

What is the reason behind this result? Since we know the strong connection between the two theories, now we want to **derive DBMF from IBMF**¹. We start from:

$$\dot{\rho}_i = -\mu\rho_i + (1 - \rho_i)q_i \quad \text{where} \quad q_i = 1 - \prod_{j=1}^N [1 - \beta A_{ij}\rho_j]$$

The Adjacency Matrix is replaced by an **Annealed Adjacency Matrix (AAM)**:

$$\bar{A}_{ij} = \frac{k_j P(k_i|k_j)}{NP(k_i)} \quad (1)$$

which for *random networks* becomes:

$$\bar{A}_{ij} = \frac{k_i k_j}{N \langle k \rangle} = \frac{k_i k_j}{2L}$$

where the probability $P(k_i|k_j)$ of picking a random node becomes k_j . The last equation represents the number of trials that we have to create the connection i, j over all the possible connections in the network. If we substitute the explicit form of the adjacency matrix in the expression of q_i :

$$q_i = 1 - \prod_{j=1}^N \left[1 - \beta \frac{k_j P(k_i|k_j)}{NP(k_i)} \rho_j \right]$$

From individual nodes to degree classes:

$$\dot{\rho}_k = -\mu\rho_k + (1 - \rho_k) \left[1 - \prod_{k'=1}^N \left[1 - \beta \frac{k P(k'|k)}{N_{k'}} \rho_{k'} \right]^{N_{k'}} \right]$$

Assuming $\beta\rho_k \ll 1$, we can approximate the product with:

$$\dot{\rho}_k = -\mu\rho_k + \beta k (1 - \rho_k) \sum_{k'} P(k'|k) \rho_{k'}$$

and remembering that $\Theta_k = \sum_{k'} P(k'|k) \rho_{k'}$:

$$\dot{\rho}_k = -\mu\rho_k + \beta k (1 - \rho_k) \Theta_k$$

which is the formula obtained for DBMF. Hence, in the DBMF we are implicitly assuming that $\beta\rho_k \ll 1$. This is the reason why DBMF is accurate only around the epidemic threshold. At the end, we are able to pass from IBMF to DBMF and actually we are explaining the difference in the accuracy between the two models.

0.0.1 IBMF and Pair approximation

Let us make a very brief introduction in what it means cut down the chain to pair approximation. Until now, all the models that we saw were cut at the individual level which means that $\text{Prob}[\sigma_i(t) = 0]$ and $\text{Prob}[\sigma_j(t) = 0]$ are statistically independent:

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

¹We repeat that in annealed networks we are not considering a single network but an average of all the possible random networks that you can generate from a degree distribution. Instead, in the quenched network, we take a particular one and we want the result for this specific network.

Now, if we cut at the link (i, j) level we have to consider $\text{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1]$. This involves three nodes terms and so on:

$$\frac{d}{dt}\rho(i, t) = -\mu\rho(i, t) + \beta \sum_j A_{ij}\rho(j, t) - \beta \sum_j A_{ij}E[X_i(t)X_j(t)]$$

where $E[X_i(t)X_j(t)]$ is the two nodes expectation of being infected. We need an expression for the $\frac{N}{2}$ equations for $E[X_i(t)X_j(t)]$. The idea is:

$$\begin{aligned} \frac{d}{dt}E[X_i(t)X_j(t)] &= -2\mu E[X_i(t)X_j(t)] + \beta \sum_k A_{ik}E[X_j(t)X_k(t)] \\ &\quad + \beta \sum_k A_{jk}E[X_i(t)X_k(t)] - \beta \sum_k (A_{ik} + A_{jk})E[X_i(t)X_j(t)X_k(t)] \end{aligned} \tag{2}$$

and the most common possible closures are:

$$E[X_i(t)X_j(t)X_k(t)] = E[X_i(t)X_j(t)]E[X_k(t)]$$

or

$$E[X_i(t)X_j(t)X_k(t)] = \frac{E[X_i(t)X_j(t)]E[X_j(t)X_k(t)]}{E[X_j(t)]}$$

where the second is similar to the first but we are considering the two extremes and then the probability that j is infected.

1

Epidemic spreading on networks: advanced models

In this chapter, we are gonna study non-Markovian epidemic spreading. In the literature, it is not seen, but if you want to implement a realistic model it is very important.

