Spreading of Persistent Infections in Heterogeneous Populations

J. Sanz,[1] L. M. Floría,[1,2] and Y. Moreno[1,3, *]

[1]*Institute for Biocomputation and Physics of Complex Systems (BIFI), University of Zaragoza, Zaragoza 50009, Spain*
[2]*Departamento de Física de la Materia Condensada, Universidad de Zaragoza, Zaragoza E-50009, Spain*
[3]*Departamento de Física Teórica. Universidad de Zaragoza, Zaragoza E-50009, Spain*
(Dated: March 22, 2010)

Up to now, the effects of having heterogeneous networks of contacts have been studied mostly for diseases which are not persistent in time, i.e., for diseases where the infectious period can be considered very small compared to the lifetime of an individual. Moreover, all these previous results have been obtained for closed populations, where the number of individuals does not change during the whole duration of the epidemics. Here, we go one step further and analyze, both analytically and numerically, a radically different kind of diseases: those that are persistent and can last for an individual's lifetime. To be more specific, we particularize to the case of Tuberculosis' (TB) infection dynamics, where the infection remains latent for a period of time before showing up and spreading to other individuals. We introduce an epidemiological model for TB-like persistent infections taking into account the heterogeneity inherent to the population structure. This sort of dynamics introduces new analytical and numerical challenges that we are able to sort out. Our results show that also for persistent diseases the epidemic threshold depends on the ratio of the first two moments of the degree distribution so that it goes to zero in a class of scale-free networks when the system approaches the thermodynamic limit.

I. INTRODUCTION

Disease spreading has been the subject of intense research since long time ago [1–3]. Our current knowledge comprises mathematical models that have allowed to better understand how an epidemic spreads and to design more efficient immunization and vaccination policies [1–3]. These models have gained in complexity in recent years capitalizing on data collections which have provided information on the local and global patterns of relationships in the population [4–6]. In particular, with the advent of modern computational resources and tracking systems, it is now feasible to contact-trace the way the epidemic spreads or at least to predict the paths that a given pathogen might follow. In this way, some of the assumptions at the basis of the theoretical models that were difficult to test -the backbone through which the diseases are transmitted- are now more accurately incorporated into epidemiological models [7–11].

Strikingly, the systems on top of which diseases spread show common nontrivial topological and statistical properties [12, 13]. A large number of networks of contacts in real-world social, biological and technological systems have been found to be best described by the so-called scale-free (SF) networks. In SF networks, the number of contacts or connections of a node with other nodes in the system, the degree (or connectivity) k, follows a power law distribution, $P(k) \sim k^{-\gamma}$. Recent studies have shown that the SF topology has a great impact on the dynamics and function of the system under study [12–15]. The reason is that, at variance with homogeneous or regular networks, SF architectures are a limiting case of heterogeneity where the connectivity fluctuations diverges if $2 < \gamma \leq 3$ as the system size tends to infinity (the thermodynamic limit). This means that there are nodes in the network which has an eventually unbounded number of connections compared to the average degree. Examples of such networks include the Internet [16, 17], the world-wide-web (WWW) [18], food-webs, and metabolic or protein networks [13, 19].

In the context of disease spreading, SF networks lead to a vanishing epidemic threshold in the limit of infinite population when $\gamma \leq 3$ [20–23]. This is because the ratio $\langle k \rangle / \langle k^2 \rangle$ determines the epidemic threshold above which the outbreak occurs. When $2 < \gamma \leq 3$, $\langle k \rangle$ is finite while $\langle k^2 \rangle$ goes to infinity, that is, the transmission probability required for the infection to spread goes to zero. Moreover, the previous result holds both for the Susceptible-Infected-Susceptible (SIS) and Susceptible-Infected-Removed (SIR) epidemiological models [20–22].

In this paper, we will deal with a different kind of diseases -those that are persistent in time and shows a latent period that can be as large as an individual's lifetime. Our first aim is to enlarge the epidemiological framework for complex networks reported previously for the SIS model [20] and proposed as well for the SIR model [21, 22] by integrating the spreading dynamics of persistent diseases within it. With this purpose, we consider a variation of the Susceptible-Exposed-Infected-Removed model [24] on complex heterogeneous networks. As we will see, this kind of dynamics introduces new challenges essentially different

to other successfully treated before. In particular, as the latent period is high enough, we shall work with an open system in which new individuals are born and others dead for causes not directly related to the spreading of the disease. We present novel analytical and numerical methods that allow us to obtain the epidemic threshold for this kind of disease in heterogeneous populations. Moreover, we introduce a numerical method that is well-suited to deal with the kind of problems we face. Our results point out that also for persistent infections the virtually unbounded fluctuations of the degree distribution have an important enhancing effect on the epidemic spreading.

II. THE MODEL

To be more precise and without loss of generality, we will particularize our model to one case of persistent infection -the most threatening one which is tuberculosis (TB). TB is an old disease whose world-wide prevalence had been diminishing even before vaccination and prophylaxis strategies were firstly accomplished [25–27]. Its recent return in developing countries, mainly in Southeast Asia, have attracted renewed interest in it. The current world estimate of prevalence is about 33% while the number of deaths per year that it is causing reaches more than 3 million people [28]. Depending of the kind and the intensity of immune response that the host immune system performs after initial infection with M. tuberculosis bacillus, the individual can suffer latent infection, (in which the bacteria are under a growth-arrest regime and the individual neither suffer any clinical symptom nor becomes infectious) or active infection, where the host suffers clinical symptoms and can transmit the pathogen by air [29, 30]. Latently infected individuals can, generally after an immune-depression episode, reach the active phase. Estimating the probability of developing direct active infection after a contact, or alternatively, the lifetime's risk for a latent infected individual to evolve into the active phase, are not easy tasks. However, it is generally accepted that only 5-10% of the infections directly produce active TB [29, 30], while the ranges concerning the estimation of typical "half-life" of latent state rounds about 500 years [24].

