Limited data on infectious disease distribution exposes ambiguity in epidemic modeling choices

Laura Di Domenico, Eugenio Valdano, Vittoria Colizza

Limited data on infectious disease distribution exposes ambiguity in epidemic modeling choices

Laura Di Domenico [1], Eugenio Valdano [2] and Vittoria Colizza [2,3,*]

[1] *Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland*

[2] *Sorbonne Université, INSERM, Institut Pierre Louis d'Epidemiologie et de Santé Publique (IPLESP), Paris, France*

[3] *Department of Biology, Georgetown University, Washington, District of Columbia, USA*

(Received 26 January 2024; accepted 20 May 2024; published 10 June 2024)

Traditional disease transmission models assume that the infectious period is exponentially distributed with a recovery rate fixed in time and across individuals. This assumption provides analytical and computational advantages, however, it is often unrealistic when compared to empirical data. Current efforts in modeling nonexponentially distributed infectious periods are either limited to special cases or lead to unsolvable models. Also, the link between empirical data (the infectious period distribution) and the modeling needs (the definition of the corresponding recovery rates) lacks a clear understanding. Here we introduce a mapping of an arbitrary distribution of infectious periods into a distribution of recovery rates. Under the Markovian assumption to ensure analytical tractability, we show that the same infectious period distribution at the population level can be reproduced by two modeling schemes that we call *host-based* and *population-based*, depending on the individual response to the infection, and aggregated empirical data cannot easily discriminate the correct scheme. Besides being conceptually different, the two schemes also lead to different epidemic trajectories. Although sharing the same behavior close to the disease-free equilibrium, the *host-based* scheme deviates from the expected epidemic when reaching the endemic equilibrium of a susceptible-infectious-susceptible transmission model, while the *population-based* scheme turns out to be equivalent to assuming a homogeneous recovery rate. We show this through analytical computations and stochastic epidemic simulations on a contact network, using both generative network models and empirical contact data. It is therefore possible to reproduce heterogeneous infectious periods in network-based transmission models, however, the resulting prevalence is sensitive to the modeling choice for the interpretation of the empirically collected data on the length of the infectious period. In the absence of higher resolution data, studies should acknowledge such deviations in the epidemic predictions.

DOI: 10.1103/PhysRevResearch.6.023265

I. INTRODUCTION

The infectious period has a key role in the progression of an infectious disease. It is the time interval during which an infected host can transmit the pathogen to other susceptible individuals, therefore it is closely linked to the ecological persistence of the disease and the challenges of its eradication. The infectious period depends both on disease natural history and on the interventions possibly put in place: treatment, for instance, can be effective in reducing it. Collecting and characterizing this type of data is quite challenging, as demonstrated by statistical studies on measles [1] and long-lived infections such as HIV [2]. Recent examples about COVID-19 [3] and monkeypox [4] show that a common feature of empirical infectious period distributions is their overdispersion or underdispersion, deviating from an exponential distribution. This is in contrast with the traditional analytical approach adopted in the physics community, based on modeling the infectious period with a fixed recovery rate. By this approach, each infectious host (or node i in a network) recovers at a rate $\mu_i = \mu$. This paradigm has a twofold advantage. First, it is easy to handle both analytically and numerically; second, it maps the process into a spontaneous one-body reaction, allowing us to borrow solutions from other fields of physics, like reaction-diffusion processes and decays. However, this approach uniquely constrains the distribution of the infectious period τ to an exponential distribution having expected value $\langle\tau\rangle = \mu^{-1}$. Therefore, fitting $\langle\tau\rangle$ from real data leaves no extra degrees of freedom to model the dispersion of the data.

Efforts to overcome this problem already exist, e.g., through additional infectious compartments [5–7] or by splitting hosts into epidemiologically relevant groups [8,9]. An alternative approach is to directly plug heterogeneous recovery rates into the model [10,11]. All three approaches have limitations. Adding compartments limits the choice of possible distributions, partitioning individuals is limited to specific epidemiological contexts, and using distributed recovery rates is not justified by a clear link with a corresponding distribution of the infectious periods. On the other hand, parametrizing models with an explicit distribution of the infectious period makes them harder to solve [12].

In this paper, we develop a theory to include arbitrary distributions of infectious periods, so they can be informed by real data. We treat both the infectious period τ and the recovery rate μ as stochastic variables, and determine the mapping between their probability distributions. We treat

*vittoria.colizza@inserm.fr

TABLE I. Mapping from the infectious period distribution $g(\tau)$ to the recovery rate distribution $f(\mu)$, and corresponding average infectious period $\langle\tau\rangle$, for some commonly used infectious period distributions. δ is the Dirac delta distribution. H is the Heaviside step function. For gamma-distributed g, the displayed f is valid for $\kappa < 1$. For power-law distributed g, $h > 2$ ensures the existence of both f and $\langle\tau\rangle$.

Infectious period distribution $g(\tau)$		Recovery rate distribution $f(\mu)$		$\langle\tau\rangle$
Exponential	$r\,e^{-r\tau}$	Dirac delta	$\delta(\mu - r)$	r^{-1}
Gamma (κ, θ)	$\frac{\tau^{\kappa-1}e^{-\tau/\theta}}{\Gamma(\kappa)\theta^{\kappa}}$	Power law	$\frac{\sin(\kappa\pi)}{\pi}(\theta\mu - 1)^{-\kappa}H(\theta\mu - 1)$	$\kappa\theta$
Power law (h, ϵ)	$(h-1)\epsilon^{h-1}(\tau+\epsilon)^{-h}$	Gamma $(h-1, \epsilon^{-1})$	$\frac{\epsilon^{h-1}\mu^{h-2}e^{-\epsilon\mu}}{\Gamma(h-1)}$	$\frac{\epsilon}{(h-2)}$

recovery of an individual node i as a spontaneous process occurring at a given rate μ_i, but we sample that rate from a distribution f appropriately chosen so the resulting distribution of infectious periods τ at the population level follows a desired distribution g, possibly fitted on data. Our mapping allows to analytically derive $f(\mu)$ from $g(\tau)$.

This mapping also provides a clear understanding of the link between empirical data (the infectious disease distribution) and the modeling needs (the definition of the corresponding recovery rate). The infectious period distribution $g(\tau)$ is usually reconstructed from population data collected through surveillance and observational studies. The recovery rate μ, instead, is a variable defined by the modeling scheme at the individual level. This implies a degeneracy of models that may assign individual rates differently, but produce the same $g(\tau)$. We study this by defining the host-based scheme, by which each host i is given a recovery rate μ_i that is fixed in time and sampled from f, and the population-based scheme, by which every time a host recovers it resamples its recovery rate from f (i.e., μ_i varies with time). The latter scheme models a scenario in which the chance of recovery for a specific host changes after each reinfection, for instance, because of a difference in the immunity response. We show that both schemes recover the same distribution g computed from f through our mapping. However, we prove that, while the two schemes share their critical behavior (close to the epidemic threshold), they significantly differ in the endemic equilibrium. This has far-fetching implications, as the choice of the correct scheme may be nonunivocal, depending on the epidemic context, but data collection often does not allow us to empirically discriminate between the host-based and the population-based schemes.

In Sec. II, we build and discuss the analytical mapping from g to f. We also define the host-based and population-based schemes. We then characterize the epidemic threshold (Sec. III) and the endemic equilibrium (Sec. IV). We then apply our methodology to two real-world scenarios, i.e., the spread of nosocomial infections in health-care settings using empirical contact data collected from sensors (Sec. V) and the spread of livestock disease over an empirical network of cattle trade movements (Sec. VI). In Sec. VII, we discuss the implications of our results for the modeling community.

