LECTURE NOTES OF LIFE DATA EPIDEMIOLOGY

Collection of the lectures notes of professors Chiara Polletto and Sandro Meloni.

Edited By

Alice Pagano

The University of Padua

Contents

Part I

Meloni's Lectures

1

Basics Definitions and Compartmental Models

Some models are wrong, but most of them are useful.

– Unknown author

Lecture 2.
Friday 2nd October, 2020.
Compiled: Tuesday 27th October, 2020.

Models in science have two different roles: **understanding** what happens and **prediction**. There are two types of models: one more simple and one more complex. In the simplest one you just consider the minimal number of parameters and events involved: this allow to understand what are the main mechanism of a phenomena.

In this course, we gonna start with very simple models in which we firstly assume that there are no structures in the population. This is not accurate, but allows us to understand at a first glance some underlying mechanism. Then, we will consider social structures and contact network models. We will also take into account the interaction of different populations and we analyze how data move from one population to the other. At last, there is another class of models called "Agent Based" models, but we will only brief introduct it.

1.1 Comportamental models

We introduce the **comportamental models** in which most based epidemiologist theories are based on. In reality, there are different levels of understanding how the diseases diffuse, as for instance at a biological level or a more simplest one. It is impossible to insert all the details of a process in a models. We need to summarize all the biological process in few parameters which are the average of what you see inside the population. This is the same principle behind the statistical mechanics in which we concentrate for large scale effects.

Let us consider a population of any individuals and try to characterize them. For

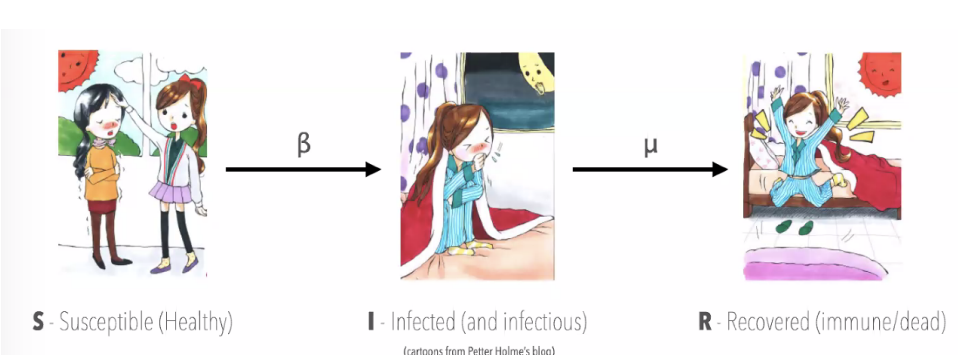

Figure 1.1: Classification of infected population in three different stage of the disease.

instance we have three compartments and we want to classify people according to the state of their diseases, as in Fig. 1.1. We have also transition from one state to the other with rates as β and μ which describes the dynamic of the diseases The only problems it is that this approximation is quite strong, because by fixing the rates we are assuming that the process underlying is a Markovian one. In reality, you do not have a exponential distribution but a decay. This will be seen in the course when we will talk about "non-Markovian" epidemics. For instance the β can be the "per contact" infectous rate, so you are just counting the number of contacts. As examples we can consider different models which depend on the type of disease as SI, SIS, SIR and so on and so forth.

There are difference between medical status and infection status. We do not care about medical status but only about the disease and how the immuning system is reacting. We have four main stages of the disease: starting from a helthy state, the individual can contract the disease and then it become infectious until it recover (Fig. 1.2). The most important things is that the compartments are not the same of the medical status.

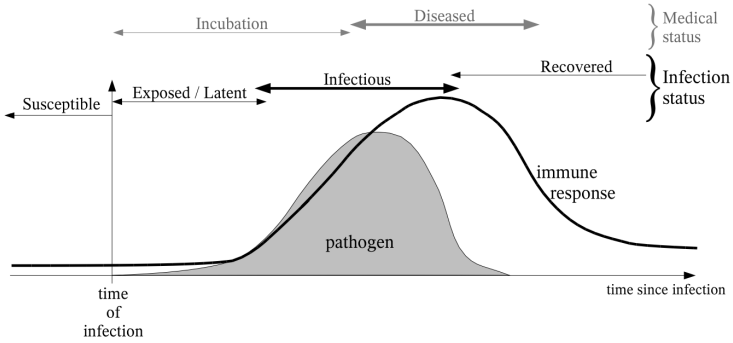

Figure 1.2: A caracature of the time-line of infection, showing the dynamics of the pathogen (gray area) and the host immune response (black line) as well as labeling the various infection classes: susceptible, exposed, infectious, and recovered. Note that the diseased period, when symptoms are experienced, is not necessarily correlated with any particular infection class.

Now, let us introduce the **Basic Reproductive Number** R_0 (pronunced R naught) which is a measure of the infection of the population. We put one guy inside a group for instance for three days and at the end of these days we count the number of secondary cases that we have. This is the main idea of R_0. This number determines wheter a disease will spread or not:

$$\begin{cases} R_0 < 1 \\ R_0 = 1 \\ R_0 > 1 \end{cases} \tag{1.1}$$

Let us consider the plot of Fig. 1.3, we have a sort of second order phase transition at the point $R_0 = 1$. Note that the R_0 number for the SARS is higher thant the one of COVID-19, but we did not experienced an outbreak of this disease. For computing R_0 we are assuming that the population is totally susceptible, but this assumption is valid only at the very early stages. Then, we care ombining also epidemiological and demographical features. The conclusion is that this number could vary from one population to another. Since we are doing a coarse-graining of the dynamics, the number R_0 represents the mean of possibly different distributions. It can give the wrong idea of similar R_0, means similar outbreaks. The distribution of infections can be quite heterogeneous, hence the mean could be quite representive in homogeneous populations. For instance, let us consider the plot in Fig. 1.4. We see that SARS is

etherogeneous, while Spanish Flu was a homogeneous one. The COVID-19 is most likely somewhere in the middle.