1.1 Markovian Models

All the models we have seen till now assume that β (infection process) and μ (recovery process) are constant rates. This means that movement between compartments takes place at constant rate, or equivalent the probability of jumping from one compartment to another does not depend on the time spent in the compartment. Essentially, we are considering a memoryless process. The jump are memoryless, i.e. we are running a Markov chain.

The fundamental property of a Markov chain is called **Markov property**: *the jump probability at time $t + \Delta t$ does not depend on elapsed time, but only on t* , i.e. we have not to take into account the time that we spent in that compartment. This property is very useful for mathematical treatments.

Jumps made at a constant rate (β and μ) imply that the time spent inside each compartment τ follows a exponential distribution:

$$P(x) = \tau e^{-\tau x} \quad (1.1)$$

with mean $\tau = \frac{1}{\mu}$, i.e. τ is the infectious period (average time spent in I).

What are the implication of this property? The most important one is that an exponentially distributed infectious period implies that the most probable duration of the disease is 0. Indeed, the probability decreases with time. More particularly, it depends on the mean, but in any case the most probable jump (time in which I am making the jump) is at the beginning. It is something that is not realistic, indeed if you got influenza at least you will spend some time infected. And actually, if you are looking how infectious period are distributed in real life, it is something which is quite different. For a disease, you know exactly when it starts but do not you know when it ends. For instance, let us consider the left plot in Fig. 1.1 for 2009 H1N1 Influenza. For this type of disease the plot show the distribution of the infectious period, which has as probable value 2 days and an half. The most important things is that it is not 0. On the right, we have also the estimates for Covid-19. One process we use to measure the infectious period is the **serial interval**, i.e. from symptoms to symptom. Or we have also the **generation time**, i.e. from infection to infection. Obviously, these are approximations but can give you some means.

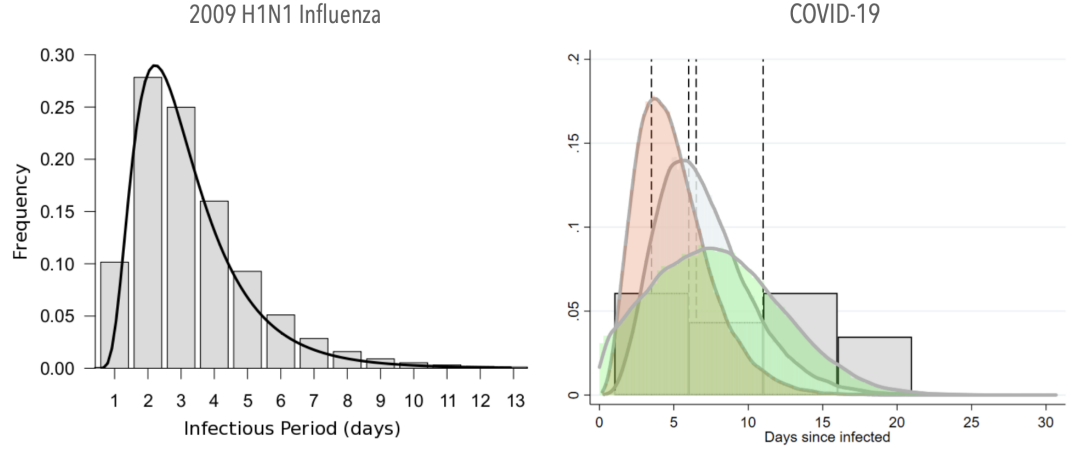


Figure 1.1: **Left:** Histogram of 1009 H1N1 Influenza. **Right:** distribution of Covid-19.

The problem is that, all these results demonstrate that these kind of diseases are non-markovian: transition probability depends on the time spent in a compartment. Indeed, patients usually spend some time infected before starting to recover.

1.2 Non-Markovian Epidemic Spreading

How are we gonna model this non markovianity? What are the distribution that better describe what we saw in the data?

Patients usually spend some time infected before starting to recover. This situation is better approximated by a **Gamma distribution**:

$$P(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}} \quad (1.2)$$

where k is the *shape*, θ the *scale* and $\Gamma(x)$ is the *gamma function*:

$$\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$$

The gamma distribution has mean $k\theta$ and variance $k\theta^2$. This shape start to be somehow what we saw in the data.