II. MAPPING INFECTIOUS PERIODS INTO RECOVERY RATES

We consider the susceptible-infectious-susceptible (SIS) model [13,14]. A susceptible individual becomes infected at rate λ upon contact with an infected host. It then recovers spontaneously to the susceptible state at rate μ. Recovery confers no immunity. The transmission rate λ is constant, while the recovery rate μ is a stochastic variable with distribution f. A varying λ with constant μ was studied in Ref. [15] for other purposes. Let τ be the stochastic variable representing the infectious period, with distribution g. In the case of a constant μ, the distribution g would be the exponential distribution.

Just as in a standard reaction-diffusion process, recovery is Markovian once the recovery rate is fixed, i.e., $\tau|\mu \sim \mathrm{Exp}(\mu)$. Under this assumption, we now write $g(\tau)$ as a marginal probability, borrowing methods usually adopted in the field of superstatistics [16,17]. We can write

$$g(\tau) = \int_0^\infty \mathrm{d}\mu f(\mu)\mu e^{-\mu\tau} = -\frac{\mathrm{d}}{\mathrm{d}\tau}\mathcal{L}[f](\tau), \tag{1}$$

where $\mathcal{L}$ is the Laplace transform operator. By integrating in τ and solving for f, we obtain

$$f(\mu) = \mathcal{L}^{-1}[\hat{G}](\mu), \tag{2}$$

where $\hat{G}$ is the tail distribution function of τ, i.e., $\hat{G}(\tau) = \int_\tau^\infty \mathrm{d}x\, g(x)$. Equation (2) is solvable either by explicit computation of the inverse Laplace transform—when possible—or by numerical integration [18]. In the latter case, it is possible to determine beforehand if the solution exists by noticing that one can generate the moments of f by repeatedly deriving Eq. (1) with respect to τ, and then setting $\tau = 0$:

$$m_0 = 1; \tag{3}$$

$$m_n = (-1)^{n-1}\frac{\mathrm{d}^{n-1}}{\mathrm{d}\tau^{n-1}}g(\tau)\bigg|_{\tau=0} \qquad \forall n \in \mathbb{N}\setminus\{0\}, \tag{4}$$

where m_n is the n-th moment of f, i.e., $m_n = \int_0^\infty \mathrm{d}\mu\mu^n f(\mu)$. As such, determining if the solution of Eq. (2) exists maps onto a Stiltjes moment problem [19] (see Appendix A).

Table I reports the expression of $f(\mu)$ for some commonly used distributions of infectious periods: exponential, gamma, and power law.

The mapping introduced above works for both the host-based and population-based schemes. The difference relies on the fact that in the former scheme each host i samples its μ_i from f only once at the beginning, while in the latter each host resamples its μ_i every time it recovers. Therefore, in terms of infectious period, in the population-based scheme each individual follows the same infectious period distribution $g(\tau)$, while in the host-based scheme each individual is characterized by a different (exponential) infectious period

distribution, producing the distribution $g(\tau)$ when aggregated at the population level.

III. EPIDEMIC THRESHOLD

The epidemic threshold is the critical value λ_c of the transmission rate that discriminates between the disease-free state ($\lambda < \lambda_c$), and the endemic regime ($\lambda > \lambda_c$). The computation of the epidemic threshold provides an important public health metric to evaluate intervention policies [9,20].

The epidemic threshold depends on both disease features (transmission, recovery), and the topology of the underlying network of contacts along which the spreading occurs. We assume a network of N nodes with adjacency matrix A. Let $x_i(t)$ be the probability that node i is infectious at time t, with $i = 1, \ldots, N$. In the host-based scheme, node i has a fixed recovery rate μ_i. Following the microscopic Markov chain formalism [21–23], we can write the differential equations describing the evolution of the disease as a perturbation of the disease-free state [thus neglecting $O(x_i x_j)$ and higher orders]:

$$\frac{\mathrm{d}x_i(t)}{\mathrm{d}t} = -\mu_i x_i(t) + \lambda \sum_j A_{ij} x_j(t). \tag{5}$$

In matrix form, this reads $\dot{\mathbf{x}} = (-M + \lambda A)\mathbf{x}$, where $\mathbf{x} = (x_1(t), \ldots, x_N(t))$ and $M = \mathrm{diag}\{\mu_1, \ldots, \mu_N\}$ is the diagonal matrix containing all the recovery rates, which have been sampled from $f(\mu)$ at $t = 0$. The epidemic threshold λ_c then solves the equation

$$\rho(-M + \lambda_c A) = 0, \tag{6}$$

as proven in Refs. [23,24], where ρ indicates the spectral radius, i.e., the largest eigenvalue.

In the population-based scheme, we can observe that, close to the disease-free equilibrium, reinfection events are suppressed and can thus be dropped in the threshold computation as higher-order terms. This means that we can neglect the update mechanism of μ and retrieve the same epidemic threshold as in the host-based scheme.

In the standard case of exponentially distributed τ (i.e., constant μ), Eq. (6) reduces to

$$\lambda_c = \frac{\mu}{\rho(A)} = \frac{1}{\langle\tau\rangle\rho(A)}, \tag{7}$$

given that the rate of the exponential distribution coincides with the inverse of its expected value. We now solve Eq. (6) in the case of a nonexponential distribution $g(\tau)$, i.e., with heterogeneous recovery rates μ_i. In practical applications, the matrix A often comes from a generative network model designed to reproduce key topological features of the contact structure of the population under study. Reference [25] argues that generative network models are representable in terms of adjacency matrices whose rank equals the number of node features constrained. For instance, if one just fixes the expected degree of each node (the so-called configuration model [26–29]), one will get the rank-one adjacency matrix $A = KK^T/(N\langle k\rangle)$, where K is the N-dimensional vector containing the expected degree of each node and $\langle k\rangle$ is the average expected degree. The activity-driven model introduced in Ref. [30] is an example of a rank-two model (see Appendix E).

For a generic rank-r model, one can write

$$A = V\Delta V^T, \tag{8}$$

where V is an $N \times r$ matrix encoding node properties and Δ is a r-dimensional bilinear form encoding the geometry of the model (see Ref. [25]). The epidemic threshold of the generic network model requires plugging Eq. (8) into Eq. (6), and working out the calculations under the assumption that the node properties fixed by the model are uncorrelated with recovery rates (see Appendix B for explicit computations). Unexpectedly, this leads to the following expression of the epidemic threshold:

$$\lambda_c = \frac{1}{\langle\tau\rangle\rho(A)}, \tag{9}$$

which is the same as Eq. (7), when expressed in terms of the average infectious period. This result shows that only the average infectious period impacts the epidemic threshold. The distribution $g(\tau)$ may be arbitrarily complex, but its first moment is enough to discriminate between disease extinction and endemicity. A model with fixed μ is therefore sufficient to study the critical behavior of disease spreading, and this is beneficial in two aspects: (i) such a model is analytically and numerically the simplest possible, and (ii) estimating $\langle\tau\rangle$ from data is easier than fitting the full distribution, especially if the available sample is small. However, Eq. (9) also warns us about the misuse of the recovery rate. It rigorously proves that the relevant observable is indeed the average infectious period and not the average recovery rate. Replacing $1/\langle\tau\rangle$ with $\langle\mu\rangle$ in Eq. (7) would lead to an overestimation of the epidemic threshold, because the identity

$$\langle\tau^j\rangle = j!\langle\mu^{-j}\rangle, \quad \forall j \in \mathbb{N} \tag{10}$$

[provable from Eq. (1)], combined with Jensen's inequality [31], implies that $\langle\mu\rangle \geqslant \langle\tau\rangle^{-1}$. Overestimating the epidemic threshold is potentially harmful, as it leads to underestimation of the risk of the disease becoming endemic.

