k^{n-\gamma}\,\mathrm{d}k \tag{6}$$

It converges only if $\gamma - 1 > n$. This gives an hint on how the average of the degree scales as the size of the network. This is a very important result. Considering $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$, we have that:

- if $\gamma < 2$, both $\langle k \rangle$ and $\langle k^2 \rangle$ diverge with $N \to \infty$;

- if $2 < \gamma < 3$, the average degree $\langle k \rangle \to c$ but $\langle k^2 \rangle \to \infty$ as $N \to \infty$ and $\sigma^2 \to \infty$.

Remember that most real networks have $\gamma \leq 3$, hence the variance of the degree also diverges and so we have extremely heterogeneous networks not homogeneous ones. This means that all the models we have been used before in which we assumed that all the people in the population were equal, does not hold anymore.

0.1.3 Barabási-Albert Model

The last thing we need is an algorithm to create a random scale-free network. We can realy on the Barabasi-Albert model. This is the second paper which starts the

field. We have to rethink all the models we have been used right now and include these things.

The idea behind this paper is extremely simple. They analized real networks and then they assumed $P(K) \sim k^{-3}$ in order to create a model to reproduce this behavior. Their model is based on the concept of growing random networks. We start with a small number of nodes, at each time-step a new node enters the network and connects with pre-existing nodes but with a preferential attachment. Hence, at each step the network growth.

The principle behind the preferintial attattchement which is based on a very simple concept: rich get richer. The more connected a node is, the more likely it is to receive new links. The probability to attract a new link at time t proportional to degree at time t:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{7}$$

In the case of influencers, if I have a lot of followers the probability of increase my connections is increasing too. Actually, this idea is not even so new and it is something we have been known. This model hence is just a modification of the Price model: if I published a paper and someone think that it is interesting, more attention I get more attention I will get in the future.

So, I am drawing the link at random but not uniformely. Let us briefly summarize the main steps of the algorithm:

- we start with a clique of m_0 nodes;

- at each time step t, we add a new node to the network;

- we create m (i.e. $m = 2$) links between the new node and the existing ones according to the preferential attachment (remember to update the connection probability after each link);

- repeat until size N is reached.

In particular, let us consider Fig. 5. We start with a small number of nodes connected in a link. In the first time step I am gonna adding a new node, and I am gonna connectecting to the other nodes. Let us assume that every time I add a node, I am adding two links. I calculate the probability of getting a new node which in this case is equal for all the nodes. Then I pick up one at random and I connect to it. Then I have to update the probabilities. We see that the node on the left got an higher probability of getting new connections. Then I start with a ned node untile size N is reached.
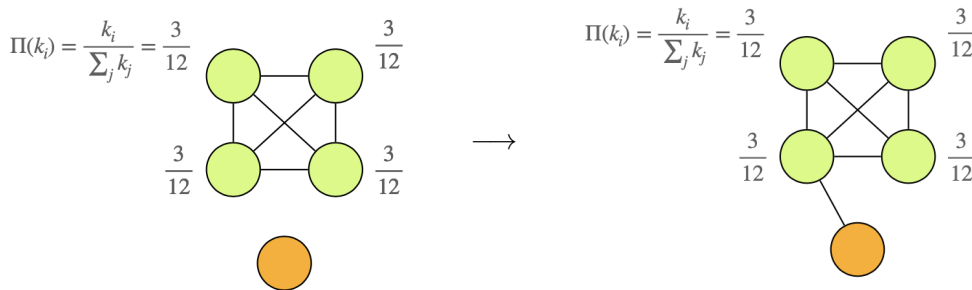
Figure 5: Example of Barabási-Albert algorithm.

This algorithm is able to create netowrks with some interesting properties. The simple idea is that the degree distribution is:

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \sim k^{-3}$$

for large k, where m is the number of links you are adding at each step (this is a parameter which drives you the minimal degree of the network). We obtained that we obtained $\gamma = 3$ independent from m and m_0. Hence, the maximum degree of the network scales as $k_{max} \sim N^{1/2}$. Moreover, we have that $\langle k \rangle \to c$, but $\langle k^2 \rangle \to \infty$ with N, as before. The length of the network is:

$$\langle l \rangle \sim \frac{\ln(N)}{\ln(\ln(N))}$$

hence we have small-world.

1

Epidemic Spreding on Networks

The idea is that we pass from the classical well-mixed population to the contact networks. So we assume that we have some sort of connection between individuals. The idea is that:

- all individuals are equivalent;

- we remove the assumption tha all individuals have the same number of contacts and we assume that individuals do not interact at random: friends, family and so on. You have somehow constraing how the disease will spread. The fact that we have repeated contacts has strong effect on the dynamics.

1.1 SIS model in a network

Let us try to build a general model for a general network. For doing that the idea is, let us start by writing down the equations for an SIS dynamics on a general network. First of all, we define a binary variable for each node i: $\sigma_i(t)$. This variable can only have two values:

- $\sigma_i(t) = 0$, if the individual is susceptible;

- $\sigma_i(t) = 1$, if the individual is infected.

