0.0.1 Heterogeneous Networks

Now, we want to understand what is the effect of **heterogeneity** in the spread of the disease. That is to say that we drop the following assumption $k_i \sim \langle k \rangle$: all **nodes are not equal** any more.

Let us consider now the **heteroegeneous mean-field approximation**. Let us use a **DBMF model** and let us cut the chain at an **individual level**. This last assumption means that we consider the probability for a single individual to get the infection. Let us follow the thread of paper *"Epidemic Spreading in Scale-Free Networks"*, written by Pastor-Satorras and Vespignani. It actually provides a **SIS model on scale-free networks**. The main idea behind this paper is the following. Since nodes are not equal anymore, *the probability of getting the infection strongly depends on their position (i.e. degree) in the network.* Authors' intuition is that **nodes** with **the same degree behave in the same way**. In order to do that, we need to divide the network in **degree classes**: that is to say we group together all the nodes with the same degree.

In order to write down the equations, we need to consider the number of compartments we have and introduce a density for each of them:

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

where s_k and ρ_k are the fractions of susceptible and infected nodes of degree k in the network. We have that N_k represents the number of nodes with degree k. As before, we introduced the fractions of susceptible and infected individuals (s_k, ρ_k) in the system, but in this case depending on each degree k. Obviously, the total fraction of ρ and s in the system is given by the sums:

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

The **equation** that describes how the **probability of being infected** changes in time for the nodes that belong to the **same 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 can distinct as usual a "recovery" term and an "infection" term. In particular, the probability of a contact between a susceptible individual that has degree k and an infected one is highlighted in green. This product consists in two terms: the probability of being infected $(1 - \rho_k(t))$ and the probability of having contact with an infected $\Theta_k(t)$.

We want now to dwell deeper and explain better this last term. The probability that a generic node with degree 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 possible degree classes k'. In this way we expect to obtain the probability of connecting with any one of them, multiplied by the probability for that specific node to be infected. Note, however, that we are making no assumption about the function $P(k'|k)$, which may change according k. In principle, it could be anything, in the sense that it strongly depends on the structure of the network.

However, in order to simplify the problem and derive some results, there are cases where we can make some assumptions on the structure of the latter.

For **random networks**, e.g. picking a node at random, the probability to be connected to a node of degree k' given the node degree we start from is k, is the following:

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

where we simply applied the definition of conditional probability. Note that $P(k')$ is the generic probability of getting a connection at random, times k', which is the number of connection that we pick up (namely the degree k). Finally we normalize over all possible degrees of the network [1]. What we obtain is the probability that a generic node in the network is linked to k'. Note as $P(k'|k)$ does not depend on k.

After replacing this last result in 3:

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

Let us take a look closer to the different terms. In the numerator: there is the product between the probability that a link, randomly picked, points to k', times probability of being infected, Finally, we then we sum over all the possible degrees. While, expression on denominator is related only to the structure of the network. In addition, one should note that $\Theta_k(t)$ **does not depend on** k anymore. Since we are just picking up at random it should be the same for all the nodes.

The method that we are going to exploit to **solve** the differential equation $\frac{\mathrm{d}}{\mathrm{d}t}\rho_k(t)$ is similar to the ones previously used in other models. The first assumption is to be 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 then to substitute the expression for ρ_k, obtained thanks to 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.

However, in order to solve this last equation, we need some workaround. First of all one should note, as what happens in statistical mechanics, this expression has different solutions depending on the value of Θ:

- the **trivial solution** $\Theta = 0$, that of course is not in our interest;

- the **non trivial solution**. We can rewrite the self consistent equation as follows:
$$\Theta = \frac{1}{\langle k \rangle}\sum_k \frac{k^2 P(k)\beta\Theta}{\mu + \beta k \Theta} = f(\Theta)$$

  Hence, the solutions are the values for which it holds $\Theta \equiv f(\Theta)$. These, geometrically, are the interceptions between the line Θ and the function $f(\Theta)$ and have to be found graphically (or using computational algorithms).