IV. ENDEMIC PREVALENCE

Above the epidemic threshold, the SIS model converges to an endemic equilibrium characterized by a certain disease prevalence (i.e., fraction of population infected at a given time) [32]. Quantifying the endemic prevalence, alongside the epidemic threshold, is relevant from a public health perspective, as it allows to anticipate the impact of the disease spreading in the long-term.

The endemic equilibrium is typically harder to derive analytically concerning the epidemic threshold: no closed-form solution exists beyond homogeneous mixing even in the case of exponentially distributed τ. Here we focus on homogeneous mixing, i.e., a sequence of Erdős–Rényi networks [33], and compute the corresponding endemic prevalence in the population-based and host-based scheme. Appendix C contains a generalization to the configuration model for the population-based scheme.

To proceed, we divide compartments I and S into subclasses according to the recovery rate. Classes $I_j(t)$ and $S_j(t)$ represent, respectively, the number of infected and susceptible individuals at time t with recovery rate equal to μ_j. Recovery

rate thus gets formally discretized in an arbitrarily large number of values. The spreading equations are

$$\frac{\mathrm{d}}{\mathrm{d}t}I_j(t) = -\mu_j I_j(t) + \lambda \frac{1}{N} S_j(t) \sum_h I_h(t), \tag{11}$$

where the average connectivity of the homogeneous network is absorbed in the transmission rate λ. At time $t = 0$, we have a fraction $f(\mu_j)\mathrm{d}\mu_j$ of the total number of individuals N that have rate μ_j. In the population-based scheme, as time passes, large values of μ are replaced sooner, as they generate, on average, shorter infectious periods. Likewise, smaller values of μ persist longer. This implies that the fraction of hosts with rate μ_j at a given time deviates from the initial fraction $f(\mu_j)\,\mathrm{d}\mu_j$ as time passes. Notwithstanding, if we look exclusively at compartment $S_j(t)$, we can state that $S_j(t) = f(\mu_j)\mathrm{d}\mu_j\, S(t)$ because a new μ_j is assigned after recovery, and the interevent time between recovery and reinfection does not depend on the recovery rate μ_j of the susceptible individual. By inserting this in Eq. (11), we can then set the right-hand side (rhs) equal to zero and sum over j to get rid of the discretization and obtain the endemic prevalence at equilibrium:

$$x_{\mathrm{eq}} = \frac{I_{\mathrm{eq}}}{N} = 1 - \frac{1}{\langle\tau\rangle\lambda}. \tag{12}$$

So we find that, as for the epidemic threshold, the endemic equilibrium in the population-based scheme depends only on the average infectious period, regardless of its distribution, and coincides with the equilibrium obtained assuming a homogeneous recovery rate $\mu = \langle\tau\rangle^{-1}$.

It is a different matter for the host-based scheme. We find another formula to analytically derive endemic prevalence x_{eq} for homogeneous mixing as a solution of the following equation:

$$\mathcal{L}[g](\lambda x_{\mathrm{eq}}) = 1 - x_{\mathrm{eq}}. \tag{13}$$

Details of the derivation can be found in Appendix D. In general, Eq. (13) is solvable numerically. In some cases, it leads to an analytic expression for x_{eq}. One such case is obviously when τ is exponentially distributed, giving the same result as in Eq. (12). If τ is gamma distributed (see Table I), Eq. (13) becomes relatively simple: $(1 + x_{\mathrm{eq}}\lambda\langle\tau\rangle/\kappa)^{-\kappa} = 1 - x_{\mathrm{eq}}$, where we used the parametrization $\langle\tau\rangle = \kappa\theta$. Then, further assuming $\kappa = 1/2$, gives $x_{\mathrm{eq}} = 1 - \frac{1+\sqrt{1+8\langle\tau\rangle\lambda}}{4\langle\tau\rangle\lambda}$. This example explicitly shows how different the endemic equilibrium can be from the exponentially distributed case [Eq. (12)].

In Sec. III, we showed that the average infectious period $\langle\tau\rangle$ alone completely determines the epidemic threshold. Equation (13) instead shows that higher moments of τ have an impact on the endemic equilibrium, in the case of the *host-based* scheme. In Fig. 1, we keep $\langle\tau\rangle$ fixed and explore different levels of dispersion around it in case of gamma-distributed and power-law-distributed infectious periods. Comparison with exponentially distributed τ (at same $\langle\tau\rangle$) shows that in the *host-based* scheme: (i) higher variance gives consistently lower endemic prevalence and (ii) at fixed variance, gamma-distributed τ leads to lower prevalence than power-law-distributed τ.

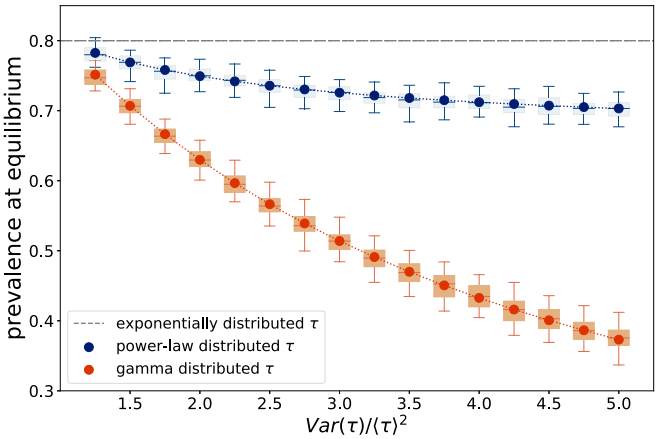

FIG. 1. Endemic prevalence in the host-based scheme as a function of the relative variance in $g(\tau)$. The gray line represents the endemic equilibrium for exponentially distributed τ (constant μ). Red and blue dots represent the numerical solution of Eq. (13), respectively, for gamma-distributed and power-law-distributed τ (see also Table I). Box-plots represent values of the endemic equilibrium obtained from 100 SIS stochastic simulations on an Erdős–Rényi network of $N = 10^3$ nodes and average degree 5. Disease parameters were fixed at $\langle\tau\rangle = 300$ time steps, $\lambda = 5\lambda_c$.

Figure 1 shows results obtained assuming a homogeneous network. We also performed numerical simulations of disease spreading on a heterogeneous network (namely, the activity-driven network introduced in Ref. [30]). We include the results in Appendix E. The results are qualitatively similar to the ones discussed above.

V. APPLICATION TO MRSA DIFFUSION IN HEALTHCARE SETTINGS

Methicillin-resistant *Staphylococcus aureus* (MRSA) is responsible for severe bacterial infections. Its acquired resistance to antimicrobial treatment makes it one of the most dreaded infections occurring in health-care settings [34]. Patients can get colonized through direct contact with asymptomatic carriers (including health-care workers). Outbreaks of MRSA infection increase mortality, hospitalization times, and are difficult and costly to contain [34].

