0.1   Global Invasion Threshold

Following the formalism introduced for the SIR metapopulation model and marko-vian mobility, we want now to analytically derive the **epidemic threshold** for metapopulations. The latter is the *threshold* for a pathogen that, when overcomed, allows it to spread among the entire population. In other words, we want to find the **conditions** for a **local outbreak** to **spread** at a **global** scale.

Figure 1:   Dynamics of spatial spread at subpopulation level and the generation time n they contract the disease.
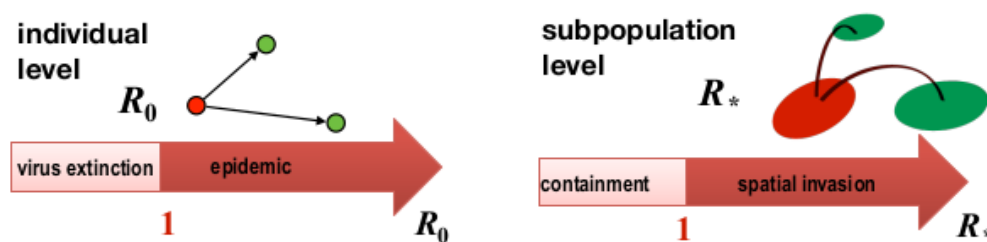
In order to pursue our goal, we make the so called **coarse graining**: starting from the formalism typical of a single subpopulation we decrease the number of degrees of freedom by making some averages or finding some more general rules. Quantities must therefore be scaled accordingly: R_* is the correspondent to R_0 for metapopulations (if $R_* > 1$ we observe an outbreak), and the typical **timescale** is now the *duration of* an *outbreak* in a patch. Hence, we follow the spread from one subpopulation to another, by the mean of **mapping**[1] [2] the **spreading dynamics among subpopulation** into the spreading on a **network**.

Figure 2:   **Left:** Dynamics of spread on a individuals-in-a-network level. **Right:** Dynamics of spreading among subpopulation, once we made a *coarse graining*. Graphically, they look like similar, despite the change in value, and meaning, of parameters as output of the mapping. R^* is the analogous of the basic reproductive ratio R_0 at metapopulation level.

We are indeed approximating as a **branching process** and try to follow the invasion dynamics at the subpopulation level, denoting by D^n the diseased subpop-ulations at the n-th generation. Hence, we do not follow the dynamics in term of times, but in terms of *generation*.

---

[1] Colizza & Vespignani, PRL 2007, JTB 2008
[2] Cross, et al. JRSoc Interface 2007

0.1.1 Homogeneous networks

Let us assume that we are dealing with **homogeneous systems**, for the sake of simplicity, and with only 2 patches: namely j and i as one can see in Fig. 3.
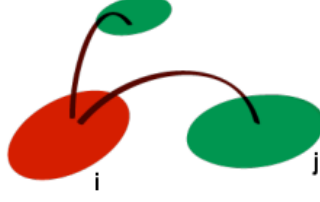

Figure 3: Dynamics of spatial spread at subpopulation level, but considering only two patches.

The variables one may want to introduce now are the following ones: w_0 is the **number of travellers** along each link, whereas $\langle k \rangle$ is on average the **number of connection** of each subpopulation, which coherently with our notation has **population** N. Let us introduce now α, that is the **epidemic attack rate**: it is the *"final size"* of the epidemic within each patch. We recall that the latter depends on the basic reproductive ratio R_0.

Using these quantities we can compute the probability of early extinction of the disease, that is the probability for the disease to be not able to spread to other patches once it started in patch i:

$$P_{ext} = \left(\frac{1}{R_0} \right)^{\lambda_{ij}} \tag{1}$$

where R_0 is the usual term and λ_{ij} is the total number of infectious individuals sent from i to j during the local outbreak:

$$\lambda_{ij} = \frac{w_0}{N} \frac{\alpha N}{\mu} \tag{2}$$

Note that it is the product of two terms: the first factor is the probability to travel, while the second is the ratio between the total number of people that get the infection rescaled by the recovery rate μ. We recall taht α is the maximum value of the prevalence $i = I/N$ within a patch.