The spreading dynamics of TB like diseases has been studied in recent years. However, to the best of our knowledge, these works assume the homogeneous mixing hypothesis, that is, a perfectly homogeneous system in which all individuals are dynamically equivalent. As mentioned above, many of the systems on top of which diseases spread, are better described by scale-free connectivity patterns. Therefore, in what follows, our main objective will be to assemble a basic model fitted for tuberculosis spreading that firstly takes into account the heterogeneity in the distribution of the networks of contacts. We note that the increasingly alarming situation about TB epidemiology evidences the need to increase the effort in TB research in a global way. In the context of the study of its epidemiology, new models must be developed in order to gain predictive skills, incorporating the recent theoretical advances referring to disease spreading on complex heterogeneous substrates as well as meta-population approaches and new computational tools for numerical analysis and simulation. In this sense, ours is a first contribution that addresses one of the most important parameters in epidemiological description: the epidemic threshold.

Let us then introduce our model. We consider that individuals in the population are compartmentalized into three groups: healthy -$U(t)$-, infected but not infectious -*or latently infected* $L(t)$- and sick individuals $T(t)$ which are infected and are infectious as well. The transition between these subpopulations proceeds in such a way that a healthy individual acquires the bacteria through a contact with an infectious subject with probability λ. In its turn, this newly infected individual may develop the disease directly with probability p. However, the most common case is the establishment of a dynamic equilibrium between the bacillus and the host's immune system, which allows the survival of the former inside the latter. When this happens, newly infected individuals become latently infected, because despite harboring the bacteria in blood, neither becomes sick nor is able to infect others.

On the other hand, after a certain period of time (which may be several years) and usually following an episode of immuno-suppression, the balance between the bacterium and its host can be broken. In this case, the bacteria grow and the individual falls ill beginning to develop the clinical symptoms of the disease. In addition, if the infection attacks the lungs (pulmonary TB), the bacillus is present in the sputum, making the guest infectious.

The dynamics of the disease, in a well mixed population, is then described by the following system of nonlinear differential equations:

$$\frac{dU(t)}{dt} = bN(t) - \lambda\beta U(t)\frac{T(t)}{N(t)} - \mu U(t),$$

$$\frac{dL(t)}{dt} = (1-p)\lambda\beta U(t)\frac{T(t)}{N(t)} - (\mu + r)L(t), \tag{1}$$

$$\frac{dT(t)}{dt} = p\lambda\beta U(t)\frac{T(t)}{N(t)} + rL(t) - (\mu + \mu_{tb})T(t),$$

in which:

$N(t) = U(t) + L(t) + T(t)$ represents the total population at time t,

β is the number of contacts per time unit,

λ is the probability that the bacteria is transmitted to a new host after a contact with an infectious subject,

b is the birth rate per capita and per unit time,

μ is the natural death rate per capita and unit time,

μ_{tb} is the rate of disease-related deaths per capita and unit time,

r is the transition frequency of latent infection (i.e., the probability that a latently infected individual becomes infectious),

with the closure relationship:

$$\frac{dN(t)}{dt} = (b - \mu)N(t) - \mu_{tb}T(t). \tag{2}$$

The model above is a variation of the archetypal SIR model to which a fourth class has been added (latency class L). This kind of model has been largely treated in the literature in its well-mixed version, and it is frequently referred as SEIR model. As a first step, in this work, we will identify the removed individuals mostly with dead ones, and therefore we do not consider the possibility of natural or medical recovery (this simplification is in part justified by the large latency period of infected individuals and the constant flow of newborns into the system). A more refined model would consist of introducing such eventual recovery fluxes in the model, as well as the possibility of further relapses (the so called endogenous reactivation). These phenomenologies might be important mainly for diseases (like TB) for which the only feasible treatment in many areas consists of supplying large series of antibiotics. Thinking on the tuberculosis case, further refinements, like the inclusion of varieties of less infectious extra-pulmonary diseases [31], could also have consequences on the disease's dynamics.

III. STRUCTURED POPULATIONS

A. Dynamics

The previous system of differential equations describes the dynamics of the epidemics in the well-mixed case. However, as argued above, the number of contacts of a given individual in a population can vary, which is reflected in an heterogeneous distribution of the number of contacts in the system. To account for this fact, we next consider a structured population described by a connectivity distribution $P(k)$. The system of Eqs (1) has to be modified accordingly. Assuming that all individuals with the same number of contacts, i.e., belonging to the same connectivity class k, are dynamically equivalent, the new system of differential equations are formulated for each degree class. Therefore, for a structured population, we have that:

$$N_k(t) = P(k)N(t), \tag{3}$$

with:

$$U_k(t) + L_k(t) + T_k(t) = N_k(t). \tag{4}$$

Moreover, it is convenient to express the previous equations in terms of densities, also defined within each connectivity class:

$$u_k(t) = \frac{U_k(t)}{N_k(t)},$$

$$l_k(t) = \frac{L_k(t)}{N_k(t)}, \tag{5}$$

$$t_k(t) = \frac{T_k(t)}{N_k(t)},$$

so that the following closure relation for any value of k is verified:

$$u_k(t) + l_k(t) + t_k(t) = 1 \qquad \forall(k, t). \tag{6}$$

On the other hand, the probability Θ that any given link points to an infectious individual is given by:

$$\Theta(t) = \frac{\sum_k k T_k(t)}{\sum_k k N_k(t)} = \frac{\sum_k k P(k) t_k(t)}{\langle k \rangle}, \tag{7}$$

which leads to the following set of equations that describes the dynamics within each connectivity class:

$$\frac{dU_k(t)}{dt} = bP(k)N - \lambda k \Theta(t) U_k(t) - \mu U_k(t),$$

$$\frac{dL_k(t)}{dt} = (1-p)\lambda k \Theta(t) U_k(t) - (\mu + r) L_k(t), \tag{8}$$

$$\frac{dT_k(t)}{dt} = p\lambda k \Theta(t) U_k(t) + r L_k(t) - (\mu + \mu_{tb}) T_k(t).$$

