To summarize what we have seen last lecture: it holds for most real networks that the average path length scales as:

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

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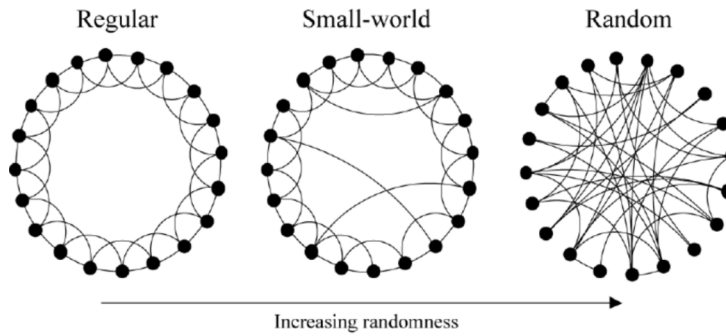

Figure 1: Idea of Watts and Strogatz model.

Let us focus on the first regular ring in Fig. 1, in which we have that each node is connected with its two nearest neighbours in both sides. The structure as we can see is totally regular. If we want to measure the **longest distance** that we can find in the network:

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

But what actually happens if we rewire only a single link? We therefore want to connect it with another random node in the network as in the picture in the middle "small-world" in Fig. 1. It can be seen that, by doing a single rewiring, the size of the system reduces in an incredible way. On the other hand, if we extend this argument and choose a probability p for rewiring (i.e. we increase randomness), what happens is that every time we rewire a connection, the average distance is reduced by a factor 2. Repeating this process several times, we observe a **logarithmic scaling**. Finally, the random network we obtain scales as:

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

And it is represented by the random circle in Fig. 1.

0.1 Degree distribution over networks

Now the question is how degrees are distributed for different type of networks. Let us consider a **small network**, its degree distribution will be really resembling to the plot on the left of Fig. 2. However, now we want to understand how this quantity distributes in **real networks**. In order to build a real network, the first assumption that we can make is building the connections *at random*, so with a probability p. Consequently the degree distribution is one of the kind as in 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, according to a predefined
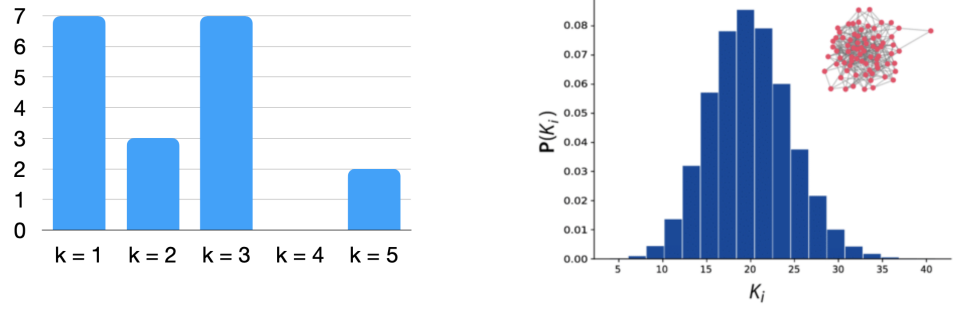
Figure 2: **Left:** degree distribution in a small network. **Right:** degree distribution in a network with random connections.

probability p. Before 1959 (the year of the publication of Erdös and Rényi's paper) people were actually assuming that connections were regular, so no randomness at all. However, since randomness in real world is a deal, thanks to E.R. random connections were taken into account for the first time. In particular, the algorithm for creating such a network is:

- 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 link**. If links are drawn at random with probability p, the probability p_k that a node has k neighbors 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 (i.e. linearly!).

The problem of this distribution is that it is difficult to be dealt with analytically, specially as N increases, indeed:

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

which becomes narrower as N becomes larger, therefore some sort of **approximation** needs to be introduced.