The number of diseased subpopulation at generation n, namely D^n, is computed iteratively from the number of diseased patches at generation D^{n-1}:

$$D^n = (\langle k \rangle - 1)(1 - P_{ext}) \left(1 - \sum_{m=0}^{n-1} \frac{D^m}{V} \right) D^{n-1} \tag{3}$$

where V is the number of patches. Moreover, in the factor $(\langle k \rangle - 1)$, the -1 is present in order to ignore the patch we have got the infection from. In addition, the one highlighted is the probability that the patch is disease free at $n-1$-th generation. The last factor D^{n-1} is the number of diseased patches at the previous generation: as one can imagine, if the remaining factor is **larger** than 1, then the epidemic has an outbreak, otherwise it gets extinct. Indeed:

$$D^n = R_* \left(1 - \sum_{m=0}^{n-1} \frac{D^m}{V} \right) D^{n-1}$$

where the condition for a global outbreak is $R_* > 1$.

Accordingly to what we have just said, the factor R_* is therefore defined *at the beginning of the infection*, i.e. when all patches are susceptible and so the sum gives no contribution and can be *neglected*, as:

$$R_* = (\langle k \rangle - 1)(1 - P_{ext}) \tag{4}$$

Moreover, if R_0 is really low[3], the probability of having an outbreak is:

$$1 - P_{ext} = 1 - \left(\frac{1}{R_0}\right)^{\lambda_{ij}} \simeq \lambda_{ij}(R_0 - 1) = \frac{\alpha w_0}{\mu}(R_0 - 1) \tag{5}$$

which actually simplifies our expression for R_*. Indeed, recalling that according to our assumption every patch has same number of connections and travellers, we obtain:

$$R_* = (\langle k \rangle - 1)\frac{\alpha w_0}{\mu}(R_0 - 1) \tag{6}$$

We want now to study the **dependencies** of the **invasion potential**. It is indeed a *growing* function of: R_0, both mobility related quantities *overall traffic rescaling* w_0 and *average number of connections* $\langle k \rangle$. In addition, it is *inversely proportional* wrt recovery rate μ or, in other words, it is a *growing* function of **infectious duration**. The more we stay infected the larger becomes R_* and if the latter is larger than $R_* > 1$ the epidemic is able to spread to other patches, which actually makes sense.

0.1.2   Heterogeneous networks

Let us introduce a more realistic model and consider that networks are actually different from the homogeneous we have seen up to now. Indeed, **real systems** are **highly heterogeneous**. As we have already pointed out (see Sec. **??** and Fig. **??** and **??**): the number of connections and travellers along the connections is heterogeneous[4]. This implies that average quantities, such as $\langle k \rangle$, are not representative of the properties of the patches: **homogeneous approximation is bad**. However, we were able to find some **scaling relations**: these are some approximate laws that make computations feasible and are related to the **degree-block** description (coarse-graining).
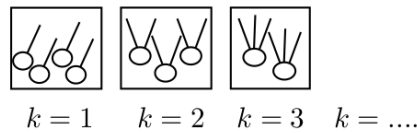

$$k = 1 \quad\quad k = 2 \quad\quad k = 3 \quad\quad k = ...$$

Figure 4:   Patches are divided into different degree blocks, according to the number of connections k they have.

With **degree-block** description we mean that we group patches according to their degree, and consider patches within the same degree-classes homogeneous. In other words, we make the approximation, similar to the one we have made with networks, that patches with **same number of connections behave in the same way**, like when we used the *Mean Field Degree Approximation* with networks.

We can rewrite some quantities using some empirical laws found when analyzing air traveling data: catchment populations, number of connections for every airport etc. The number of individuals resident in the patch k is:

$$N_k = N_0 k^\phi$$

---

[3]Actually this last approximation does not hold for $COVID19$, since it was able to spread given that its R_0 was high.

