Lecture 8.
Thursday 22nd October, 2020.
Compiled: Tuesday 27th October, 2020.

0.0.1 Heterogeneous Networks

What is the effect of heterogeneity in the spread of the disease? This assumption $k_i \sim \langle k \rangle$ does not hold anymore, so we cannot assume that all the nodes are equal.

Let us consider **heteroegeneous mean-field approximation**, suppose a **DBMF model** and let us cut the chain at an **individual level** (we consider the individual probability of getting the infection).

Let us start with the paper "Epidemic Spreading in Scale-Free Networks". It provides a SIS model on scale-free networks. The idea is that since nodes are not equal anymore, the probability of getting the infection strongly depends on their position (i.e. degree) in the network. The Pastor-Satorras and Vespignani's intuition is that nodes with the same degree behave in the same way. We are gonna divide the network in degree classes: we group togheter all the nodes with the same degree.

To write down the equation we need to multiply the number of compartments:

$$s_k = \frac{S_k}{N_k}, \qquad \rho_k = \frac{I_k}{N_k}$$

where s_k and ρ_k are the fraction of suscpetible/infected nodes of degree k in the network. We have that N_k represent the number of nodes with degree k in the network. So, we are defined as before the number at degree k of suscpetible and infected in the system. The total fraction of ρ and s in the system is given by:

$$\rho = \sum_k P(k)\rho_k, \qquad s = \sum_k P(k)s_k \tag{1}$$

Now, we write the equation for each degree class:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_k(t) = -\mu\rho_k(t) + \beta k\,(1-\rho_k(t))\Theta_k(t) \tag{2}$$

where we have as usual a recovery and infection part. In particular, we the probability of a contact between a suscpetible of degree k and an infected is represented in green. The idea behind it is that we have the probability of being infected $(1-\rho_k(t))$ and the probability of having contact with an infected $\Theta_k(t)$. In particular, the probability that a node with dehree k has an infected neighbor can be expressed as:

$$\Theta_k(t) = \sum_{k'} P(k'|k)\rho_{k'} \tag{3}$$

where we sum over all the possibile degree classes and we are gonna see the probability of connecting with one of them, hence this is the probability that another node is infected. Note that we are making no assumption about the function $P(k'|k)$ which will change with k. It could be in principle anything, in the sense that it depends on the structure of the networ. However, there are some cases in which we can do some assumptions on the structure of the network.

For random networks, e.g. picking a node at random:

$$P(k'|k) = \frac{k'P(k')}{\sum_{k'} k'P(k')} = \frac{k'P(k')}{\langle k \rangle} \tag{4}$$

where $P(k')$ is the probability of getting a connection at random. Then, we multiply it by k' which is the number of connection that we pick up. Then we normalize over the average degree of the network. Hence, at the end it is the probability that a point in the network points to k'. Note that $P(k'|k)$ does not depend on k.

We obtain:

$$\Theta_k(t) = \frac{\sum_{k'} P(k')\rho_{k'}(t)}{\langle k \rangle} = \Theta(t)$$

In the numerator: we take the probability that a link taken at random points to k', then we multiply by the probability of being infected and then we sum over all the possible degrees. We note that $\Theta_k(t)$ does not depend on k anymore. We are just picking up at random, so it should be the same for all the nodes.

The method that we are gonna use to solve the differential equation $\frac{\mathrm{d}}{\mathrm{d}t}\rho_k(t)$ is similar to the ones used before for the other models. First of all, we assume that we are in the steady state:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_k(t) = 0 \qquad \rightarrow \qquad \rho_k = \frac{\beta k \Theta}{\mu + \beta k \Theta}$$

The next step is to substitute the expression for ρ_k obtained inside Θ:

$$\Theta_k(t) = \frac{\sum_{k'} k' P(k')\rho_{k'}(t)}{\langle k \rangle} = \Theta(t) \qquad \rightarrow \qquad \Theta = \frac{1}{\langle k \rangle} \sum_k \frac{k^2 P(k) \beta \Theta}{\mu + \beta k \Theta}$$

this is the self consistent equation for Θ.