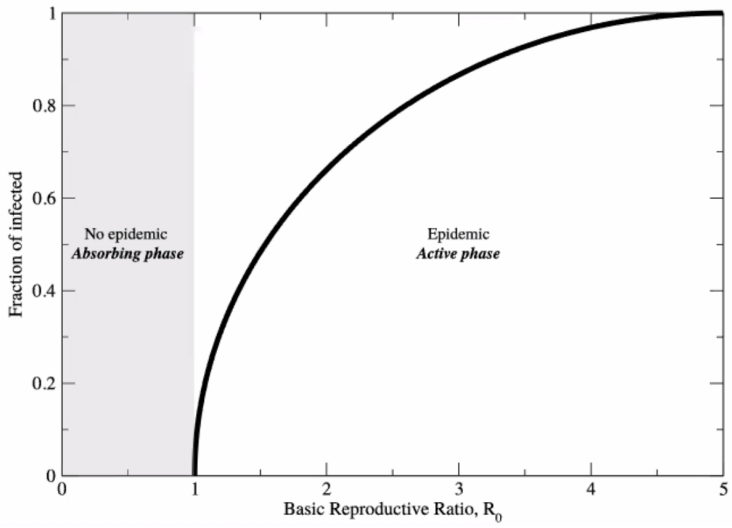

Figure 1.3: Fraction of infected vs basic Reproductive Ratio, R_0.

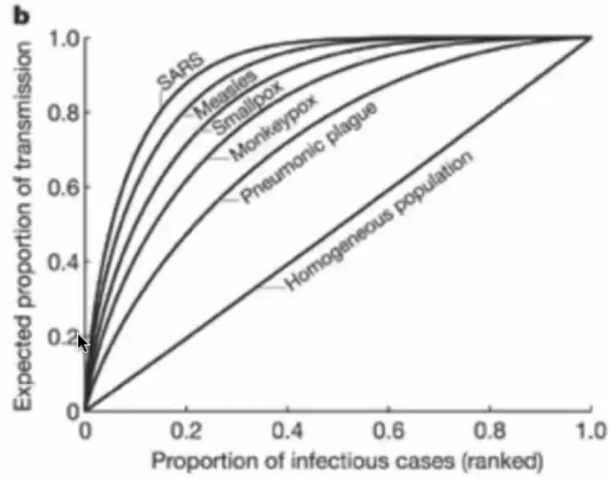

Figure 1.4: Figure from: Lloyd-Smith et al. Nature 438, 355–359 (2005).

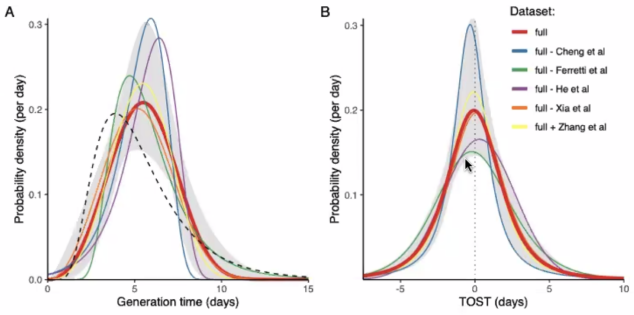

Figure 1.5: Figure from: Ferretti et al. `https://www.medrxiv.org/content/10.1101/2020.09.04.20188516v1`

Now, we introduce the **Effective Reproductive Number** $R(t)$, which is the same of R_0 but it varies in time. Hence, it is the average number of secondary

cases a case produces in a population at time t. Another important quantity is the **Infectious period**:

$$\tau = \frac{1}{\mu}, \quad \tau = \frac{1}{(\alpha + \mu)} \tag{1.2}$$

Other relevant quantities are the **Incubation period** (from infection to symptoms), the **Generation time** (from infection to infection), the **Serial interval** (from symptoms to symptoms) or the **TOST** (measures from symptoms to infection). The problems is that TOST in many cases can be negative (see Fig. 1.5 for more details).

Lecture 3.
Thursday 8^{th} October, 2020.
Compiled: Tuesday 27^{th} October, 2020.

1.2 Basic models

We are gonna introduce some of the basic models we will deal for the entire course. We are assuming that we are in **well-mixed populations**, or homogeneous mixing. Mathematically, it is what is called mean field approximation. In the well-mixed population assumptions, we are assuming that:

- all individuals are equivalent, hence every one has the same probability of getting the infection;
- every individual has the same number of contacts $N - 1$, or on average $\langle k \rangle$;
- another important assumption is that we are in a closed population. Hence, the sum of the density distribution of the individuals is 1 and we have no deaths or births. In practice, we are assuming that our time scale is so little that we can consider the population constant.

1.2.1 SI model

This simple model is the **SI** (susceptible infected). You can get the infection and once you get it you cannot recover (you stay infected forever). The transition is:

$$S + I \xrightarrow{\beta} I + I \tag{1.3}$$

The β is the "per contact" infection rate and dictates the speed of the spreading. We can write down the equation and solve it deterministically:

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= -\beta \langle k \rangle si \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \beta \langle k \rangle si \end{aligned} \tag{1.4}$$

where $\langle k \rangle$ represents the contacts, the i means the fraction of infected in the population ($i = I/N$), while s the fraction of susceptible in the population ($s = S/N$). The product si is the probability of contacts, and βsi is the proabability of having one more infected.

To solve it analitically, we should remember that our population is closed hence $s + i = 1$, we only have one equation with $s = 1 - i$. We have that:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta i(1 - i) \quad \rightarrow \frac{1}{\beta i(1 - i)} \mathrm{d}i = \mathrm{d}t \quad \rightarrow \frac{1}{\beta(1 - i)} \mathrm{d}i + \frac{1}{\beta i} \mathrm{d}i = \mathrm{d}t$$

Integrating both sides:

$$-\log|1 - i| + \log|i| = \beta(t + C) \rightarrow \frac{i}{1 - i} = e^{\beta(t+C)} = Ae^{\beta t}$$

with $A = i_0/(1 - i_0)$. The result is:

$$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}} \tag{1.5}$$

which is a sigmoid function (Fig. 1.6) which always saturates at 1. We have the first part where we have the exponential growth (which is the one we have seen in the media for covid-19), then at a certain point you are slowing down. The reason of slowing down is because of the term si, the probability of funding new supsceptible is going down. Finally, you reach 1 after a very long time. The value of β is the one which drives the spreading. Increasing β we have a faster exponential growth. This was the simplest model.