What is similar to the gamma distribution is the **Erlang distribution**, where we have the factorial instead of the Gamma function:

$$P(x) = \frac{\lambda^k x^{k-1} e^{-\lambda x}}{(k-1)!} \quad (1.3)$$

or the **Weibull distribution**:

$$P(x) = \begin{cases} \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (1.4)$$

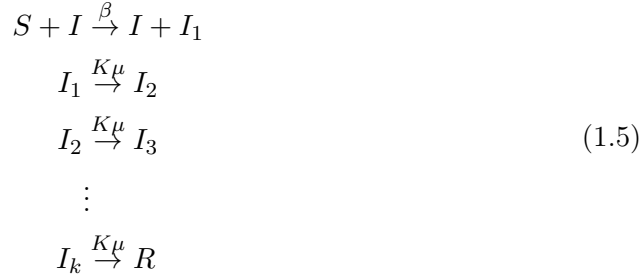
All of these distribution are able to reproduce real histograms.

However, the question still remains: how to include non-markovian elements in classical epidemiological models (with the assumption of markovian...)? We use a trick for the indectionious period: *the sum of exponential random variables obeys a Gamma distribution*. How we are gonna incorporate that in our model? Instead of having just one transition at a coasntant rate (expoential distribution), what we need to have to have a gamma distribution?

The trick is the fact that instead of having just one transition, we are gonna include more and more transitions. Instead of having just one single infectious state, individuals move from one compartment to the other, such that they spend at least some time infectious before starting the recovery. We obtain again a Markovian model.

1.2.1 SIR Model with Multiple Infectious Stages

To repeat, the solution is using **multiple infectious compartment**, i.e. SIR in well mixed populations becomes $SI_1I_2 \dots I_kR$. For instance, we are imposing that these transitions are sequential:



Hence, if I want to get recovered I need to spent some times infectious, but the model is still markovian! More precisely, the equations are:

$$\begin{aligned}
 \frac{ds}{dt} &= -\beta si \\
 \frac{di_1}{dt} &= \beta si - K\mu i_1 \\
 \frac{di_2}{dt} &= K\mu i_1 - K\mu i_2 \\
 &\vdots \\
 \frac{dr}{dt} &= K\mu i_k
 \end{aligned} \tag{1.6}$$

where the rate of each I transition is $K\mu$ and with $i = \sum_{k=1}^K i_k$. We got that this is the infectious period distribution:

$$P(\tau) = \frac{(\mu K)^K}{\Gamma(K)} \tau^{K-1} e^{-\mu K \tau} \tag{1.7}$$

where the mean is still $1/\mu$, but the shape is totally different. which is the gamma function. We have two special cases:

- if $K = 1$, we obtain an exponential distribution;
- if $K \rightarrow \infty$ fixed, we obtain a delta distribution.

Other quantities are:

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

and the **final epidemic size**:

$$r_\infty = 1 - e^{-R_0 r_\infty}$$

Moreover, we have an early growth:

$$i(t) \simeq i_0 e^{\lambda t}$$

instead of $i(t) \simeq i_0 e^{(\beta-\mu)t}$. Hence, the disease is growing faster and has a shorter duration.

With λ as the solution of:

$$R_0 = \frac{\lambda}{\mu \left(1 - \left(\frac{\lambda}{K\mu} + 1 \right)^{-K} \right)}$$

The $SI_1 I_2 \dots I_k R$ model has several limitations:

- it is defined only for well-mixed populations;
- focus only on the infectious period distribution. We have that infections are still Markovian and this model only reproduces as Gamma distribution.

1.2.2 Generalized SIS Model

Now, we are going to present something which is more general where we can include non-markovian both in recovery and infections. Is it possible to write down a general model on networks? The answer is yes, it is a bit more complicated and we still need some kind of approximation at some point. In particular, we need a mean-field approximation.

We have to change our point of view. We are gonna use a slightly different approach, i.e. instead of probabilities we are gonna talk about events. The idea is that we are gonna modelling in this case the infections and recoveries with two random numbers which we extract from distribution and are as general as possible.

The ingredients are:

- a random number $R_i(t)$: recovery time of node i when infected;
- a random number $M_{ij}(t)$: infection times at which node i tries to infect node j .

In order:

- node i get the infection at time t ;
- we extract the random number $R_i(t)$ which represents the time in which node i is gonna recovery (or, the time for which it stay infected);
- then, we extract the random number $M_{ij}(t)$ which represents the number of trials that i try to infect node j while infected;
- we generate a sequence of times

$$T_{ij}^{(1)} \leq \dots \leq T_{ij}^{(M_{ij}(t))} \leq R_i(t)$$

in which node i try to infect node j . For instance, $T_{ij}^{(1)}$ is the first time that node i try to infect node j then we have the second time and so on;

- I am gonna repeat the last step for all my neighbours.