Finally, the number of individuals with connectivity k evolves according to:

$$\frac{dN_k(t)}{dt} = (b - \mu)N_k(t) - \mu_{tb} T_k(t) = (b - \mu - \mu_{tb} t_k) N_k(t). \tag{9}$$

At this point, and building on the previous equation, it is important to point out a feature of the model: the influence of the infection dynamics on the connectivity distribution $P(k)$. First, if we add the above equation for all k, we obtain that the total population evolves as:

$$\frac{dN(t)}{dt} = (b - \mu)N(t) - \mu_{tb} \sum_k T_k(t) = \left(b - \mu - \mu_{tb} \sum_k P(k) t_k \right) N(t). \tag{10}$$

However, if we substitute $N_k(t) = P(k)N(t)$ directly into Eq. (9) and assume $P(k)$ to be constant, we would arrive to:

$$P(k)\frac{dN(t)}{dt} = P(k)\left(b - \mu - \mu_{tb} t_k\right) N(t).$$

The last expression is only compatible with Eq. (10) under the unrealistic assumption that all connectivity classes have the same proportion of sick individuals. We must therefore assume that the distribution of connectivity is also a function of time: $P(k,t)$, and therefore:

$$\frac{dN_k(t)}{dt} = \frac{d[P(k,t)N(t)]}{dt} = N(t)\frac{dP(k,t)}{dt} + P(k,t)\frac{dN(t)}{dt}, \tag{11}$$

so, if we substitute Eq. (11) into Eq. (9) we get:

$$N(t)\frac{dP(k,t)}{dt} + P(k,t)\frac{dN(t)}{dt} = P(k)\left(b - \mu - \mu_{tb} t_k\right) N(t),$$

expression from which, if we replace $dN(t)/dt$ from Eq. (10), we get the temporal evolution of $P(k,t)$ as:

$$\frac{dP(k,t)}{dt} = -P(k,t)\mu_{tb}\left[t_k(t) - \langle t_k \rangle(t)\right], \tag{12}$$

where

$$\langle t_k \rangle(t) = \sum_k P(k,t) t_k(t). \tag{13}$$

Reformulating the equations in terms of densities using the definitions of the densities given above, Eqs. (8) become

$$\frac{du_k(t)}{dt} = b - u_k(t)(b + \lambda k \Theta(t) - \mu_{tb} t_k(t)),$$

$$\frac{dl_k(t)}{dt} = (1-p)\lambda k \Theta(t) u_k(t) - (b + r)l_k(t) + \mu_{tb} l_k(t) t_k(t), \tag{14}$$

$$\frac{dt_k(t)}{dt} = p\lambda k \Theta(t) u_k(t) + r l_k(t) - (b + \mu_{tb})t_k(t) + \mu_{tb} t_k(t)^2.$$

B. Evolution of the degree distribution

At this point it is appropriate to point out one aspect that will hinder any numerical characterization of the epidemic threshold. Our main goal will be to calculate both numerically and analytically the critical value λ_c beyond which the population presents an endemic proportion of sicks individuals. However, we expect λ_c to be dependent on the ratio $\frac{\langle k \rangle}{\langle k^2 \rangle}$, which is in turn a function of the connectivity distribution $P(k)$. The degree distribution, as previously shown, changes in time as the dynamics of infection progresses. As we should see, we can handle this time dependence analytically, but we should be forced to design a simulation method to account for the rate of births and deaths and the effects of these two processes on the degree distribution.

The aforementioned features might lead to a situation in which the infection dynamics would modify the underlying structure of the network through which the disease is being spread. Therefore, λ_c could also vary as one expects it to be intrinsically related to the first two moments of a seemingly time-dependent degree distribution. The reason why we consider the distribution of contacts per unit time as heterogeneous, even for the current airborne-transmitted disease is based on the observation that the number of contacts a person can have per unit of time is subjected to two sources of heterogeneity. Firstly, what we can call *geo-demographic, macroscopic* heterogeneity, in which the number of contacts depends on the population density in the region in which an individual inhabits. Secondly, at a more *individual, microscopic* level, the heterogeneity arises because the number of contacts depends, in a region of constant population density (i.e., a town or neighborhood in a city), on the daily activity pattern of the individual within that region. These two factors define, ultimately, the function $P(k)$. Having that said, the assumption implicitly incorporated in the first equation of system (8) does not hold. Note that this equation implies that the connectivity of individuals is hereditary and therefore that the number of births within each k class equals the birth rate times the number of individuals within each k class, $N_k = P(k, t)N$.

The above situation would be equivalent to assume that the dynamics of the disease being studied is the only one that influences the demographic structure of a population, which is not true since it is clear that there are countless cultural, economic and social factors that ultimately define the above two levels of heterogeneity. We therefore assume in what follows that the newborns of each generation are distributed among the k classes according to an invariant distribution function, which we further assume to be the initial degree distribution of the original network: $P(k, t_o)$. As we shall see, this assumption, besides being more plausible, has the advantage that makes the connectivity distribution to be roughly stable, and so will be the critical value λ_c.