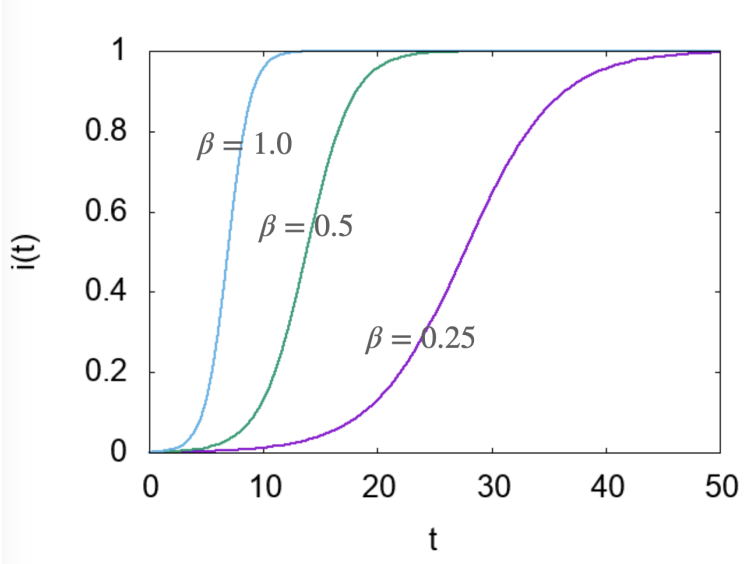

Figure 1.6: Plot of the solution of the SI model for different β.

Remark. In the course we are gonna use capital letter for integer numbers and small lecter for densities.

1.2.2 SIS model

Now, let us go to the **SIS** model. This model starts to be more complicated. We have two different transitions:

$$\begin{aligned} S + I &\xrightarrow{\beta} I + I \\ I &\xrightarrow{\mu} S \end{aligned} \tag{1.6}$$

whose second one is spontaneous. This model is used for diseases that do not confer immunity. **Endemic state** means that the disease circulates in the population for very large times. The important things is that it is the simplest models in which a dynamical equilibrium can be reached. An individual could recover after the disease but it do not get immunity, indeed there are always people infected that can propagate the disease. The μ is the recovery rate which determines the time-scale of the infection. Dividing β by μ you can rescale all the dynamics. The equations are exactly the same of before except for a term:

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= -\beta \langle k \rangle si + \mu i \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \beta \langle k \rangle si - \mu i \end{aligned} \tag{1.7}$$

and you can solve them in the way of before. Also the shape of the solution is exactly the same:

$$i(t) = i_0 \frac{(\beta - \mu)e^{(\beta-\mu)t}}{\beta - \mu + \beta i_0 e^{(\beta-\mu)t}} \tag{1.8}$$

If we plot it we have the same form but with the difference that we are not reaching one, but $\frac{\beta-\mu}{\beta}$. Hence, as said, we have some sort of dynamical equilibrium: the new infected are the same of the new recovery that you are getting. The population will fluctuate around this value $\frac{\beta-\mu}{\beta}$ and enlarging μ will give to larging fluctuations (Fig. 1.7).

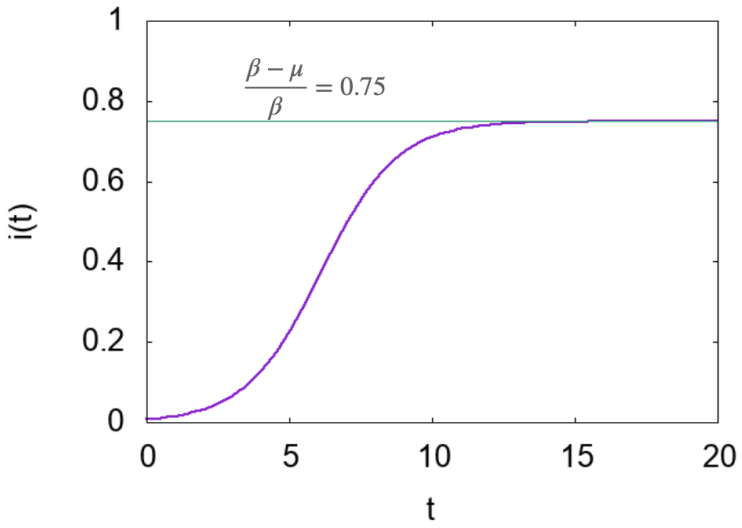

Figure 1.7: Plot of the solution of the SIS model.

It could be more instructive to study what happens at the beginning for this model. At the beginning I can assume that almost my population is composed by my susceptible ($s \sim 1$) and the number of infected is very little ($i \ll 1$). Hence, we can rewrite the differential equations as:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta \langle k \rangle si - \mu i \sim \beta \langle k \rangle i - \mu i \rightarrow i(t) \sim i_0 e^{(\beta\langle k \rangle - \mu)t}$$

We have that if $\beta \langle k \rangle < \mu$ I have not spreading at this point, while if $\beta \langle k \rangle > \mu$ the exponential becomes positive and I have the exponential growing at the beginning (Fig. 1.8).

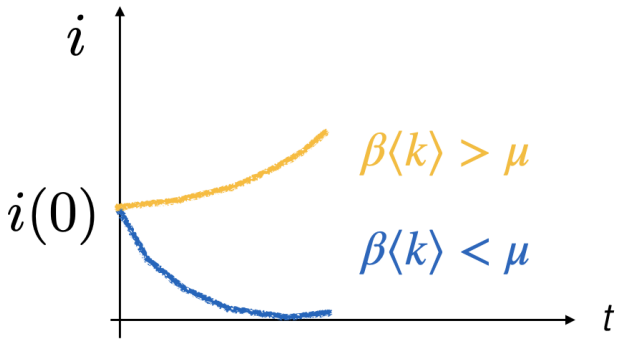

Figure 1.8: Initial transient for the SIS model.

The very important thing is that if I consider what it is happening I have two

choices for the steady state:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = 0 \rightarrow \begin{cases} i = 0 & \beta \langle k \rangle < \mu \\ i > 0 & \beta \langle k \rangle > \mu \end{cases}$$

and we have that:

$$i > 0 \iff \beta > \beta_c = \frac{\mu}{\langle k \rangle} \tag{1.9}$$

which is the **epidemic threshold**. This is telling you if the disease is gonne spread. The epidemic threshold is the minimum value of the infection probability for which the disease survives. This is what in physics is called a second order phase transition (Fig. 1.9). In this case the critical exponents are the same of the Ising model (they are in the same class of universality). This is one of the most important quantities we are gonna study.

