Using multitype branching processes to quantify statistics of disease outbreaks in zoonotic epidemics

Sarabjeet Singh[*]
Theoretical and Applied Mechanics, Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14853, USA

David J. Schneider[†]
Robert W. Holley Center for Agriculture and Health, Agricultural Research Service, United States Department of Agriculture, and Department of Plant Pathology and Plant-Microbe Biology, Cornell University, Ithaca, New York 14853, USA

Christopher R. Myers[‡]
Laboratory of Atomic and Solid State Physics, Department of Physics, and Institute of Biotechnology, Cornell University, Ithaca, New York 14853, USA

(Received 22 November 2013; published 4 March 2014)

Branching processes have served as a model for chemical reactions, biological growth processes, and contagion (of disease, information, or fads). Through this connection, these seemingly different physical processes share some common universalities that can be elucidated by analyzing the underlying branching process. In this work we focus on coupled branching processes as a model of infectious diseases spreading from one population to another. An exceedingly important example of such coupled outbreaks are zoonotic infections that spill over from animal populations to humans. We derive several statistical quantities characterizing the first spillover event from animals to humans, including the probability of spillover, the first passage time distribution for human infection, and disease prevalence in the animal population at spillover. Large stochastic fluctuations in those quantities can make inference of the state of the system at the time of spillover difficult. Focusing on outbreaks in the human population, we then characterize the critical threshold for a large outbreak, the distribution of outbreak sizes, and associated scaling laws. These all show a strong dependence on the basic reproduction number in the animal population and indicate the existence of a novel multicritical point with altered scaling behavior. The coupling of animal and human infection dynamics has crucial implications, most importantly allowing for the possibility of large human outbreaks even when human-to-human transmission is subcritical.

DOI: 10.1103/PhysRevE.89.032702 PACS number(s): 87.23.Cc, 05.40.−a, 87.15.Zg, 87.10.Mn

I. INTRODUCTION

Coupled reaction processes taking place across spatial domains or populations with complex structure can exhibit rich dynamics and phase transitions. Reaction-diffusion models have been used to study wave propagation and pattern formation [1], and epidemic models have been useful in understanding the spread of infectious diseases, rumors, computer viruses, and fads in populations [2]. In the context of stochastic formulations, branching processes often form the basis of such models. In describing epidemics, the susceptible-infected-recovered (SIR) model has been extensively studied in fully mixed populations [3] and more recently on complex networks [4,5]. The SIR model of disease dynamics is an important cornerstone of mathematical epidemiology where systematic analysis has been possible through the use of branching processes [3,6]. In the limit of large population, the SIR process converges to a linear birth-death process—a special case of continuous-time branching processes [7]. The linearized process is amenable to analytical treatment that yields some important results about the original nonlinear process, such as the existence of a second-order phase transition at a critical threshold. Below the threshold, all outbreaks are small [with size $o(N)$ as population size $N \to \infty$], whereas above the threshold some outbreaks can be large [with size $\mathcal{O}(N)$ as $N \to \infty$]. At the critical point, the distribution of outbreak sizes shows a power-law scaling of $P(n) \sim n^{-3/2}$ [where $P(n)$ is the probability of an outbreak of size n] with the average outbreak size scaling as $N^{1/3}$ [8].

Whereas the structure and statistics of epidemics in a single population are extremely well characterized, the structure of coupled epidemics in metapopulations have received much less attention. An exceedingly important example of such coupled outbreaks are zoonotic infections that spill over from animal populations to humans representing a major challenge in public health [9–11]. A zoonotic disease system typically involves one or more animal species with humans as the end hosts where cross-species transmission (*spillover*) is facilitated by direct or vector-mediated interactions between animals and humans. Although recent work has sought to characterize and classify the salient features of zoonoses [11–13], such classification schemes are largely descriptive, and the basic phenomenology of cross-species infection has not been addressed in sufficient mathematical detail. Models that explicitly incorporate spillover dynamics are exceedingly rare, despite the fact that such events are the defining characteristic of zoonotic infection [11]. Among models that do exist, stochastic treatments of spillover dynamics are much less common than deterministic models, a fact echoed in a recent survey [14].

[*]ss2365@cornell.edu
[†]dave.schneider@ars.usda.gov
[‡]c.myers@cornell.edu

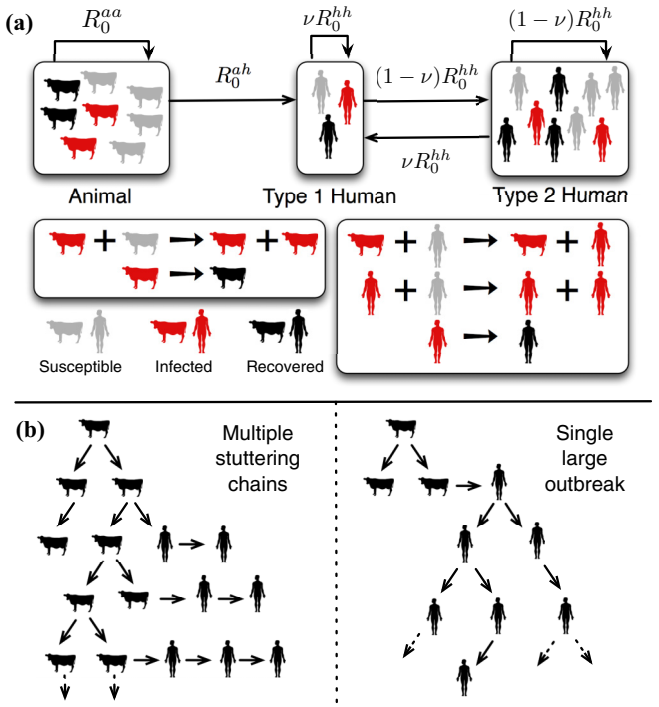

FIG. 1. (Color online) (a) Schematic of our zoonoses model. The labeled arrows denote $\mathcal{R}_0$'s for inter- and intrapopulation transmission. (b) Schematic depicting two possible mechanisms for zoonotic outbreaks in human populations. Left: Infection spreads efficiently in the animal population but inefficiently in humans, with each introduction into humans leading to a stuttering chain that goes extinct. Right: An initial spillover leads to a large outbreak sustained by human-to-human transmission.

In a zoonotic system, it is important to investigate how the nature of the critical threshold and the statistics of human outbreaks, which are well known for the simple SIR model, change with the addition of spillover dynamics. We address this important question in this paper using a simple two-species model of zoonoses (see Fig. 1). The model at its core represents two SIR processes that are coupled through cross-species transmission. The two host populations, animal and human, are fully mixed within their respective species, with a partial overlap between the species. The partial overlap or the "mixing fraction," ν, represents the fraction of human hosts that are fully mixed with the animal hosts and are thus at risk for direct infection from animals. The three types of possible infection transmission reactions are animal-to-animal, animal-to-human, and human-to-human. This model describes zoonoses where the infection prevalence in the animal population changes rapidly on time scales of interest. This might occur with the introduction of a disease into a new amplifier animal host population [15,16] or with the evolutionary emergence of a new, more virulent, strain of an existing animal pathogen [17,18]. Due to the dynamical nature of spillover driven by the animal SIR process, the degree of animal-to-animal transmission becomes an important determinant of human outbreaks alongside the degree of human-to-human transmission.

While the statistics of human outbreaks are crucial for developing a systematic understanding, there are additional questions that one may ask specific to a cross-species disease system. These include the probability of spillover, the distribution of time to spillover (from the point when the infection starts in the animal population), and the distribution of infected animals at the moment of spillover. These questions are relevant from the perspective of parameter estimation and control interventions, as we address in this paper.

An outline of our methodology and results is as follows. The coupled SIR process (Fig. 1) converges to a multitype linear birth-death process in the limit of large system size for fixed ratio of animal and human population size. The limiting linear process allows us to use the same techniques that have been employed to analyze the simple SIR in the limit of large system size. The results presented here are divided into two main parts. First, we calculate the joint distribution of a subset of state variables as a function of time using probability generating functions (PGF). We use the PGF to calculate the probability of spillover, the distribution of first passage times, and the distribution of infected animals at the time of spillover. Second, we calculate the PGF for the distribution of outbreak sizes in humans, which we then use to calculate the critical threshold, the scaling laws for the distribution function at the critical point, as well as finite-size scaling for the average outbreak size. Last, we calculate the probability of a large human outbreak.

The basic reproduction number $\mathcal{R}_0$ for an epidemic is defined as the average number of new infections produced by an infectious host in a fully susceptible population. In our analysis we find that the critical threshold is a function of the basic reproduction numbers in both animals and humans, $\mathcal{R}_0^{aa}$ and $\mathcal{R}_0^{hh}$, and identify a parameter regime where large outbreaks are possible in human populations—*sustained by repeated introductions from the animal population*—even if human-to-human transmission is subcritical (i.e., when $\mathcal{R}_0^{hh} < 1$). Information only about infection in the human population is insufficient to distinguish such a scenario from one involving a single primary introduction followed by extensive human-to-human transmission [see Fig. 1(b)]. Our systematic characterization of the spectrum of possible behaviors helps to augment and clarify the previously proposed classification frameworks [11–13]. We see our work as a stepping stone toward more complex and realistic models that might help address the spatial and ecological aspects of zoonotic emergence, the evolution of virulence, and public health interventions in the form of dynamic control strategies.