  Since Θ is a probability, it holds that $0 < \Theta \leq 1$. This means that, it is required for a non trivial solution to exist, the slope of $f(\Theta)$ must be greater than 1. Mathematically, it means that:
$$\frac{\mathrm{d}}{\mathrm{d}\Theta}\left[\frac{1}{\langle k \rangle}\sum_k \frac{k^2 P(k)\beta\Theta}{\mu + \beta k \Theta}\right]_{\Theta=0} \geq 1$$

---

[1] One should keep in mind that $\sum_k kP(k) = \sum_{k'} k'P(k')$.

that leads to the following 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 the **existence** of an **endemic state**. Since the network has become more complex, also the structure for the condition of the endemic state acquires in complexity. Indeed, for the epidemic threshold:

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

which is pretty similar to the one previously found, but also includes a term that increases its complexity.

The first check one can make is to verify whether this last result holds also in the case of homogeneous networks. For such networks $\langle k^2 \rangle = \langle k \rangle^2$, therefore:

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

which is exactly the expression we previously found.

Recalling what we were discussing last lectures, 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 network becomes larger also its variance increases, that is:

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

hence the **epidemic threshold vanishes** for $N \to \infty$. This is a quite important result because,**if** our **network is big enough**, **every disease will spread, no matter its infectivity** (see Fig. 1). The converse is still valid: if we have disease with a very low infection rate in a small part of the network, it will not disappear if the network is large enough[2]! That is to say we **always** find ourselves in an **endemic state**, while the threshold becomes very small. These results are actually valid for the most real epidemic models, given the networks are large enough.
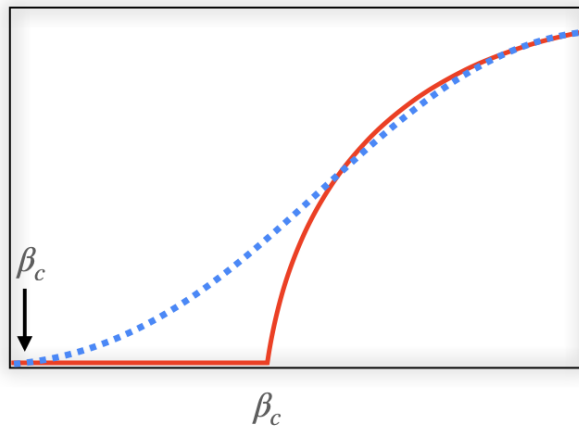

Figure 1: In scale-free networks (and many heavy-tailed distributions) the epidemic threshold vanishes in the thermodynamic limit.

---

[2]Physically, we refer to this as taking the thermodynamic limit.

Obviously, **real networks** are not infinite: therefore we need some **finite-size corrections**. For example, we may want to derive an expression for epidemic threshold when the size of the system does not diverge.

Let us consider the degree distribution for **scale-free networks**: since the degree cannot go to infinity, it is convenient to introduce an **exponential cut-off** at some point. For instance, let us consider the air transportation network: we see that until a certain point a certain trend is followed, but then the slope of the curve starts to change and resembles to an exponential. This implies that we cannot have an infinite number of connections: the line starts out as a power law and then ends up introducing some sort of exponential cut-off. The behavior is similar to the one in Fig. 2.
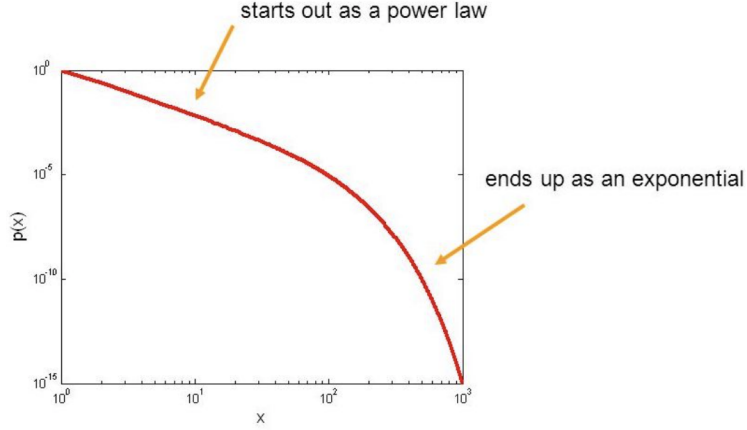
Figure 2: Power-law with an exponential cut-off.