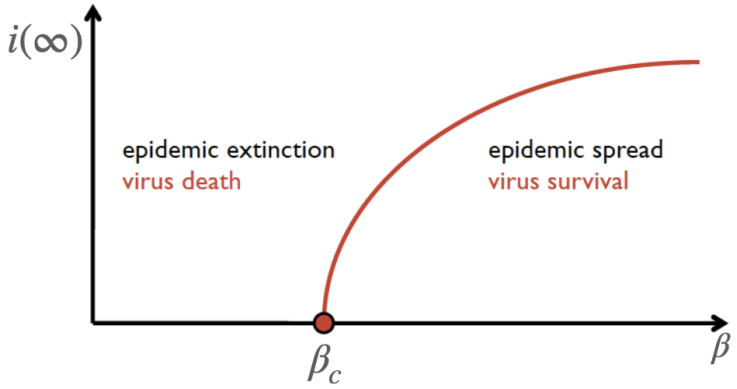

Figure 1.9: Epidemic diagram.

What is the relation between R_0 and the epidemic threshold? Obviously, they are strongly correlated. We are saying that we have a critical value and below it we have no spreading, while above we have a fraction of infected people. The epidemic threshold is given you the condition under which you have the spreading. Mathematically, giving a speficic model, its critical version is giving the value for which $R_0 = 1$, which means that if you are above the threshold you need a minimum of infected people which is 1. In the case of the SIS model:

$$R_0 = \frac{\beta \langle k \rangle}{\mu} = 1 \tag{1.10}$$

1.2.3 SIR model

The idea is the same of the SIS, but we are adding a new state which accounts for long lasting immunity. Hence, once you got the disease you can have a long immunity. However, the density of the population is still fixed to 1. The transitions of this model are:

$$\begin{aligned} S + I &\xrightarrow{\beta} I + I \\ I &\xrightarrow{\mu} R \end{aligned} \tag{1.11}$$

and we have no endemic state. The differential equations are:

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= -\beta \langle k \rangle si \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \overbrace{\beta \langle k \rangle si}^{\text{New infections}} - \overbrace{\mu i}^{\text{Recovery}} \\ \frac{\mathrm{d}r}{\mathrm{d}t} &= \mu i \end{aligned} \tag{1.12}$$

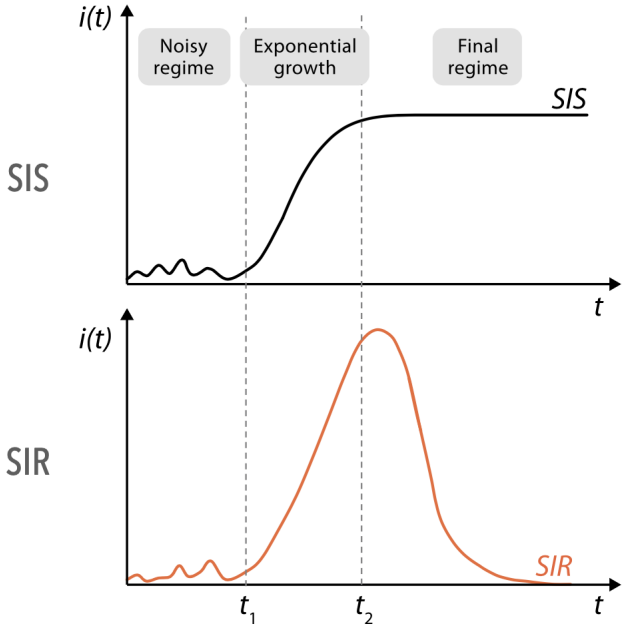

Figure 1.10: Epidemic regimes.

It is a good point to introduce the different regimes that you have during a spreading and which are represented in Fig. 1.10 for the SIS and SIR models. Initially, at the beginning of each spreading, you have the **noisy phases** where the numbers are too small to cause a spreading, hence you have a sort of stocastic fluctuations. In most of the cases, you can end up without infected. If you stop a guy in this noisy phase, you are able to stop the disease (if it is hetherogeneous). Then, we have the **exponential growth**. Then, the disease is slowing down. Finally, you reach the steady state for the SIS (endemic state), while for the SIR the disease disappear (absorbing state).

To calculate the epidemic threshold in the case of the SIR the calculations are the same of before. In particular, we assume that $r \ll 1$ so that:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta \langle k \rangle si - \mu i \sim \beta \langle k \rangle i - \mu i \rightarrow i(t) \sim i_0 e^{(\beta \langle k \rangle - \mu)t}$$

and the result is again:

$$\beta > \beta_c = \frac{\mu}{\langle k \rangle} \tag{1.13}$$

Since we can get an analitic expression for S and I in this SIR model, we want to study what is the behavior at the end for $t = \infty$. We get that:

$$\frac{\mathrm{d}s}{\mathrm{d}r} = \frac{-\beta \langle k \rangle s}{\mu}$$

If we assume $r_0 = 0$ and we integrate the above expression with respect to r, we obtain:

$$s(t) = s_0 e^{-r(t)\frac{\beta\langle k\rangle}{\mu}}$$

As said, we cannot solve this equation directly, but we can study the behavior in the long term. At $t = \infty$, we have that $i(\infty) = 0$ and thus $s(\infty) = 1 - r(\infty)$:

$$1 - r(\infty) - s_0 e^{-r(\infty)\overbrace{\frac{\beta\langle k\rangle}{\mu}}^{R_0}} = 0$$

This is a transcendental equation which cannot be solved analytically but it gives important hints on the behavior of the disease. Note that $R_0 = \beta\langle k\rangle/\mu$ and this explains why R_0 drives the exponential growth of the disease. Moreover, we note that the initial fraction s_0 of susceptible plays a role in shaping the final fraction of recovered. In particular, if $s_0 \ll 1$ the disease cannot spread. We can obtain herd immunity.

1.3 Extensions of the SIR model

We want to modify the SIR to include something that we want in our model.

1.3.1 SIR with Demography

In the last models, we were assuming that the population was totally closed and so it always sum up to 1. This is one thing that we want to remove because it is unrealistic. We will assume that there could be births and deaths. If we consider the demography, we see that every year there are new child that are infected by disease as Measles and Chickenpox. We expect that usually they die out over weeks.

Hence, the simplest assumption is: similar to the infectious period, individuals have a lifespan $1/\alpha$ years (where lifespan is much greater than the infectious period, deaths are not due to the disease). We are assuming that α is the death rate in all the classes. Moreover, α is also the crude birth rate and we assume that births happens only in suscpetible individuals.