The point is: if we want to solve this expression, we need some sort of trick. First of all, what happens is that as usual this expression has different solutions:

- the first one is the trivial solution $\Theta = 0$, but we are interested in the non trivial one;
- to obtain the non trivial solution let us note that:

$$\Theta = \frac{1}{\langle k \rangle} \sum_k \frac{k^2 P(k) \beta \Theta}{\mu + \beta k \Theta} = f(\Theta)$$

Hence, the solution are the values of Θ equal to $f(\Theta)$. This is the interception between the line Θ and the function $f(\Theta)$. Since Θ is a probability, we have that $0 < \Theta \leq 1$. This means that the slope of $f(\Theta)$ should be greater than 1. Mathematically, it menas that:

$$\frac{\mathrm{d}}{\mathrm{d}\Theta}\left[\frac{1}{\langle k \rangle} \sum_k \frac{k^2 P(k) \beta \Theta}{\mu + \beta k \Theta}\right]_{\Theta=0} \geq 1$$

leading to the condition:

$$\frac{\beta}{\mu \langle k \rangle} \sum_k k^2 P(k) \geq 1 \qquad \rightarrow \qquad \frac{\beta \langle k^2 \rangle}{\mu \langle k \rangle} \geq 1 \tag{5}$$

which is the condition for an endemic state. Since the network has becoming more complex, also the structure for the condition of the endemic state is becoming complex. For the epidemic treshold:

$$\frac{\beta \langle k^2 \rangle}{\mu \langle k \rangle} = 1 \qquad \rightarrow \qquad \beta_c = \frac{\mu \langle k \rangle}{\langle k^2 \rangle} \tag{6}$$

which is pretty similar to the one found before but has a term which increase its complexity.

We have to check if it works also for an homogeneous network: this is the first check that we can make. For an homogeneous network $\langle k^2 \rangle = \langle k \rangle^2$, recovering:

$$\beta_c = \frac{mu \langle k \rangle}{\langle k^2 \rangle} = \frac{\mu}{\langle k \rangle}$$

which is exactly the expression that we saw before, so things are working well.

Recalling what we say last week, in scale-free networks with $2 < \gamma \leq 3$, we have $\langle k \rangle \to c$ and $\langle k^2 \rangle \to \infty$ as $N \to \infty$. As the netowork is becoming larger, also its variance is becoming larger. This means that:

$$\beta_c = \frac{\mu \langle k \rangle}{\langle k^2 \rangle} \to 0$$

so the epidemic threshold vanishes for $N \to \infty$. Obviously, this is a quite important because, if our network is big enough, every disease will spread, no matter its infectivity (see Fig. 1). When we have disease with a very low infection in a little part of population, they do not disappear because we are in a large network! Hence, we have no more an endemic state and the epidemic threshold is really really small for most real epidemic network. This result work as in a thermodynamic limit.

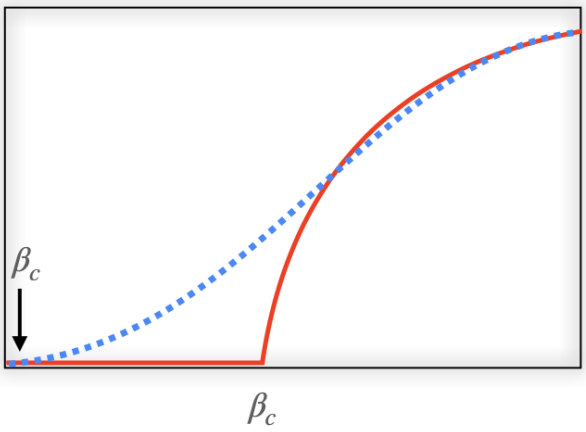

Figure 1: In scale-free networks (and many heavy-tailed distributions) the epidemic threshold vanishes in the thermodynamic limit.

Obviously, this cannot happen in real network. What happens is that we need some finite-size correction. For example, an expression for epidemic thresold when size cannot reach infinity. If we use scale free distribution, at some point since the degree cannot go to infinity, it is convenient to introduce an exponential cut-off.

For instance, let us consider an air trasportation network, we see that until a certain point we have a line, then the curve start to change. We cannot have an infinite number of connection. we have a line and then we will see some sort of exponential degree. The behavior is similar to the one in Fig. 2.

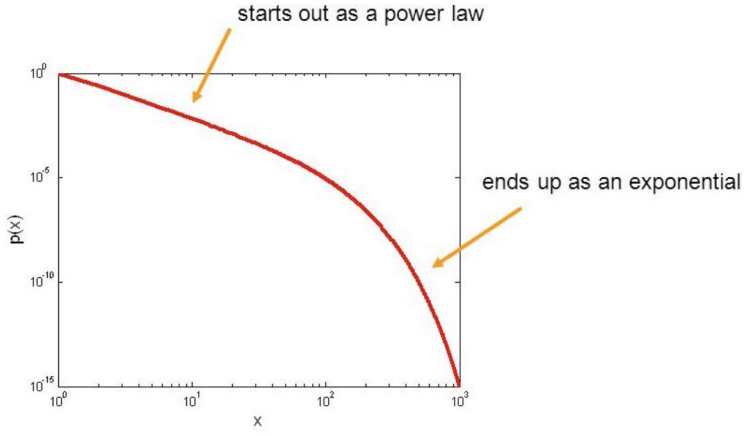

Figure 2: Power-law with an exponential cut-off.