MRSA carriage duration is nonexponentially distributed. We analyzed data on time to observed clearance [35] to reconstruct the distribution of carriage time $g(\tau)$. We fitted the data through maximum likelihood using exponential and gamma distributions. The Akaike information criterion (AIC) selected the gamma distribution as the best-fitting model (see Fig. 2). This supports, therefore, the application of the approach presented here, as accurate model predictions are needed to improve surveillance and response against MRSA diffusion.

We use the estimated $g(\tau)$ to simulate the spread of MRSA carriage on a real network of contacts among 590 individuals (both patients and health-care workers) collected through wearable sensors in a long-term and rehabilitation facility in Northern France [36,37]. We simulated both the host-based and population-based schemes. We also considered the scenario of constant recovery rate (exponentially distributed τ) as

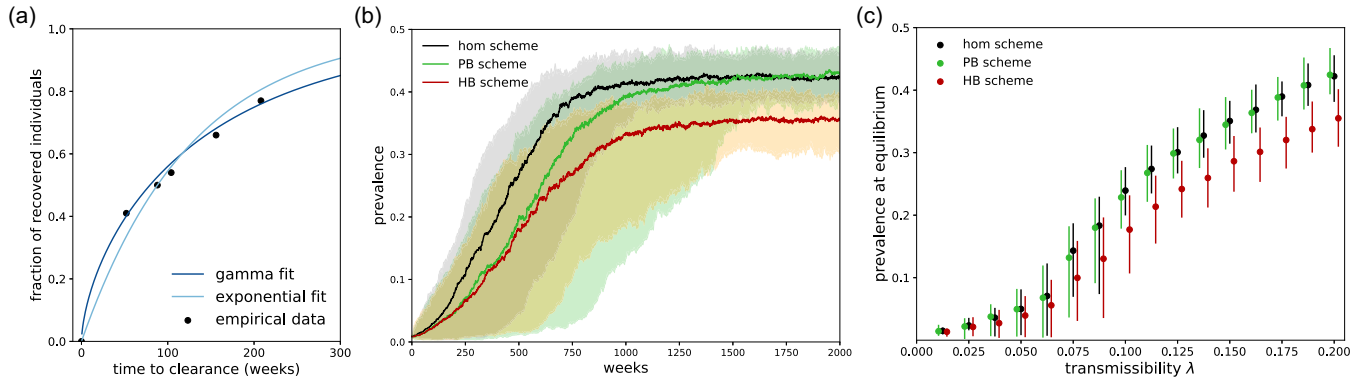

FIG. 2. MRSA infection spreading over an empirical contact network within a hospital. (a) Gamma vs exponential estimated cumulative distribution for MRSA colonization duration. Black dots represent empirical data on time to clearance from Ref. [35]. Values of parameters estimated at maximum likelihood: $r = 0.007$, $\kappa = 0.62$, $\theta = 238.04$ (see Table I for the definition). AIC = 316 and 319 for gamma and exponential distributions, respectively. (b) Colonization prevalence over time, for the host-based (red) and population-based schemes (green) with a gamma-distributed infectious period in comparison with the homogeneous scheme (black), assuming an exponential distribution and a constant recovery rate. Here λ was fixed at 0.2. (c) Colonization prevalence at equilibrium for different values of transmissibility λ. Results are averaged over 100 stochastic runs. Uncertainty bars and shaded areas indicate 95% probability ranges. Parameters of the gamma-distributed τ are the fitted estimates. The rate of the exponential distribution is fixed to have the same $\langle\tau\rangle$ as the gamma-distributed one. Contact data are aggregated at a weekly time scale, getting a time-evolving, weighted network where weights encode contact duration over a specific week.

benchmark, with the same $\langle\tau\rangle$. The results of the simulations are displayed in Figs. 2(b) and 2(c).

Epidemic trajectories in Fig. 2(b) show that heterogeneity in the recovery rates has an effect in slowing down disease spread with respect to the homogeneous scheme. When approaching the endemic equilibrium, Fig. 2(c) confirms that the homogeneous modeling scheme with a fixed rate and the population-based scheme share the same endemic prevalence, even in the case of a realistic temporal contact network. Instead, in the host-based scheme the predicted endemic prevalence turns out to be smaller. As the transmission rate λ decreases, the values for the three schemes converge, supporting the idea that they all hold the same epidemic threshold.

VI. APPLICATION TO LIVESTOCK DISEASE AND CATTLE TRADE NETWORK

Livestock infectious diseases threaten animal health and welfare, and ineffective epidemic control can have severe consequences for the economy of a country [38]. Livestock diseases can persist long enough in a given area, becoming endemic. For example, bovine tuberculosis has been affecting the cattle population in Italy for decades, and still circulates in southern regions [39]. Here, we considered outbreak data and cattle trade movements in Sicily, the Italian region with the highest prevalence of tuberculosis.

In a network perspective, livestock disease spreading are often modeled at the farm level, where a node is a farm and

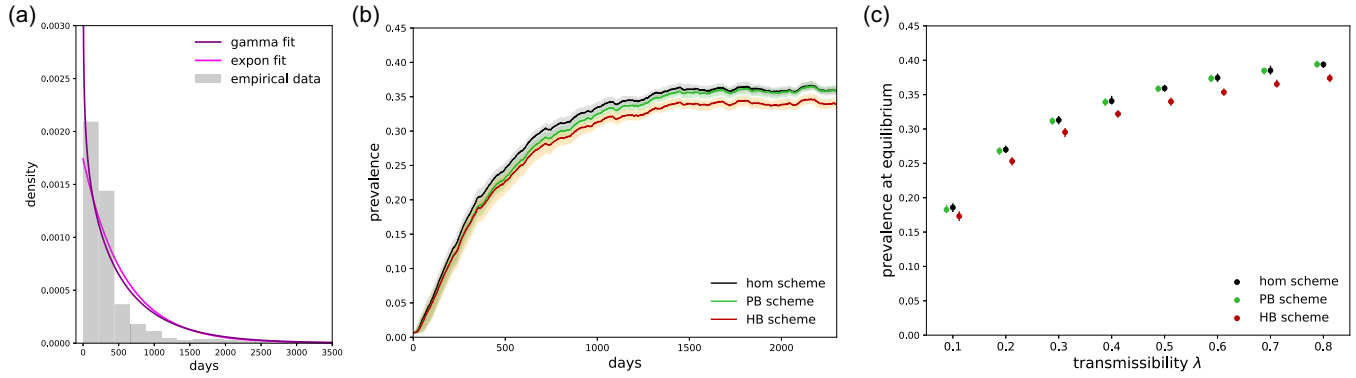

FIG. 3. Bovine tuberculosis spreading over an empirical network of farms connected through cattle-trade movements. (a) Gamma vs exponential estimated probability distributions for duration of an outbreak in a farms. Grey bars represent empirical data on outbreak duration observed in Sicily (Italy). Values of parameters estimated at maximum likelihood: $r = 0.002$, $\kappa = 0.82$, $\theta = 699.9$ (see Table I for the definition). AIC = 21 553 and 21 592 for gamma and exponential distribution, respectively. (b) Fraction of infected farms over time for the host-based (red) and the population-based scheme (green) with a gamma-distributed infectious period in comparison with the homogeneous scheme (black), assuming an exponential distribution and a constant recovery rate. Here λ was fixed at 0.5. (c) Fraction of infected farms at equilibrium for different values of transmissibility λ. Results are averaged over 100 stochastic runs. Uncertainty bars and shaded areas indicate 95% probability ranges. Parameters of the gamma-distributed τ are the fitted estimates. The rate of the exponential distribution is fixed to have the same $\langle\tau\rangle$ as the gamma-distributed one.

links are represented by livestock movements between farms due to cattle trade [9,40]. A farm is considered infected if there is an ongoing outbreak, hence duration of the outbreak in a farm represents the infectious period of the node. We used a database on tuberculosis outbreaks managed by IZSAM (Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise). The database contains records of the duration of 1468 tuberculosis outbreaks in cattle occurring in Sicily between 1983 and 2015. Fitting an exponential and a gamma distribution, we found that a gamma distribution best described the data in terms of AIC (Fig. 3). We thus applied our framework to model a gamma-distributed infectious period distribution. In the context of livestock diseases, the host-based scheme can model a scenario in which there are farms with faster recovery rates than others because of better surveillance systems and better compliance to public health recommendations.

