Spreading dynamics on heterogeneous populations: Multitype network approach

Alexei Vazquez
The Simons Center for Systems Biology, Institute for Advanced Study, Einstein Drive, Princeton, New Jersey 08540, USA
(Received 28 June 2006; published 21 December 2006)

I study the spreading of infectious diseases in heterogeneous populations. The population structure is described by a contact graph where vertices represent agents and edges represent disease transmission channels among them. The population heterogeneity is taken into account by the agent's subdivision in types and the mixing matrix among them. I introduce a type-network representation for the mixing matrix, allowing an intuitive understanding of the mixing patterns and the calculations. Using an iterative approach I obtain recursive equations for the probability distribution of the outbreak size as a function of time. I demonstrate that the expected outbreak size and its progression in time are determined by the largest eigenvalue of the reproductive number matrix and the characteristic distance between agents on the contact graph. Finally, I discuss the impact of intervention strategies to halt epidemic outbreaks. This work provides both a qualitative understanding and tools to obtain quantitative predictions for the spreading dynamics of heterogeneous populations.

DOI: 10.1103/PhysRevE.74.066114 PACS number(s): 89.75.Hc, 05.70.Ln, 87.19.Xx, 87.23.Ge

I. INTRODUCTION

The globalization of human interactions has created a fertile ground for the fast and broad spread of infectious diseases, potentially leading to worldwide epidemics. We are thus forced to understand the spreading of infectious diseases within a global scenario. Yet the study of worldwide epidemics is challenging given the heterogeneity of the populations involved [1–5]. The first sign of heterogeneity is given by the variability of the reproductive number within or across populations [6–8]. The reproductive number is defined as the number of secondary cases generated by a primary infected case within a population of susceptible individuals. In the case of sexually transmitted diseases the reproductive number is proportional to the rate of sexual partner acquisition [1,9] and exhibits wide changes across individuals [1,6,10–12]. In network-based approaches the reproductive number is proportional to the node's degree [13–15] and exhibits wide variations as well [16]. In the absence of biases among the connections between agents this heterogeneity is completely taken into account by the reproductive number distribution [14,15].

There are other properties beyond the reproductive number requiring the subdivision of a population in different classes or types. This includes but is not limited to age, geographical location, social status, and sexual behavior. In general these heterogeneities cannot be characterized by a single probability distribution. They require a multitype approach with probability distributions characterizing each type and a mixing matrix describing the patterns of transmission among them. Multitype models are difficult to deal with and are generally tackled using multiagent simulations [2,4,5,17–19]. The advantage of multiagent simulations is that we can consider several details and study their impact on the spreading dynamics. On the other hand, given the large number of variables and model parameters it is difficult to understand which are the key parameters driving the system's dynamics. Therefore, exact or approximate calculations are required to funnel the multiagent simulations into specific regions of the parameter space.

The theory of multitype branching processes provides a framework to characterize the spreading dynamics in heterogeneous populations [20,21]. In multitype branching processes the concept of reproductive number is generalized to the reproductive number matrix, with elements giving the expected number of secondary cases of a given type generated by a primary case of a given type. In turn, the long-time behavior is determined by the largest eigenvalue ρ of this matrix [20,21]. The number of infected individuals decays exponentially when $\rho<1$ and growth exponentially otherwise. Although rarely mentioned, these results are valid provided the spreading process goes on for a large number of generations. In recent studies of the single-type case [22–24] I have shown that this approximation does not hold for most real networks underlying the spreading of infectious diseases among humans and computer malwares.

In this work I study the spreading of infectious diseases in multitype networks. I take as a starting point the static problem formulation developed by Newman [25] and the theory of continuous-time multitype branching processes [20]. I reexamine these mathematical approaches to accommodate distinctive properties of real networks that have not been previously considered. In Sec. II, I introduce the basic framework. Focusing on the population structure I consider the contact graph characterizing the detailed interactions among agents and the type network characterizing the interactions among agent's types. Through some simple examples I illustrate the properties of the mixing matrix and its type-network representation. This section ends defining a branching process which models a spanning tree from an index agent to all other agents in the contact graph. In Sec. III, I characterize the local spreading dynamics from an agent to its contacts, taking the susceptible-infected-removed (SIR) model as a case study. Bringing together the underlying network structure and the local transmission dynamics in Sec. III, I define a branching process that models the disease spreading dynamics. In Sec. IV, I extend the iterative approach for a single type [22–24] to accommodate the particularities of the multitype case. Focusing on the expected behavior I obtain general equations determining the progression of the expected number of cumulative and new infections. Starting

from these equations I analyze some limiting cases. First, I derive the final expected outbreak size and, second, I analyze the time progression of the expected outbreak size for the homogeneous transmission case. In Sec. IV C, I discuss the impact of the population heterogeneity on intervention strategies. I emphasize the role of the characteristic distance between agents to quantify the impact of intervention strategies on small-world populations. I also illustrate interventions targeting specific agent's types using a bipartite population as a case study. Finally, in Sec. V, I provide an overview of the main results and discuss future directions.

II. POPULATION STRUCTURE

Consider a population of N agents that are susceptible to an infectious disease. By *agent* I mean any entity that could host and transmit the disease. For instance, if we are interested in the transmission of infectious diseases among humans, an agent is a human in the first place. For vector-borne diseases such as malaria we could have in addition agents representing the intermediary host while for airborne diseases an agent could also represent a public place. Another application is the spreading of computer malwares, and in this case agents represent computer users. The agents are assumed to be heterogeneous, meaning that there are different agent classes or *types* according to their pattern of connectivity to other agents and/or to the speed at which they could potentially transmit an infectious disease. For instance, humans can be divided according to their age, social status, and geographical location. Furthermore, in the case of vector- and air-borne diseases there is an additional type given by the nonhuman intermediary.

Let us assume that the agent population is divided into M types and there are N_a agents of type $a=1,\ldots,M$ satisfying the normalization condition

$$\sum_{a=1}^{M} N_a = N. \tag{1}$$

Note that within this work I use the indices $a,b,\ldots$ for the agent's type. In the following I introduce two representations of the population structure at the agent and type levels, respectively.

A. Contact graph

The contact graph takes precisely into account who could potentially transmit the disease to whom [1,26–29], more precisely given by the following definition.

Definition II.1. The contact graph is a colored graph where vertices represent agents, edges represent the potential disease transmission channels among them, and vertices are colored according to the agent's type.

The contact graph represents the population mixing at the agent's level. Since there is a one-to-one relation between vertices and the corresponding agents I use these two terms interchangeably. All the information necessary to characterize a given graph is provided by its adjacency matrix. Yet given the large size of real populations and their change in time, the only way to achieve such a detailed description relies on agent-based simulations. My scope is to bypass this detailed description and focus on statistical properties that do not depend on the population structure details or their change in time.