[4]Colizza & Vespignani, PRL 2007, JTB 2008

Whereas the flux between two nodes k and k':

$$w_{kk'} = w_0(kk')^\theta$$

therefore the **probability** of **travelling** from a node of degree k to a node of degree k' is:

$$p_{kk'} = \frac{w_0}{N_0} \frac{(kk')^\theta}{k^\phi} \tag{7}$$

where the ϕ, θ are empirical values. Once we have as **input** the **degree distribution** $P(k)$, we can try to find the number of diseased subpopulations at generation n, with k mobility connections, namely D_k^n.
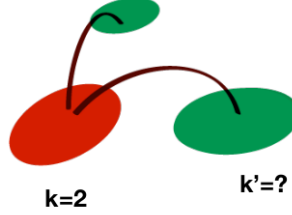

Figure 5: A patch that has $k = 2$ connections was infected. We want to compute what is the probability that it is connected to a disease-free patch of degree k'.

We want now to derive the invasion equation as before, but this time in a heterogeneous mean field approach framework. The question to be replied is therefore the following: *what is the probability that an infected patch with degree k' is connected to a disease-free patch of degree k.* The answer to this question is the following:

$$D_k^n = \sum_{k'} D_{k'}^{n-1} (k' - 1)P(k|k') \left(1 - \sum_{m=0}^{n-1} \frac{D_k^m}{V_k}\right) (1 - P_{ext}(\lambda_{k'k})) \tag{8}$$

which looks like 3, except for the sum over all possible degrees k' that was introduced since we are dividing our patches in same-degree classes. Let us analyze every term: the $k' - 1$ is the number of mobility connections through which the seeding might potentially occur. This is multiplied by the probability that contact has degree k, given we are starting from a node with degree k': explicitly *in networks* when we pick a node at random $P(k|k') = k\frac{P(k)}{\langle k \rangle}$. In this step, we have assumed that the network is uncorrelated. This however implies that when we are making connections at random we are more likely to connect to hubs, leading to the so called *friendship paradox*, a.k.a. "rich gets richer", where nodes that have already many connections tend to accumulate more. These last two terms are given by the **topology** of the network. The second last term (red) is equal to the one previously obtained: it is indeed the probability that the contact patch belonging to k class is disease-free. Last term (orange) is, as before, the probability for the epidemic to not get extinct before the global outbreak. One should note that, as previously done, this last term can be approximated to $\lambda_{k'k}(R_0 - 1)$ assuming this is a *branching process*, and moreover for heterogeneous networks:

$$\lambda_{k'k} = \frac{w_0(kk')^\theta}{N_0 k^\phi} \frac{\alpha}{\mu} (N_0 k^\phi) = w_0(kk')^\theta \frac{\alpha}{\mu}$$

Finally, 8 can be rewritten as it follows:

$$D_k^n = (R_0 - 1)\frac{\alpha w_0}{\mu} \frac{k^{1+\theta}P(k)}{\langle k \rangle} \sum_{k'} D_{k'}^{n-1}(k' - 1)k'^\theta \tag{9}$$

If we define the highlighted factor in red as Θ^{n-1} for a more compact writing, we can multiply both rhs and lhs of Eq. 9 by $\sum_k (k-1)k^\theta$ thus obtaining:

$$\Theta^n = (R_0 - 1)\frac{\alpha w_0}{\mu}\frac{\left\langle k^{2+2\theta}\right\rangle - \left\langle k^{1+2\theta}\right\rangle}{\langle k \rangle}\Theta^{n-1} \tag{10}$$

In this way we obtain a function that is monotone as well and has the property that the epidemic spreads if Θ^n is greater than Θ^{n-1}. But, actually, this how R_* works, so we have found an expression for the **invasion potential**:

$$R_* = (R_0 - 1)\frac{\alpha w_0}{\mu}\frac{\left\langle k^{2+2\theta}\right\rangle - \left\langle k^{1+2\theta}\right\rangle}{\langle k \rangle} > 1 \tag{11}$$