II. THE COUPLED SIR PROCESS

The three-type metapopulation model considered here consists of animals, type 1 humans, and type 2 humans. Figure 1 shows a schematic of the model and the reaction equations are as follows:

$$(S_a, I_a, \ldots) \xrightarrow{\beta_{aa} S_a I_a / N_a} (S_a - 1, I_a + 1, \ldots)$$

$$(\ldots, I_a, R_a, \ldots) \xrightarrow{\gamma_a I_a} (\ldots, I_a - 1, R_a + 1, \ldots)$$

$$(\ldots, S_{h1}, I_{h1p}, \ldots) \xrightarrow{\beta_{ah} S_{h1} I_a / N_a} (\ldots, S_{h1} - 1, I_{h1p} + 1, \ldots)$$

$$
\begin{aligned}
(\ldots,S_{h1},I_{h1s},\ldots) &\xrightarrow{\beta_{hh}S_{h1}I_h/N_h} (\ldots,S_{h1}-1,I_{h1s}+1,\ldots)\\
(\ldots,I_{h1p},R_{h1p},\ldots) &\xrightarrow{\gamma_h I_{h1p}} (\ldots,I_{h1p}-1,R_{h1p}+1,\ldots)\\
(\ldots,I_{h1s},R_{h1s},\ldots) &\xrightarrow{\gamma_h I_{h1s}} (\ldots,I_{h1s}-1,R_{h1s}+1,\ldots)\\
(\ldots,S_{h2},I_{h2},\ldots) &\xrightarrow{\beta_{hh}S_{h2}I_h/N_h} (\ldots,S_{h2}-1,I_{h2}+1,\ldots)\\
(\ldots,I_{h2},R_{h2}) &\xrightarrow{\gamma_h I_{h2}} (\ldots,I_{h2}-1,R_{h2}+1). \qquad (1)
\end{aligned}
$$

The full state vector is

$$(S_a,I_a,R_a,S_{h1},I_{h1p},I_{h1s},R_{h1p},R_{h1s},S_{h2},I_{h2},R_{h2}),$$

but for brevity, only those state variables that change are shown in each reaction above. There are N_a animal hosts and N_h human hosts in the system. A fraction ν of all human hosts are of type 1 that can receive both primary transmission from animals and secondary transmission from other humans. Type 2 human hosts, which are a fraction $1-\nu$ of the total human population, can only receive a secondary transmission. Type 1 humans are fully mixed with both animals and type 2 humans. $(S_\star,I_\star,R_\star)$ denote the population in the susceptible, infectious, and recovered compartments of each group such that $S_\star + I_\star + R_\star = N_\star$. The subscripts (or superscripts) aa, ah, and hh stand for the animal-animal, animal-human, and human-human transmission reactions, respectively. β_{aa}, β_{ah}, and β_{hh} are the rates of infectious contact per infectious host, and γ_a and γ_h represent the rates of recovery. The subscripts "p" and "s" (such as in I_{h1p} and I_{h1s}) distinguish between primary and secondary infections, respectively. The total number of infected human hosts is given by $I_h = I_{h1p} + I_{h1s} + I_{h2}$.

The basic reproduction numbers associated with the aa, ah, and hh transmissions are as follows:

$$\mathcal{R}_0^{aa} = \frac{\beta_{aa}}{\gamma_a}, \quad \mathcal{R}_0^{hh} = \frac{\beta_{hh}}{\gamma_h}, \quad \mathcal{R}_0^{ah} = \frac{\nu\beta_{ah}}{\rho\gamma_a} \equiv \frac{\hat{\beta}_{ah}}{\gamma_a}. \qquad (2)$$

$\mathcal{R}_0^{ij}$ represents the average number of new infections produced by a single infected host of type i in a fully susceptible population of type j. The animal epidemic follows the simple SIR process with $\mathcal{R}_0^{aa}=1$ as the critical threshold. For $\mathcal{R}_0^{aa}<1$, all animal outbreaks will be small. For $\mathcal{R}_0^{aa}>1$, the probability of having a large outbreak [of size $\mathcal{O}(N_a)$] is $1-1/\mathcal{R}_0^{aa}$.

III. THE MULTITYPE LINEAR BIRTH-DEATH PROCESS

We investigate the model in the limit of $N_a,N_h\to\infty, N_a/N_h\to\rho$. In this limit, the epidemic does not saturate and the depletion of susceptible pool in finite time can be ignored. The original nonlinear process reduces to a multitype linear birth-death process [19] with the following rate equations:

$$
\begin{aligned}
(I_a,\ldots) &\xrightarrow{\beta_{aa}I_a} (I_a+1,\ldots)\\
(\ldots,I_a,R_a,\ldots) &\xrightarrow{\gamma_a I_a} (\ldots,I_a-1,R_a+1,\ldots)\\
(\ldots,I_{h1p},\ldots) &\xrightarrow{\nu\beta_{ah}I_a/\rho} (\ldots,I_{h1p}+1,\ldots)\\
(\ldots,I_{h1s},\ldots) &\xrightarrow{\nu\beta_{hh}I_h} (\ldots,I_{h1s}+1,\ldots)\\
(\ldots,I_{h1p},R_{h1p},\ldots) &\xrightarrow{\gamma_h I_{h1p}} (\ldots,I_{h1p}-1,R_{h1p}+1,\ldots)\\
(\ldots,I_{h1s},R_{h1s},\ldots) &\xrightarrow{\gamma_h I_{h1s}} (\ldots,I_{h1s}-1,R_{h1s}+1,\ldots)\\
(\ldots,I_{h2},\ldots) &\xrightarrow{(1-\nu)\beta_{hh}I_h} (\ldots,I_{h2}+1,\ldots)\\
(\ldots,I_{h2},R_{h2}) &\xrightarrow{\gamma_h I_{h2}} (\ldots,I_{h2}-1,R_{h2}+1). \qquad (3)
\end{aligned}
$$

The linear process can be represented by a smaller set of equations if we ignore the type labeling. Before doing so, we introduce a new set of state variables for the human hosts that will be convenient in the forthcoming analysis. Let $Z_\star(t) = I_\star(t) + R_\star(t)$, where $\star$ stands for particular subscripts used in what follows. $Z_{h,p}(t)$ denotes the number of primary human infections and $Z_{h,s}(t)$ denotes the number of secondary human infections irrespective of the human host type. The total number of infected human hosts is then $Z_h(t) = Z_{h,p}(t) + Z_{h,s}(t)$, regardless of type. The reduced set of equations describing the multitype linear birth-death process are as follows:

$$
\begin{aligned}
(I_a,R_a,I_h,Z_{h,p},Z_{h,s}) &\xrightarrow{\beta_{aa}I_a} (I_a+1,R_a,I_h,Z_{h,p},Z_{h,s})\\
(I_a,R_a,I_h,Z_{h,p},Z_{h,s}) &\xrightarrow{\gamma_a I_a} (I_a-1,R_a+1,I_h,Z_{h,p},Z_{h,s})\\
(I_a,R_a,I_h,Z_{h,p},Z_{h,s}) &\xrightarrow{\hat{\beta}_{ah}I_a} (I_a,R_a,I_h+1,Z_{h,p}+1,Z_{h,s}) \qquad (4)\\
(I_a,R_a,I_h,Z_{h,p},Z_{h,s}) &\xrightarrow{\beta_{hh}I_h} (I_a,R_a,I_h+1,Z_{h,p},Z_{h,s}+1)\\
(I_a,R_a,I_h,Z_{h,p},Z_{h,s}) &\xrightarrow{\gamma_h I_h} (I_a,R_a,I_h-1,Z_{h,p},Z_{h,s}),
\end{aligned}
$$

where $\hat{\beta}_{ah} \equiv \nu\beta_{ah}/\rho$. The description of the process dynamics can be expressed in the form of probability $P_{i,j,k,l,m}(t)$ of the state variables $(I_a,R_a,I_h,Z_{h,p},Z_{h,s})$ being in the state (i,j,k,l,m) at time t given that the process starts with a single infectious animal host. Similarly, let $Q_{n,p}(t)$ be the probability of the state variables $(I_h,Z_{h,s})$ being in the state (n,p) at time t starting from a single infectious human host at time 0. The set of all possible transitions that involve a single infectious animal host includes the following: production of another infectious animal host, moving into recovery, or production of an infectious human host. Let the probability of being in the state (i,j,k,l,m) at time t from these one step transitions be

$$P_{i,j,k,l,m}^{a\to aa}(t), \quad P_{i,j,k,l,m}^{a\to\emptyset}(t) \quad \text{and} \quad P_{i,j,k,l,m}^{a\to ah}(t),$$

respectively. Similarly, for the transitions that begin with a single infectious human host, let $Q_{n,p}^{h\to hh}$ and $Q_{n,p}^{h\to\emptyset}$ be the probability of reaching the (n,p) state based on infection or recovery taking place at the first step. The probability $P_{i,j,k,l,m}^{a\to\emptyset}(t)$ is 1 when $i=k=0$ and 0 otherwise. Similarly, $Q_{n,p}^{h\to\emptyset}$ is 1 when $n=0$ and 0 otherwise. These probabilities would satisfy the following Kolmogorov backward equation [7]:

$$
\begin{aligned}
\frac{dP_{i,j,k,l,m}}{dt} &= \beta_{aa}P_{i,j,k,l,m}^{a\to aa} + \gamma_a\delta_{i,0}\delta_{k,0} + \beta_{ah}P_{i,j,k,l,m}^{a\to ah}\\
&\quad -(\beta_{aa}+\gamma_a+\beta_{ah})P_{i,j,k,l,m} \qquad (5)\\
\frac{dQ_{n,p}}{dt} &= \beta_{hh}Q_{n,p}^{h\to hh} + \gamma_h\delta_{n,0} - (\beta_{hh}+\gamma_h)Q_{n,p}.
\end{aligned}
$$