So, we have the following reformulation of the system of differential equations (8):

$$\frac{dU_k(t)}{dt} = bP(k, t_o)N - \lambda k\Theta(t)U_k(t) - \mu U_k(t),$$

$$\frac{dL_k(t)}{dt} = (1 - p)\lambda k\Theta(t)U_k(t) - (\mu + r)L_k(t), \tag{15}$$

$$\frac{dT_k(t)}{dt} = p\lambda k\Theta(t)U_k(t) + rL_k(t) - (\mu + \mu_{tb})T_k(t),$$

with the definition of the number of individuals in each class of connectivity:

$$N_k(t) = N(t)P(k, t), \tag{16}$$

and inside each class:

$$U_k(t) = N_k(t)u_k(t),$$

$$L_k(t) = N_k(t)l_k(t), \tag{17}$$

$$T_k(t) = N_k(t)t_k(t).$$

Now the total population within each connectivity class verifies:

$$\frac{dN_k(t)}{dt} = bNP(k, t_o) - \mu NP(k, t) - \mu_{tb}T_k(t), \tag{18}$$

so that, if we add in k, the last modification has no effect on the variation of the total volume of the population. The temporal evolution of the degree distribution is now given as:

$$\frac{dP(k, t)}{dt} = b\left[P(k, t_o) - P(k, t)\right] - P(k, t)\mu_{tb}\left[t_k(t) - \langle t_k \rangle(t)\right]. \tag{19}$$

Finally, writing the equations in terms of the densities we get

$$\frac{du_k(t)}{dt} = b\frac{P(k,t_o)}{P(k,t)}\left(1 - u_k(t)\right) - u_k(t)\left(\lambda k\Theta(t) - \mu_{tb}t_k(t)\right),$$

$$\frac{dl_k(t)}{dt} = (1-p)\lambda k\Theta(t)u_k(t) - \left(b\frac{P(k,t_o)}{P(k,t)} + r\right)l_k(t)+$$

$$+ \mu_{tb}l_k(t)t_k(t), \tag{20}$$

$$\frac{dt_k(t)}{dt} = p\lambda k\Theta(t)u_k(t) + rl_k(t) - \left(b\frac{P(k,t_o)}{P(k,t)} + \mu_{tb}\right)t_k(t) + \mu_{tb}t_k(t)^2.$$

C. Characterization of the equilibrium points.

The previous set of differential equations tells us how the different densities of interest evolves within each connectivity class. Their corresponding macroscopic quantities are defined as

$$\langle u\rangle(t) = \sum kP(k,t)u_k(t),$$

$$\langle l\rangle(t) = \sum kP(k,t)l_k(t), \tag{21}$$

$$\langle t\rangle(t) = \sum kP(k,t)t_k(t),$$

where $\langle u\rangle(t)$, $\langle l\rangle(t)$ and $\langle t\rangle(t)$ are the mean densities of healthy, latent and sick individuals, respectively.

Let us now go one step further and characterize the equilibrium points. The magnitudes of interest are the average densities, so that an equilibrium point $(\langle u\rangle^*, \langle l\rangle^*, \langle t\rangle^*)$ must verify by definition:

$$\left(\frac{d\langle u\rangle^*}{dt}, \frac{d\langle l\rangle^*}{dt}, \frac{d\langle t\rangle^*}{dt}\right) = (0,0,0).$$

We also impose a further constraint which is that the degree distribution of the network is stationary, that is:

$$\frac{dP(k)^*}{dt} = 0 \qquad \forall k.$$

At this point one must ask whether macroscopic stability also implies stability within each connectivity class. The answer is yes, if we also demand stability of the degree distribution. Admittedly, if we equate expression (19) to zero and solve for the stationary $P(k,t)^*$ we get:

$$P(k,t)^* = \frac{bP(k,t_o)}{b + \mu_{tb}(t_k^* - \langle t\rangle^*)},$$

which shows that this value depends on the microscopic scale t_k. Therefore, the stability of the degree distribution imposes a stationary condition on t_k for all k, which in its turns extends to the other densities l_k^* and u_k^*. Hence, we have:

$$\left(\frac{du_k^*}{dt}, \frac{dl_k^*}{dt}, \frac{dt_k^*}{dt}\right) = (0,0,0) \qquad \forall k.$$

The above condition is trivially satisfied for the solution $(u_k^*, l_k^*, t_k^*) = (1,0,0)\ \forall k$, which leads to a degree distribution exactly as the initial distribution. We next analyze the stability of this solution, which shall allow us to characterize the epidemic threshold.

D. Epidemic threshold

As stated before, in this section we will study the stability of the solution $(u_k^*, l_k^*, t_k^*) = (1,0,0)\ \forall k$. At this point, as no latent or infected individuals are introduced in the network, the degree distribution does not change in time; so that $P(k)^* = P(k,t) = P(k,t_o)$. This situation allows to work with the system of differential equations given by (14) instead of working with the more general case given by the system (20).

1. Case $p = 1$

For simplicity and to gain some preliminary insight into the problem, we first study the case $p = 1$, which means that there is no latent phase (i.e., the latent subpopulation disappears, $l_k = 0 \; \forall k$). Using $u_k + t_k = 1$ we get,

$$\frac{du_k}{dt} = b - u_k(b + \lambda k \Theta - \mu_{tb}) - \mu_{tb} t_k^2,$$

where we have omitted temporal dependences, as we will do from now on. Looking for the stationary solution, we have that the condition $\frac{du_k}{dt} = 0$ implies:

$$u_k = -\left(\frac{1}{2\mu_{tb}}\right)\left(b + \lambda k \Theta - \mu_{tb} \pm \sqrt{(b + \lambda k \Theta - \mu_{tb})^2 + 4b\mu_{tb}}\right),$$

from which the meaningful solution is the one with the negative sign. The previous expression is consistent with the meaning of u^* since we recover the expected result $u^* = 1$ when $\Theta = 0$. Moreover, if we calculate the derivative with respect to Θ we get:

$$\frac{du_k^*(\Theta)}{d\Theta} = \frac{\lambda k}{2\mu_{tb}}\left(-1 + \frac{b + \lambda k \Theta - \mu_{tb}}{\sqrt{(b + \lambda k \Theta - \mu_{tb})^2 + 4b\mu_{tb}}}\right) < 0,$$