To have a constant population, we need to assume:

$$\frac{\mathrm{d}s}{\mathrm{d}t} + \frac{\mathrm{d}i}{\mathrm{d}t} + \frac{\mathrm{d}r}{\mathrm{d}t} = 0 \tag{1.14}$$

Our equations become:

$$\begin{aligned}
\frac{\mathrm{d}s}{\mathrm{d}t} &= \alpha - \beta si - \alpha s \\
\frac{\mathrm{d}i}{\mathrm{d}t} &= \beta si - \mu i - \alpha i \\
\frac{\mathrm{d}r}{\mathrm{d}t} &= \mu i - \alpha r
\end{aligned} \tag{1.15}$$

The infectious period is:

$$\tau = \frac{1}{\alpha + \mu} \tag{1.16}$$

on average, individuals spend less infected because some of them die while infected. However, it is a small change since lifespan is much greater than the infectious period.

Also R_0 is reduced due to mortality:

$$R_0 = \frac{\beta}{\alpha + \mu} \tag{1.17}$$

We want to study the equilibrium points of the dynamic. Assuming:

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \frac{\mathrm{d}i}{\mathrm{d}t} = \frac{\mathrm{d}r}{\mathrm{d}t} = 0$$

we want to find the equilibrium values s^*, i^* and r^*. We have that:

$$\frac{\mathrm{d}i}{\mathrm{d}t} = 0 = \beta si - \mu i - \alpha i \quad \rightarrow \beta s^* i^* - (\mu + \alpha) i^* = 0$$

and finally we obtain the equation:

$$i^*[\beta s^* - (\mu + \alpha)] = 0 \tag{1.18}$$

which is not differential anymore. The two solutions are $i^* = 0$ (**disease free state**) or $s^* = \frac{\alpha+\mu}{\beta} = \frac{1}{R_0}$ which is the **endemic state**. Hence, the important result is that the SIR model with demography shows an endemic state.

Substituting $s^* = \frac{1}{R_0}$ in $\frac{\mathrm{d}s}{\mathrm{d}t} = \alpha - \beta si - \alpha w$, we get:

$$i^* = \frac{\alpha}{\mu}\left(1 - \frac{1}{R_0}\right) = \frac{\alpha}{\beta}(R_0 - 1)$$

Finally, the three values of the fraction of infected, suscpetible and recovered in the endemic state are:

$$(s^*, i^*, r^*) = \left(\frac{1}{R_0}, \frac{\alpha}{\beta}(R_0 - 1), 1 - \frac{1}{R_0} - \frac{\alpha}{\beta}(R_0 - 1)\right) \tag{1.19}$$

This exists only if $R_0 > 1$. Moreover, via linear stability analysis it can be demonstrated that this equilibrium is stable and is reached through damped oscillations

Lecture 6.
Thursday 15th October, 2020.
Compiled: Tuesday 27th October, 2020.

1.3.2 SIRS Model

During the years the immune system may lose the ability to recognize a known pathogen acquired via both a passed disease or a vaccine. Moreover, there could be the possibility that viruses mutate as the seasonal influenza. Hence, let us build a model in which after an individual is recovered, can become again suscpetible after a period of time.

The SIRS Model allows to interpolate between SIR ($w = 0$) and SIS ($w \to \infty$). We have again an absorbing or endemic state. The transitions of this model are:

$$\begin{aligned} S + I &\xrightarrow{\beta} I + I \\ I &\xrightarrow{\mu} R \\ R &\xrightarrow{w} S \end{aligned} \tag{1.20}$$

In particular, the differential equations which describes the model are:

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= \alpha + wr - \beta si - \alpha s \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \beta si - \mu i - \alpha i \\ \frac{\mathrm{d}r}{\mathrm{d}t} &= \mu i - wr - \alpha r \end{aligned} \tag{1.21}$$

The endemic state can be found by putting the derivatives equal to zero.

Since the $R \to S$ does not affect the I, we have that the infectious period is:

$$\tau = \frac{1}{\alpha + \mu} \tag{1.22}$$

while the R_0 factor is:

$$R_0 = \frac{\beta}{\alpha + \mu} \tag{1.23}$$

Moreover, the values of s^*, i^* and r^* can be obtained easily following the same arguments as of the SIR with demography.

1.3.3 SEIR Model

In reality people do not become instantaneously infectious, but there is a **latent period** which is the time between infection and becoming infectious. Indeed, the pathogen replication takes time, i.e. viral load too low to transmit the infection. Moreover, it is extremely heterogenous since it can take from few hours to years.

It is important to remind that the latent period is not the same of the incubation period (see Fig. 1.11). An individual can be infectiuous before symptoms. For instance it has a pre-syntomatic infection as in the case of Covid-19.

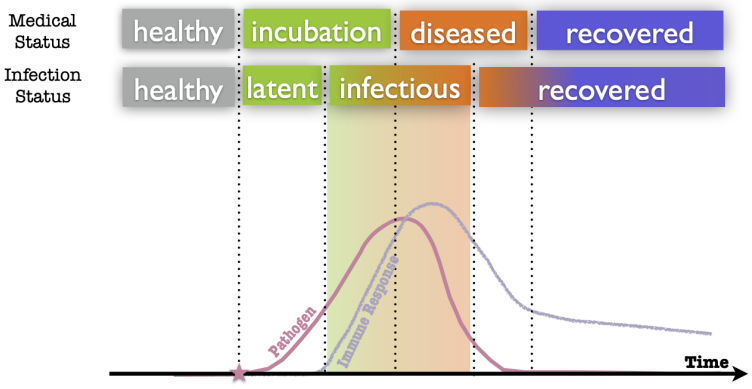

Figure 1.11: Difference between infection status and medical status.

The transition of the SEIR model are:

$$\begin{aligned} S + I &\xrightarrow{\beta} I + E \\ E &\xrightarrow{\sigma} I \\ I &\xrightarrow{\mu} R \end{aligned} \tag{1.24}$$

with the equations:

$$\begin{aligned} \frac{\mathrm{d}s}{\mathrm{d}t} &= \alpha - \beta si - \alpha s \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \beta si - (\alpha + \sigma)e \\ \frac{\mathrm{d}i}{\mathrm{d}t} &= \sigma e - (\alpha + \mu)i \\ \frac{\mathrm{d}r}{\mathrm{d}t} &= \mu i - \alpha r \end{aligned} \tag{1.25}$$

Hence, the spreading is delayed due to the time in E.

The **endemic state** is:

$$\begin{aligned} s^* &= \frac{(\alpha + \mu)(\alpha + \sigma)}{\beta\sigma} = \frac{1}{R_0} \\ e^* &= \frac{\alpha(\alpha + \mu)}{\beta\sigma}(R_0 - 1) \\ i^* &= \frac{\alpha}{\beta}(R_0 - 1) \end{aligned} \tag{1.26}$$

For very short latent time ($\sigma \to \infty$) we recover the endemic state of the SIR.