The set of backward equations (5) can be solved using probability generating functions (PGFs) [6,7]. Let $G_a(x,y,u,z,w;t)$

be the PGF for the joint distribution of the dynamic variables when a single animal host was infected at time 0,

$$G_a(x,y,u,z,w;t) = \sum_{i,j,k,l,m} P_{i,j,k,l,m}(t) x^i y^j u^k z^l w^m. \quad (6)$$

Similarly, let $G_h(u,w;t)$ be the PGF for the joint distribution of $(I_h(t), Z_{h,s}(t))$ where a single human host is infected at time 0,

$$G_h(u,w;t) = \sum_{n,p} Q_{n,p}(t) u^n w^p. \quad (7)$$

The PGF for $P^{a\to aa}_{i,j,k,l,m}(t)$ is simply G_a^2, reflecting the independence of branching process emanating from two individuals. Similarly, the PGF for $P^{a\to ah}_{i,j,k,l,m}(t)$ is $G_a G_h$ and that for $Q^{h\to hh}_{n,p}$ is G_h^2. Using the PGF representation we now write down the following backward equation:

$$\begin{aligned}\frac{\partial G_a}{\partial t} &= \beta_{aa} G_a^2 + \gamma_a y + \hat{\beta}_{ah} G_a G_h z - (\beta_{aa} + \hat{\beta}_{ah} + \gamma_a) G_a \\ \frac{\partial G_h}{\partial t} &= \beta_{hh} G_h^2 w + \gamma_h - (\beta_{hh} + \gamma_h) G_h. \end{aligned} \quad (8a)$$

The initial conditions for the PDEs are as follows:

$$\begin{aligned} G_a(x,y,u,z,w;0) &= x \\ G_h(u,w;0) &= u, \end{aligned} \quad (8b)$$

which encode the information that there is a one infected animal host at time 0. The equation for G_h can be solved exactly. The solution is provided in Refs. [6,7] and we reproduce it here as follows:

$$G_h(u,w;t) = \frac{A_h(B_h-u) + B_h(u-A_h) e^{-\beta_{hh} w (B_h - A_h) t}}{(B_h-u) + (u-A_h) e^{-\beta_{hh} w (B_h - A_h) t}}, \quad (9a)$$

where $A_h(w)$ and $B_h(w)$ are solutions of the following quadratic equation such that $0 < A_h < 1 < B_h$,

$$\mathcal{R}_0^{hh} w s^2 - \left(\mathcal{R}_0^{hh} + 1\right) s + 1 = 0. \quad (9b)$$

The PGF G_h quantifies the distribution of a single chain of infections that originates from a single primary infection. The more interesting aspect of the zoonoses dynamics is captured by the first equation (for G_a). A full analytical solution to this process has recently been reported [20], but extracting information specifically about the first spillover event into humans—which is our primary focus here—is complicated to derive from that general result. To address first passage time phenomena more directly, we introduce a simpler subset process below and derive its solution.

A. Analytical solution for a subset process

The distribution of $(I_a, R_a, Z_{h,p})$ is governed by a reduced set of reaction equations:

$$\begin{aligned} (I_a, R_a, Z_{h,p}) &\xrightarrow{\beta_{aa} I_a} (I_a + 1, R_a, Z_{h,p}) \\ (I_a, R_a, Z_{h,p}) &\xrightarrow{\gamma_a I_a} (I_a - 1, R_a + 1, Z_{h,p}) \\ (I_a, R_a, Z_{h,p}) &\xrightarrow{\hat{\beta}_{ah} I_a} (I_a, R_a, Z_{h,p} + 1). \end{aligned} \quad (10)$$

Let $\mathcal{G}_a(x,y,z;t)$ represent the PGF for the distribution of the above process. The PDE for the PGF is given by the following:

$$\frac{\partial \mathcal{G}_a}{\partial t} = \beta_{aa} \mathcal{G}_a^2 + \gamma_a y - (\beta_{aa} + \hat{\beta}_{ah} + \gamma_a) \mathcal{G}_a + \hat{\beta}_{ah} \mathcal{G}_a z, \quad (11a)$$

with the initial condition

$$\mathcal{G}_a(x,y,z;0) = x. \quad (11b)$$

Following the methods outlined in Ref. [6], we obtain the following solution to the PDE:

$$\mathcal{G}_a(x,y,z;t) = \frac{A_a(B_a-x) + B_a(x-A_a) e^{-\beta_{aa}(B_a - A_a)t}}{(B_a-x) + (x-A_a) e^{-\beta_{aa}(B_a - A_a)t}}. \quad (12a)$$

$A_a(y,z)$ and $B_a(y,z)$ are roots of the following quadratic equation such that $0 < A_a < 1 < B_a$:

$$\mathcal{R}_0^{aa} s^2 - \left(\mathcal{R}_0^{aa} + 1 + \mathcal{R}_0^{ah}(1-z)\right) s + y = 0. \quad (12b)$$

The distribution reported here has been solved before in the context of a human-only epidemic process with two types of hosts [21]. In subsequent sections, we shall require the value of roots at the point $z = 0$. Adopting notation from [22], we define the following:

$$\begin{aligned} V_0(y) &= A_a(y,0), \quad v_0 = A_a(1,0) \\ V_1(y) &= B_a(y,0), \quad v_1 = B_a(1,0). \end{aligned} \quad (12c)$$

B. First passage time and probability of spillover

We define the time to spillover as the first passage time T for human infection, i.e., as the time when the first primary infection occurs in the human hosts. The distribution of first passage times is given by the following:

$$\begin{aligned} P[T \leqslant t] &= P[Z_{h,p}(t) > 0] \\ &= 1 - \mathcal{G}_a(1,1,0;t) \\ &= 1 - \frac{v_0(v_1-1) + v_1(1-v_0) e^{-\beta_{aa}(v_1 - v_0)t}}{(v_1-1) + (1-v_0) e^{-\beta_{aa}(v_1 - v_0)t}}. \end{aligned} \quad (13)$$

This distribution is shown for various parameter values in Fig. 2, along with results of discrete event simulation drawn from the underlying set of reactions. Simulations were done using Gillespie's direct method [23] for reaction kinetics. It can be seen from Fig. 2 that the distribution is defective since the disease can go extinct in the animal population before the primary transmission occurs in the human population. The distribution can be used to calculated moments of the first passage time conditioned on the occurrence of spillover (see Appendix A for derivation):

$$\mathbb{E}[T^n \mid T < \infty] = \frac{n!}{(\beta_{aa})^n (v_0 - 1)(v_1 - v_0)^{n-1}} \mathrm{Li}_n\left(\frac{v_0 - 1}{v_1 - 1}\right), \quad (14)$$

where Li_n is the polylogarithm function of order n. The mean and the standard deviation of the first passage time are plotted as a function of $\mathcal{R}_0^{aa}$ and $\mathcal{R}_0^{ah}$ in Fig. 3. The conditional nature of the distribution leads to a nonmonotonic dependence on $\mathcal{R}_0^{aa}$. First passage times for $\mathcal{R}_0^{aa} < 1$ are limited by the time scale for the eventual extinction in the animal population: spillover must occur quickly if it is going to happen at all. The expected time to extinction in the animal population diverges

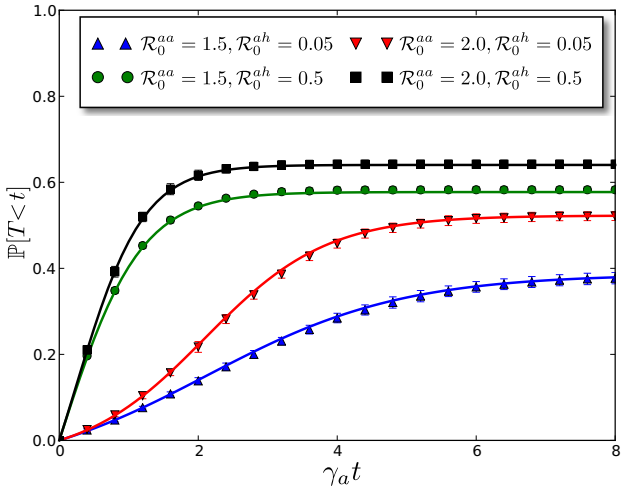

FIG. 2. (Color online) First-passage time distribution for spillover into the human population $P[T < t]$, comparing the analytical calculation given by Eq. (13) (solid line) with the results of discrete event simulation (using Gillespie's direct method, for finite system size $N_a = N_h = 10^3$). The x axis is time normalized by the mean infectious period $(1/\gamma_a)$, of the animal species. The markers represent the mean of 8000 simulation runs.

as $\mathcal{R}_0^{aa} \to 1$, leading to an increase in the mean first passage time. The mean also decreases with increasing $\mathcal{R}_0^{ah}$ because of the increasing rate of animal-to-human transmission.