We introduce our considerations into our model 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 we just added will become the dominant term and what happens is that, for large k_c and $2 < \gamma < 3$, the epidemic threshold can be approximated as:

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

we are not going to prove the computations. However, in the lab, we will compare the epidemic thresholds for a random and for a scale-free networks in order to see how they differ. This was the last consideration about the study of the SIS model in a network.

0.1 SIR model in a network

0.1.1 Degree-based mean-field theories (DBMF)

The same as before equations can be derived for the SIR model under the assumption of **heterogeneous mean-field**. The main difference is that we need **one more equation** to take into account also the compartment related to **recovered** individuals. Their densities are $\rho_k^S(t)$, $\rho_k^I(t)$ and $\rho_k^R(t)$, and it holds that $\rho_\infty^R = \lim_{t \to \infty} \sum_k P(k) \rho_k^R(t)$. Equations take the form:

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

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 before. Actually it represents the link from which the infection arrived to that node, however we will not show how to derive this expression beside one small consideration: the $\frac{k'-1}{k}$ term that is the main difference from the SIS model. It is present due to the fact that we cannot infect a node that has already transmitted us the disease: either because it has already recovered or because it is still infected. In this way we are taking into account that the disease is coming "from one side", therefore for us is forbidden to spread the infection towards that specific direction: recovered (or already infected) individuals cannot be infected twice.

The **epidemic threshold for random networks** results:

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

and the important thing to notice is that $\beta_c^{SIS} \neq \beta_c^{SIR}$. This is the first time so far that the **epidemic thresholds** for these two models **differ**!

0.1.2 Individual-based mean-field theories (IBMF)

Up to now we were assuming that all the nodes with the same degree were equal. Now, since we are going to study the **individual based mean-field** theories, we will not consider a specific instance of the network, but an average over all the possible networks we can obtain **given** that **degree distribution**. That is to say, that under the **Heterogenous Mean-Field framework** we are solving the epidemics problem for an **ensemble of networks** whose common feature is the degree distribution $P(k)$[3].

In the degree based approach we previously assumed that all the nodes with the same degree to be equal. We were therefore analyzing not a specific instance of networks, but its *average*. This is actually what in physics we refer as **annealed networks**. On the opposite, we call **quenched networks** when we consider a *particular realization* of one network. The idea is really simple: instead of considering the average, we consider a particular instance network. This is the main difference between a degree based (i.e. annealed networks) or an individual based approach (i.e. quenched networks).

Let us write down the equations for the **quenched mean-field**. We are going to introduce a **discrete time** framework in order to make equation simpler. However, nothing prevents us to use differential equations, where time is a continuos variable.

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

At the following time-step, 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 susceptible multiplied by the probability of contracting the disease (yellow term).

We now need an expression for $q_i(t)$, that is the **probability** for node i to **be infected**

---

[3]the so called "ensemble" of networks!

by, at least, one neighbour. The basic idea for doing this is:

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

Let us consider Fig. 3, in green we have susceptible nodes, which include node i itself, and in red its infected neighbours. The probability of getting infected, at least, by a generic node j is:

$$\beta A_{ij}\rho_j(t)$$

Its complementary to 1 is the probability of *NOT* get the infection by node j.

$$[1 - \beta A_{ij}\rho_j(t)]$$

Repeating this argument for all neighbors that are actually infected, we can obtain the probability of *NOT* contracting the disease from *ANY* neighbor, namely:

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

Again, we previously introduced $q_i(t)$ as the probability of getting infected by at least one neighbor. Hence, the probability of getting infected the complementary to one probability of not getting infected by any neighbor:

$$q_i(t) = 1 - \prod_{j=1}^{N} [1 - \beta A_{ij}\rho_j(t)] \tag{14}$$
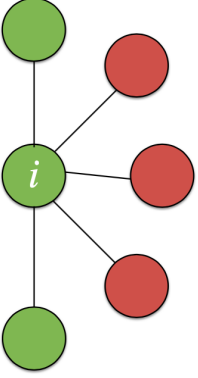
Figure 3: In green suscpetible nodes, while in red the infected neighbours.