The R_0 factor is:

$$R_0 = \frac{\beta\sigma}{(\alpha+\mu)(\alpha+\sigma)} \tag{1.27}$$

Since latent time is way shorter than demography, usually $\frac{\sigma}{\sigma+\alpha} \simeq 1$ thus $R_0 = \frac{\beta}{\alpha+\mu}$ as in the SIR with demography.

Since the endemic state, the infectious period and R_0 are similar between SEIR and SIR, adding exposed individuals it may seem an unnecessary complication but, if we look at the time evolution at the early stages there is a huge difference between SEIR and SIR model:

$$\begin{aligned} i_{SEIR}(t) &\approx e^{\left(\sqrt{4(R_0-1)\sigma\mu+(\sigma+\mu)^2}-(\sigma+\mu)\right)t/2} \approx i_0 e^{\left(\sqrt{R_0}-1\right)\mu t} \\ i_{SIR}(t) &\approx i_0 e^{(R_0-1)\mu t} \end{aligned} \tag{1.28}$$

Even if the behavior at the steady state is similar the temporal evolution of SEIR is slower than SIR. This has important implications for policy making.

The SEIR can be the starting point for modeling realistic diseases: i.e. Covid-19 (see Fig. 1.12).

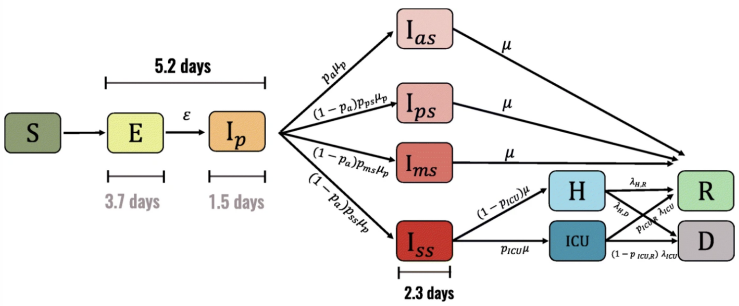

Figure 1.12: Model for Covid-19.

1.4 Summary of compartmental models in well-mixed populations

Let us summarize all the comportmental models in well-mixed populations explained:

- we solved the **SI model** analitically and we observe that the growth is as a sigmoid:

  $$i(t) = \frac{i_0 e^{\beta t}}{1 - i_0 + i_0 e^{\beta t}}$$

  In the early phases we have an exponential growth governed by β, and it always saturates at 1;

- in the **SIS model** things starts to change. We have and endemic (meta-stable) state:

  $$i = \frac{\beta - \mu}{\beta}$$

  and we reach a sort of dynamic equlibrium. We can define an epidemic treshold:

  $$\beta > \beta_c = \frac{\mu}{\langle k \rangle}$$

- for the **SIR model** we cannot solve the equations analitically. We have no endemic state and the epidemic treshold is again:

$$\beta > \beta_c = \frac{\mu}{\langle k \rangle}$$

- then, in the **SIRS model** we include waning immunity. This model interpolates between SIR and SIS model. We have the endemic state and the infectious period is:

$$\tau = \frac{1}{\alpha + \mu}$$

and moreover:

$$R_0 = \frac{\beta}{\alpha + \mu}$$

- at the end, we have the **SEIR model** in which we include a latent period. We have that:

$$R_0 = \frac{\beta\sigma}{(\alpha + \mu)(\alpha + \sigma)}$$

and the exposed slow down the spreading.

2

Basics of Network Science

2.1 Basics definition

The most important thing is the concept of **network** (graph) $G(V, E)$ which is just an object composed by a set of nodes (vertices) V and a set of links (edges) E:

- the **nodes** represents the *entities* $V = [\dots, i, j, k, \dots]$ of the relationships, the entries, the people in a social network and so on. The number of nodes is $N = |V|$;
- the **links** represents the relationship between entities $E = [\dots, (i, j), (i, k), \dots]$. The number of links is $L = |E|$.

Links can be of different types and so networks: the basics distinction is between **undirected** and **directed** links. Another important distinction is between **unweighted** and **weighted** links.

The **network density** (connectance) is the fraction of links over all the possible pairs:

$$d = \frac{L}{N(N-1)} \tag{2.1}$$

Real networks usually have a very low density, so are sparse systems ($L \ll N^2$).

Another things that we will use a lot is the way of representing a graph in a mathematical sense. We will use the **adjacency matrix** A of the network, where:

- $a_{ij} = 1$, if a link between nodes i and j exists;
- $a_{ij} = 0$ otherwise.

We can exploit many useful mathematical properties such as the spectrum of the matrix. Moreover, the matrix is symmetrical for undirected/unweighted graphs, i.e. $a_{ij} = a_{ji}$.

Since real networks are usually sparse, the adjiacency matrix is inefficient for storing graphs in a computer, is better to use adjiacency lists, etc.

Two nodes that share a link are defined as connected, adjacent, neighbors. In particular, the **neighborhood** of node i is the set of nodes connected to i. The number of neighborhood k_i of each node i is what is called the **degree** of the node. This is the basic measure that we are gonna use a lot. Once you define the degree, the next step is defining what is the average of the degree over the entire network:

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i, \qquad \text{or} \qquad \langle k \rangle = \frac{2L}{N} = d(N-1) \tag{2.2}$$

The next definition is the concept of **path**, which is a sequence of links which permit to go from node i to node j following links. Another important part is what is called the **shortest path** between i and j. This gives us the idea of how big the network is. In particular, the distance l_{ij} represents the length of the shortest path between i and j. There could be multiple shortest paths between i and j.

The network is **connected** if every possible couple of nodes is reachable trough a path. Otherwise, each connected part is defined as a connected **component**.

The shortest path of maximum length in the network is defined as **diameter**:

$$l_{max} = \max_{ij} l_{ij}$$

Another measure that is quite important is the **average (shortest) path length**:

$$\langle I \rangle = \frac{\sum_{ij} l_{ij}}{N(N-1)}$$

Now, let us see some example as "The Oracle of Bacon", the "Erdos Number" and so on. The question which arises is: why such short distances in such large networks? In particular, we note that in the last example of "Six degrees of separation" real networks are smaller (shorter) than one would expect. This is what is called the small world phenomena. If you study the average path length which scales linearly as the scales of the network it scales as the logarithm of the network or in some case the logarithm of the logarithm of the network. This is huge important in the spreading of diseases.