The distribution of first passage times also gives us a way to calculate the probability of spillover as

$$P[\text{spill}] = P[T < \infty] = 1 - v_0. \tag{15}$$

Using the law of total probability, the probability of spillover can be decomposed as follows:

$$P[\text{spill}] = P[\text{spill} \mid \text{small outbreak}]P[\text{small outbreak}] + P[\text{spill} \mid \text{large outbreak}]P[\text{large outbreak}], \tag{16}$$

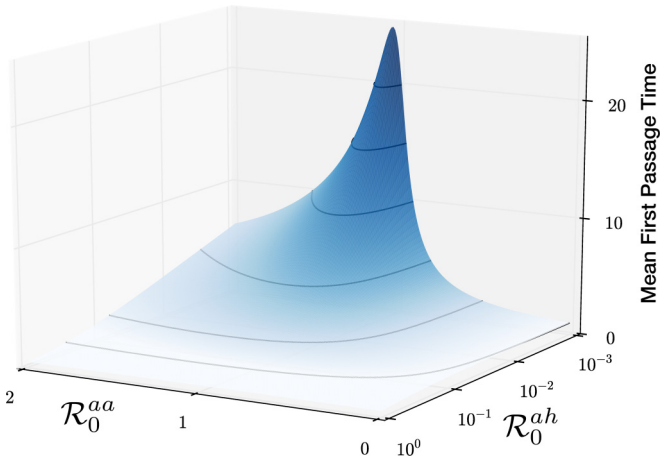

FIG. 3. (Color online) The mean time to spillover (in units of the mean infectious period of animal hosts) as a function of $\mathcal{R}_0^{aa}$ and $\mathcal{R}_0^{ah}$. Solid lines on the surface represent contours for the mean values. Gradient in the shade represents the standard deviation of the distribution (dark:high, light:low spanning the range [0.4,18.8] on a log scale).

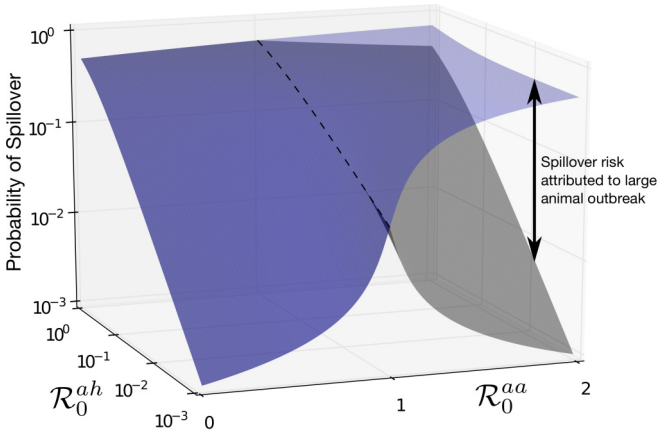

FIG. 4. (Color online) Probability of spillover (blue, upper surface) and the conditional probability of spillover given a small outbreak in the animal population (gray, lower surface). The dashed line marks the separation between the two surfaces at $\mathcal{R}_0^{aa} = 1$. The difference between the two surfaces gives the contribution of large animal outbreaks to spillover risk.

where the conditioning is done on the state of the outbreak in the animal population. For $\mathcal{R}_0^{aa} \leqslant 1$, all animal outbreaks are small. For $\mathcal{R}_0^{aa} > 1$, the probability of a large animal outbreak is nonzero. In the infinite size limit, it is implicitly assumed that $P[\text{spill} \mid \text{large outbreak}] = 1$. Using this result and $P[\text{large outbreak}] = 1 - 1/\mathcal{R}_0^{aa}$ in Eq. (16), we can calculate $P[\text{spill} \mid \text{small outbreak}]$ where $\mathcal{R}_0^{aa} > 1$, as follows:

$$P[\text{spill} \mid \text{small outbreak}] = \begin{cases} 1 - \mathcal{R}_0^{aa} v_0 & \text{if } \mathcal{R}_0^{aa} \leqslant 1, \\ 1 - v_0 & \text{if } \mathcal{R}_0^{aa} > 1. \end{cases} \tag{17}$$

The probability of spillover as a function of relevant model parameters [Eq. (15)] is shown in Fig. 4 as the upper surface; also shown is the probability of spillover given that there is a small outbreak in the animal population [Eq. (17)]. We also calculate finite-size corrections to the probability of spillover in Appendix B. The results of the calculation are shown in Fig. 5.

Stochastic models offer a stark contrast to deterministic models that invariably associate spillover events with large outbreaks in the animal population. While large outbreaks do enhance the risk of spillover, small outbreaks also contribute as can be seen from Fig. 4. This result indicates that some spillovers may be almost impossible to trace back in the animal population if they arise from a small outbreak where only a few animal hosts were infected and no contact tracing data are available.

C. Prevalence in the animal population at spillover

The distribution of infectious and recovered hosts in the animal population at the first passage time can be calculated by methods outlined in Ref. [22]. By interpreting the spillover process as a linear birth-death-killing (BDK) process, the distribution of infectious hosts at spillover is the same as the distribution of killing positions in the BDK process—geometrically distributed with parameter $1 - 1/v_1$, where v_1 was defined in Eq. (12c). The calculation can be extended

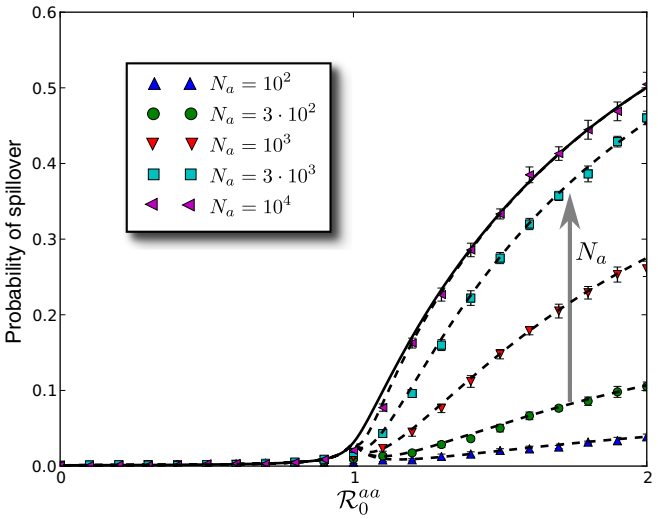

FIG. 5. (Color online) Finite-size corrections to the probability of spillover. Dashed lines represent the analytical solution [Eq. (B3)] for different values of N_a. The solid line represents the solution from the linear birth-death process [Eq. (15)]. Colored markers represents values calculated from 10 000 simulation runs done using Gillespie's direct method. All results are for fixed $\mathcal{R}_0^{ah} = 10^{-3}$.

to include recovered hosts as well. The joint distribution of infectious and recovered animal hosts at first passage time is generated by the following PGF:

$$J(x,y) = \frac{x\hat{\beta}_{ah}}{1-v_0}\int_0^\infty \frac{\partial \mathcal{G}_a(x,y,0;t)}{\partial x}dt$$
$$= \frac{x\hat{\beta}_{ah}}{\beta_{aa}(1-v_0)(V_1(y)-x)}$$
$$= \frac{x(v_1-1)}{V_1(y)-x}. \tag{18}$$

The simplification in the last step of Eq. (18) comes from the fact that v_0 and v_1 are roots of Eq. (12b), i.e.,

$$(v_1-1)(1-v_0) = \frac{\hat{\beta}_{ah}}{\beta_{aa}} = \frac{\mathcal{R}_0^{ah}}{\mathcal{R}_0^{aa}}. \tag{19}$$

The mean number of infectious and recovered animal hosts at the first passage time T are given by the following:

$$\langle I_a(T)\rangle = \left.\frac{\partial J}{\partial x}\right|_{(1,1)}, \quad \langle R_a(T)\rangle = \left.\frac{\partial J}{\partial y}\right|_{(1,1)}. \tag{20}$$

A plot of the mean number of infectious animal hosts at first passage time is shown as a function of $\mathcal{R}_0^{aa}$ and $\mathcal{R}_0^{ah}$ in Fig. 6. We can sample analytically the distributions represented in Eq. (18) and compare with the results of stochastic simulations for finite system sizes, as is shown in Fig. 7. As seen in the Fig. 7(a) for the expected number of infected animal hosts, the tail of the analytical distribution overestimates the prevalence slightly because of epidemic saturation that occurs in finite populations.