If the last condition holds, then the epidemic will spread. As one can easily note, R_* is a growing function of R_0, the overall traffic rescaling and average number of connections (w_0), epidemic attack rate α, infectious duration ($\tau = \mu^{-1}$) and, finally, of the **moments of the degree distribution and its fluctuations** which are very large for random networks: $\left\langle k^{2+2\theta}\right\rangle - \left\langle k^{1+2\theta}\right\rangle \approx 7 \cdot 10^4$ while $\langle k \rangle = 10$. In this way, the network topology is indeed *helping* and *favoring* the spread of the disease (see Fig. 6)! One should keep in mind that we are dealing with travelling individuals as integers, and the stochasticity relies in the process of travelling, which might either happen or not (Bernoulli process). However, if they are treated as integers we are not sure that outbreak will happen for sure. Indeed, this would happen when dealing with continuous variables (i.e. fraction of individual) which would be quite unrealistic.
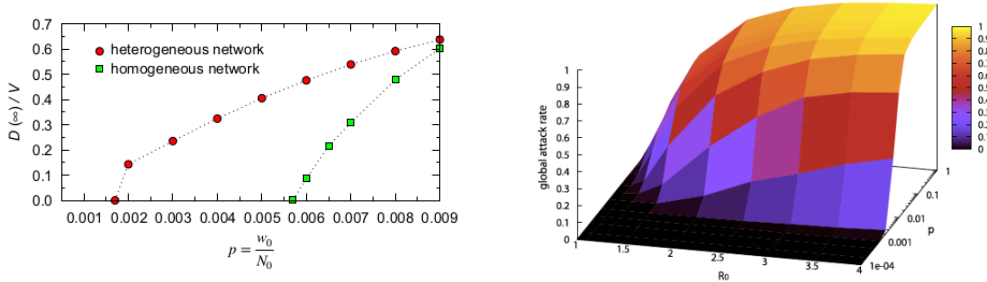

Figure 6: **Left:** Network topology favors the spreading of the disease and its global outbreak. **Right:** Attack rate in function of different traveling probabilities and the basic reproductive ratio R_0.
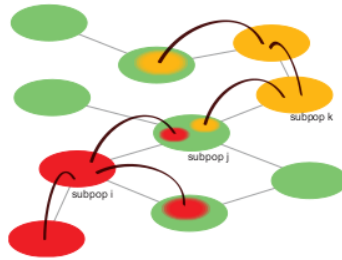
0.2 Spatial spread of competing diseases

Figure 7: Two strains present in the same metapopulation network.

We will present now an application to understand better what we have explained so far: let us assume that there are two **competing pathogens** that are in the

same metapopulation system. The **compartmental** model[5] that we obtain is the one in figure 8: there are two strains of a certain disease present, and one individual might be infected by either one of them. Once he recovers, he acquires the **full-cross immunity**, that is to say that he cannot be infected by both of them any more. The main difference between the two is that the infectious period of one is actually **faster** than the others ($\tau_{slow} > \tau_{fast}$), whereas they share the **same** R_0.
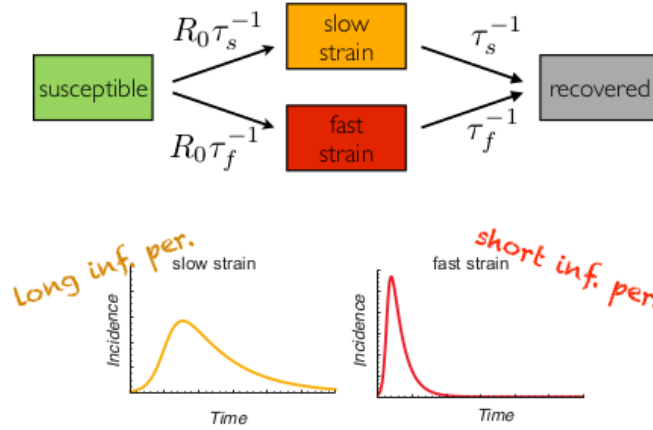

Figure 8:
Top. Compartmental model for two competing disease with same R_0 and different infectious time τ. Once one recovers, he acquires the full cross immunity. **Bottom.** Different epidemic timescales for the fast and slow strain of the disease.