We can model this kind of things by adding an exponential term:

$$P(K) \sim k^{-\gamma} e^{-k/k_c} \tag{7}$$

where k_c is a characteristic degree. At some point the term added will becomes the dominant term. What happens? For large k_c and $2 < \gamma < 3$ the epidemic threshold reads as:

$$\beta_c \simeq \left(\frac{\mu k_c}{k_{min}} \right)^{\gamma - 3} \tag{8}$$

we will not show the calculation. In the lab, we will compare the epidemic treshold in an random network and in a scale-free network to see their differences. This were the results for the SIS model in a network.

0.1 SIR model in a network

0.1.1 Degree-based mean-field theories (DBMF)

We can write the same equations for the SIR model under the same assumption of heteroegeneous mean-field. The difference is that we need one more equation to take into account also the equation for recovered individuals. The densities are $\rho_k^S(t)$, $\rho_k^I(t)$ and $\rho_k^R(t)$ with $\rho_\infty^R = \lim_{t\to\infty} \sum_k P(k)\rho_k^R(t)$. We have that:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\rho_k^I(t) &= -\mu\rho_k^I(t) + \beta k \rho_k^S(t)\Gamma_k(t) \\ \frac{\mathrm{d}}{\mathrm{d}t}\rho_k^R(t) &= \mu\rho_k^I(t) \end{aligned} \tag{9}$$

with $\rho_k^S(t) = 1 - \rho_k^I(t) - \rho_k^R(t)$ and where:

$$\Gamma_k(t) = \sum_{k'} \frac{k'-1}{k'} P(k'|k)\rho_{k'} \tag{10}$$

is the probability of a contact with an infected node and plays exactly the same role of Θ of before. It represents the link from which the infection arrived at the node. We will not show how to obtain its form. It is little different with respect of the SIS model because of the structure of the SIR. There is a neibhobour that can trasmit the infection to me, but it can recover. We need to take into account that the disease is coming from one side, because that side of the network for us is forbidden in the sense that we have recovered that cannot be infected again.

The epidemic treshold for random networks results:

$$\beta_c = \frac{\mu \langle k \rangle}{\langle k^2 \rangle - \langle k \rangle} \tag{11}$$

and the important things is that $\beta_c^{SIS} \neq \beta_c^{SIR}$. This is the first time in the course that the epidemic treshold for these two models is not the same.

0.1.2 Individual-based mean-field theories (IBMF)

Before we were assuming that all the nodes with the same degree were equal. Now, we are gonna studying the individual based mean-field theories for individuals. We are not considering a single network, but an average over all the possible network we can create given that degree distribution. Hence, under the Heterogenous Mean-Field framework we are solving the epidemics problem for an ensemble of networks whose common feature is the $P(k)$: ensemble of networks.

In the degree based approach we assumed that all the nodes with the same degree are equal, so we are not looking at the single network but at their average. This is what in physics is called **annealed networks**. In the opposite, we call **quenched networks** when we consider a particular realization of one network. The idea is: instead of considering the average we consider a particular network. This is the main difference between a degree based or an indivudual based approach.

Let write down the equation for the **quenched mean-field**. We are gonna use the discrete time formulation of equation because it is a bit simpler (obiously, we can also write the same equations by means of differential equations).

Let us consider $\rho_i(t)$ as the probability of a node of being infected at time t. The total fraction of infected is given by $\rho(t) = \sum_i \rho_i(t)$. The probability of being infected at time $t+1$ is:

$$\rho_i(t+1) = \rho_i(t)(1-\mu) + (1-\rho_i(t))q_i(t) \tag{12}$$

which is the sum of the probability of being infected and not get cured (green term) and the probability of being suscpetible times the probability of getting the disease (yellow term). We have that $q_i(t)$ is the probability of node i of getting infected by, at least, one neighbour.

We need an expression for $q_i(t)$. The basic idea is that:

$$q_i(t) = 1 - \prod_{j=1}^{N} [1 - \beta A_{ij}\rho_j(t)] \tag{13}$$

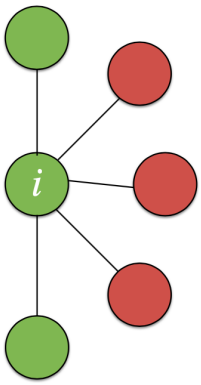

Figure 3: In green suscpetible nodes, while in red the infected neighbours.