Given a prevalence of n infected animal hosts at spillover (and no information about recovered hosts), the maximum likelihood estimate for the parameters yields the equation $v_1 = n/(n-1)$. From Eqs. (12b) and (12c) we arrive at the

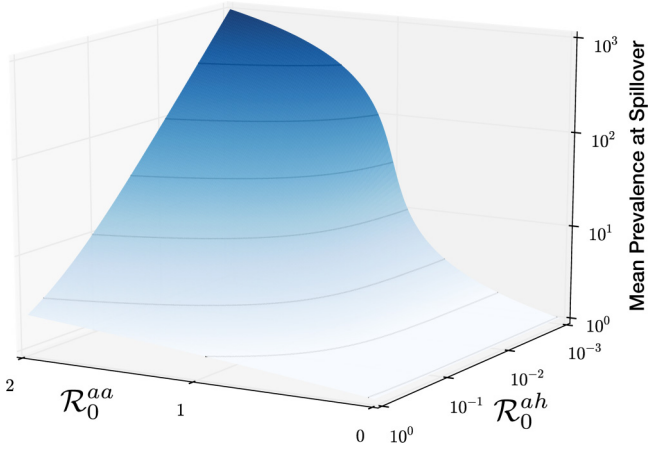

FIG. 6. (Color online) The expected number of infectious animal hosts at the time of first primary human infection as a function of $\mathcal{R}_0^{aa}$ and $\mathcal{R}_0^{ah}$. Solid lines on the surface represent contours for the mean values. Gradient in the shade represents the standard deviation of the distribution (dark:high, light:low spanning the range $[10^{-2}, 10^3]$ on a log scale).

following relationship between the model parameters:

$$\bar{\mathcal{R}}_0^{aa} = (n-1)\left(\bar{\mathcal{R}}_0^{ah} + n^{-1}\right), \tag{21}$$

where $\bar{\mathcal{R}}_0^{aa}$ and $\bar{\mathcal{R}}_0^{ah}$ are estimators for the corresponding parameters.

While the first passage time reveals the time scale of spillover, the disease prevalence reveals the state of the system at spillover; the different dependence on transmission parameters for these two quantities, however, has interesting implications for parameter estimation and disease control (compare Figs. 3 and 6). Our results indicate that the fluctuations in the prevalence at spillover increase with $\mathcal{R}_0^{aa}$, in contrast to the first passage time which has the highest fluctuations near $\mathcal{R}_0^{aa} = 1$. In the absence of animal surveillance, the first spillover into humans is usually the point at which the disease is first detected and control interventions are initiated [18]. For $\mathcal{R}_0^{aa}$ substantially larger than 1, the spillover is likely to happen relatively early, but the disease prevalence may be quite large, making control difficult. In contrast, for $\mathcal{R}_0^{aa}$ close to 1, the disease is likely to be detected late, but there will be a low prevalence in the animal population at that time. This is encouraging for public health interventions aimed at controlling the disease in the animal population, although the long delay before detection introduces greater uncertainties as to whether other factors might need to be included in a more complicated model (such as demographic changes in the animal and human populations or evolution of pathogen virulence).

IV. BRANCHING PROCESSES

In the limit of $t \to \infty$, the SIR process is isomorphic to a Galton-Watson branching process where the offspring are the new infections produced by an infected host. The distribution of outbreak sizes can be calculated by making the *treelike* approximation for small outbreaks.

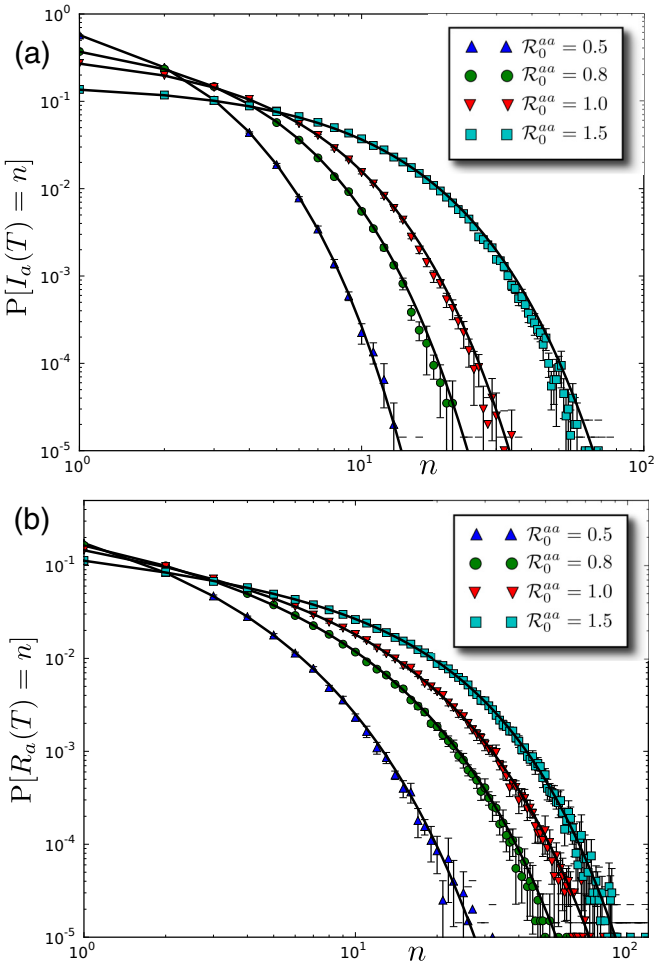

FIG. 7. (Color online) The marginal distribution of the number of infectious animal hosts (a) and the number of recovered animal hosts (b) at first passage time T for finite system size ($N_a = N_h = 1000$). Solid lines represents the analytical solution obtained by sampling from the PGF in Eq. (18). Colored markers represents values calculated from 2×10^5 simulation runs done using Gillespie's direct method. All results are for fixed $\mathcal{R}_0^{ah} = 0.1$.

A. Distribution of outbreak sizes

For the animal population, let $H_a(z)$ be the PGF for the distribution of outbreak sizes. From Eq. (12a), we obtain the following:

$$H_a(z) = \mathcal{G}_a(1,z,1;\infty) = A_a(z,1) = \frac{\mathcal{R}_0^{aa} + 1 - \sqrt{\left(\mathcal{R}_0^{aa}+1\right)^2 - 4\mathcal{R}_0^{aa} z}}{2\mathcal{R}_0^{aa}}. \quad (22)$$

Let $H_{h,p}(x)$ be the PGF for the distribution of primary infections in the human population. Then, from Eq. (12a), we obtain the following:

$$H_{h,p}(x) = \mathcal{G}_a(1,1,x;\infty) = A_a(1,x) = \frac{\mathcal{R}_0^{aa}+1+\mathcal{R}_0^{ah}(1-x)-\sqrt{\left(\mathcal{R}_0^{aa}+1+\mathcal{R}_0^{ah}(1-x)\right)^2-4\mathcal{R}_0^{aa}}}{2\mathcal{R}_0^{aa}}. \quad (23)$$

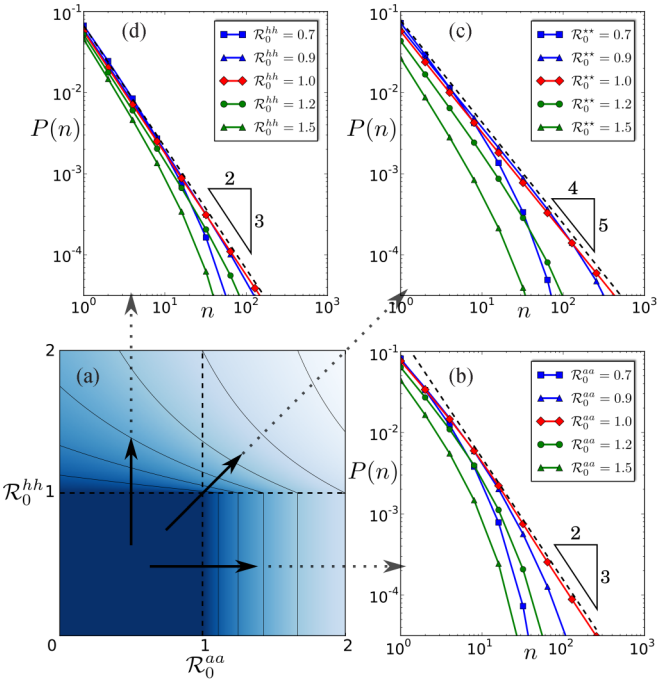

FIG. 8. (Color online) The distribution of sizes of small human outbreaks. (a) Colormap for the probability that an outbreak in the human hosts is small spanning the range [0.36 (light), 1.0 (dark)]. Solid lines represent constant-probability contours in the colormap. (b)–(d) Probability of having a small outbreak of size n at different crossings of the threshold boundary. All results are for fixed $\mathcal{R}_0^{ah} = 0.1$.