First, we need to specify the time scale on which the statistical properties are measured. An agent may go over different stages associated with the disease, including latent, infectious, and recovered, but only during the infectious stage he/she transmits the disease. Thus, the time scale that matters for the spreading process is the infectious period. The duration of the infectious period is determined by several factors. Intrinsic factors are those resulting from the host-disease interaction while extrinsic factors are those associated with intervention strategies such as patient isolation. At this point I intentionally exclude the extrinsic factors. Their influence is taken into account when defining the disease-spreading dynamics (see Sec. III). It is also worth mentioning that the infectious period is a random variable. Therefore, the statistical properties introduced below are the expectation after averaging over the infectious period distribution (see example below).

Once the time scale is specified, the contact graph represents the interactions among agents within that time scale. More precisely, each edge in the contact graph represents the occurrence of the relevant contact among the corresponding agents during the specified time window. The degree of a node is the total number of edges emanating from it regardless the node type at the other end. Let $p_k^{(a)}$ be probability distribution that a type a node has degree k and denote by

$$\langle k\rangle_a = \sum_{k=1}^{\infty} p_k^{(a)} k \tag{2}$$

its mean. Note that by allowing k to take values larger than 1 we are already taking into account the existence of concurrency [30–32]—i.e., the fact that one agent may contact more than one other agent. To characterize the spreading process it is also relevant to determine the same distribution but for a vertex found and the end of an edge selected at random. This sampling introduces a bias towards nodes with a higher degree, resulting in the probability distribution

$$q_k^{(a)} = \frac{k p_k^{(a)}}{\sum_{s=1}^{\infty} s p_s^{(a)}}, \tag{3}$$

with average excess degree

$$\langle k\rangle_a^{(\mathrm{excess})} = \sum_k q_k^{(a)} (k-1), \tag{4}$$

where the -1 subtracts the edge from where the node was reached. Associated with these two probability distributions we introduce the generating functions

$$U_a(x) = \sum_{k=0}^{\infty} p_k^{(a)} x^k, \tag{5}$$

$$V_a(x) = \sum_{k=1}^{\infty} q_k^{(a)} x^{k-1}. \tag{6}$$

From the derivatives of $U_a(x)$ and $V_a(x)$ we obtain the moments of $p_k^{(a)}$ and $q_k^{(a)}$, respectively: for instance,

$$\dot{U}_a(1) = \langle k \rangle_a, \tag{7}$$

$$\dot{V}_a(1) = \langle k \rangle_a^{(\text{excess})}. \tag{8}$$

Since the agent population is finite, there is a typical distance D between every two agents in the contact graph. Social experiments such as the Kevin Bacon and Erdős numbers [33] or the Milgram experiment [34] reveal that social actors are separated by a small number of acquaintances, often known as the small-world property [35]. Computer networks are characterized by the small-world property as well [36–39]. These observations are supported furthermore by theoretical approaches demonstrating that D grows at most as $\log N$ in random graphs [40–43]. More recently it has been shown that for several real networks D actually decreases or remains constant as the networks evolve and increases its size [44]. Thus, I explicitly take into account that D is finite.

Example II.2 (Poisson contact process). Let us assume that type-a agents establish connections with other agents at a constant rate λ_a and that the infectious period is constant and equal to T. In this case we obtain a Poisson distribution for the agent's degree:

$$p_k^{(a)}(T) = \frac{(\lambda_a T)^k e^{-\lambda_a T}}{k!}. \tag{9}$$

Now let us relax the assumption that the infectious period is a constant and consider the case where it follows the exponential distribution $F(T)=1-e^{\nu T}$, where ν^{-1} is the average infectious period. Taking the average of Eq. (9) over $F(T)$ we obtain

$$p_k^{(a)} = \frac{1}{\lambda_a}\left(\frac{\lambda_a}{\lambda_a + \nu}\right)^{k+1}. \tag{10}$$

This example shows that the contact graph is determined by both the interaction among individuals and the infectious period. The former is intrinsic to the agent population while the last one is disease dependent.

B. Type network

At the type level the population structure is determined by the mixing patterns among types. Given a type-a agent and one of its edges I define the *mixing matrix* elements e_{ab} as the probability that the agent at the other end is of type b. From the mixing matrix we can construct the type network.

Definition II.3. Type network: In the type network a node represents a type, an arc is drawn from type a to b if $e_{ab} > 0$, and the arc's weights are given by e_{ab}.

Figure 1 shows some simple type networks. The single-type case is represented by a node with a loop [Fig. 1(a)]. A bipartite population is represented by two nodes with an incoming and an outgoing arc [Fig. 1(b)]. This example models a heterosexual population with no other distinction than gender or a metapopulation given by people and public places [19]. A fully mixed population is represented by a complete network [Fig. 1(c)]. A less intuitive example is the type network shown in Fig. 1(d), representing a population divided into two cities and the commuters between them.

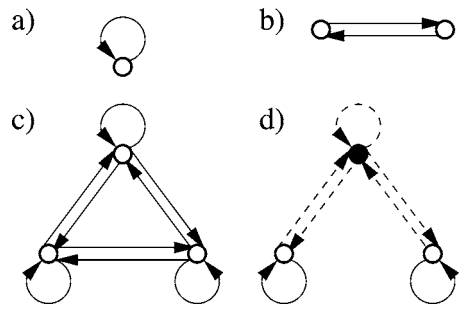

FIG. 1. Type-network representation of simple mixing matrices. (a) Single-type population. (b) Bipartite population. (c) Fully mixed population with three types. (d) Two cities (open circles) and the commuters among them (solid circle). The solid (dashed) lines represent intracity (intercity) connections.

C. Annealed spanning tree

Given a contact graph, let us consider an epidemic outbreak starting from a single agent, the *index case*. In the worst case scenario the disease propagates to all the agents that could be reached from the index case using the network connections. In this case the outbreak is represented by a spanning or causal tree from the index case to all reachable agents. The *generation* of an agent in this tree corresponds with the topological or hopping distance from the index case. This picture motivates the introduction of the following branching process.

Definition II.4. Multitype annealed spanning tree.

Consider a colored contact graph characterized by $\{N_a, p_k^{(a)}\}$ and the type network $\{e_{ab}\}$. The multitype annealed spanning tree (AST) is the branching process satisfying the following properties.

(i) The process starts from an index case of type $a \in \{1,\ldots,M\}$ at generation $d=0$. The index case generates k sons with probability distribution $p_k^{(a)}$. Each son is of type b with probability e_{ab}.

(ii) Each son at generation $1 \leqslant d < D$ generates $k-1$ sons with probability distribution $q_k^{(a)}$. Each son is of type b with probability e_{ab}.

(iii) A son at generation $d=D$ does not generate new sons.

The term annealed means that we are not analyzing the true (quenched) spanning tree on the graph but a branching process with similar statistical properties. This approximation is particularly good if the contact graph is continuously changing in time in spite of the constancy of its statistical properties. The sharp cutoff at $d=D$ is an approximation. In reality there is a gradual decay of the number of nodes found after reaching the average distance between nodes. Nevertheless, the study of the single-type case indicates that the sharp

cutoff is a very good approximation, resulting in good agreement with the numerical simulations [22].