Note as the system of $(\rho_i(t+1))$ equations can be solved numerically by iteration. This results to be precise for the entire epidemic diagram, and faster than numerical simulations: there is no need of averages and reproduces individual nodes probabilities. Indeed, in this framework we will obtain two equations for each of the nodes: we have 2^N equations, where N is the size of the system.

Remark. One should have noted that this last approach differs from the degree based mean field theories by the fact that now we are including adjacency matrix A_{ij}, while before we took only the average.

We can also **solve analytically** the system at the **steady state** in order to estimate the **epidemic threshold**. Assuming that we find ourselves in the steady state:

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

it follows that:

$$\mu\rho_i^* = (1 - \rho_i^*)q_i^* \quad \rightarrow \quad q_i^* = 1 - \prod_{j=1}^{N} \left[1 - \beta A_{ij}\rho_j^*\right] \tag{15}$$

Now, if we think about what happens when we are in **proximity of the epidemic threshold** (*epidemic onset*), it happens that ρ_i^* can be assumed to be small for all the nodes $\rho_i^* = \varepsilon_i^* \ll 1$. Therefore, the product in q_i^* can be approximated by a sum:

$$q_i^* = 1 - \prod_{j=1}^{N} \left[1 - \beta A_{ij}\varepsilon_j^*\right] \simeq \beta \sum_{j=1}^{N} A_{ij}\varepsilon_j^* \tag{16}$$

Substituting what we have just found in the lhs of 15 we obtain:

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

that is a linear system where the interaction is given by the adjacency 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 that:

$$\frac{\mu}{\beta}\varepsilon_i^* = \sum_{j=1}^{N} A_{ij}\varepsilon_j^* \tag{18}$$

This linear system has solution only if $\frac{\mu}{\beta}$ is an **eigenvalue** of the **adjacency matrix** A_{ij}. Here we should understand why last lecture we stated that the spectrum of the adjacency matrix is something we may be interested in. Hence:

$$\beta = \frac{\mu}{\Lambda_i} \tag{19}$$

where Λ_i is a generic eigenvalue of the adjacency matrix A_{ij}. However, since we are interested in the **smallest** possible **value** of β for which there exists solution, we need to take the **largest eigenvalue** of the adjacency matrix A:

$$\beta_c = \frac{\mu}{\Lambda_{max}} \tag{20}$$

The last one is the **expression** for the **epidemic threshold**, and it is a **general result** that is valid not only while using this approximation, but for a more general framework in a generic network.

0.1.3 DBMF vs IBMF: Epidemic treshold

One may wonder now what is the relation between the two values for the epidemic thresholds we have found for the different mean-field theories, that is DBMF and IBMF. We have found that:

- for **DBMF**:
$$\beta_c^{DBMF} = \frac{\mu\langle k\rangle}{\langle k^*\rangle}$$

- for **IBMF**:
$$\beta_c^{IBMF} = \frac{\mu}{\Lambda_{max}}$$

For **scale-free networks** $P(k) \sim k^{-\gamma}$ it holds that:

$$\Lambda_{max} \sim \max\left(\sqrt{k_{\max}}, \frac{\langle k^*\rangle}{\langle k\rangle}\right) \tag{21}$$

And in particular:

$$\beta_c \sim \begin{cases} \mu/\sqrt{k_{\max}} & \gamma > 5/2 \\ \mu\langle k\rangle/\langle k^2\rangle & 2 < \gamma < 5/2 \end{cases} \tag{22}$$

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