which guarantees that u^* will always be less than unity and therefore is a real, valid solution. The study of the value of Θ in the steady state help us to identify the epidemic threshold. We write:

$$\Theta^* = \frac{1}{\langle k \rangle}\sum_k k P(k) t_k^* = 1 - \frac{1}{\langle k \rangle}\sum_k k P(k) u_k^*,$$

which, after substituting u_k^* for its value, leads to:

$$\Theta^* = f(\Theta) = \frac{1}{2} - \frac{b}{2\mu_t b} + \frac{\lambda \langle k^2 \rangle}{2\mu_{tb}\langle k \rangle}\Theta -$$

$$- \frac{1}{2\mu_{tb}\langle k \rangle}\sum_k k P(k)\sqrt{(b + \lambda k \Theta - \mu_{tb})^2 + 4b\mu_{tb}}.$$

The graphical interpretation of the above equation indicates that the existence of an equilibrium point in which $\Theta^* > 0$ is equivalent to the existence of a point at which $f(\Theta)$ crosses the bisector of the first quadrant. Evaluating the second derivative of $f(\Theta)$ one gets:

$$\frac{d^2 f(\Theta)}{d\Theta^2} = \frac{-2b\lambda^2}{\langle k \rangle}\sum_k \frac{P(k)k^3}{[(b + \lambda k \Theta - \mu_{tb})^2 + 4b\mu_{tb}]} < 0,$$

which ensures that the condition for the existence of such intersection is reduced to:

$$\left(\frac{df(\Theta)}{d\Theta}\right)_{\Theta = 0} = \frac{\lambda \langle k^2 \rangle}{\langle k \rangle}\frac{1}{b + \mu_{tb}} = 1,$$

from which the epidemic threshold is derived as:

$$\lambda_c = \frac{(b + \mu_{tb})\langle k \rangle}{\langle k^2 \rangle}.$$

Note that apart from the factor $(b + \mu_{tb})$, the previous result, formally coincides with the epidemic threshold of the SIR model.

2. Case $p \neq 1$

This is a somewhat more involved case. For structured populations, the resolution of the system of differential equations (20) cannot be done explicitly. We next find the epidemic threshold for the case $p \neq 1$ using two approaches. On one hand, we study the time derivative of Θ. On the other hand, we will also make use of the singularity of the Jacobian at the point $(u_k, l_k, t_k) = (1, 0, 0)$ to argue that the expression for the critical threshold is given by:

$$(\lambda)_c = \frac{\langle k \rangle}{\langle k^2 \rangle}\frac{(r + b)(\mu_{tb} + b)}{pb + r}. \tag{22}$$

a. Time evolution of Θ in populations with a low number of sicks. A first approach to characterize the epidemic threshold in heterogeneous networks when $p \neq 1$ is to study the sign of the derivative of Θ at the onset of an epidemic outbreak. We consider an initially healthy population in which a small proportion of infectious individuals is introduced so that $t_k << 1 \; \forall k$. The derivative of Θ is:

$$\left(\frac{d\Theta}{dt}\right)_{\Theta\sim0} = \frac{\sum_k P(k)k\frac{dt_k}{dt}}{\langle k \rangle} + \frac{\sum_k t_k k\frac{dP(k)}{dt}}{\langle k \rangle} -$$

$$- \left[\frac{\sum_k P(k)kt_k}{\langle k \rangle}\right]\left[\frac{\sum_k k\frac{dP(k)}{dt}}{\langle k \rangle}\right]$$

which, after substitution of the values of the derivatives of $P(k,t)$ and $t_k(t)$ leads to:

$$\left(\frac{d\Theta}{dt}\right)_{\Theta\sim0} = \frac{\sum_k P(k)kl_k}{\langle k \rangle} + p\lambda\Theta\frac{\sum_k P(k)k^2 u_k}{\langle k \rangle} - (b + \mu_{tb})\Theta + \mu_{tb}\Theta^2.$$

At this point we make two simplifications. The first and most easily justifiable is to neglect the term Θ^2. The second is related to the presence of l_k in the above equation, that we have to transform in a dependency with respect to t_k. Specifically, we assume to be sufficiently close to the stationary point $(u_k, l_k, t_k) = (1, 0, 0)$ as to be able to assume that the three derivatives vanish. In other words, and focusing our attention on latent and sick classes, we assume that:

$$\left(\frac{dl_k}{dt}\right)_{\Theta\sim0} = (1-p)\lambda k\Theta u_k - (b+r)l_k + \mu_{tb}l_k t_k \simeq 0,$$

$$\left(\frac{dt_k}{dt}\right)_{\Theta\sim0} = p\lambda k\Theta u_k + rl_k - (b+\mu_{tb})t_k + \mu_{tb}t_k^2 \simeq 0,$$

from which:

$$l_k = \frac{(1-p)(b+\mu_{tb})\,t_k - (1-p)\mu_{tb}t_k^2}{r + pb - \mu_{tb}pt_k} = \frac{(1-p)(b+\mu_{tb})}{r + pb}t_k + O(t_k^2),$$

which allows to express the derivative of Θ as:

$$\left(\frac{d\Theta}{dt}\right)_{\Theta\sim0} = \Theta\left[\frac{r\,(1-p)(b+\mu_{tb})}{r + pb} - (b+\mu_{tb})+\right.$$

$$\left.+p\lambda\frac{\sum_k P(k)k^2 u_k}{\langle k \rangle}\right].$$

In the limit $u_k \simeq 1 \; \forall k$ the third term within brackets is the ratio $\langle k^2 \rangle/\langle k \rangle$, from which the epidemic threshold condition may be derived as:

$$\left(\frac{d\Theta}{dt}\right)_{\Theta\sim0} = \Theta\left[\frac{r\,(1-p)(b+\mu_{tb})}{r + pb} - (b+\mu_{tb}) + p\lambda_c\frac{\langle k^2 \rangle}{\langle k \rangle}\right] = 0,$$

finally leading to the expected expression for the threshold:

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}\frac{(r+b)(\mu_{tb}+b)}{pb+r}. \tag{23}$$

b. Analysis of the Jacobian While for well-mixed populations the condition of singularity of the Jacobian allows to get the epidemic threshold in a straightforward way, for heterogeneous populations the analysis of the Jacobian is a difficult task because Θ is a function of each and every one of the t_k's. This translates into the need of calculating a determinant whose order is three times the number of connectivity classes. What we can reasonably do is to verify is the threshold condition is verified for systems in which there are two or three different connectivity classes. In the first case in which only two different classes of
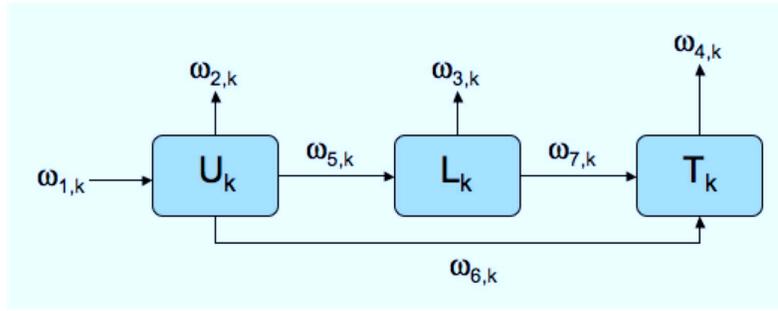
FIG. 1: Set of allowed transitions in the epidemic model within each connectivity class.

connectivity exist, the Jacobian is just a quite distasteful 6x6 determinant that, after some cumbersome and lengthy algebra, can be reduced to the expression:

$$J = b^2(b+r)(b+\mu_{tb}) \left[(b+r)(b+\mu_{tb}) + \frac{\lambda \langle k^2 \rangle}{\langle k \rangle}(pb+r) \right],$$

which equated to zero leads again to the previously obtained expression for the epidemic threshold. If we instead consider a population with three degree classes, the algebraic complexity of the problem largely increases as we now have to solve a determinant of size 9x9. However, we can proceed as before getting the following expression for the Jacobian:

$$J = b^3(b+r)^2(b+\mu_{tb})^2 \left[(b+r)(b+\mu_{tb}) + \frac{\lambda \langle k^2 \rangle}{\langle k \rangle}(pb+r) \right].$$

This leads us to the sensible conjecture that increasing the number of connectivity classes does not add new roots to the Jacobian, but it only would increase the degeneracy of the non-interesting solutions $b = 0$, $b = -r$ and $b = -\mu_{tb}$.

E. Numerical simulations

When designing numerical simulations to inspect the dynamics of the system under study, we have two difficulties not previously addressed in the literature. These numerical issues with which we have to deal come from the fact that we have a system that is simultaneously open and structured. As a result of dealing with an open system, new individuals are being added to the population at a rate given by the birth rate. Additionally, these new individuals must enter the network of contacts with a predefined connectivity. While deciding how many nodes our new individuals connect to is not a problem, it certainly is to decide what are those nodes the newcomers will be linked to, as this will impact the degree distribution in a nontrivial way. This is an unavoidable numerical complication that we should face relentlessly if the analytical calculations are to be compared with Monte Carlo simulations.

To this end, we have adapted a simulation method based on transition probabilities first proposed in [32] for SIR models in complex networks. The numerical approach considers all transitions between states that take place during the dynamical evolution of the subpopulations, defined by the system of differential equations (20). When dealing with structured populations, these transition rates depend, in general, on the connectivity class within which they occur. Moreover, within each k-class seven transitions are possible (see Fig. 1):

- Birth of healthy individuals,

- Natural death of healthy individuals,

- Natural death of latently infected individuals,

- Natural or disease-related death of sick individuals,

- Transition from a healthy to the latent state,

- Transition from a healthy to the sick (infectious) state,

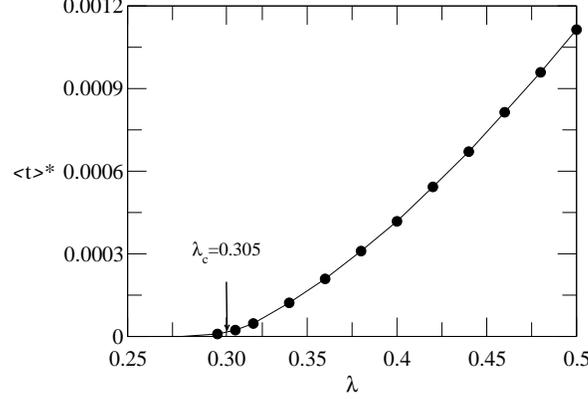- Transition from a latent to the sick (infectious) state.

FIG. 2: Stationary proportion of sick individuals as a function of λ ($\in [0.25, 0.5]$) for $p = 0.07$, $r = 0.002$, $\mu_{tb} = 0.2$, $\mu = 0.009$ and $b = 0.01$. The arrow marks the position of the epidemic threshold ($\lambda_c = 0.305$) as given by analytical calculations using the previous values for the parameters and the degree distribution. Error bars are smaller than symbol size. The initial size of the network, characterized by a degree distribution $P(k, t_o) \sim k^{-3}$, is $N_0 = 10^6$. Each point is an average over at least 1000 values of $\langle t \rangle^*$.