Each primary infected host in the human population acts as the progenitor for a branching process comprising of secondary infections. Let $\hat{H}_{h,s}(x)$ be the PGF for the distribution of secondary infections emanating from a primary progenitor. Then, from Eq. (9a), we obtain the following:

$$\hat{H}_{h,s}(z) = G_h(1,z;\infty) = A_h(z) = \frac{\mathcal{R}_0^{hh} + 1 - \sqrt{\left(\mathcal{R}_0^{hh}+1\right)^2 - 4\mathcal{R}_0^{hh} z}}{2\mathcal{R}_0^{hh} z}. \quad (24)$$

The PGF for the joint distribution of primary and secondary infections can be written as follows:

$$H_h(x,z) = H_{h,p}(x\hat{H}_{h,s}(z)). \quad (25)$$

The PGF for the total number (irrespective of whether the infection was primary or secondary) is given by the following:

$$H_h(z) = H_{h,p}(zH_{h,s}(z)). \quad (26)$$

The probability of an outbreak being small is simply $H_h(1)$. This probability is shown as a function of $\mathcal{R}_0^{aa}$ and $\mathcal{R}_0^{hh}$, for fixed $\mathcal{R}_0^{ah}$, in Fig. 8(a). Last, the PGF for secondary infections is given by the following:

$$H_{h,s}(z) = H_h(1,z). \quad (27)$$

Following Ref. [24], we can extract the probability of n human hosts getting infected using the Cauchy integral formula,

$$P[Z_h(\infty) = n] = \frac{1}{2\pi i}\oint \frac{H_h(z)}{z^{n+1}}\,dz, \quad (28)$$

where the integral is done over the unit circle $|z| = 1$ in the complex plane. Similarly, the joint probability distribution can be extracted by extending the Cauchy integral formula to higher dimensions,

$$P[Z_{h,p}(\infty) = m, Z_{h,s}(\infty) = n] = \frac{1}{(2\pi i)^2} \oint\oint \frac{H_h(x,z)}{x^{m+1} z^{n+1}}\, dx\, dz, \quad (29)$$

where the integrals are over two unit circles in the x and z complex planes.

B. Critical threshold

The critical threshold is defined as the point in parameter space where the average outbreak size diverges [4,5] and the probability of a large outbreak becomes greater than 0. We introduce the following notation for the average outbreak sizes.

$$\langle n \rangle_a \equiv \mathbb{E}[R_a(\infty)] \quad \langle n \rangle_h \equiv \mathbb{E}[R_h(\infty)]. \quad (30)$$

For the animal population,

$$\langle n \rangle_a = H_a'(1) = \frac{1}{1 - \mathcal{R}_0^{aa}}, \quad (31)$$

which yields the condition $\mathcal{R}_0^{aa} = 1$ as the critical threshold. For the human population,

$$\langle n \rangle_h = H_h'(1) = H_{h,p}'(1)\{1 + H_{h,s}'(1)\} = \frac{\mathcal{R}_0^{ah}}{\left(1 - \mathcal{R}_0^{aa}\right)\left(1 - \mathcal{R}_0^{hh}\right)}. \quad (32)$$

From the above expression, the critical threshold for the human population is given by $\max(\mathcal{R}_0^{aa}, \mathcal{R}_0^{hh}) = 1$. Thus, large outbreaks in the human population are possible even if $\mathcal{R}_0^{hh} < 1$, emphasizing the potential importance of *spillover-driven large outbreaks*.

C. Asymptotic scaling near the critical threshold

The scaling of the outbreak sizes near the threshold boundary can be investigated through the singularity analysis of the associated generated function (see Appendix C for details). The threshold boundary can be divided into three parts: (1) $\mathcal{R}_0^{aa} = 1, \mathcal{R}_0^{hh} < 1$; (2) $\mathcal{R}_0^{aa} < 1, \mathcal{R}_0^{hh} = 1$; and (3) $\mathcal{R}_0^{aa} = 1, \mathcal{R}_0^{hh} = 1$. (Subscripts 1, 2, and 3 below refer to these three parts of the boundary, respectively.) Along the two lines of the threshold boundary excluding the multicritical point, the distribution of outbreak sizes scale as follows:

$$P_i(n) \sim \zeta_i^{-n}\, n^{-3/2}, \quad i = 1,2, \quad (33)$$

whereas near the multicritical point the distribution scales as follows:

$$P_3(n) \sim \zeta_3^{-n} n^{-5/4}. \quad (34)$$

The variables ζ_i are a function of the distance from the threshold boundary. The ζ's are defined in terms of $\Delta_a = 1 - \mathcal{R}_0^{aa}$ and $\Delta_h = 1 - \mathcal{R}_0^{hh}$,

$$\zeta_1 = 1 + \frac{\Delta_h \Delta_a^2}{4\mathcal{R}_0^{ah}} + \mathcal{O}\left(\Delta_h \Delta_a^3\right)$$
$$\zeta_2 = 1 + \frac{\Delta_h^2}{4} \quad (35)$$
$$\zeta_3 = 1 + \frac{\Delta_h^2}{4}, \quad \text{where} \quad \Delta_h = \frac{\Delta_a^2}{2\mathcal{R}_0^{ah}} + \mathcal{O}\left(\Delta_a^3\right).$$

The scaling laws with exponential cutoffs are shown in Fig. 8 along different crossings of the threshold boundary. Note that ζ_3 is a valid exponential cutoff only on a parabolic curve near the multicritical point. The problem of estimating the corrections away from this curve is nontrivial and rigorous results are still an open problem. But intuitively we know that in this case, the generating function will have two singularities which are coalescing at the multicritical point. In such a scenario, there will be a crossover regime where the power-law exponent switches from 3/2 to 5/4 depending on the distance from the threshold boundary.

The appearance of a different scaling exponent at the multicritical point points to a different universality class emerging from the simultaneous divergence of two individual SIR processes in our system. We note that the same $n^{-5/4}$ scaling—arising from one critical process driving another—has been reported recently in a different, albeit related, multitype critical branching process intended to model multistage SIR infections [25].

D. Finite-size scaling at critical threshold

Using the heuristic arguments presented in Ref. [25], we can calculate how the average outbreak size scales with system size at the threshold boundary $\max(\mathcal{R}_0^{aa}, \mathcal{R}_0^{hh}) = 1$. Let M_a be the "maximal" size of an outbreak in the animal population, when $\mathcal{R}_0^{aa} = 1$, such that an outbreak cannot exceed this size due to depletion of susceptible hosts [25]. The effective $\mathcal{R}_0^{aa}$ for a finite-size system reduces to the following:

$$\hat{\mathcal{R}}_0^{aa} = 1 - M_a/N_a. \quad (36)$$

Using Eq. (31), we obtain the following estimate for the scale of the average outbreak size:

$$\langle n \rangle_a \sim N_a/M_a. \quad (37)$$

From the 3/2 scaling law for single-type SIR [25], we obtain a second estimate for the average outbreak size,

$$\langle n \rangle_a = \sum_{n=1}^{M_a} n n^{-3/2} \sim \sqrt{M_a}. \quad (38)$$

Equating the two estimates and imposing self-consistency, one obtains the following scaling laws (see Ref. [25]):

$$M_a \sim N_a^{2/3}, \quad \langle n \rangle_a \sim N_a^{1/3}. \quad (39)$$

One can also calculate the scaling window that represents the distance from the threshold boundary within which the scaling law will hold [8]. For the animal SIR, the scaling window is given by the following:

$$\left|1 - \mathcal{R}_0^{aa}\right| \sim N_a^{-1/3}. \quad (40)$$

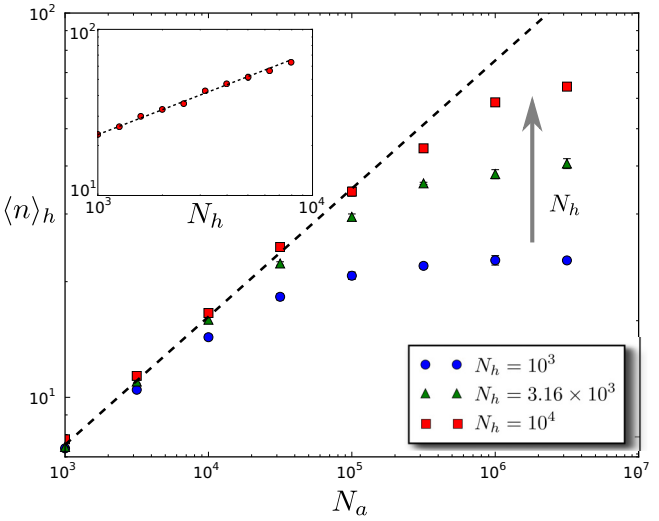

FIG. 9. (Color online) Finite-size scaling at the threshold boundary $\mathcal{R}_0^{aa} = 1, \mathcal{R}_0^{hh} < 1$. The plot shows the scaling law for average outbreak size in humans $\langle n\rangle_h \sim N_a^{1/3}$ and crossover to $N_h^{1/2}$ when $N_h \sim N_a^{2/3}$ on a log-log plot. The points are the average of 7×10^4 stochastic realizations. The dashed line has slope 1/3. Inset: The average outbreak size $\langle n\rangle_h$ plotted against N_h on a log-log scale for fixed $N_a = 10^7$. The dashed line has slope of 1/2. The points are the average over 10^5 stochastic realizations. All results for $\mathcal{R}_0^{aa} = 1, \mathcal{R}_0^{ah} = 0.5, \mathcal{R}_0^{hh} = 0.1$.