III. LOCAL SPREADING DYNAMICS

To proceed further I now specify how the disease is transmitted from an agent to its neighbors in the contact graph. Let r_{ab} be the probability that an infected agent of type a infects a susceptible neighbor of type b. I assume that if $e_{ab}>0$, then $r_{ab}>0$, while the absence of transmission between two types is taken into account setting $e_{ab}=0$. Given an agent i of type a (primary case) and one of its neighbors j of type b (secondary case), we define the generation time $X_{ij}^{(a,b)}$ as the time elapsed from the infection of the primary case to the infection of the secondary case provided it happens. I assume that the generation times are independent random variables with distribution function

$$G_{ab}(\tau) = \text{Prob}(X_{ij}^{(a,b)} \leqslant \tau), \tag{11}$$

parametrized by the type of the primary and secondary cases.

Example III.1 (SIR model). In the SIR model agents can be in the three exclusive states of susceptible, infected, and removed. A *susceptible* agent is one that has not become infected but is susceptible to acquire the infection. An *infected* agent is one that has already acquired the disease and can potentially transmit the disease. A *removed* agent is one that has been previously infected but is already excluded from the spreading process. I make a distinction between removal because of death or natural recovery and removal because of intervention strategies resulting in the isolation or cure of infected individuals. The death or natural recovery of infected agents was the basics to define the characteristic time scale in which the contact graph is defined (Sec. II A). The impact of intervention strategies is analyzed below. I introduce this distinction to facilitate the analysis of intervention strategies without modifying the structure of the contact graph.

Consider an agent i of type a and one of its neighbors j of type b. Let $Y_{i,j}^{(a,b)}$ be the infection time of agent j by i in the absence of intervention strategies and let $G_{\text{I}}^{(a,b)}(\tau) = \text{Prob}(Y_{i,j}^{(a,b)} \leqslant \tau)$ be its distribution function. Furthermore, let $Z_i^{(a)}$ be the removal time of agent i in the presence of intervention strategies and let $G_{\text{R}}^{(a)}(\tau) = \text{Prob}(Z_i^{(a)} \leqslant \tau)$ be its distribution function. The probability that agent j is infected by agent i by time t is given by

$$b_{ab}(t) = \int_0^t dG_{\text{I}}^{(a,b)}(\tau)[1 - G_{\text{R}}^{(a)}(\tau)]. \tag{12}$$

From this magnitude we obtain the probability that agent j gets infected by agent i no matter when

$$r_{ab} = \lim_{t\to\infty} b_{ab}(t) \tag{13}$$

and the distribution of generation times

$$G_{ab}(\tau) = \frac{1}{r_{ab}} b_{ab}(\tau). \tag{14}$$

The SIR model could be further generalized taking immunization into account. In this case noninfected agents are divided into susceptible and immune. If s_a is the probability that a type-a agent is immune, then the probability that agent j is infected by agent i by time t reads

$$b_{ab}(t) = (1 - s_b)\int_0^t dG_{\text{I}}(\tau)[1 - G_{\text{R}}(\tau)]. \tag{15}$$

Furthermore, the transmission probability r_{ab} and the generation time distribution $G_{ab}(\tau)$ are obtained by substituting this equation into Eqs. (13) and (14), respectively.

These examples illustrate how to calculate the transmission probability r_{ab} and the generation time distribution $G_{ab}(\tau)$ from standard models characterizing the spreading of infectious diseases. More importantly, by encapsulating the model details into r_{ab} and $G_{ab}(\tau)$ we obtain general results that are model independent.

Continuous-time multitype AST

At this point the local spreading dynamics has been completely specified and we can superimpose it on the multitype AST.

Definition III.2. Continuous-time multi-type AST.

A continuous time multitype AST is composed of two elements: a multitype AST II.4 and a local spreading dynamics defined by $\{r_{ab}, G_{ab}(\tau)\}$. The global dynamics is then specified by the following rules.

(i) The process starts with an infected agent of type $a \in \{1,\ldots,M\}$ while all other agents are susceptible.

(ii) An infected agent of type a infects each of its neighbors of type b with probability r_{ab} and generation time distribution $G_{ab}(\tau)$.

From the mathematical point of view the continuous-time multitype AST belongs to a more general class of multitype branching processes [20]. The novelty of the present work resides in the application of this framework to accommodate crucial properties of real networks that were not considered before.

IV. EXPECTED BEHAVIOR

Given an infected agent of type a the expected number of secondary infections of type b it generates is given by

$$R_{ab} = \langle k\rangle_a e_{ab} r_{ab} \tag{16}$$

if it is the index case and by

$$\tilde{R}_{ab} = \langle k\rangle_a^{(\text{excess})} e_{ab} r_{ab} \tag{17}$$

otherwise. The matrices R and $\tilde{R}$ are extensions of the basic reproductive number to the multitype case. In the following it becomes clear that $\tilde{R}$ is more relevant and therefore I refer to it as the reproductive number matrix.

Lemma IV.1. Consider an ensemble of continuous-time multitype AST III.2 with index case of type a. Let $N_{ab}(t)$ be the mean total number of infected type-b agents at time t and

let $I_{ab}(t)dt$ be the mean number of type-b agents that are infected between time t and $t+dt$. Then

$$N_{ab}(t) = \sum_{d=1}^{D} (H \star \tilde{H}^{\star(d-1)})_{ab}(t), \tag{18}$$

$$I_{ab}(t) = \sum_{d=1}^{D} \frac{d}{dt}(H \star \tilde{H}^{\star(d-1)})_{ab}(t), \tag{19}$$

where

$$H_{ab}(t) = R_{ab} G_{ab}(t), \tag{20}$$

$$\tilde{H}_{ab}(t) = \tilde{R}_{ab} G_{ab}(t), \tag{21}$$

and the multiplication symbolized by $\star$ involves a matrix multiplication and a convolution in time—i.e.,

$$(H \star \tilde{H})_{ab}(t) = \sum_{c=1}^{M} \int_0^t dH_{ac}(\tau) \tilde{H}_{cb}(t-\tau), \tag{22}$$

$$(\tilde{H}^{\star 2})_{ab}(t) = (\tilde{H} \star \tilde{H})_{ab}(t). \tag{23}$$

The proof of this result is given in Appendix A. This lemma provides explicit equations for the expected progression of an epidemic outbreak. In some particular cases these equations may be further expressed in terms of elementary functions, allowing a straightforward interpretation. More generally these equations can be evaluated numerically in cases in which further reduction is not possible. In addition, theorem IV.1 is a starting point for calculations addressing some limiting cases, which are the subject of the following subsections.

A. Final outbreak size

The final outbreak size is obtained taking the limit $t \to \infty$ in Eq. (18), resulting in

$$N_{ab}(\infty) = \sum_{d=1}^{D} (R \tilde{R}^{d-1})_{ab}. \tag{24}$$

When $\tilde{R}$ can be diagonalized we can write $\tilde{R} = P \Delta P^{-1}$, where P is the matrix composed of the eigenvectors of $\tilde{R}$, Δ is the diagonal matrix constructed from the corresponding eigenvalues $(\rho_a,\ a=1,\ldots,M)$, and P^{-1} is the inverse of P. Thus Eq. (24) is reduced to

$$N_{ab}(\infty) = (RP\tilde{N}P^{-1})_{ab}, \tag{25}$$

where $\tilde{N}$ is a diagonal matrix with diagonal entries

$$\tilde{N}_{aa} = \begin{cases} \dfrac{\rho_a^D - 1}{\rho_a - 1}, & \text{for } \rho_a \neq 1, \\ D, & \text{for } \rho_a = 1. \end{cases} \tag{26}$$

The eigenvalues of the reproductive number matrix appear to the power of D. Therefore we expect that the term containing the largest eigenvalue gives the major contribution. The following theorem shows that this is indeed the case.