Lecture 7.
Friday 16th October, 2020.
Compiled: Tuesday 27th October, 2020.

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**.

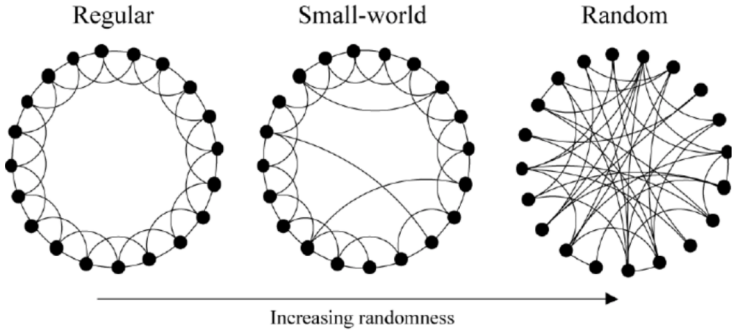

Figure 2.1: Idea of Watts and Strogatz model.

Let us focus with the first regular ring in Fig. 2.1 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. 2.1. 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. 2.1.

2.2 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. 2.2 represents the distribution of the degrees in this small network. How is this quantity distributed in real networks?

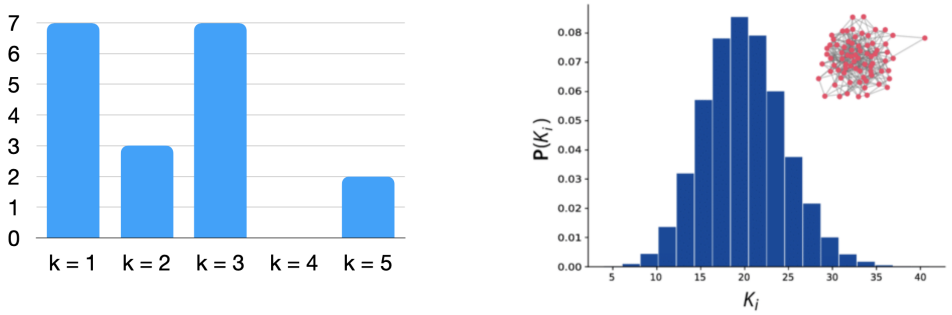

Figure 2.2: Left: degree distribution in a small network. **Right:** degree distribution in a network with random connections.

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. 2.2.

2.2.1 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:

- 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:

$$p_k = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{2.3}$$

The average and variance of such a distribution are:

$$\langle k \rangle = p(N-1), \qquad \sigma_k^2 = p(1-p)(N-1) \tag{2.4}$$

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:

$$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle}{k!} \tag{2.5}$$

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.

2.2.2 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. 2.4 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. 2.4, 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

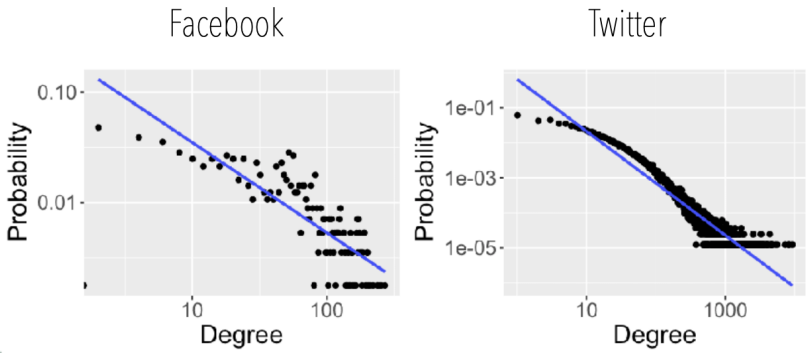

Figure 2.3: Real network of Facebook and Twitter.

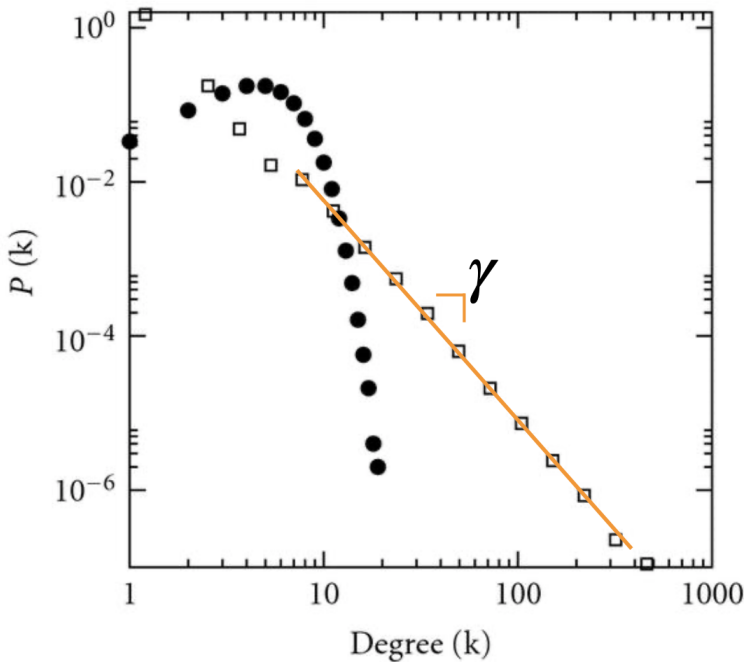

Figure 2.4: Difference between random networks and scale-free networks.

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:

$$P(k) = C_0 k^{-\gamma} \qquad \text{with} \quad C_0 = (\gamma - 1) k_{min}^{\gamma - 1} \tag{2.6}$$

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

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

Thus, when:

$$k_{max} = k_{min} N^{\frac{1}{\gamma - 1}} \tag{2.7}$$

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:

$$\langle k^n \rangle = \int_{k_{min}}^{\infty} k^n P(k)\, \mathrm{d}k = \int_{k_{min}}^{\infty} C_0 k^{n-\gamma}\, \mathrm{d}k \tag{2.8}$$

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.

2.2.3 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 attatchement 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:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \tag{2.9}$$

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. 2.5. 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.