We then constructed a daily temporal network starting from the data set of cattle trade movements in Sicily from January 2006 to December 2012, provided by IZSAM [40]. The network includes over 14 000 nodes. We accounted for the number of traded animals to define the risk of transmission between farms. Transmission may occur from an infected farm i to a susceptible one j upon trading animals at time t with a probability $1-(1-\lambda)^{W_{t,ij}}$, where $W_{t,ij}$ is the number of animals moved and λ indicates the per-animal risk of transmission, as done in Ref. [40]. We then simulated disease spreading over the network for different values of λ, with gamma-distributed infectious periods under the host-based and population-based schemes, in comparison with an exponential distribution with the same mean. We found qualitatively similar results with respect to the case of MRSA spreading in hospital. The endemic prevalence of tuberculosis in the population-based scheme coincides with the one obtained with exponentially distributed infectious periods, while the host-based scheme leads to a lower endemic equilibrium (Fig. 3). With respect to the previous case study, the fitted gamma-distribution is closer to an exponential distribution (shape parameter $k = 0.82$ for tuberculosis outbreak duration while $k = 0.62$ for MRSA colonization), hence the differences between the host-based and the population-based schemes are less evident, but they are still significant.

VII. DISCUSSION

A fixed recovery rate across individuals and throughout the epidemic outbreak fails in reproducing realistic distributions of infectious periods. Yet, it is the key modeling ingredient of traditional approaches because it treats recovery as a Markovian process, i.e., a spontaneous decay. This assumption allows analytical calculations, and largely simplifies numerical implementations. Based on a mapping of the infectious period distributions into a recovery rate distribution, we introduced a modeling framework that can capture and integrate arbitrary infectious period distributions that can be informed by empirical data, while remaining analytically and numerically treatable.

Collected empirical data usually provide information on the infectious period at the population level, not at the individual level, through a distribution of values. This lack of resolution at the host level opens the path to two possible modeling schemes, assuming either an immutable recovery rate per host (host-based scheme) or a rate that can be updated at each infection episode because it is altered by factors affecting the immune response of the individual (population-based scheme). When data on reinfection of the same individual are too scarce to estimate a distribution for each host (as is often the case, with a few exceptions [41]), the two schemes become empirically equivalent, but they conceal significant differences in terms of predictions.

We analytically prove that the epidemic threshold, in the case of any generative network model for hosts interactions, does not depend on the scheme chosen. We also prove that such threshold depends only on the average infectious period, making the standard assumption of constant recovery rate sufficient to correctly model the behavior around the disease-free state. Differences emerge when the system moves away from the critical point. The endemic disease prevalence—a predictor of how easy it is to eradicate a disease in a population—is quantitatively different in each scheme, as shown in theoretical examples of disease spread and in two case studies applied to the spread of the multiresistant bacteria MRSA in health-care settings and persistence of bovine tuberculosis in Sicily.

The difference extends to the out-of-equilibrium dynamics, as disease spreading in the population-based scheme is faster with respect to the host-based scheme.

Our findings show that modeling heterogeneity within the population-based or the host-based paradigms, although apparently equivalent, has a considerable impact on the endemic disease prevalence, and caution should be taken when addressing specific epidemic contexts where data do not allow to distinguish between the schemes. This problem might disappear naturally in contexts in which individual recovery rates have a concrete meaning or can be measured directly, e.g., in information diffusion processes [42], but in the general case of biological diseases, modelers should be aware that an arbitrary choice of the scheme may represent a potential source of bias to be considered.

ACKNOWLEDGMENTS

This study was partially funded by the ANR projects SPHINX (No. ANR-17-CE36-0008-05), DATAREDUX (No. ANR-19-CE46-0008-03), ARCANE (No. ANR-23-CE45-0036-03), and the EU Horizon 2020 grant MOOD (No. H2020-874850) to V.C. We thank the IZSAM for providing access to data obtained from the Italian National Database for Animal Identification and Registration and from the Italian National Animal Health Managing Information System. This information is managed by IZSAM on behalf of the Italian Ministry of Health.

APPENDIX A: EXISTENCE OF THE FUNCTION f

The conditions on $g(\tau)$ under which there exists a probability density function $f(\mu)$ solving Eq. (1) can be described in terms of a *moment problem* [19]. Let us evaluate the nth

derivative of $g(\tau)$ in $\tau = 0$ from Eq. (1),

$$\left.\frac{\mathrm{d}^n}{\mathrm{d}\tau^n} g(\tau)\right|_{\tau=0} = (-1)^n \int_0^{+\infty} \mathrm{d}\mu f(\mu)\, \mu^{n+1} = (-1)^n m_{n+1}, \tag{A1}$$

where m_n indicates the nth moment of $f(\mu)$. Thus, a solution of Eq. (1) exists if and only if the sequence

$$\begin{aligned} m_{n+1} &= (-1)^n \left.\frac{\mathrm{d}^n}{\mathrm{d}\tau^n} g(\tau)\right|_{\tau=0} \quad \forall n \in \mathbb{N} \\ m_0 &= 1 \end{aligned} \tag{A2}$$

is a Stieljies moment sequence, i.e., it represents the sequence of moments of a measure on the interval $[0, +\infty)$. We set $m_0 = 1$ as we are looking for a probability measure. A sufficient and necessary condition for a real sequence $\{m_n\}_{n\in\mathbb{N}}$ to be a Stieljies moment sequence states that the Hankel matrices

$$H_n^{(1)} = \begin{pmatrix} m_0 & m_1 & \dots & m_n \\ m_1 & m_2 & \dots & m_{n+1} \\ \vdots & \vdots & \ddots & \vdots \\ m_n & m_{n+1} & \dots & m_{2n} \end{pmatrix}, \tag{A3}$$

$$H_n^{(2)} = \begin{pmatrix} m_1 & m_2 & \dots & m_{n+1} \\ m_2 & m_3 & \dots & m_{n+2} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n+1} & m_{n+2} & \dots & m_{2n+1} \end{pmatrix} \tag{A4}$$

need to be positive semidefinite for any $n \in \mathbb{N}$ [19]. This property is useful to assess if the framework in terms of recovery rates is applicable or not given a certain $g(\tau)$.

In Sec. II, we presented the analytical form of the distribution of recovery rates $f(\mu)$ when $g(\tau)$ is a gamma(κ, θ) with $\kappa < 1$. We can show that such distribution does not exist when $\kappa = 2$. Indeed, for $g(\tau) = \theta^{-2}\tau e^{-\tau/\theta}$, the sequence turns out to be $m_n = -(n-1)\theta^{-n}$ and the Hankel matrix $H_2^{(1)}$ is not positive semidefinite.