Each of these transitions is characterized by a characteristic transition rate $\omega_{i,k}$ that can be directly derived from the system of equations that characterizes the rate at which they occur within the class k as:

$$\omega_{1,k} = bNP(k, t_o)$$
$$\omega_{2,k} = \mu NP(k, t)u_k$$
$$\omega_{3,k} = \mu NP(k, t)l_k$$
$$\omega_{4,k} = (\mu + \mu_{tb})NP(k, t)t_k$$
$$\omega_{5,k} = (1 - p)\lambda k NP(k, t)u_k\Theta$$
$$\omega_{6,k} = p\lambda k NP(k, t)u_k\Theta$$
$$\omega_{7,k} = rNP(k, t)l_k$$

Similarly, we define the sum of all these transition rates as the average rate at which *one* transition (of any kind) occurs:

$$\Omega = \sum_{i,k} \omega_{i,k}. \tag{24}$$

This average transition rate in its turn defines the characteristic or average time τ elapsed between any two consecutive transitions, the latter being defined as the inverse of Ω:

$$\tau = \frac{1}{\Omega}. \tag{25}$$

Given the previous definitions, the Monte Carlo algorithm is implemented in such a way that at each MC step (of duration τ) one single transition takes place. Finally, the probability $\Pi_{i,k}$ that a given transition actually happens, is calculated as:

$$\Pi_{i,k} = \frac{\omega_{i,k}}{\Omega} = \tau\omega_{i,k}, \tag{26}$$

that determines which of all possible transitions is realized at each time step τ.

We have made extensive numerical simulations of the model starting from an initial population made up of $N \simeq 10^6$ individuals, whose network of contacts follows an initial degree distribution $P(k) \simeq k^{-3}$. Moreover, every newborn joins the system with a degree that verifies the same connectivity distribution. As for the values of the parameters of the dynamics, and thinking of typical values for persistent diseases, we have set the following values: $\mu = 0.009$ years^{-1}, $b = 0.010$ years^{-1}, $\mu_{Tb} = 0.200$ years^{-1}, $r = 0.002$ years^{-1}, and $p = 0.070$. Demographical parameters b and μ are roughly those of a country like Spain, while the parameters p and r are in the range of typical values for the case of tuberculosis. μ_{tb} has been chosen attending to numerical convenience (tuberculosis reaches a disease-related mortality rate that can be as large as 0.8). On the other hand, we note that a numerical criterion to define stationarity should also be adopted. In our simulations, we first let the system evolve for 4500 years and later take averages in a window of 10^6 Monte Carlo steps (which corresponds, roughly, with a

TABLE I: Dependency of the epidemic thresholds on the initial size of the network. Analytical values are obtained from Eq. (23) using the moments of the distribution generated numerically.

N_o	Analytical λ_{cA}	Numerical λ_{cN}
1000	0.52903	0.46968
10000	0.42559	0.37595
50000	0.37620	0.33088
100000	0.35854	0.31926

temporal lapse of 100 years), for the mean densities of healthy, latent and sick individuals defined in (21). This is a long enough time and ensures that all of the outputs to the left of the threshold are stabilized to the state $(1, 0, 0)$ (this is achieved almost surely already for $t \leq 4000$ years).

With the values for the parameters as specified above, the epidemic threshold is $\lambda_c = 0.305$. In Fig. 2), we have plotted the stationary proportion of sick individuals for values of the probability of transmission λ in the interval $[0, 0.5]$. As can be seen from the figure, the numerical and analytical results are in a reasonable agreement, despite the numerical challenges of simulating an open system in which the dynamics evolves on top of a complex topology. This indicates that the numerical method is accurate enough as to be used in situations where analytical predictions are not at hand.

However, it is possible to carry out numerical simulations using a variation of the algorithm above in order to improve the accuracy in the determination of the epidemic threshold λ_c. In this variation, instead of using a criterion for stationarity, we focus our attention in the vicinity of the critical value. More specifically, we first evaluate analytically the value for λ_c and start the simulation there (obviously, if we don't have an analytical hint, the simulation can be started at any value of λ). At each realization, we expect 6000 years for an eventual arrival of the system to the state $(\langle u \rangle, \langle l \rangle, \langle t \rangle) = (1, 0, 0)$. In the case that this state is not reached in that time, we assume that we are to the right of the critical point and so, we move to the left in λ just a little quantity $\delta \lambda = 0.01$. If, on the contrary, a minimum number of realizations (we used 10 in our simulations) stabilize at state $(\langle u \rangle, \langle l \rangle, \langle t \rangle) = (1, 0, 0)$, we assume to be to the left of the critical point and, consequently, we perform a $\delta \lambda$ switch to the right. Each time that such kind of flip-flop algorithm (a sort of bisection method) changes direction, we divide by two the value of $\delta \lambda$ until the desired precision in λ_c is obtained. Using this numerical approach, we have numerically calculated the values of λ_c for different system sizes. The results are reported in Table I. As expected from finite-size effects, the larger the size of the population, the smaller the absolute error between numerical and analytical thresholds is. Moreover, the larger the system size the smaller the epidemic threshold, which eventually should vanish in the thermodynamic limit.

IV.   CONCLUSIONS.

We have discussed a model for the spreading of persistent infections in complex heterogeneous populations. The framework extends the epidemiological picture proposed in previous works. Our approach is particularly suited for diseases like Tuberculosis, which shows large latency periods. The latent period results from the dynamical equilibrium that is established between the bacterium and the host's immune system, so that the host might not become infectious during its lifetime. These particular features makes it compulsory to work with an open system where newborns are continuously introduced in the population and individuals might die due to causes different from the disease itself. By assembling a model with all these ingredients, we have shown analytically that the epidemic threshold is proportional to the ratio between the first and second moments of the degree distribution. Therefore, our results point in the same direction that those obtained for the SIS and SIR model on top of the same topologies -the virtually unbounded connectivity fluctuations play a key role in the infection dynamics enhancing the epidemic incidence and lowering the epidemic threshold.