Since the two strains have actually different infectious periods, they lead to **different infectious dynamics**. Indeed, τ affects both the slope of the *exponential growth*, as well as the *peak* and the presence of the disease at longer times. For instance, keeping $R_0 \propto \beta/\mu$ **fixed**, if τ is small we stay infected for less time but the transmissibility β must be actually higher, hence the spreading explodes. On the other hand, if τ is large, we stay infected more and therefore the disease has more time to spread despite the lower β. At the end of the epidemic, we will have reached the same amount of population but ended up with very different dynamics.

If we let the two different disease spread in a **well-mixed population** with the same assumptions as before: the faster strain will infect faster much more people, being the transmissibility higher. In a certain sense, the "slow" disease will *die out*.

Let us now **introduce**, under the same assumptions of two competing diseases one faster than the others with fixed R_0, the **metapopulation network**. The probability to travel from a patch to another is p, regardless of the patches. The results of the simulation are depicted in figure 9.

At the end of simulation there might be a little chance for the two strains to coexist if they reach the same patch, however, more likely the fast strain will make the other die out when the *probability of travelling* is *higher*. On the other hand, if the *travelling probability* is *less* they even might not encounter each other, and therefore the slower strain is favored to last longer and still be present at the end of simulation time. We can find a **probability** p_c for which the two diseases show a *crossover*, having the same prevalence. One should note that, for higher traveling probability, we approach the *homogeneous mixing regime*.

Recalling what we have discussed, we can bring what is at the scale of individuals, namely the infection duration μ^{-1}, R_0 and the logistic curve $I(t) \sim e^{\mu(R_0-1)t}$ to a

---
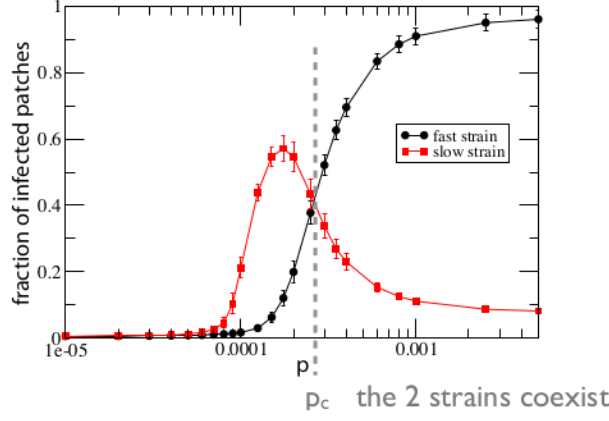
[5]Poletto, PLOS Comp Biol PRL 2007, JTB 2008.

Figure 9: Numerical simulation results for two competing strains in metapopulation networks for different value of travelling probability.

patches scale. The timescale will be the *outbreak duration* T and R_* will be:

$$R_* = (\langle k \rangle - 1)\frac{\alpha p N}{\mu}(R_0 - 1) \tag{12}$$

while the **number of diseased** patches is $D(t) \sim e^{\frac{1}{T}(R_*-1)t}$. One should note that, since we are assuming that all patches are the same, the outbreak duration T on average is the same for every patch.

According to the fact that R_* is an increasing function of μ^{-1}, we have that the invasion potential $R_*^s > R_*^f$. However, for *large p* it holds that both $R_*^s, R_*^f \gg 1$, but the faster strain is able to reach more rapidly new patches. Instead, for *small p*, we have $R_*^s > R_*^f$ and the slower strain is more able to percolate through the system.