Fortunately, since for sparse networks we have $k \ll N$, the binomial (N, p) distribution can be approximated by a **Poisson distribution** with parameter $\lambda = pN$. Indeed, given that $\langle k \rangle = p(N-1)$, if we have $k \ll N$ then it implies that $p \ll N$. Hence we can write the following:

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

and

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

Obtaining the **Poisson distribution** we were looking for:

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

As before, the average and the variance scale exactly in the same way with the size of the network ($\sim \lambda = Np$). This actually tells us that **all** the **nodes are** more or less **the same**. Indeed when we observe a bounded variance, it means that all the nodes more or less have the same degree. In particular, as p increases the graph undergoes a **transition** from disconnected to fully connected one:

- if $Np < 1$, the graph will almost surely have no connected components 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;

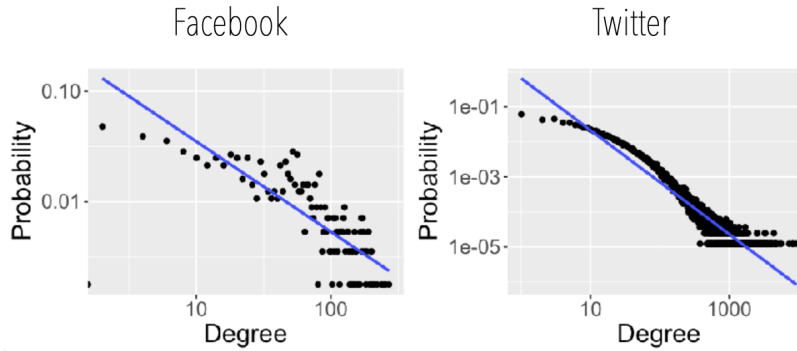- if $p > \frac{(1-\varepsilon)\ln N}{N}$, the graph will almost surely be connected.

Figure 3:   Real network of Facebook and Twitter.

0.1.2   Scale-free networks

However, so far we have not discussed how real networks look like, in particular what is their degree distribution. In the last decades we started to have really complex and large networks, whose structure really differs from the structure we usually see for a random network. In Fig. 4, as an example, we shown two real social networks we know pretty well: Facebook and Twitter. Note that both plots are in log-log scale. Generalizing, we can say that most of the real networks scales in the same way.

We now want to understand how the **degree distribution** looks like. Let us consider Fig. 4: black dots follow the Poissonian distribution that we were mentioning before, while the squares follow a power-law $P(k) \sim k^{-\gamma}$, which is **heavy tailed distribution**, in the sense that possibility for large degrees is not null. One should note that the Poissonian distribution is not able to reproduce the heterogeneity we can see in the data, while the power-law is. Hence, in most contexts real networks are **highly heterogeneous** and degrees can span **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. Since we observe similar structures for different
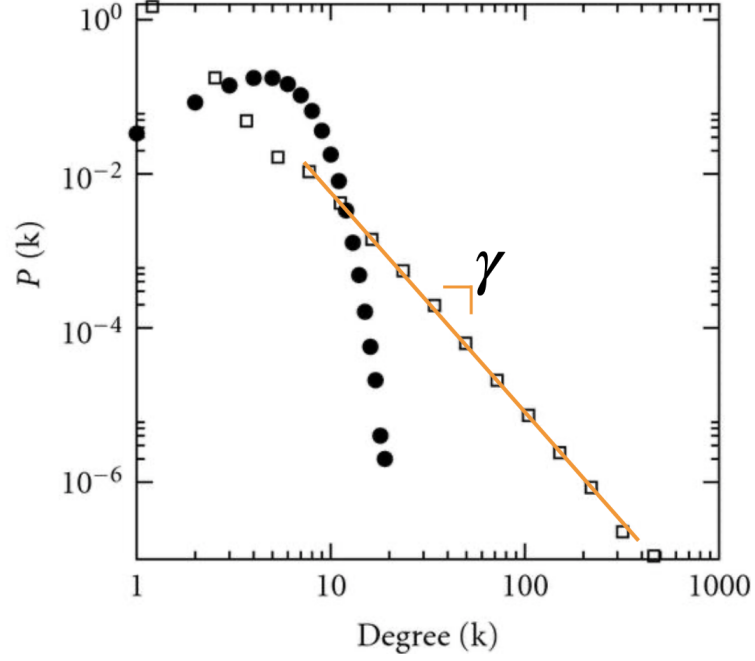
Figure 4: Difference between random networks and scale-free networks.

scales, these networks are said to be **scale-free** networks. In most real networks γ has small values, i.e. $\gamma \leq 3$.

Heterogeneity means that almost all nodes have a very low connectivity, way less than a random net. However, the probability of having very large degrees is not zero (**hubs**): even for relatively small networks we can observe large hubs. One should take into account that this is something really important for the spreading of diseases: thanks to these large hubs we can see shortcuts for spreading, or the so called **super-spreaders**.