Theorem IV.2 (complete type network). Consider a complete type network and let ρ be the largest eigenvalue of $\tilde{R}$, Eq. (17). Then

$$N_{ab}(\infty) = u_{ab} \frac{\rho^D - 1}{\rho - 1}, \tag{27}$$

where u_{ab} is independent of D.

Proof. The mixing matrix of a complete type network is positive definite and, therefore, R, Eq. (16), and $\tilde{R}$, Eq. (17), are positive definite as well. From the Perron-Frobenius theorem [45] it follows that the largest eigenvalue of $\tilde{R}$ is simple and all the entries of its corresponding left eigenvector $\vec{v}$ are different from zero and have the same sign. In particular we choose all the components of $\vec{v}$ to be positive. Since $v_c \neq 0$ for all c, we can write

$$(R\tilde{R}^{d-1})_{ab} = \sum_{c=1}^{M} R_{ac} (\tilde{R}^{d-1})_{cb} = \sum_{c=1}^{M} \frac{R_{ac}}{v_c} v_c (\tilde{R}^{d-1})_{cb}. \tag{28}$$

Taking into account that $\Sigma_c v_c \tilde{R}_{cb} = \rho v_b$ and $v_c > 0$ we obtain the inequalities

$$u_{ab}^{(\min)} \rho^{d-1} \leqslant (R\tilde{R}^{d-1})_{ab} \leqslant u_{ab}^{(\max)} \rho^{d-1}, \tag{29}$$

where

$$u_{ab}^{(\min)} = \min_c R_{ac} \frac{v_b}{v_c}, \tag{30}$$

$$u_{ab}^{(\max)} = \max_c R_{ac} \frac{v_b}{v_c}. \tag{31}$$

From Eqs. (27) and (24) we obtain

$$1 + u_{ab}^{(\min)} \frac{\rho^D - 1}{\rho - 1} \leqslant N_{ab}(\infty) \leqslant 1 + u_{ab}^{(\max)} \frac{\rho^D - 1}{\rho - 1}. \tag{32}$$

Finally, from this equation we obtain Eq. (27) with

$$0 < u_{ab}^{(\min)} \leqslant u_{ab} \leqslant u_{ab}^{(\max)} < \infty, \tag{33}$$

where the inequality $u_{ab} > 0$ follows from Eq. (30). ■

The overall outbreak size is thus determined by the largest eigenvalue of the reproductive number matrix ρ and the average distance between nodes, D. The connection with previous works [20] is obtained taking the limit $D \gg 1$. In this limit the outbreak size is negligible with respect to the population size when $\rho < 1$ and when $\rho > 1$ it represents a finite fraction of the population. Figure 2 illustrates that these extreme behaviors gradually disappear with decreasing D. When D is small the expected outbreak size changes smoothly with increasing ρ, including the region around $\rho = 1$.

The prefactor u_{ab} in Eq. (27) quantifies the contribution of each type to the overall outbreak size. u_{ab} is determined by the reproductive number matrix of patient zero and the eigenvector associated with the largest eigenvalue, Eqs. (30) and (31). From the analysis of these quantities we can determine which population types are at more risk depending on the type of patient zero.

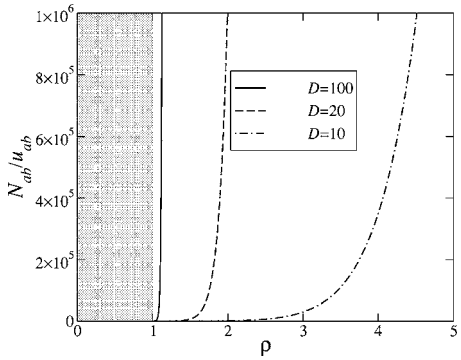

FIG. 2. Expected outbreak size as a function of the largest eigenvalue of the reproductive number matrix for different values of D. The region $\rho<1$ is indicated by the shadowed region.

Theorem IV.2 is generalized in Appendix B to strongly connected type networks—i.e., type networks where there is a path from every type a to every type b. Apart for some technicalities, similar conclusions are obtained.

B. Spreading dynamics with constant transmission rate

A general study considering both the topological and dynamical heterogeneities is hard to approach. As a starting point I consider the particular case when the local spreading dynamics is type independent—i.e., $G_{ab}(\tau)=G(\tau)$. From Eq. (19) we obtain the incidence

$$I_{ab}(t)=\sum_{d=1}^{D}(R\tilde{R}^{d-1})_{ab}G^{\star d}(t). \tag{34}$$

Regarding the matrix elements this expression has the same structure as Eq. (24). When $\tilde{R}$ can be diagonalized we rewrite Eq. (34) as

$$I_{ab}(t)=(RP\tilde{I}(t)P^{-1})_{ab}, \tag{35}$$

where $\tilde{I}(t)$ is a time-dependent diagonal matrix with diagonal entries

$$\tilde{I}_{aa}(t)=\sum_{d=1}^{D}\rho_a^{d-1}\frac{d}{dt}G^{\star d}(t). \tag{36}$$

For each a, $\tilde{I}_{aa}(t)$ has the same form as for the single-type case [22]. In other words each eigenvalue contribution is interpreted as a single type and the overall behavior is obtained after the superposition of these modes (34). As for the final outbreak size, some asymptotic properties are determined by the contribution given by the largest eigenvalue, resulting in the following theorem.

Theorem IV.3. Consider a complete type network and an agent-independent exponential distribution of generation times $G_{ab}(\tau)=1-e^{-\lambda\tau}$, where λ is the transmission rate. Let ρ be the largest eigenvalue of $\tilde{R}$, Eq. (17), and let

$$\theta=\frac{D-1}{\rho}. \tag{37}$$

$\theta\gg 1$: If $\rho>1$ and $1\ll\lambda t\ll\theta$, then

$$I_{ab}(t)\sim e^{(\rho-1)\lambda t}. \tag{38}$$

$\theta\ll 1$: If $\lambda t\gg\theta$, then

$$\frac{I_{ab}(t)}{N_{ab}(\infty)}=\frac{\lambda(\lambda t)^{D-1}e^{-\lambda t}}{(D-1)!}\left[1+O\left(\frac{t_0}{t}\right)\right], \tag{39}$$

where

$$t_0=\frac{\theta}{\lambda} \tag{40}$$

and the symbol $O(t_0/t)$ indicates that this is an asymptotic result when $t\gg t_0$ with correction terms of the order of t_0/t.

The case $\theta\gg 1$ provides the connection between this work and continuous-time multitype branching processes with an infinite number of generations. Indeed, Mode has already demonstrated the exponential growth regime for the case $D=\infty$ (see [20], Chap. 3). Theorem IV.3 shows that in the other limit $\theta\ll 1$ the spreading dynamics is instead characterized by a γ distribution. The demonstration of this result for the more general case of a strongly connected type network is given in Appendix C.