The calculation for human outbreaks is separated into three cases [as highlighted in Figs. 8(b), 8(c), and 8(d)]. For $\mathcal{R}_0^{aa} = 1$, $\mathcal{R}_0^{hh} < 1$, the average outbreak size is given by substituting $\hat{\mathcal{R}}_0^{aa}$ in Eq. (32) as follows:

$$\langle n\rangle_h \sim N_a/M_a = N_a^{1/3}. \tag{41}$$

The second estimate is obtained by using the scaling law of 3/2 derived in Eq. (C8b),

$$\langle n\rangle_h = \sum_{n=1}^{M_h} n^{-1/2} \sim \sqrt{M_h}. \tag{42}$$

Equating the two estimates reveals $M_h \sim N_a^{2/3}$. If $\mathcal{O}(N_a) \gg \mathcal{O}(N_h^{3/2})$, the scaling relation leads to the maximal outbreak exceeding the system size, which is physically inconsistent. Thus, the maximal outbreak scale needs to be capped at N_h, i.e.,

$$M_h \sim \min\left(N_a^{2/3}, N_h\right). \tag{43a}$$

From (43a), we can estimate that the crossover regime between the two scales in the min function is given by $N_h \sim N_a^{2/3}$. The scaling of average outbreak size is given by $\sqrt{M_h}$, i.e.,

$$\langle n\rangle_h \sim \min\left(N_a^{1/3}, N_h^{1/2}\right). \tag{43b}$$

The analytical result of Eq. (43b) is validated in Fig. 9. The scaling window near the boundary $\mathcal{R}_0^{aa} = 1, \mathcal{R}_0^{hh} < 1$ is given by the following:

$$\left|1 - \mathcal{R}_0^{aa}\right| \sim \max\left(N_a^{-1/3}, N_h^{-1/2}\right). \tag{44}$$

The case of $\mathcal{R}_0^{aa} < 1, \mathcal{R}_0^{hh} = 1$ results in the same calculations as for a single-type SIR. Thus, the scaling laws are the same as in Eq. (39),

$$M_h \sim N_h^{2/3}, \quad \langle n\rangle_h \sim N_h^{1/3}. \tag{45}$$

with the scaling window same as that for the simple SIR,

$$\left|1 - \mathcal{R}_0^{hh}\right| \sim N_h^{-1/3}. \tag{46}$$

At the multicritical point, the effective basic reproduction numbers are as follows:

$$\hat{\mathcal{R}}_0^{aa} = 1 - M_a/N_a, \quad \hat{\mathcal{R}}_0^{hh} = 1 - M_h/N_h.$$

From (32), we arrive at the first estimate,

$$\langle n\rangle_h \sim \frac{N_a}{M_a}\frac{N_h}{M_h} = \frac{N_a^{1/3} N_h}{M_h}. \tag{47}$$

The second estimate is derived from Eq. (C14a) as follows:

$$\langle n\rangle_h = \sum_{n=1}^{M_h} n^{-1/4} \sim M_h^{3/4}. \tag{48}$$

Equating the two estimates provides the scaling for the maximal outbreak size,

$$M_h \sim \left(N_a^{1/3} N_h\right)^{4/7}. \tag{49}$$

Since the maximal outbreak size cannot exceed the system size, the following holds true:

$$M_h \sim \min\left(N_h, \left(N_a^{1/3} N_h\right)^{4/7}\right). \tag{50a}$$

The scale of the average outbreak size is given by the following:

$$\langle n\rangle_h \sim \min\left(N_h^{3/4}, \left(N_a N_h^3\right)^{1/7}\right). \tag{50b}$$

The crossover region in the multicritical case is $N_h \sim N_a^{4/9}$. The scaling window in this case would depend on both $\mathcal{R}_0^{aa}$ and $\mathcal{R}_0^{hh}$, i.e.,

$$\left|\left(1 - \mathcal{R}_0^{aa}\right)\left(1 - \mathcal{R}_0^{hh}\right)\right| \sim \max\left(N_h^{-3/4}, \left(N_a N_h^3\right)^{-1/7}\right). \tag{51}$$

The finite-size scaling laws have important implications for determining whether a critical outbreak is spillover driven or intrinsically driven. Note that while the two lines of the threshold boundary have the same scaling in the distribution of outbreak sizes, the average outbreak size scales differently along those lines. Whereas an intrinsically driven critical outbreak ($\mathcal{R}_0^{aa} < 1, \mathcal{R}_0^{hh} = 1$) scales only with the abundance of the human host population, a spillover driven critical outbreak ($\mathcal{R}_0^{aa} = 1, \mathcal{R}_0^{hh} < 1$) can depend on the abundance of the animal population. In addition, for $N_h \ll N_a^{2/3}$, a spillover-driven outbreak has a greater extent of $\mathcal{O}(N_h^{1/2})$ as compared to an intrinsically driven outbreak (simple SIR) which is capped at $\mathcal{O}(N_h^{1/3})$.

E. Probability of large outbreak

For the animal population, the probability of large outbreak is calculated as follows:

$$P[R_a(\infty) = \infty] = 1 - H_a(1) = 1 - \frac{1}{\mathcal{R}_0^{aa}}. \tag{52}$$

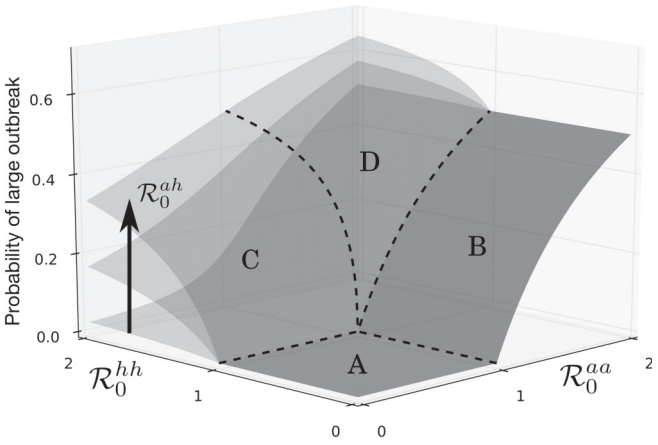

FIG. 10. Probability of a large outbreak in humans for increasing values of $\mathcal{R}_0^{ah} = [0.05, 0.4, 1.0]$. The upper surface is partitioned into four sections: (A) where all outbreaks are small, (B) where spillover-driven large outbreaks are possible, (C) where large outbreaks can only be sustained by human-to-human transmission, and (D) where sustained spillover and human-to-human transmission result in a large outbreak.

Similarly, the probability of a large human outbreak is calculated as follows. Assuming $\mathcal{R}_0^{ah} > 0$,

$$P[R_h(\infty) = \infty] = 1 - H_h(1)$$
$$= 1 - H_{h,p}(1, H_{h,s}(1))$$
$$= \begin{cases} 0 & \text{if } \mathcal{R}_0^{hh} \leqslant 1 \text{ and } \mathcal{R}_0^{aa} \leqslant 1, \\ 1 - \frac{1}{\mathcal{R}_0^{aa}} & \text{if } \mathcal{R}_0^{hh} \leqslant 1 \text{ and } \mathcal{R}_0^{aa} > 1, \\ 1 - A_a\left(1, \frac{1}{\mathcal{R}_0^{hh}}\right) & \text{if } \mathcal{R}_0^{hh} > 1. \end{cases} \quad (53)$$

The probability is shown as a surface plot in Fig. 10.

If $\mathcal{R}_0^{hh} \leqslant 1$, an outbreak in the human population can be large if and only if the outbreak in the animal population is large. In such a case, the probability of a large human outbreak is equal to the probability of a large outbreak in the animal population, which is a function of only $\mathcal{R}_0^{aa}$ (see Fig. 12). On the other hand, if $\mathcal{R}_0^{hh} > 1$, a large human outbreak can occur even if the animal outbreak is small. Figures 11 and 12 compare the analytical results with results from stochastic simulation. Away from the phase transitions at $\mathcal{R}_0^{aa} = 1$ and $\mathcal{R}_0^{hh} = 1$, the results from stochastic simulations show good agreement with the theory. Near the phase transition, the simulation results should converge to the theory for increasing N. Since the definition of a large outbreak becomes precise only in the limit of large system size, there are no finite-size corrections that can be derived in this case.