Let us consider Fig. 3, in green we have the node i, and in red the infected neighbours. We see that $\beta A_{ij}\rho_j(t)$ is the probability of getting infected by node j (at least). We have that $[1-\beta A_{ij}\rho_j(t)]$ is the probability of NOT getting infected by node j. I have to repeat this calculation for all my infected nehbours, hence $\prod_{j=1}^{N}[1-\beta A_{ij}\rho_j(t)]$ is the probability of NOT getting infected by any neighbor. Finally, $q_i(t)$ represents the probability of getting infected by at least one neighbor. Hence, the probability of getting infected is one minus the probability of not getting infected by any neighbor.

The system $(\rho_i(t+1))$ can be solved numerically by iteration. This is precise for the entire epidemic diagram, faster than numerical simulations, there is no need of averages and reproduces individual nodes probabilities. Indeed, in this case we are gonna have an equation for each node: we have 2^N equation where N is the size of the system.

Remark. The difference with the degree based mean field theories is that actually A_{ij} we are including all the adjiacency matrix, while before only the average.

We can also solve analytically the system at the steady state to estimate the epidemic threshold. Assuming the steady state:

$$\lim_{t\to\infty} \rho_i(t) = \rho_i^* \quad \to \quad \rho_i(t+1) = \rho_i(t) = \rho_i^*$$

Hence:

$$\mu\rho_i^* = (1-\rho_i^*)q_i^* \quad \to \quad q_i^* = 1 - \prod_{j=1}^{N} \left[1 - \beta A_{ij}\rho_j^*\right] \tag{14}$$

Assuming being close to the epidemic threshold (epidemic onset) means that ρ_i^* can be assumed to be small for all the nodes $\rho_i^* = \varepsilon_i^* \ll 1$. This implies that the product in q_i^* can be approximated by a sum:

$$q_i^* = 1 - \prod_{j=1}^{N} \left[1 - \beta A_{ij}\varepsilon_j^*\right] \simeq \beta \sum_{j=1}^{N} A_{ij}\varepsilon_j^* \tag{15}$$

Substituting we have:

$$\mu\varepsilon_i^* = \beta(1-\varepsilon_i^*)\sum_{j=1}^{N} A_{ij}\varepsilon_j^* \tag{16}$$

We have a linear system where the interaction is represented by the adjiacency matrix:

$$\mu\varepsilon_i^* = \beta\sum_{j=1}^{N} A_{ij}\varepsilon_j^* - \beta\varepsilon_i^*\cancel{\sum_{j=1}^{N} A_{ij}\varepsilon_j^*}$$

Neglecting second order terms we have:

$$\frac{\mu}{\beta}\varepsilon_i^* = \sum_{j=1}^{N} A_{ij}\varepsilon_j^* \tag{17}$$

This linear system has solution only if $\frac{\mu}{\beta}$ is an eigenvalue of the adjiacency matrix A_{ij}. This is why last week we say that the spectrum of the adjiacency matrix is important. Hence:

$$\beta = \frac{\mu}{\Lambda_i} \tag{18}$$

where Λ_i is an eigenvalue of the adjiacency matrix A_{ij}.

Since, we are interested in the smallest possible value of β for which we have a solution, we take the largest eigenvalue of the adjiacency matrix A:

$$\beta_c = \frac{\mu}{\Lambda_{max}} \tag{19}$$

This is the expression for the epidemic treshold, and it is a general results that it is valid not only by this approxmation but for a generic case for a generic network.

0.1.3 DBMF vs IBMF: Epidemic treshold

What is the relation between the epidemic treshold for DBMF and IBMF? We have that:

- in DBMF:
$$\beta_c^{DBMF} = \frac{\mu\langle k\rangle}{\langle k^*\rangle}$$
- in IBMF:
$$\beta_c^{IBMF} = \frac{\mu}{\Lambda_{max}}$$

For scale-free networks $P(k)\sim k^{-\gamma}$ we have that:

$$\Lambda_{max} \sim \max(\sqrt{k_{\max}}, \frac{\langle k^*\rangle}{\langle k\rangle}) \tag{20}$$

Specifically:

$$\beta_c \sim \begin{cases} \mu/\sqrt{k_{\max}} & \gamma > 5/2 \\ \mu\langle k\rangle / \langle k^2\rangle & 2 < \gamma < 5/2 \end{cases} \tag{21}$$

We can conclude that IBMF is more accurate than DBMF. Due to the approximation, the DBMF is accurate only in the proximity of the epidemic threshold while IBMF is accurate for the entire epidemic diagram.