C. Impact of intervention strategies

The expected outbreak size is a monotonic increasing function of ρ, playing the role of the basic reproductive number in homogeneous populations [1,46]. Therefore, the aim of intervention strategies is to reduce the characteristic reproductive number ρ. On the other hand, intervention strategies imply an economical cost including but not limited to the development of new vaccines and their deployment through vaccination campaigns. Our task is to design optimal intervention strategies that minimize the expected outbreak size with a feasible economical cost.

To be more precise let us consider a scenario where the disease is transmitted at constant rate λ from infected to susceptible agents, infected agents are isolated at a rate μ, and a fraction s of the population is immune to the disease. In this case the infection and removal times follow the exponential distribution functions $G_I(\tau)=1-e^{-\lambda\tau}$ and $G_R(\tau)=1-e^{-\mu\tau}$, respectively. Thus, from Eq. (13) we obtain $r_{ab}=1-\beta$ where

$$\beta=1-\frac{\lambda}{\lambda+\mu}(1-s) \tag{41}$$

is the blocking fraction—i.e., the fraction of potential disease transmissions that are blocked either because of immunization or patient isolation. Since $r_{ab}=1-\beta$ is independent of the primary and secondary case types, we can write the reproductive number matrices (16) and (17) as $R_{ab}=(1-\beta)K$ and $\tilde{R}=(1-\beta)\tilde{K}$, respectively, where

$$K_{ab}=\langle k\rangle_a e_{ab}, \quad \tilde{K}_{ab}=\langle k\rangle_a^{(\text{excess})}e_{ab}. \tag{42}$$

In turn, the largest eigenvalue of $\tilde{R}$ is given by

$$\rho=(1-\beta)\kappa, \tag{43}$$

where κ is the largest eigenvalue of $\tilde{K}$. κ represents the free spread reproductive number in the absence of intervention

strategies. In turn, ρ, Eq. (43), is the effective reproductive number considering the impact of intervention strategies.

From the analysis made in Sec. IV A it follows that there are two different scenarios. When $D \gg 1$ the target of intervention strategies is $\rho=1$. The blocking fraction to achieve this is obtained from Eq. (43), resulting in

$$\beta_c = 1 - \frac{1}{\kappa}. \tag{44}$$

This result has been already reported, at least for the case of two types [1]. Most real networks are characterized, however, by small D values. In this case the expected outbreak size is a smooth function of ρ (see Fig. 2) and the concept of epidemic threshold makes no particular sense. In this context the target of intervention strategies should be a predefined expected outbreak size.

So far we have considered homogeneous intervention strategies. Now let us assume that the rate of patient isolation and the immunized fraction are different for each agent's type and given by μ_a and s_a, respectively. In this case the blocking fraction is given by

$$\beta_{ab} = 1 - \frac{\lambda}{\lambda + \mu_a}(1 - s_b) \tag{45}$$

and $r_{ab}=1-\beta_{ab}$, which depends on the type of both the primary and secondary case. From the Perron-Frobenius theorem it follows that ρ is a continuous increasing function of all the entries of the $\widetilde{R}$ [47]. Since $\widetilde{R}_{ab}=(1-\beta_{ab})\widetilde{K}_{ab}$, then ρ is a continuous decreasing function of β_{ab} for all (a,b). The goal is to determine which strategy leads to the largest reduction of ρ.

Example IV.4. Consider the spread of HIV in a heterosexual population with no further distinction beyond gender. In this case the type network is bipartite [see Fig. 1(b)]. Let k_1 and k_2 be the average excess degree for the connections from women to men and vice versa. Let us also assume that the rate of patient isolation is zero and that we could immunize a fraction s of the overall population, distributed between a fraction xs and $(1-x)s$ of immunized women and men, respectively. The question is to determine the value of x representing the best intervention strategy. In this case the reproductive number matrix is given by

$$\widetilde{R} = \begin{bmatrix} 0 & [1-(1-x)s]k_2 \\ (1-xs)k_1 & 0 \end{bmatrix} \tag{46}$$

and it has the largest eigenvalue

$$\rho = \sqrt{[1-s+x(1-x)s^2]k_1k_2}. \tag{47}$$

It results that ρ is minimum for $x=0$ or $x=1$; i.e., the best intervention strategy is to direct all the immunization resources to only one subpopulation.

V. DISCUSSION

These calculations are an application of the theory of multitype branching processes to model the spread of infectious diseases on heterogeneous populations. Using this modeling approach I bypass the detailed structure of the contact graph underlying the disease spread and focus on its statistical properties. At the type's level, the network representation of the mixing matrix provides a very intuitive interpretation of the mixing patterns. This representation allows us to make use of the Fobrenious-Perrón theorem [45] to obtain the leading behavior. In this way the impact of the model parameters becomes relevant through the largest eigenvalue of the reproductive number matrix ρ and the average distance between nodes, D.

The relevance of ρ is a well-known fact since the work of Mode [20]. The novelty resides in the modulation of the ρ dependence by D. This result is not merely a theoretical observation. Given the small-world character of social and electronic communication networks it is of practical importance as well. The good news is that in spite of this modulation by D the target of intervention strategies is still the characteristic reproductive number. More precisely, reducing the characteristic reproductive number we reduce the final outbreak size. This observation may explain why intervention strategies have been successful although assuming $D \gg 1$. The bad news is that to quantify the impact of the intervention strategies we need to estimate D. Therefore, to achieve a realistic epidemic outbreak forecast and efficient intervention strategies we most take into account the small-world property of real networks.

There are different paths to estimate D. First, we can use a direct approach such as Milgram's experiments [34]. Second, we can measure other network properties such as the degree distribution and then try to estimate D using network models [40–44,48]. Finally, I have shown that the progression of the expected number of new infections is modulated by D (see [22–24] and Sec. III). More precisely, in small-world populations the incidence is expected to grow as a power law and we can estimate D from the power-law exponent.

Further work is required to test the validity of the type-network approach. This can be done by running agent-based simulations having a strict control of the different statistical properties characterizing the population structure. These statistical properties can be then plugged into the multitype network approach to obtain quantitative predictions that can be compared with the simulation results.

In conclusion, this work opens new avenues to future research on the spreading of infectious diseases on heterogeneous populations. It allows for a qualitative understanding through the analysis of the type-network representation of the mixing matrix. More important, it leads to general results that can be tackled case by case using exact or approximate calculations and numerical computations.

ACKNOWLEDGMENTS

I thank J. Koopman and G. Szabo for several comments and discussions.

APPENDIX A: ITERATIVE APPROACH

Consider a branch of the AST rooted in a type-a agent, at generation d, which was infected at time zero. Let $P_n^{(d,a,b)}$

$\times(t)$ be the probability distribution to find n infected type-b agents at time t in that branch. In particular $P_n^{(0,a,b)}(t)$ is the probability distribution of the total number of infected type-b agents at time t on the whole AST, given that the index case was of type a. Based on the tree structure we can develop an iterative approach to compute $P_n^{(d,a,b)}(t)$ recursively.