We define another variable $\rho(i, t)$:

$$\rho(i, t) \equiv \text{Prob}[\sigma_i(t) = 1]$$

which represent the probability of that node i is infected at time t. So from here, we can write a general equation for the SIS in a nwwtork:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(i, t) = \overbrace{-\mu\rho(i, t)}^{\text{Recovery}} + \beta \sum_j A_{ij} \overbrace{\text{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1]}^{\text{Infection}} \tag{1.1}$$

The problematic part is that we have to find an expression for the two nodes infection probability (in green). Since we are in a network, the probability of being infected depends on my neighbours: the (i, j) infection probability depends on the status of all the other neighbors l of j and i and so on. Hence, we have to follow the entire chain of connection, so we cannot obtain a closed form for this probability, since it depends on the three nodes probability and so on. Again we repeat that we have to follow the entire network if we want to know what is gonna happening and we have to write down the entire time evolution of the system. This approach is feasible only for small graphs (i.e. 4/5 nodes) and few compartments.

The idea is that we need to cut down this probability chain at some point. So at some point we need a closure of our equation, by means of approximations. We are going to not taking into account all the structure of the network, but at some point we will take the average. After that we will be able to solve the problem. In physics this are called **mean-field approximations**. Since we are not able to solve many body problems at a certain point, we will consider a random field which acts on the entire system and at some point we will consider the average effects on the system.

What happens in our problem is that we are susbtituting in some way this probability $\text{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1]$ with some average probability. Obviously, depenging on the assumption we are making for this approximation we will obtain different results. We have different types of approximations based on:

- Network structure:

    - Homogeneous mean-field (all the nodes are equal);

    - Heterogeneous mean-field;

- Coarsening level:

    - Degree-based mean-field theories (DBMF) in which we assume that all the nodes of the same degree are equal;

    - Individual-based mean-field theories (IBMF) in which we assume that all the nodes are different and that we will take individual connections between individuals;

- Where to cut the chain:

    - Individual level;

    - Pair approximations;

    - Triangles, etc...;

1.1.1   Homogeneous networks

Let us start with the simplest approximation. We assume homogeneous networks, DBMF and we cut the chain at an individual level.

We are considering networks where nodes's degree is bounded, hence:

- we have that $k_i \simeq \langle k \rangle$;

- we have also that the standard deviation is bounded $\frac{\sigma_k}{\langle k \rangle} = \sqrt{\frac{1-p}{p(N-1)}} \xrightarrow{N \to \infty} 0$.

All the nodes can be assumed to be equal, so their position on the network does not matter anymore. The spatial homogeneity becomes $\rho(i, t) \equiv \rho(t)$.

Cutting at the individual level means that $\text{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1]$ (one being infected and the other one supsceptible) are statistically independent. Hence, we can decompose this one with a product of individual probability:

$$\text{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1] \qquad \rightarrow \qquad \text{Prob}[\sigma_i(t) = 0] \cdot \text{Prob}[\sigma_j(t) = 1]$$

Remembering that:

$$\rho(t) = \text{Prob}[\sigma(t) = 1]$$

is the density of infected at time t. If we put everything togheter, we get the equation:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\mu\rho + \beta \sum_j A_{ij}(1-\rho)\rho \qquad \rightarrow \qquad \frac{\mathrm{d}\rho}{\mathrm{d}t} = -\mu\rho + \beta(1-\rho)\rho \sum_j A_{ij}$$

If we remember that:

$$\sum_j A_{ij} = k_i \simeq \langle k \rangle \tag{1.2}$$

is the degree of the network, we get exactly the same expression that we got before for SIS model in a well-mixed population:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \beta \langle k \rangle (1 - \rho)\rho - \mu\rho \tag{1.3}$$

This is a very important result. Actually, we are considering all the nodes statistically independent and we are recovering exactly the same result of well-mixed population. The only difference is that when we were considering well-mixed population, we assumed that the probabilities where *exactly* statistically independent. Now, this is just an approximation.

Obviously, all the results for SIS in well-mixed populations are valid, as the epidemic treshold.

Remark. Let us recap what we saw at the end of this lecture. We moved from well-mixed populations to contact networks, so we add more complexity and the model is more realistic. We also wrote down what is a generic model for SIS dynamics on a general network. We consider the adjiacency matrix and so on. We cannot write down a closed equation for this, because we have two nodes infection probability. The problem is that we do not have an exact expression for this probability. The expression for this probability take into account the probability of three nodes ijk. This is unfeasible for all the models and all the possible graph, in the literature we have just 4/5 nodes. The idea is that we ned a way to approximate this probability, to cut this infinite chain to a certain value. We will use mean-field approximation. We have a quantity that depends on all different quantities but which in this approximation is reduced to its average. We are switching from a many body problem to a one body problem. We start to see a model of SIS for homogeneous network in which all the nodes are equal. There is the same probability of getting infection and we can consider the probabilities statistically independent. We derived all the equations. The solutions we found are the same of before for well-mixed population, the difference is that now it is not exact but only an approximation.