We want now to study the **limiting cases** of these scale-free networks. For instance, we want to see how the **average degree** behaves, or prove that the **largest degree** scales with the size of the network. Let us consider the power-law:

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

To understand how k_{max} scales with N, we have to study the case where:

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

Thus, when:

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

Since in most of networks $\gamma \sim 2 - 3$, so it is easily to understand that k_max scales **sub-linearly** with N, but still way faster than random graphs. This is valid for previous plots, such as in Fig. 3 as well.

Recalling the definition for the general n^{th} moment of a distribution:

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

We note as it converges only if $\gamma - 1 > n$. This gives an hint on how the **average degree** scales as the size of the network: a very important result. If instead we consider the variance $\sigma^2 = \langle k^2 \rangle - \langle k \rangle^2$, we it holds 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 $\sigma^2 \to \infty$.

Remembering that most real networks have $\gamma \leq 3$, hence the **variance of the degree also diverges**. The result is that we have extremely **heterogeneous networks** and not homogeneous ones. This is indeed coherent to our observations. It has indeed a very strong **implication**: all the models we have been using before, in which we assumed that **all the people** in the population were **equal**, does **not hold** anymore.

0.1.3 Barabási-Albert Model

So far we have discussed about scale-free networks, but actually we have not created a single one yet. Therefore, an **algorithm** to create such network we can rely on, is the **Barabasi-Albert model**. This topic is discussed in a paper that is the second, chronologically speaking, that gave birth to modern Network Science.

The **idea** behind this paper is extremely simple: once some real networks had been analyzed they assumed that the degree distribution $P(K) \sim k^{-3}$, in order to create a model to reproduce the behaviors observed. Moreover, their model was based on the concept of **growing** for random networks. We start with a small number of nodes, named **clique**, and, at each time-step, a new node enters the network and connects with pre-existing nodes but according to a **preferential attachment**. Therefore, at each step the network grows in size.

The principle on which **preferential attachment** is based on is a very simple concept: *rich gets richer*. That is to say: the more connected a node is, the more likely it is for it to receive new links. The probability for a node i to attract a new link at time t, is proportional to its degree k_i at time t:

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

If we speak about **influencers**, having them a lot of followers, the probability for them to increase their connections is very high. Actually, this idea is not even new, and it is something already known. Indeed this model is just a modification of the *Price model*: if we published a paper and more than someone has found it interesting, it will be more likely for it to receive much more attention in the future.

Specifically for this model, we are drawing links at random, according to some probability that indeed is not uniform. 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 the desired size N is reached.

In particular, let us consider Fig. 5. We start with a small number of nodes connected via some links. At the *first time step* we add a new node, and then we need to draw connections to the other nodes. Let us assume that every time we add a node, we are adding two links. First, we need to compute the set of probabilities of connecting to each node and, at the first time-step, is equal for all the nodes. Then we pick

up one node at random and we draw the link. The following step is to update the set probabilities for each node, according to their degree. We see that the node on the left has got an higher probability of getting new connections since the last node inserted has linked to it. Then, we iterate this procedure by introducing a new node and draw connections following the same procedure, until we end up with a total number of N nodes.
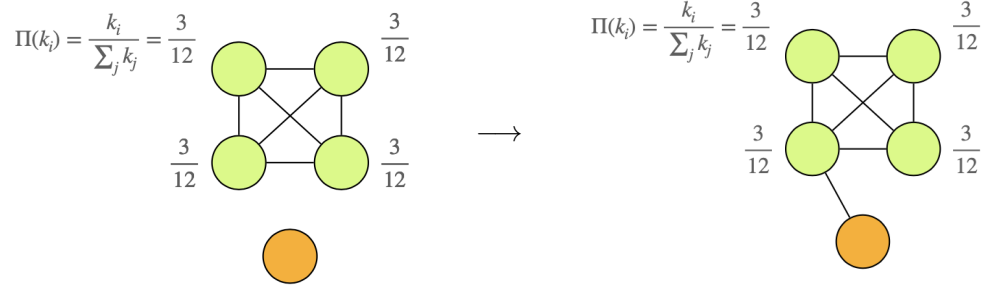

Figure 5: Example of Barabási-Albert algorithm.

This algorithm is indeed able to create networks with some **interesting properties**. Indeed we can approximate the **degree distribution** as:

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

where m is the number of links we are adding at each step. Note that m is a parameter that is related the minimal degree of the network. However, this approximation is valid for **large** k.