Lemma A.1. Consider a type-a infected agent at generation d of a continuous-time multitype AST. This agent has degree k with probability $p_k^{(a)}$ for $d=0$ and excess degree $k-1$ with probability $q_k^{(a)}$ for $0<d<D$. Let us index by α its neighbors in the next generation $d+1$, where $\alpha\in\{1,\ldots,k\}$ for $d=0$, $\alpha\in\{1,\ldots,k-1\}$ for $0<d<D$, and $\alpha\in\{\emptyset\}$ for $d=D$. Then

$$P_n^{(0,a,b)}(t)=p_0^{(a)}[\delta_{ab}\delta_{n1}+(1-\delta_{ab})\delta_{n0}]+\sum_{k=1}^{\infty}p_k^{(a)}\sum_{n_1=0}^{\infty}\cdots\sum_{n_k=0}^{\infty}\delta_{\sum_{\alpha=1}^{k}n_\alpha+\delta_{ab},n}\prod_{\alpha=1}^{k}\sum_{c=1}^{M}e_{ac}\times\left[r_{ac}\int_0^t dG_{ac}(\tau)P_{n_\alpha}^{(1,c,b)}(t-\tau)+\delta_{n_\alpha,0}[1-r_{ac}G_{ac}(t)]\right],\tag{A1}$$

$$P_n^{(d,a,b)}(t)=q_1^{(a)}[\delta_{ab}\delta_{n1}+(1-\delta_{ab})\delta_{n0}]+\sum_{k=2}^{\infty}q_k^{(a)}\sum_{n_1=0}^{\infty}\cdots\sum_{n_{k-1}=0}^{\infty}\delta_{\sum_{\alpha=1}^{k-1}n_\alpha+\delta_{ab},n}\prod_{\alpha=1}^{k-1}\sum_{c=1}^{M}e_{ac}\times\left[r_{ac}\int_0^t dG_{ac}(\tau)P_{n_\alpha}^{(d+1,c,b)}(t-\tau)+\delta_{n_\alpha,0}[1-r_{ac}G_{ac}(t)]\right],\tag{A2}$$

$$P_n^{(D,a,b)}(t)=\delta_{ab}\delta_{n1}+(1-\delta_{ab})\delta_{n0},\tag{A3}$$

where $\delta_{ij}=1$ for $i=j$ and zero otherwise.

Proof. Let n be the number of infected type-b agents in a branch rooted at type-a agent, and let n_α be the infected type-b agents on the branches rooted at each of its neighbors α. Then

$$n=\delta_{ab}+\sum_{\alpha}n_\alpha,\tag{A4}$$

where δ_{ab} takes into account if the root agent is or it is not of type b. The probability distribution of n is given by the sum of all possible combinations of the random variables n_α that satisfy Eq. (A4). Now, the root agent and its neighbors lie on a tree and therefore n_α are independent random variables. Furthermore, all agents at generation $d+1$ have the same statistical properties; i.e., n_α are identically distributed random variables. Therefore, the probability of each combination is decomposed into the product of the probability distribution of the number of infected agents of type b on the subbranches rooted at each neighbor. Thus, taking into account that each neighbor is of type c with probability e_{ac} we obtain

$$P_n^{(0,a,b)}(t)=p_0^{(a)}[\delta_{ab}\delta_{n1}+(1-\delta_{ab})\delta_{n0}]+\sum_{k=1}^{\infty}p_k^{(a)}\sum_{n_1=0}^{\infty}\cdots\sum_{n_k=0}^{\infty}\delta_{\sum_{\alpha=1}^{k}n_\alpha+\delta_{ab},n}\times\prod_{\alpha=1}^{k}\sum_{c=1}^{M}e_{ac}Q_{n_\alpha}^{(1,a,c,b)}(t),\tag{A5}$$

$$P_n^{(d,a,b)}(t)=q_1^{(a)}[\delta_{ab}\delta_{n1}+(1-\delta_{ab})\delta_{n0}]+\sum_{k=2}^{\infty}q_k^{(a)}\sum_{n_1=0}^{\infty}\cdots\sum_{n_{k-1}=0}^{\infty}\delta_{\sum_{\alpha=1}^{k-1}n_\alpha+\delta_{ab},n}\times\prod_{\alpha=1}^{k-1}\sum_{c=1}^{M}e_{ac}Q_{n_\alpha}^{(d+1,a,c,b)}(t),\tag{A6}$$

where $Q_{n_\alpha}^{(d+1,a,c,b)}(t)$ is the probability distribution of n_α which we proceed to calculate.

Let us focus on one neighbor and assume that it is of type c. With probability $1-r_{ac}$ this agent is not infected at any time and with probability $r_{ac}[1-G_{ac}(t)]$ it is not yet infected at time t given it will be infected at some later time, resulting in

$$Q_0^{(d+1,a,c,b)}(t)=1-r_{ac}G_{ac}(t).\tag{A7}$$

On the other hand, with probability r_{ac} the neighbor is infected at some time τ, with distribution function $G_{ac}(\tau)$, and the spreading dynamics continues to subsequent generations. Once the neighbor is infected, the number of infected agents of type b on that subbranch is a random variable with probability distribution $P_n^{(d+1,c,b)}(t-\tau)$. Therefore, for $n>0$,

$$Q_n^{(d+1,a,c,b)}(t)=r_{ac}\int_0^t dG_{ac}(\tau)P_n^{(d+1,c,b)}(t-\tau).\tag{A8}$$

Finally, substituting Eqs. (A7) and (A8) into Eqs. (A5) and (A6) we obtain Eqs. (A1) and (A2). The demonstration of Eq. (A3) is straightforward. For $d=D$ the process stops and therefore there is only one infected agent, the root itself, which is or it is not of type b, resulting in Eq. (A3). ■

Associated with the probability distribution $P_n^{(d,a,b)}(t)$ we introduce the generating function

$$F^{(d,a,b)}(x,t)=\sum_{n=0}^{\infty}P_n^{(d,a,b)}(t)x^n.\tag{A9}$$

Using the recursive relations for the probability distribution, Eqs. (A1)–(A3), we obtain the following recursive relations for the generating function:

$$F^{(0,a,b)}(x,t)=x^{\delta_{ab}}U_a\left(\sum_{c=1}^{M}e_{ac}\left[1-r_{ac}G_{ac}(t)+r_{ac}\int_0^t dG_{ac}(\tau)F^{(1,c,b)}(x,t-\tau)\right]\right),\tag{A10}$$

$$F^{(d,a,b)}(x,t) = x^{\delta_{ab}} V_a\left(\sum_{c=1}^{M} e_{ac}\left[1 - r_{ac}G_{ac}(t) + r_{ac}\int_0^t dG_{ac}(\tau)F^{(d,c,b)}(x,t-\tau)\right]\right), \tag{A11}$$

$$F^{(D,a,b)}(x,t) = x^{\delta_{ab}}. \tag{A12}$$

Using these recursive equations we can proof lemma IV.1.