APPENDIX B: EPIDEMIC THRESHOLD OF THE GENERIC NETWORK MODEL

The generic rank-r network model is defined by its metadegrees (r properties for each node), encoded in the $n \times r$ matrix V, and the signature of the nonsingular metric Δ. See Ref. [25] for further details. The adjacency matrix of such model is the rank-r matrix $A = V\Delta V^T$. Let M be a diagonal matrix containing the recovery rate μ_j of node j in its jth diagonal entry. Then, the linearized evolution of the disease close to the disease-free state follows the vector equation $\dot{\mathbf{x}} = (-M + \lambda V\Delta V^T)\mathbf{x}$, where $x_j(t)$ is the probability that node j is infectious at time t. Finding the epidemic threshold means finding the lowest value of λ for which $-M + \lambda V\Delta V^T$ as a zero eigenvalue. We can compute the characteristic polynomial of this matrix using Ref. [43]: $p(t) = \det\{\Delta^{-1} - \lambda V^T(t + M)^{-1}V\}$. The condition of the zero eigenvalue is then $\det\{1 - \lambda V^T M^{-1} V\Delta\} = 0$. The threshold condition that follows is $\lambda_c = 1/\rho(V^T M^{-1} V\Delta)$, where ρ is the spectral radius.

The last step consists of proving the following: $V^T M^{-1} V\Delta = \langle\mu^{-1}\rangle V^T V\Delta$. Let us call $Z = V^T M^{-1} V\Delta$ and compute its entry $Z_{\alpha\beta}$,

$$\begin{aligned} Z_{\alpha\beta} &= \sum_{i,j=1}^{N}\sum_{\gamma=1}^{r} V_{i\alpha}\delta_{ij}\mu_i^{-1}V_{j\gamma}\Delta_{\gamma\beta} \\ &= \sum_{\gamma=1}^{r}\Delta_{\gamma\beta}\sum_{i=1}^{N} V_{i\alpha}\mu_i^{-1}V_{i\gamma} \\ &= \sum_{\gamma=1}^{r}\Delta_{\gamma\beta}\sum_{i=1}^{N}(v_\alpha)_i\mu_i^{-1}(v_\gamma)_i \\ &= \sum_{\gamma=1}^{r}\Delta_{\gamma\beta}\langle v_\alpha\mu^{-1}v_\gamma\rangle N \\ &= \langle\mu^{-1}\rangle\sum_{\gamma=1}^{r}\Delta_{\gamma\beta}\langle v_\alpha v_\gamma\rangle N \\ &= \langle\mu^{-1}\rangle\sum_{\gamma=1}^{r}\sum_{i=1}^{N} V_{i\alpha}V_{i\gamma}\Delta_{\gamma\beta} \\ &= \langle\mu^{-1}\rangle B_{\alpha\beta}, \end{aligned}$$

under the assumption $\langle v_\alpha\mu^{-1}v_\gamma\rangle = \langle\mu^{-1}\rangle\langle v_\alpha v_\gamma\rangle$, where v_α is the αth column of the matrix V, i.e., the αth metadegree, and B is the matrix $V^T V\Delta$, obtained from A after rank reduction, following the notation in Ref. [25]. As the matrixes A and B share spectral properties, we can conclude that

$$\lambda_c = \frac{1}{\rho(V^T M^{-1} V\Delta)} = \frac{1}{\langle\mu^{-1}\rangle\rho(B)} = \frac{1}{\langle\tau\rangle\rho(A)}. \tag{B1}$$

APPENDIX C: ENDEMIC EQUILIBRIUM IN THE POPULATION-BASED SCHEME

In this Appendix, we derive the endemic prevalence in the population-based scheme and show that it coincides with the one obtained when using a homogeneous recovery rate. We consider a contact network with a fixed degree distribution $P(k)$, the so-called configuration network model [44].

Within the homogeneous modeling scheme, a constant recovery rate μ is assigned to each individual in the population, so the infectious period τ is exponentially distributed with mean $\langle\tau\rangle = \mu^{-1}$. Let $x_k(t)$ be the probability that a node with degree k is infected at time t. According to the degree-based mean-field approximation [45,46], the equation describing the evolution of the SIS model is

$$\frac{\mathrm{d}x_k(t)}{\mathrm{d}t} = -\mu x_k(t) + \lambda(1 - x_k(t))k\sum_{k'} P(k'|k)x_{k'}(t),$$

where $P(k'|k)$ is the probability that a node with degree k is connected with a node of degree k'. To derive the number of infected individuals at equilibrium, one must solve for I_k the following equation:

$$I_k = \frac{1}{\mu}\left[\frac{\lambda}{N}(N_k - I_k)k\sum_{k'}\frac{P(k'|k)}{P(k')}I_{k'}\right] \tag{C1}$$

where I_k is the number of infected at equilibrium that have degree k.

Now we assume the population-based scheme and consider a general distribution $g(\tau)$ for the infectious period and the corresponding distribution $f(\mu)$ for the recovery rates derived from Eq. (2). Each individual is assigned a recovery rate μ_j from $f(\mu)$, updated by resampling after each recovery. We can still reason in terms of classes of degree k, but we need to further divide each class according to the recovery rate. Let us define $x_{jk}(t)$ as the probability that a node with recovery rate μ_j and degree k is infected at time t. Then we can write

$$\frac{\mathrm{d}x_{jk}(t)}{\mathrm{d}t} = -\mu_j x_{jk}(t) + \lambda(1-x_{jk}(t))k\sum_{k'}\sum_{h} P(\mu_h, k'|\mu_j, k)x_{hk'}(t),$$

where $P(\mu_h, k'|\mu_j, k)$ is the probability that a node with recovery rate μ_j and degree k has a link to a node with recovery rate μ_h and degree k'. We assume that recovery rate and degree are uncorrelated, i.e., $P(\mu_h, k'|\mu_j, k) = P(\mu_h, k'|k) = P(\mu_h)P(k'|k)$, so we obtain

$$\frac{\mathrm{d}x_{jk}(t)}{\mathrm{d}t} = -\mu_j x_{jk}(t) + \lambda(1-x_{jk}(t))k\sum_{k'} P(k'|k)x_{k'}(t),$$

since $\sum_{k'} P(k'|k)\sum_h P(\mu_h)x_{hk'}(t) = \sum_{k'} P(k'|k)x_{k'}(t)$. In terms of number of infected I_{jk} with recovery rate μ_j and degree k, we obtain

$$\frac{\mathrm{d}I_{jk}(t)}{\mathrm{d}t} = -\mu_j I_{jk}(t) + \frac{\lambda}{N}S_{jk}(t)k\sum_{k'}\frac{P(k'|k)}{P(k')}I_{k'}(t).$$

The quantity $S_{jk}(t)$ is the number of susceptible nodes at time t with recovery rate μ_j and degree k. This is equal to $S_k(t)$ multiplied by the probability of being assigned the recovery rate μ_j at the time of the last recovery, i.e., $S_{jk}(t) = P(\mu_j)S_k(t)$:

$$\frac{\mathrm{d}I_{jk}(t)}{\mathrm{d}t} = -\mu_j I_{jk}(t) + \frac{\lambda}{N}P(\mu_j)S_k(t)k\sum_{k'}\frac{P(k'|k)}{P(k')}I_{k'}(t).$$

Solving for the equilibrium, and summing over the index j, we obtain

$$\begin{aligned} I_k &= \sum_j I_{jk} \\ &= \sum_j \frac{P(\mu_j)}{\mu_j}\left[\frac{\lambda}{N}(N_k - I_k)k\sum_{k'}\frac{P(k'|k)}{P(k')}I_{k'}\right] \\ &= \langle\mu^{-1}\rangle\left[\frac{\lambda}{N}(N_k - I_k)k\sum_{k'}\frac{P(k'|k)}{P(k')}I_{k'}\right], \end{aligned}$$

which is equal to Eq. (C1) provided that the mean infectious period $\langle\tau\rangle = \langle\mu^{-1}\rangle$ is the same as the one assumed in the homogeneous modeling scheme. In conclusion, within the *population-based* scheme, the endemic prevalence depends only on the average infectious period, and not on its distribution.