An important result is that $\gamma = 3$ and it is **independent** of m and m_0. Hence, the **maximum degree** of the network scales as $k_{max} \sim N^{1/2}$. Moreover, it holds that $\langle k \rangle \to c$, but $\langle k^2 \rangle \to \infty$ with N, as we have seen before. Finally, the **average length** of the network is:

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

which tells us that the small-world property holds as well.

1

Epidemic Spreading on Networks

Now it is time to drop the assumption of the well-mixed population, and start taking into account **contact networks**. In other words we are considering that **individuals can be connected in different ways** one another. The main idea is that:

- **all** individuals are **equivalent**;

- we remove the assumption that all individuals have the same number of contacts and we assume that each node **do not interact at random**. This reflects the reality, since we usually have more contacts with some people (friends, family, colleagues...) rather than others. The fact that we may have repeated contacts with someone else has strong effects on the dynamics: we are somehow constraining the way how the disease will spread.

1.1 SIS model in a network

Let us try to build a general model for a general network, without making any assumption on the latter. In order to do that, we start by introducing the equations of SIS model for a generic network.

The first step is to define a **binary variable** for each node i: $\sigma_i(t)$. This variable can only take two values:

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

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

As one can easily see, this variable describes the state of a generic i-th node at time t. Defining another variable $\rho(i,t)$:

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

which represents the **probability** of that node i is infected at time t. Using this formalism, we can recall the general equation for the SIS in a network:

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

The most problematic part is to compute 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 forth. Therefore we would have to follow the entire

chain of connections, but this would turn out to be a problem: we cannot obtain a closed form for this expression, since it actually depends on the probabilities of all its neighbors. In turn, they would depend on their neighbors probability and so and so forth.

We want to stress one more time that if we want to predict what is going to happen in the system, we would need to consider the entire network and the time evolution for all the nodes. This approach however is **feasible** only for **small graphs** (i.e. 4/5 nodes) and **few compartments**.

This argument reminds us that we may need some sort of an **approximation**: indeed we need to **cut down** this **probability chain**. That is to say that, at some point, we require a closure of our equations, by the mean of approximation: we are not going to take into account the entire structure of the network. At some point we will take the **average**, and after that we will be able to solve the problem. In physics this kind of arguments 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 we will consider its average effects on the system.

Tailoring this procedure to our specific problem, we are substituting in some way the probability $\text{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1]$ with some average probability. Obviously, depending on the assumption we are making for this approximation, we will obtain different results.

There are actually many different types of approximations based on different features:

- **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 by taking the simplest approximation: we assume **homogeneous network**, **DBMF** and we cut the chain at an **individual** level.

It means that we are considering networks where **nodes 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. This implies the **spatial homogeneity** it holds that: $\rho(i, t) \equiv \rho(t)$.

In addition, cutting at the individual level means that the two terms of the **joint probability** of one being infected and the other one being susceptible $\mathrm{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1]$ are **statistically independent**. This implies that the joint probability can be factorized as follows:

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

But now we recall that:

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

is the density of infected at time t. Hence, putting everything together, we derive the equation:

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

Actually, this last term is the degree of the network:

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

and by replacing it, we can obtain the same expression that we derived 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. One should keep in mind that now we are considering all the **nodes statistically independent** and we are back again to 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 derived for SIS model in well-mixed populations are still valid, for instance the epidemic threshold.

Remark. Let us recap what we have seen at the end of this lecture. We moved from well-mixed populations to contact networks, so we added more complexity in order to make the model is more realistic. We also derived the equations for SIS dynamics on a generic network and then considered its adjacency matrix. Since for us was impossible to write down a closed equation for this model, given the expression for the infection joint probability that involves two nodes, we were not able to compute exactly the probability for a single node of being infected (ρ_i). It would take into account the probability of three nodes i, j, k at the same time. This is actually unfeasible for all the models and all the possible graphs: it has been done in the literature up to only 4/5 nodes. Hence we end to somehow approximate this probability, in order to cut this infinite chain to a certain value. This is exactly why we introduce mean-field approximation: in this way we take into account the effects of all terms on a specific quantity, not individually, but on average therefore reducing the complexity of our problem. We are switching from a many body problem to a one body problem. The simplest approximation we have seen is the one of homogeneous network in which all the nodes are equal, used on SIS model. According to this argument, for each node there is the same probability of getting infected, so we can approximate the probabilities to be statistically independent. After, we derived all the equations. Their solutions were the same as the ones we had found for well-mixed population. However, in that case the solutions found were *exact*, while now are the result of an approximation.