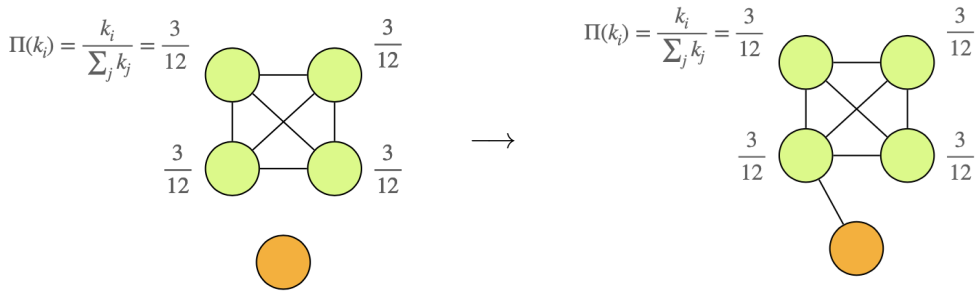

Figure 2.5: Example of Barabási-Albert algorithm.

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.

3

Epidemic Spreding on Networks

3.0.1 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:

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

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

3.0.2 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.

3.0.3 Heterogenous netowrks

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

3.1 Summary

Let us recap what we saw at the end of the last 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.

Today, we will see different ways to deals with the probability $Prob[\sigma_i = 0, \sigma_j = 1]$.

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

probabilityies statistically independent. We derived all the equations. The solutions we found are the same of before, the difference is that now it is not exact.

4

Epidemic Spreading on Networks

Today we are going start to analyze what happens when we consider hetherogeneous networks.

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

4.1 Degree-based Mean-Field theories (DBMF)

We start with the simple level which is the individual one, in which we consider the individual probability of getting the infection.

Let us start with the paper "Epidemic Spreading in Scale-Free Networks". They provide a model for SIS on scale-free networks. We cannot use the homogeneous approximation, because the network is hethereogeneous. The intuition of this paper are:

- the nodes are not equal. The probability of getting the infection strongly depends on their position (i.e. degree) in the network;
- nodes with the same degree behave in the same way;

We are gonna divide the network in degree classes. We are grouping 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.

From that we can write the equation:

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

we have as usual a recovery and infection part. In particular, we have the probability of a contact between a suscpetible of degree k and an infected 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 $\Theta_k(t)$. We have that:

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

is the probability that a node with degree k as an infected neighbor. We want to 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. We are gonna make an assumption about the structure of this thing. It could be in principle anything, in the sense that it depends on the structure of the network. However, there are some cases in which we can do some assumptions on the structure of the network.

We can assume that network are hethereogenous but I am making connection at random. Hence:

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

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 I 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'.

Hence:

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

In the numerator: we take the probability that a link taken at random points to k', then I am multiplying by the probability of being infected and then I am summing to all the possible degree. We note that this probability 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 used to solve the differential equation of $\rho_k(t)$ is pretty similar to the ones used before. First of all, we assume that we are in the steady state:

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_k(t) \to \rho_k \frac{\beta k \Theta}{\mu + \beta k \Theta}$$

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

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

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 as different solution. The first one is the trivial solution $\Theta = 0$, but we are interested in the non trivial one. 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 were the two values are equal. This is the interception between the line Θ and the function $f(\Theta)$. We want that since Θ is a probability $0 < \Theta \leq 1$. This means that the slope of $f(\Theta)$ should be greater than 1.

The next step, is seeing what means mathematically that we want a slope larger than one:

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

which means

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

this 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. If we assume that:

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

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

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

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

this is exactly the expression that we saw before. Things are working well.

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

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

the epidemic threshold is going to zero. Obviously, this is quite important because if my network is big enough, every disease will spread, no matter its infectivity! 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 epidem threshold is really really small for most real epidemic network. This result work as in a 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 infity, we introduce a cut-off. For instance, for a social network we can reach some number of followers but the system will have a cut off.

We intdroduce an exponential cut-off for this distribution. For instance in 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.

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

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

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:

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

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

4.1.1 SIR model

We can write the same equations for the SIR model under the same assumption. We need more equations to take into account also the equation for R.

$$\frac{\mathrm{d}}{\mathrm{d}t} \rho_k^I(t) ...$$

where we have:

$$\Gamma_k(t)$$

which 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 its form. It is little different because of the structure of the SIR. There is a neibhobour that can trasmit the infection to me, but it can recover. Hence, this individual can be suscpetible again. I need to take into account that the disease is coming from one side, because that side of the network for me is forbidden in the sense that I have recovered that connot be infected again.

We found:

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

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.

4.2 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 indivudual based mean-field theories for individuals. We are not consider a single network, but an average over all the possible network we can create given that degree distribution.

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

Let write down the equation for the quenced mean field. We are gonna use the discrete time formulation of equation because it is a bit simpler (we can write also the same things with differential equations).

We are consider $\rho_i(t)$ as the probability of a node of being infected at time t. What is my probability of being infected at time $t+1$ is:

$$\rho_i(t+1) = \rho_i(t)(1-\mu) + (1-\rho_i(t))q_i(t)$$

which is the probability of being infected and not get cured and the second is the normal term when I am getting the infection (the probability of being suscpetible and the probability of getting the disease). We have that $q_i(t)$ is the probability of node i of getting infetced by at least one neighbour.

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

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

We have the node i in green, and the infected neighbour in red. We see that $\beta A_{ij}\rho_j(t)$ is the probability of getting infetced 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 1 minus the probability of not getting infected by any neighbor.

Hence, once we have this things, first of all we can solve it numerically, which means for instance by iteration. In this case we are gonna have an equation for each

of the node. We have 2^N equation where N is the size of the system. The equation for $\rho_i(t+1)$ can be solved by iteration. I am gonna repeat this equation many times until I reach the steady state.

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 teh steady state to estimate the epidemic threshold. I am assuming that if I am in the epidemic treshold, what happens is that ρ is very small for all the nodes. If $\rho_i^* = \varepsilon_i^* \ll 1$, we can use an approximation for q_i^*:

$$q_i^* = 1 - \prod_{j=1}^{N}[1 - \beta A_{ij}\varepsilon_j^*] \sim \beta \sum_{j=1}^{N} A_{ij}\varepsilon_j^*$$

Substituting we have:

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

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

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

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 was important. Hence:

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

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:

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

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.

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

$$\beta_c^{DBMF} = \frac{\mu\,\langle k\rangle}{\langle k^*\rangle}, \qquad \beta_c^{IBMF} = \frac{\mu}{\Lambda_{max}}$$

The relation amogn them for scale-free networks is:

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

Specifically:

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

We can conclude that IBMF is more accurate than DBMF.

Part II

Polletto's Lectures

Bibliography