Hence, the transmissibility of the disease is seen as how many trials I am gonna make to infect. One important thing is that $R_i(t)$ and $M_{ij}(t)$ can be drawn from any distribution and not only from the exponential one. How do we extract the T_{ij} is not important at this point, because we are only gonna focus on the distribution of $R_i(t)$ and $M_{ij}(t)$.

Now, let us make some assumptions to make the model more reasonable and then treat it analytically. We assume that:

- $R_i(t)$ and $M_{ij}(t)$ do not depend on time, i.e. $R_i(t) \equiv R_i$ and $M_{ij}(t) \equiv M_{ij}$;
- $R_i(t)$ and $M_{ij}(t)$ do not depend on i and j , i.e. same distribution for all the nodes $R_i \equiv R$ and $M_{ij} \equiv M$.

hence, we are assuming that these numbers should not depend on time and for instance they are typical for the disease. It is valid both for the recovery and for the infections. However, if we consider restrictions as lockdown and so on these number should change, but for the let us consider the simplest model without such restrictions. Indeed, with these assumptions we are reducing the complexity.

We call:

- $E[R]$, the expected value of R ;
- $E[M]$, the expected value of M ;
- v_i , the probability that node i is infected in the steady state.

Now, let us build the model:

1. let us suppose that we are in the steady state of the system (all the transient are passed). In a large time interval $[0, S]$ the number of times node j has been infected is proportional to S (it is linear). Since the length of each infected period is $E[R]$, the **number of infected periods** experienced by a node j (number of times a node has been infected in the interval $[0, S]$) can be written as:

$$\frac{v_j S}{E[R]}$$

2. during each infected period, node j will try to infect i an average $E[M]$ number of times. So, the **total number of infection attempts** from node j to i in a large period of time are:

$$\frac{v_j S E[M]}{E[R]}$$

the number of times j has been infectious multiplied by the number of infection attempts per each time;

3. then, we make the **mean-field assumption**: the conjunct probability that j is infected and i is susceptible is:

$$\text{Prob}[\sigma_i(t) = 0, \sigma_j(t) = 1] \rightarrow \underbrace{\text{Prob}[\sigma_i(t) = 0]}_{(1-v_i)} \underbrace{\text{Prob}[\sigma_j(t) = 1]}_{v_j}$$

4. we can write the **total number of successful infection attempts** for j to i as:

$$S \frac{E[M]}{E[R]} v_j (1 - v_i)$$

that is the total number of attempts multiplied by the probability that i was not infected.

Summing over all the neighbors of i we have the the **total number of successful infections** i will receive during interval $[0, S]$ is:

$$S \sum_{j=1}^N a_{ij} \frac{E[M]}{E[R]} v_j (1 - v_i)$$

5. in the *steady state* the number of successful infections ($S \sum_{j=1}^N a_{ij} \frac{E[M]}{E[R]} v_j (1 - v_i)$) is asymptotically equal to the number of infected periods experienced by i : $v_i S / E[R]$. Thus we have:

$$\cancel{S} \sum_{j=1}^N a_{ij} \frac{E[M]}{\cancel{E[R]}} v_j (1 - v_i) = v_i \frac{\cancel{S}}{\cancel{E[R]}} \quad \rightarrow \quad v_i = E[M] (1 - v_i) \sum_{j=1}^N a_{ij} v_j \quad (1.8)$$

hence the probability of i being infected depends by the sum over all its neighbours, times the term $E[M]$ which is the average infection attempts that it is gonna experience during the evolution.

Do the last expression sound familiar? This is exactly the same expression of the IBMF but with a generic infection term $E[M]$:

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

The implications are:

- the definition of M implicitly includes the recovery term: $T_{ij}^{(1)} \leq \dots \leq T_{ij}^{(M_{ij}(t))} \leq R_i(t)$;
- exponential case: $E[M]$ is the expected number of infection events in a Poisson process with intensity β within an exponential recovery time with expectation $1/\mu$. Thus: $E[M] = \beta/\mu$.
- epidemic treshold (lower bound):

$$m_c = E[M_c] = \frac{1}{\Lambda_{max}}$$

Hence, this is the form in which we can generate a generic model from.