We want now to understand how the probability p_c for the **crossover** changes with respect of R_0. We need to solve the following equation:

$$\frac{D_s(t)}{D_f(t)} \sim e^{\left(\frac{R_*^s-1}{T_s} - \frac{R_*^f-1}{T_f}\right)} = 1 \tag{13}$$

The solution is plotted in the left image of Fig. 10. It is given by replacing the value for R_*, keeping also in mind that in homogeneous mixing $T_s = T_f$. In this way, it becomes reasonable to be solved, otherwise a graphical approach would have been needed.
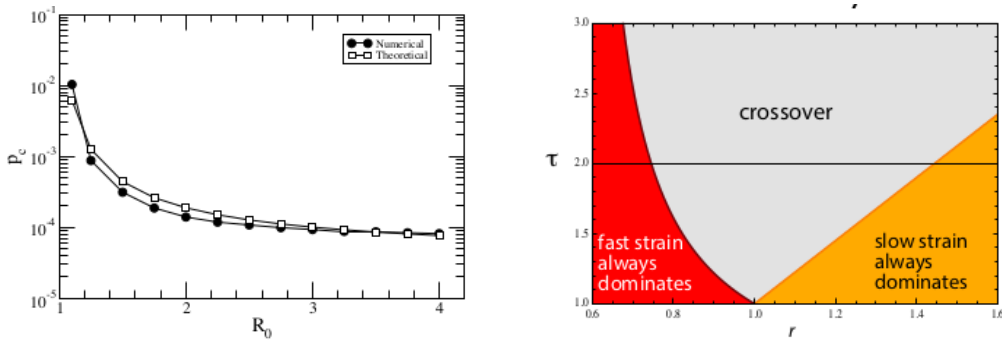

Figure 10:
Left: Crossover probability for different Basic reproductive rate R_0.
Right: Two competing strains in homogeneous metapopulations with different τ and different R_0.

Now let us **drop** the assumption that R_0 is **fixed**, and allow for different R_0[6]:

$$R_0^s = r R_0^f \tag{14}$$

where r is the ratio between the two R_0 and τ (see Fig. 10) is the ratio between the two infectious periods. Now, things start to change, indeed there are some regions where **regardless** of the level of **mobility** either one of the two strain dominates: this happens quite always for limiting cases wrt r. For *small r* fast strain *always* dominates regardless the time of infection, while for *large r* slow strain always dominates when τ is not so large. There is in addition a **crossover** region where the two may coexist and **mobility does** matter to determine which one spreads more than the other.

Analytically, if we introduce the exponential growth in homogeneous mixing as $G = \mu(R_0 - 1)$, we see that there might be two different cases:

$$G^s > G^f \implies R_*^s > R_*^f \quad or \quad R_*^f > R_*^s \implies G^f > G^s \tag{15}$$

that define our colored areas where either one strain dominates (the one that comes with larger R_*).

However it is present a third case: if $G^f > G^s$ *and* $R_*^s > R_*^f$ we may observe **crossover**. More particularly:

$$r > 1 \to \quad G^f > G^s \quad \Rightarrow \quad r R_0^f - 1 < \tau \left(R_0^f - 1 \right)$$

$$r < 1 \to \quad R_*^s > R_*^f \quad \Rightarrow \quad \tau \alpha_s \log(R_0^s) > \alpha_f \log \left(R_0^f \right)$$

Summarizing: depending on the relation between the pathogens' traits, mobility can play a determinant role or be non-influential for the outcome of the competition. In the space of epidemiological parameters, there exists a region for which lowering the traveling probability induces a cross over from the fast strain dominance to the slow strain one. This behaviour is determined by the trade-off between epidemic growth and potential for spread at the spatial level, the former being an advantage in a well mixed population while the second being relevant in a sparse environment with intermediate or low mobility coupling.

---

[6]Poletto et al. Sci Rep 2015.