Let us start from the definition of small world networks. We have seen tha the average path length scales with the logarithm of the number of nodes in the network, not just with the number of nodes. How is it possible? A paper which explain it is "collective dynamics of small world networks". Their idea is what is called the Watss and Strogatz model. Let us focus with the first regular ring in which we have each connected connected with the neighbour in their 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 (small-world). 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 I am rewiring 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:

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

Now the question is: how degrees are distributed over the newtork? Let us consider a small network, the plot on the right represent the distribution of the degrees in this small network. How this quantity is distributed in real networks? The first assumption that we can make is what happens if the connection are built at random. So there is no rule behind it, but a certain probability. Let us consider Erdos and Renyi model. In their paper we have the evolution of a graph where links between nodes are drawn at random. What are the properties of this graph? Let us call a graph G with $G(N, p)$ where N are the number of nodes and p is the probability of connection. Before 1959 people assumed that connection were regular, but this is the first paper in which random connections are been considered.

Given that we are drawing connection at random? What is the degree distribution at the end? Each link is drawed with a probability p. What is the probability that a node has k connections?

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

From that we can reocver that $\langle k \rangle = p(N-1)$ and variance $\sigma_k^2 = p(1-p)(N-1)$. 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 and we need an approximation. Fortunately, we have that for sparse networks we have $k \ll N$, the binomial (N, p) distribution can be approximated by a Poisson distribution with $\lambda = pN$. Hence:

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

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.

- 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 $O(N^{2/3})$;

The idea 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. This are examples of social networks as Facebook and Twitter. All this plots are in log-log scale. We have a lot of networks that scales in the same way.

What is the form of this connection? In black we see the Poissonian distribution that we saw before, and then we plot the function $P(k) \sim k^{-\gamma}$. We see that the Poissonian distribution is not able to reproduce the hetherogeneity you see in the data.

The power low have an important role, in particular γ is the slope of the curve in log-log scale. ?? Another important thing is that if you look at the distribution is that assuming that these two distributions have the same number of degree, the hetherogenity is much larger ?? The functional form is the same at all scales for scale-free network, but it is not the same for Poissonian distributions. These are the main characteristics.

The most important thing is that the probability of having large Hubs is not zero. Even for small networks you can have large Hubs and this is something we have to take into account for the spreading of diseases.

We can study the limiting cases of this scale-free networks. 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.

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

The most important thing is that more of the network we have $\gamma = 2,3$ (plots of before).

The general moments of the distribution are: ... The moment converges only if $\gamma - 1 > n$. This gives an hint on how the average of the degree scales as the size of the network.

- 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 $\sigma^2 \to \infty$.

This means that all the models we have been used before (where we assumed that all the people in the population are equal does not hold anymore). The variance of the degree diverges, hence we have extremely heterogeneous networks, not homogeneous ones.

The last thing we need is an algorithm to create them, we have 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 this things. The idea behind this paper is extremely simple. They analized thesen etworks and the assumed $P(K) \sim k^{-3}$ and the create a model to reproduce thsi behavior. We start with a small number of nodes and we connect the links at random but which follow a preperential attachmenet. At each step the network growth.

The principle behind the preferintial attattchement is very simple: rich get richer. In the case of influentials, if I have a lot of followers the probability of increase my connection is increasing too. Actually, this idea is not even so new. It is something we have been known. This model hence is just a modification of the Price model. I publushed a paper, someone think that it is interesting and more attention I get more attention I will get in the future.

So, I am drawing the link at random but not uniformely. In the basic algorithm, 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 gwtting new connections. Then I start with a ned node untile size N is reached. This algorithm is able to create netowrks with some interesting properties. The simple idea is that:

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

for large k. m is the number of links you are adding and this is a parameter which drives you the minimal degree of the network.

0.1 Epidemic Spreding on Networks

0.1.1 SIS model in a network

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;

- 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.

Let us try to build a general model for a general network. For doing that the idea is, let us start with the SIS dynamics and defining a binary variables in the sense that can only have two values: 0 if it is susceptibòe and 1 if it is infected. We define another variable $\rho(i,t)$ which represent the probability of that node i is infected at time t. So from here, I can write a general equation for the SIS in a nwtork:

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

The problematic part is the probability which is... we have to find an expression for it. Since we ae in a network, the probability of being ifnected depends on my neighbours. I have to follow the entire chain of connection. No closed form, it depends on the three nodes probability and so on... I have to follow the entire network if I want to know what is gonna happening. We have to write down the entire time evolution of the system.

The idea is that since I will gonna have this chain of probability, I need to cut down this chain at some point. So at some point I need a closure of my equation. I am not taking into account all the structure of the network, but at some point I will the the average. We need some sort of approximation for this probability. After that we will be able to solve the problem. In physics this are called mean-field approximation, since I am not able to solve the many body problem at a certain point I will consider a random field which acts on the entire system and at some point I will consider the average effects on the system.

What happens is that I am susbtituting in some way this probability $\mathrm{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1]$ with some average probability. Obviously, depengin on the assumption we are making for this approximation we will obtain different results. If I am assuming

that the network is homoegenous (all the nodes are equal) I have a mean-field, if it is heterogeneous I will have more equations but I will have a more reliable approximation of the system.

Coarse graining level. The distinction can be also in the level I will adopt this approximation. We can have degree based mean field theories in which we assume that all the nodes of the same degree are equal. While individual bases mean field means that all the nodes are different and that I will take individual connections between individuals. The probability of having the disease for a node can be different depending on different factors...

Where to cut the level? In individual level or pair approximation?

Let us start with the simplest approximation.

0.1.2   Homogeneous networks

The important thing is that the variance and standard deviation is bounded. I am assuming that all the nodes are equal. I can forget about phase, the position of the node on the network does not matter anymore because the nodes are equal.

I am gonna make the mean field at the individual level: I am approximation the probability $\text{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1]$ (one being infected and the other one supsceptible) are statistically independent. Hence, I can decompose this one with a product of individual probability $\text{Prob}[\sigma_i(t) = 0] \cdot \text{Prob}[\sigma_j(t) = 1]$. I define:

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

If we put everything togheter, we get the equation:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t}$$

which looks quite familiar, where we have $\sum_j A_{ij}$ which is the degree of the network. We get exactly the same expression that we got before for well-mixed population. This is important that actually what are you doing is that all the nodes are statistically independent. I am recovery exactly this result. When I was consider well-mixed population, I assumed that the probabilities where exaxtly statistically independent. Now, this is just an approximation.

All the calculation that we get are gonna hold for this thing. In the labs you will see when this assumption breaks down.

0.1.3   Heterogenous netowrks