Proof of lemma IV.1

Proof. Let

$$N^{(d,a,b)}(t) = \frac{\partial F^{(d,a,b)}(1,t)}{\partial x} \tag{A13}$$

be the mean number of infected type-b agents on the branch rooted at a type-a agent at generation d. In particular, $N_{ab}(t)=N^{(0,a,b)}(t)$. Making use of the recursive relations (A10)–(A12) we obtain

$$N^{(0,a,b)}(t) = \delta_{ab} + \dot{U}_a(1)\sum_{c=1}^{M} r_{ac}\int_0^t dG_{ac}(\tau)N^{(1,c,b)}(t-\tau), \tag{A14}$$

$$N^{(d,a,b)}(t) = \delta_{ab} + \dot{V}_a(1)\sum_{c=1}^{M} r_{ac}\int_0^t dG_{ac}(\tau)N^{(d+1,c,b)}(t-\tau), \tag{A15}$$

$$N^{(D,a,b)}(t) = \delta_{ab}. \tag{A16}$$

Iterating these recursive relations from $d=D$ to $d=0$ we obtain Eq. (18). Then differentiating with respect to time we finally obtain Eq. (19). In this step we also make use of the relation between $\dot{U}_a(1)$ and $\dot{V}_a(1)$ and the average degrees, Eqs. (7) and (8). ■

APPENDIX B: FINAL OUTBREAK SIZE, STRONGLY CONNECTED TYPE NETWORKS

Theorem IV.2 can be generalized to type networks that may not be complete but are still strongly connected; i.e., there is a path from every type a to every type b. In this case some entries of R_{ac} and $(\tilde{R}^{d-1})_{cb}$ in Eq. (28) may be zero. Intuitively this means that some types c may not be a neighbor of a and, if they are, there may not be a path from c to b (see Fig. 3). More precisely, given a type a let $\mathrm{Out}(a)$ be its set of out-neighbors—i.e., $\mathrm{Out}(a)=\{c\,|\,e_{ac}>0\}$—and given a type b let $\mathrm{In}_d(b)$ be the set of types from where b is reached after d hops on the type network—i.e., $\mathrm{In}_d(b)=\{c\,|\,(e^d)_{cb}>0\}$. Furthermore, let

$$S_d^{(a,b)} = \mathrm{Out}(a) \cap \mathrm{In}_{d-1}(b) \tag{B1}$$

denote the set of types that are out-neighbors of the index case type a and belong to at least one path of length d from a to b on the type network. For instance, in the example in Fig. 3, $S_1^{(a,b)}=\emptyset$, $S_2^{(a,b)}=\{c'\}$, $S_3^{(a,b)}=\{c''\}$, and $S_d^{(a,b)}=\emptyset$ for all $d>3$.

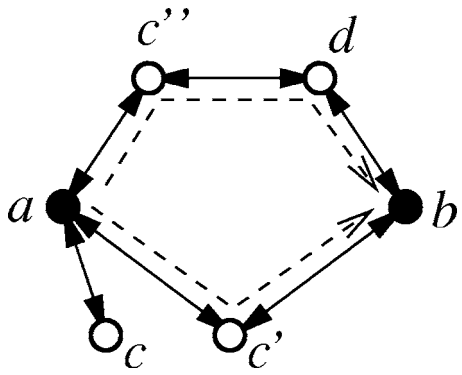

FIG. 3. Strongly connected type network with six types. The dashed lines indicate the possible paths from type a to b. Note that only types c, c', and c'' are neighbors of type a and type b can be only reached from the last two.

Theorem B.1 (strongly connected type network). Consider a strongly connected type network. Let ρ be the largest eigenvalue of $\tilde{R}$, Eq. (17), d_{ab} the distance on the type network from type a to b, $n=[D/d_{ab}]$ and $D_{ab}=nd_{ab}$. Then

$$N_{ab}(\infty) = u_{ab}\sum_{1\le d\le D|S_d^{(a,b)}\neq\emptyset} \rho^{d-1}, \tag{B2}$$

where u_{ab} is independent of D.

Proof. The conditions of the Perron-Frobenius theorem [45] are valid beyond positively defined matrices and hold for the mixing matrix representing a strongly connected network. Thus, the largest eigenvalue of $\tilde{R}$ is simple and all the entries of its corresponding eigenvector $\vec{v}$ are different from zero and have the same sign. In particular we choose all the components of $\vec{v}$ to be positive. Based on this fact we can write Eq. (28). There may be, however, some entries of e and thus of R, Eq. (16), and $\tilde{R}^{d-1}$ that are zero. Indeed we can rewrite Eq. (28) as

$$(R\tilde{R}^{d-1})_{ab} = \sum_{1\le c\le M|c\in S_d^{(a,b)}} \frac{R_{ac}}{v_c}v_c(\tilde{R}^{d-1})_{cb}. \tag{B3}$$

Thus $(R\tilde{R}^{d-1})_{ab}=0$ whenever $S_d^{(a,b)}=\emptyset$. Otherwise, we obtain the inequalities

$$u_{ab}^{(\min)}\rho^{d-1} \le (R\tilde{R}^{d-1})_{ab} \le u_{ab}^{(\max)}\rho^{d-1}, \tag{B4}$$

where

$$u_{ab}^{(\min)} = \min_{c\in\mathrm{Out}(a)} \frac{v_b}{v_c}R_{ac}, \tag{B5}$$

$$u_{ab}^{(\max)} = \max_{c\in\mathrm{Out}(a)} \frac{v_b}{v_c}R_{ac}. \tag{B6}$$

From Eqs. (27) and (24) we obtain

$$1 + u_{ab}^{(\min)} \sum_{1 \leq d \leq D | S_d^{(a,b)} \neq \emptyset} \rho^{d-1} \leq N_{ab}(\infty) \leq 1 + u_{ab}^{(\max)} \sum_{1 \leq d \leq D | S_d^{(a,b)} \neq \emptyset} \rho^{d-1}. \quad \text{(B7)}$$

From this equation we obtain Eq. (B2) with

$$0 < u_{ab}^{(\min)} \leq u_{ab} \leq u_{ab}^{(\max)} < \infty, \quad \text{(B8)}$$

where the inequality $u_{ab}^{(\min)} > 0$ follows from Eq. (B5). ■

APPENDIX C: PROOF OF THEOREM IV.3

The following theorem is a generalization of theorem IV.3 to the more general case of strongly connected type networks.

Theorem C.1. Consider a strongly connected type network and a homogeneous and exponential distribution of generation times $G_{ab}(\tau) = 1 - e^{-\lambda \tau}$, where λ is the transmission rate. Let ρ be the largest eigenvalue of $\tilde{R}$, Eq. (17), and let

$$\theta = \frac{D-1}{\rho}. \quad \text{(C1)}$$

$\theta \gg 1$: If $\rho > 1$ and $1 \ll \lambda t \ll \theta$, then

$$I_{ab}(t) \sim e^{(\rho-1)\lambda t}. \quad \text{(C2)}$$

$\theta \ll 1$: If $\lambda t \gg \theta$, then

$$\frac{I_{ab}(t)}{N_{ab}(\infty)} = \frac{\lambda(\lambda t)^{D_{ab}-1} e^{-\lambda t}}{(D_{ab}-1)!} \left[1 + O\left(\frac{\theta}{\lambda t} \right) \right], \quad \text{(C3)}$$

where D_{ab} is the same as in theorem B.1.