From another point of view, we have developed a method well suited to numerically explore the dynamics of the system under study. In particular, we have been able to deal with the new challenge of having a system where the number of individuals in the population is not constant and, moreover, are connected following a given degree distribution (the well-mixed case is not challenging). Although we have applied the numerical approach to our particular system, it is worth stressing that it is general and can be applied to any problem for which transition rates between different classes and states are known. The results obtained agree well with the analytical estimates with the additional advantage that the whole phase diagram can be explored.

Finally, it is also worth mentioning that the model discussed here is probably the simplest one may devise for the spreading of persistent infections in structured populations. However, despite the recent progresses in modeling disease contagion dynamics and pandemic outbreaks, the kind of spreading phenomena analyzed here is one of the issues that have remained less explored. Our aim is to take a first step towards more realistic modeling of persistent diseases. We have left for future investigation possible extensions of the current model that take into account the influence on the dynamics of vaccination, prophylaxis and recovery rates, as well as the effects of genetic heterogeneity in the pathogen but also in the host [24].

Acknowledgments

---

[1] R. M. Anderson, R. M. May, & B. Anderson, *Infectious diseases of humans: Dynamics and Control* (Oxford University Press, UK, Oxford, 1992).

[2] D. J. Daley, & J. Gani, *Epidemic Modelling* (Cambridge University Press, UK, Cambridge, 1999).

[3] J. D. Murray, *Mathematical Biology* (Springer-Verlag, Germany, Berlin, 2002).

[4] L. Hufnagel, D. Brockmann, & T. Geisel, *Proc. Nat. Acad. Sci. USA*, **101**, 15124-15129 (2004).

[5] R. Guimera, S. Mossa, A. Turtschi, & L. A. N. Amaral, *Proc. Nat. Acad. Sci. USA*, **102**, 7794-7799 (2005).

[6] V. Colizza, A. Barrat, M. Barthélemy, & A. Vespignani, *Proc. Nat. Acad. Sci. USA*, **103**, 2015-2020 (2006).

[7] S. Eubank, H. Guclu, V. S. Anil-Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, & N. Wang, *Nature* **429**, 180 (2004).

[8] V. Colizza, R. Pastor-Satorras, & A. Vespignani, *Nat. Phys.* **3**, 276 (2007).

[9] J. G. Gardeñes, V. Latora, Y. Moreno, & E. Profumo, *Proc. Nat. Acad. Sci. USA* **105**, 1399 (2008).

[10] V. Colizza, & A. Vespignani, *J. Theor. Biol.* **251**, 450 (2008).

[11] S. Meloni, A. Arenas, & Y. Moreno, *Proc. Natl. Acad. Sci. USA* **106**, 16897 (2009).

[12] M. E. J. Newman, *SIAM Rev.*, **45**, 167-256 (2003).

[13] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, & D.-U. Hwang, *Phys. Rep.*, **424**, 175-308 (2006).

[14] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, & D. J. Watts, *Phys. Rev. Lett.*, **85**, 5468-5471 (2000).

[15] R. Cohen, K. Erez, D. ben Avraham, & S. Havlin, *Phys. Rev. Lett.*, **86**, 3682-3685 (2001).

[16] M. Faloutsos, P. Faloutsos, & C. Faloutsos, ACM SIGCOMM '99, Comput. Commun. Rev. **29**, 251 (1999).

[17] G. Caldarelli, R. Marchetti, & L. Pietronero, *Europhys. Lett.*, **52**, 386 (2000).

[18] R. Albert, H. Jeong, & A. L. Barabási, *Nature*, **401**, 130 (1999).

[19] S. H. Strogatz, *Nature*, **410**, 268 (2001).

[20] R. Pastor-Satorras, & A. Vespignani, *Phys. Rev. Lett.*, **86**, 3200-3203 (2001).

[21] A. L. LLoyd, & R. M. May, *Science*, **292**, 1316-1317 (2001).

[22] Y. Moreno, R. Pastor-Satorras, & A. Vespignani, *Eur. Phys. J. B*, **26**, 521–529 (2002).

[23] M. E. J. Newman, *Phys. Rev. E.*, **66**, 016128 (2002).

[24] B. M. Murphy, B.H. Singer, S. Anderson, & S. Kirschner, *Mathematical Biosciences* **180**, 161 (2002).

[25] L. Wilson, *J. Hist. Med. Allied Sci.* **45** 366 (1990).

[26] K. Styblo, J. Meijer, I. Sutherland, *Bull. Int. Union Tuberc.* **24**(1), 137 (1969).

[27] T. Daniel, J. Bates, K. Downes, History of Tuberculosis, in *Tuberculosis: Pathogenesis, protection and control*, Edited by B. R. Bloom (American Society for Microbiology, pp.13-24, USA, Washington, 1994).

[28] D. Bleed, C. Watt, C. Dye *World Health Report 2001: Global Tuberculosis Control, Technical Report*, World Health Organization, WHO/CDS/TB/2001.287, 2001. Available from http://www.who.int/gtb/publications/globrep01/index.html.

[29] D. Bleed, C. Watt, C. Dye, *Am. Rev. Respirat. Dis.* **125**, 8 (1982).

[30] K. Styblo, Tuberculosis control and surveillance, in *Recent Advances in Respiratory Medicine*, Edited by D. Flenley, T. Petty (Vol. 4, pp. 77-108, UK, London, 1986).

[31] Blower, S. M., McLean, A.R., Porco, T. C., Small, P. M., Hopewell, P. C., Snchez, M. A., Moss A. R., *Nature Medicine*, **1(8)**, 815-821 (1995).

[32] Y. Moreno, J. B. Gómez, A. F. Pacheco, *Phys. Rev. E* **68** 035103(R) (2003).