APPENDIX D: ENDEMIC EQUILIBRIUM IN THE HOST-BASED SCHEME

In this Appendix, we derive the endemic prevalence in the host-based scheme, in the case of homogeneous mixing, as stated in the main text in Eq. (13).

Let $x_j(t)$ be the probability that an individual with recovery rate μ_j is infected at time t. Then,

$$\frac{\mathrm{d}}{\mathrm{d}t}x_j(t) = -\mu_j x_j(t) + \frac{\lambda}{N}(1-x_j(t))\sum_h x_h(t), \tag{D1}$$

where the average connectivity of the homogeneous network is absorbed in the transmission rate λ. By setting $\mathrm{d}x_j(t)/\mathrm{d}t = 0$, we find

$$x_j = \lambda x_{\mathrm{eq}}/(\mu_j + \lambda x_{\mathrm{eq}}), \tag{D2}$$

where x_{eq} is the endemic prevalence. In continuous terms, by integrating on both sides over all possible values of μ, we get the following equation for the endemic prevalence x_{eq}:

$$x_{\mathrm{eq}} = \int_0^\infty \mathrm{d}\mu f(\mu)\left[1 + \frac{\mu}{\lambda x_{\mathrm{eq}}}\right]^{-1}. \tag{D3}$$

We rewrite it as follows:

$$1 = \lambda\int_0^\infty \mathrm{d}\mu f(\mu)\mu^{-1}\left[1 + \frac{\lambda x_{\mathrm{eq}}}{\mu}\right]^{-1}. \tag{D4}$$

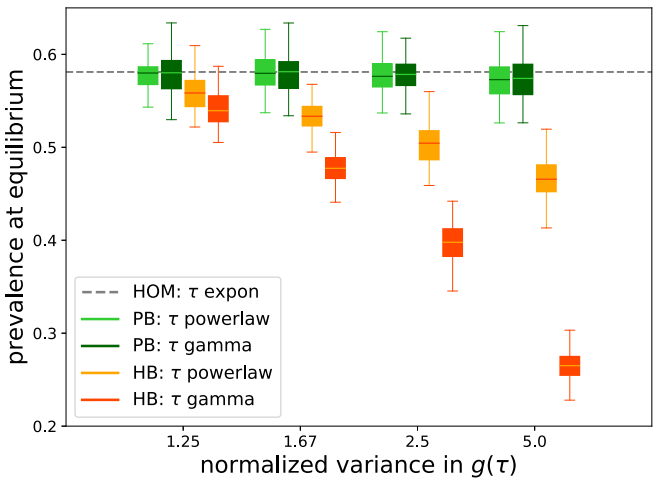

FIG. 4. Results of simulations of the spreading process performed on a temporal activity-driven network. Parameters for the activity-driven model: $N = 2000$, $\epsilon = 10^{-3}$, $m = 250$, $\eta = 1$, $\gamma = 2.8$, following the notation in Ref. [30]. Parameters for the epidemic spreading: $\langle\tau\rangle = 300$, $\lambda = 5\lambda_c$. The gray line indicates the endemic equilibrium for the exponentially-distributed τ (homogeneous scenario). In the heterogeneous scenario (nonexponentially distributed τ), the endemic prevalence for the population-based and host-based are shown in shades of green and orange. respectively. We show results as a function of the variance in $g(\tau)$ assuming two different distributions for $g(\tau)$: a gamma distribution (darker color) and a power-law distribution (lighter color). Box plots summarize values of the endemic equilibrium obtained from 100 stochastic simulations.

Expanding the inside of the integral as a geometric series, we get

$$\lambda \sum_{n=0}^{\infty} (-\lambda x_{eq})^n \int_0^{\infty} \mathrm{d}\mu f(\mu) \mu^{-(n+1)} = 1. \tag{D5}$$

The integral in left-hand side (lhs) is $\langle \mu^{-(n+1)} \rangle$, so we can use Eq. (10), reindex the sum, and get to

$$\sum_{n=0}^{\infty} (-\lambda x_{\mathrm{eq}})^n \frac{\langle \tau^n \rangle}{n!} = 1 - x_{\mathrm{eq}}. \tag{D6}$$

Now the lhs is by definition the moment-generating function of g, evaluated in $-\lambda x_{\mathrm{eq}}$. Given that such argument is never positive, this is also—by definition of moment-generating function—the Laplace transform of g, evaluated in λx_{eq}. We thus get to the final form of the equation of the endemic equilibrium:

$$\mathcal{L}[g](\lambda x_{\mathrm{eq}}) = 1 - x_{\mathrm{eq}}. \tag{D7}$$

APPENDIX E: NUMERICAL SIMULATIONS ON A HETEROGENEOUS NETWORK

The activity-driven model [30] is an example of a rank-two adjacency matrix, where each node is assigned two features: activity rate a and attractiveness b. The entries of the adjacency matrix are $A_{ij} = m(a_i b_j + a_j b_i)/N\langle b \rangle$. This model allows us to generate a time-evolving network, where individuals' interactions are heterogeneous. We constructed such a network as outlined in Ref. [30]. Figure 4 displays the results of disease spreading simulations over this temporal network, both within the host-based and population-based schemes. In line with the analytical results derived in Appendix C, we found that the endemic equilibrium reached with the population-based scheme equals the one expected in the scheme of exponentially-distributed infectious period. Instead, the prevalence at equilibrium for the host-based scheme is reduced and it depends on the shape of the distribution of infectious period $g(\tau)$. These results corroborate the conclusions of the main text found for a homogeneous network (Fig. 1).

[1] N. T. Bailey, A statistical method of estimating the periods of incubation and infection of an infectious disease, Nature (London) **174**, 139 (1954).

[2] C. A. Sabin and J. D. Lundgren, The natural history of HIV infection, Current Opin. HIV AIDS **8**, 311 (2013).

[3] T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens, Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020, Eurosurveillance **25**, 2000257 (2020).

[4] F. Miura, C. E. v. Ewijk, J. A. Backer, M. Xiridou, E. Franz, E. O. d. Coul, D. Brandwagt, B. v. Cleef, G. v. Rijckevorsel, C. Swaan, S. v. d. Hof, and J. Wallinga, Estimated incubation period for monkeypox cases confirmed in the Netherlands, May 2022, Eurosurveillance **27**, 2200448 (2022).

[5] D. Anderson and R. Watson, On the spread of a disease with gamma distributed latent and infectious periods, Biometrika **67**, 191 (1980).