Proof. $\theta \gg 1$. Following the same guidelines of the theorem B.1 proof we arrive to the inequality

$$u_{ab}^{(\min)} f_{ab}(t) \leq I_{ab}(t) \leq u_{ab}^{(\max)} f_{ab}(t), \quad \text{(C4)}$$

where

$$f_{ab}(t) = \sum_{1 \leq d \leq D | S_d^{(a,b)} \neq \emptyset} \frac{\lambda(\rho\lambda t)^{d-1} e^{-\lambda t}}{(d-1)!}. \quad \text{(C5)}$$

The Laplace transform of $f_{ab}(t)$ is given by

$$\hat{f}_{ab}(\omega) = \int_0^\infty dt f_{ab}(t) e^{-\omega t} = \frac{a}{\rho} \sum_{1 \leq d \leq D | S_d^{(a,b)} \neq \emptyset} \left(\frac{\rho\lambda}{\omega + \lambda} \right)^d. \quad \text{(C6)}$$

When $D \to \infty$ this series converges only for $\omega > (\rho - 1)\lambda$. Therefore, $f_{ab}(t) \sim e^{(\rho-1)\lambda t}$ when $\lambda t \to \infty$.

$\theta \ll 1$: The demonstration of this case is straightforward. From theorem B.1 it follows that $(R\tilde{R}^{d-1})_{ab}$ is of order ρ^{d-1} for $S_d^{(a,b)} \neq \emptyset$. Therefore, for $\rho \gg D$ the sum in Eq. (34) is dominated by the $d = D_{ab}$ term. Corrections are given by the ratio between the $d = D_{ab}$ and the preceding term satisfying $S_d^{(a,b)} \neq \emptyset$, which is at most $d = D_{ab} - 1$. ■

[1] R. M. Anderson and R. M. May, *Infectious Diseases of Humans* (Oxford University Press, New York, 1991).
[2] L. Hufnagel, D. Brockmann, and G. T., Proc. Natl. Acad. Sci. U.S.A. **101**, 15124 (2004).
[3] J. Koopman, Annu. Rev. Public Health **25**, 303 (2004).
[4] T. C. Germann, K. Kadau, I. M. Longini, and C. A. Macken, Proc. Natl. Acad. Sci. U.S.A. **103**, 5935 (2006).
[5] V. Colizza, A. Barrat, M. Barthelemy, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **103**, 2015 (2006).
[6] R. M. May and R. M. Anderson, Nature (London) **326**, 137 (1987).
[7] R. M. Anderson *et al.*, Philos. Trans. R. Soc. London, Ser. B **359**, 1091 (2004).
[8] J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, and W. M. Getz, Nature (London) **438**, 355 (2005).
[9] T. M. May and R. M. Anderson, Philos. Trans. R. Soc. London, Ser. B **321**, 565 (1988).
[10] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Berg, Nature (London) **411**, 907 (2001).
[11] J. H. Jones and M. S. Handcock, Proc. R. Soc. London, Ser. B **270**, 1123 (2003).
[12] A. Schneeberger, C. H. Mercer, S. A. Gregson, N. M. Ferguson, C. A. Nyamukapa, R. M. Anderson, A. M. Johnson, and G. P. Garnett, Sex Transm. Dis. **31**, 380 (2004).
[13] H. Andersson, Ann. Appl. Probab. **8**, 1331 (1998).
[14] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).
[15] C. Moore and M. E. J. Newman, Phys. Rev. E **61**, 5678 (2000).
[16] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2001).
[17] L. A. Rvachev and I. M. Longini, Math. Biosci. **75**, 3 (1985).
[18] A. Flahault, S. Letrait, P. Blin, S. Hazout, J. Menares, and A. J. Valleron, Stat. Med. **7**, 1147 (1988).
[19] S. Eubank, H. Guclu, V. S. A. Kumar, M. Marathe, A. Srinivasan, Z. Toroczcai, and N. Wang, Nature (London) **429**, 180 (2004).
[20] C. J. Mode, *Multitype Branching Processes* (Elsevier, New York, 1971).
[21] C. J. Mode and C. K. Sleeman, *Stochastic Processes in Epidemiology* (World Scientific, Singapore, 2000).
[22] A. Vazquez, Phys. Rev. Lett. **96**, 038702 (2006).
[23] A. Vazquez, DIMACS Series on Discrete Mathematics and Theoretical Computer Science **73**, 163 (2006).
[24] A. Vazquez, Phys. Rev. E **74**, 056101 (2006).
[25] M. E. J. Newman, Phys. Rev. E **67**, 026126 (2003).
[26] S. R. Friedman *et al.*, Am. J. Public Health **87**, 1289 (1997).
[27] W. J. Edmunds, C. J. O. O'Callaghan, and D. J. Nokes, Proc. R. Soc. London, Ser. B **264**, 949 (1997).
[28] A. C. Ghani and G. P. Garnett, J. R. Stat. Soc. Ser. A (Stat. Soc.) **161**, 227 (1998).
[29] M. J. Keeling and K. T. D. Eames, J. R. Soc., Interface **2**, 297 (2005).
[30] C. H. Watts and R. M. May, Math. Biosci. **108**, 89 (1992).
[31] M. Kretzschmar and M. Morris, Math. Biosci. **133**, 165

(1996).
[32] G. P. Garnett and J. A. Johnson, AIDS **11**, 681 (1997).
[33] D. J. Watts, *Small Worlds: The Dynamics of Networks between Order and Randomness* (Princeton University Press, Princeton, 1999).
[34] S. Milgram, Psychol. Today **2**, 60 (1967).
[35] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[36] A. Vazquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).
[37] R. Pastor-Satorras, A. Vazquez, and A. Vespignani, Lect. Notes Phys. **650**, 425 (2004).
[38] J.-P. Eckmann, E. Moses, and D. Sergi, Proc. Natl. Acad. Sci. U.S.A. **101**, 14333 (2004).
[39] H. Ebel, L.-I. Mielsch, and S. Bornholdt, Phys. Rev. E **66**, R35103 (2002).
[40] B. Bollobás, *Random Graphs* (Cambridge University Press, Cambridge, England, 2001).
[41] F. Chung and L. Lu, Proc. Natl. Acad. Sci. U.S.A. **99**, 15879 (2002).
[42] B. Bollobás and O. M. Riordan, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Weinheim, 2003), pp. 1–34.
[43] R. Cohen and S. Havlin, Phys. Rev. Lett. **90**, 058701 (2003).
[44] J. K. J. Leskovec and C. Faloutsos, e-print physics/0603229.
[45] C. Godsil and G. Royle, *Algebraic Graph Theory* (Springer, New York, 2001).
[46] C. Fraser, S. Riley, R. Anderson, and N. M. Ferguson, Proc. Natl. Acad. Sci. U.S.A. **101**, 6146 (2004).
[47] F. R. Gantmacher, *Mathrix Theory* (AMS, Providence, 1990), Vol. I.
[48] A. Vazquez, Phys. Rev. E **67**, 056104 (2003).