[6] A. L. Lloyd, Realistic distributions of infectious periods in epidemic models: Changing patterns of persistence and dynamics, Theor. Popul. Biol. **60**, 59 (2001).

[7] O. Krylova and D. J. D. Earn, Effects of the infectious period distribution on predicted transitions in childhood disease dynamics, J. R. Soc. Interface. **10**, 20130098 (2013).

[8] W. Gou and Z. Jin, How heterogeneous susceptibility and recovery rates affect the spread of epidemics on networks, Infect. Dis. Model. **2**, 353 (2017).

[9] A. Darbon, D. Colombi, E. Valdano, L. Savini, A. Giovannini, and V. Colizza, Disease persistence on temporal contact networks accounting for heterogeneous infectious periods, R. Soc. Open Sci. **6**, 181404 (2019).

[10] G. F. de Arruda, G. Petri, F. A. Rodrigues, and Y. Moreno, Impact of the distribution of recovery rates on disease spreading in complex networks, Phys. Rev. Res. **2**, 013046 (2020).

[11] S. Bonaccorsi and S. Ottaviano, Epidemics on networks with heterogeneous population and stochastic infection rates, Math. Bio. **279**, 43 (2016).

[12] E. Vergu, H. Busson, and P. Ezanno, Impact of the infection period distribution on the epidemic spread in a metapopulation model, PLoS ONE **5**, e9371 (2010).

[13] W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, Proc. R. Soc. London **115**, 700 (1927).

[14] R. M. Anderson and R. M. May, *Infectious Diseases of Humans: Dynamics and Control* (Oxford University Press, Oxford, New York, 1992).

[15] M. Starnini, J. P. Gleeson, and M. Boguñá, Equivalence between non-Markovian and Markovian dynamics in epidemic spreading processes, Phys. Rev. Lett. **118**, 128301 (2017).

[16] C. Beck, Dynamical foundations of nonextensive statistical mechanics, Phys. Rev. Lett. **87**, 180601 (2001).

[17] M. Denys, T. Gubiec, R. Kutner, M. Jagielski, and H. E. Stanley, Universality of market superstatistics, Phys. Rev. E **94**, 042305 (2016).

[18] S. G. Walker, A Laplace transform inversion method for probability distribution functions, Stat. Comput. **27**, 439 (2017).

[19] K. Schmüdgen, *The Moment Problem*, Graduate Texts in Mathematics, Vol. 277 (Springer International Publishing, Cham, 2017).

[20] E. Valdano, C. Poletto, P.-Y. Boëlle, and V. Colizza, Reorganization of nurse scheduling reduces the risk of healthcare associated infections, Sci. Rep. **11**, 7393 (2021).

[21] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, Epidemic thresholds in real networks, ACM Trans. Inf. Syst. Security **10**, 1 (2008).

[22] C. Castellano and R. Pastor-Satorras, Thresholds for epidemic spreading in networks, Phys. Rev. Lett. **105**, 218701 (2010).

[23] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, Discrete-time Markov chain approach to contact-based disease spreading in complex networks, Europhys. Lett. **89**, 38009 (2010).

[24] Y. Wang, D. Chakrabarti, C. Wang, and C. Faloutsos, Epidemic spreading in real networks: An eigenvalue viewpoint, in

Proceedings of the 22nd International Symposium on Reliable Distributed Systems (IEEE, Florence, Italy, 2003), pp. 25–34.

[25] E. Valdano and A. Arenas, Exact rank reduction of network models, Phys. Rev. X **9**, 031050 (2019).

[26] R. Pastor-Satorras and A. Vespignani, Epidemic dynamics and endemic states in complex networks, Phys. Rev. E **63**, 066117 (2001).

[27] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Absence of epidemic threshold in scale-free networks with degree correlations, Phys. Rev. Lett. **90**, 028701 (2003).

[28] M. Boguñá, C. Castellano, and R. Pastor-Satorras, Langevin approach for the dynamics of the contact process on annealed scale-free networks, Phys. Rev. E **79**, 036110 (2009).

[29] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Epidemic processes in complex networks, Rev. Mod. Phys. **87**, 925 (2015).

[30] N. Perra, B. Goncalves, R. Pastor-Satorras, and A. Vespignani, Activity driven modeling of time varying networks, Sci. Rep. **2**, 469 (2012).

[31] J. L. W. V. Jensen, FR On convex functions and inequalities between expected values, Acta Math. **30**, 175 (1906).

[32] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals*, illustrated ed. (Princeton University Press, Princeton, NJ, 2007).

[33] P. Erdos and A. Rényi, On Random Graphs. I, Publicationes Mathematicae Debrecen **6**, 290 (1959).

[34] A. Hassoun, P. K. Linden, and B. Friedman, Incidence, prevalence, and management of MRSA bacteremia across patient populations'a review of recent developments in MRSA management and treatment, Critical Care **21**, 211 (2017).

[35] E. S. Shenoy, M. L. Paras, F. Noubary, R. P. Walensky, and D. C. Hooper, Natural history of colonization with methicillin-resistant *Staphylococcus aureus* (MRSA) and vancomycin-resistant Enterococcus (VRE): A systematic review, BMC Infec. Dis. **14**, 177 (2014).

[36] T. Obadia, R. Silhol, L. Opatowski, L. Temime, J. Legrand, A. C. M. Thiébaut, J.-L. Herrmann, E. Fleury, D. Guillemot, and P.-Y. Boëlle, Detailed contact data and the dissemination of Staphylococcus aureus in Hospitals, PLoS Comput. Biol. **11**, e1004170 (2015).

[37] A. Duval, T. Obadia, L. Martinet, P.-Y. Boëlle, E. Fleury, D. Guillemot, L. Opatowski, and L. Temime, Measuring dynamic social contacts in a rehabilitation hospital: Effect of wards, patient and staff characteristics, Sci. Rep. **8**, 1686 (2018).

[38] F. M. Tomley and M. W. Shirley, Livestock infectious diseases and zoonoses, Philos. Trans. R. Soc. B **364**, 2637 (2009).

[39] J. M. Abbate, F. Arfuso, C. Iaria, G. Arestia, and G. Lanteri, Prevalence of bovine tuberculosis in slaughtered cattle in Sicily, Southern Italy, Animals **10**, 1473 (2020).

[40] A. Darbon, E. Valdano, C. Poletto, A. Giovannini, L. Savini, L. Candeloro, and V. Colizza, Network-based assessment of the vulnerability of Italian regions to bovine brucellosis, Preventive Veterinary Medicine **158**, 25 (2018).

[41] F. Carrat, E. Vergu, N. M. Ferguson, M. Lemaitre, S. Cauchemez, S. Leach, and A.-J. Valleron, Time lines of infection and disease in human influenza: A review of volunteer challenge studies, Am. J. Epidemiol. **167**, 775 (2008).

[42] P. Kumar and A. Sinha, Information diffusion modeling and analysis for socially interacting networks, Soc. Network Anal. Min. **11**, 11 (2021).

[43] G. H. Golub, Some modified matrix eigenvalue problems, SIAM Rev. **15**, 318 (1973).

[44] M. Newman, *Networks: An Introduction*, 1st ed. (Oxford University Press, Oxford, 2010).

[45] M. Boguñá and R. Pastor-Satorras, Epidemic spreading in correlated complex networks, Phys. Rev. E **66**, 047104 (2002).

[46] A. Barrat, M. Barthélemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, 2008).