V. DISCUSSION

We have analyzed a stochastic model of coupled infection dynamics in an animal-human metapopulation using the theory of multitype branching processes. Our results follow from well-established theory [6,7,21,22,25] that we applied to the problem of zoonoses. We have described spillover from animal to human populations, but such a model—or a variant of it,

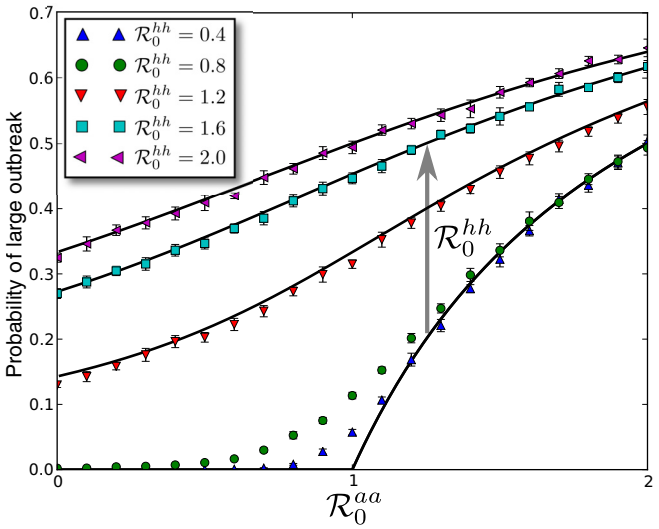

FIG. 11. (Color online) The probability of a large human outbreak for finite populations ($N_a = N_h = 10^3$). The criteria for a large outbreak was chosen as 100 or more infected human hosts. The points represent the result of 10 000 stochastic simulations. The solid lines represent the analytical solution from Eq. (53). The simulations do not agree with the analytical solution near the phase transition because of the chosen criteria for large outbreaks and finite-size effects. All results are for fixed $\mathcal{R}_0^{ah} = 1$.

perhaps with a different form of interspecies coupling—would be applicable to other cross-species infections, such as among different animal hosts. The coupling of animal and human infectious disease dynamics results in important changes to

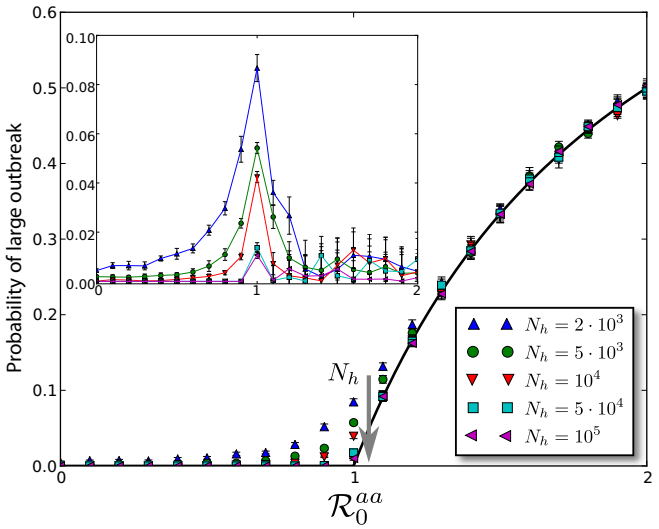

FIG. 12. (Color online) The probability of a large human outbreak for $\mathcal{R}_0^{hh} = 0.8$ and varying N_h. The criteria for a large outbreak was chosen as the number of infected hosts being greater than 1% of the total population. The points represent the result of 10 000 stochastic simulations. The solid line is the analytical solution $\max(0, 1 - 1/\mathcal{R}_0^{aa})$. Inset: The absolute difference between the analytical solution and finite-size results. All results are for fixed $\mathcal{R}_0^{ah} = 0.1$.

the structure of outbreaks in human populations as compared to those in a human-only SIR model.

The statistical quantities that we have calculated in this work provide important insights into spillover and zoonoses. For instance, the probability of spillover (Sec. III B) is strictly less than 1 because the outbreak can die out in the animal population before any primary human infections occur. The distribution of first passage times is useful for understanding the relevant timescales of spillover and their stochastic fluctuations. This serves two important purposes: first, it indicates whether demography should be factored into the model (i.e., whether spillover will take place on a time scale fast compared to demographic changes) and, more crucially, it suggests strategies for optimal surveillance in the field to pinpoint the relevant time scales and surveillance frequencies needed to identify emerging zoonotic infections. Last, the distribution of disease prevalence in the animal population at the moment of spillover (Sec. III C) highlights the intrinsic challenges to parameter estimation in order to build predictive models of cross-species spillover based on prevalence information.

The critical threshold and statistics of human outbreaks allow us to do a comparative analysis between the simple SIR and the coupled SIR presented in this work. Unlike the simple SIR, in our multispecies model the expected outbreak size diverges if either $\mathcal{R}_0^{aa}$ or $\mathcal{R}_0^{hh}$ exceeds 1 (Sec. IV B). Thus, large outbreaks in the human population are possible even if $\mathcal{R}_0^{hh} < 1$, emphasizing the potential importance of *spillover-driven large outbreaks*. This could have important ramifications for zoonotic diseases where human-to-human transmission is not the crucial determinant of the epidemic outcome such as rabies, Nipah virus, Hendra virus, and Menangle virus [13,18]. At the multicritical threshold, the outbreak sizes for the epidemics in the animal and human populations diverge simultaneously, resulting in a new universality class with a different scaling behavior. The animal-human coupling enhances the probability of longer chains [$P(n) \sim n^{-5/4}$ as compared to $P(n) \sim n^{-3/2}$], which could allow for greater opportunity for pathogen adaptation to human hosts [11,26]. Furthermore, depending on where the system is in parameter space, the scaling of the *average* outbreak in the human population sizes can vary significantly [Eqs. (43b) and (45)].

Our analysis suggests the need to be precise with terminology arising in the study of cross-species outbreaks. Various classification schemes previously proposed have delineated among different zoonotic infections based on how infectious they are in the human population, e.g., stages II, III, and IV discussed in Refs. [11,12]. But we have shown that zoonotic infections in all three stages can support large outbreaks in the human population if driven sufficiently hard by an animal outbreak. In addition, the term "stuttering chain" has been used in the literature [11,26] to describe a chain of infections starting from a single infectious host that goes extinct without affecting a significant fraction of the host population. For the single-type SIR model, the term is synonymous with "small outbreak" as we have defined here, and the epidemic threshold is the point in parameter space at which the average length of one such chain diverges. But in our multitype SIR model, the term "stuttering chain" cannot be used interchangeably with "small outbreak." Since multiple introductions can occur in the human population, an outbreak is small if and only if (1) a finite number of distinct infection chains occur in the human population and (2) all such chains stutter to extinction. A large outbreak in the human hosts occurs when any one of these conditions is violated. Specifically, a spillover-driven large outbreak occurs when the number of infection chains diverges, which can happen if $\mathcal{R}_0^{aa} > 1$. Separately, the length of any one such chain can diverge if $\mathcal{R}_0^{hh} > 1$.

The community has advocated "model-guided fieldwork" [11,27,28], as well as increased collaboration between public health scientists and ecologists in developing integrated approaches to predicting and preventing zoonotic epidemics [18]. Mathematical analysis needs to play a central role in such activities, in order to assess the implications of model assumptions. In this paper, we have endeavored to systematically characterize the behavior of a simple model system as a function of model parameters, and various extensions of this work are possible. Within the model itself, one could relax the assumption of homogeneous, full mixing within each population in order to investigate the role of heterogeneous mixing on complex contact networks [4,5,29]. Alternatively, one could develop models of processes outside of the scope of the current model, to address factors such as the ecology of interactions between wildlife and domesticated animals, the encroachment of human development into animal habitats, the evolution of virulence, the propensity for pathogens to successfully jump across species, and the efficacy of various control strategies. We envision the parameters of our cross-species infection model to be the interface to that broader class of models, which would specify how cross-species infection parameters change over time as a function of ecological, evolutionary, and immunological factors. In the current work, we have sought to identify key aspects of phenomenology, highlight the role of important processes, and suggest further inquiry into particular systems of interest in order to help frame more complex and comprehensive descriptions of zoonotic infection.

ACKNOWLEDGMENTS

The authors thank Jason Hindes, Oleg Kogan, Marshall Hayes, Drew Dolgert, and Jamie Lloyd-Smith for helpful discussions and comments on the manuscript. This work was supported by the Science & Technology Directorate, Department of Homeland Security, via Interagency Agreement No. HSHQDC-10-X-00138.

APPENDIX A: MOMENTS OF FIRST PASSAGE TIME

Here we report the calculation of moments for the first passage time distribution,

$$\begin{aligned}\mathbb{E}[T^n \,|\, T<\infty] &= \frac{\mathbb{E}[T^n \mathbf{1}_{\{T<\infty\}}]}{P[T<\infty]} \\ &= \frac{n\int_0^\infty t^{n-1}\, P[t<T<\infty]\,dt}{P[T<\infty]} \\ &= n(v_1-v_0)\int_0^\infty \frac{t^{n-1}e^{-\beta_{aa}(v_1-v_0)t}}{(v_1-1)+(1-v_0)e^{-\beta_{aa}(v_1-v_0)t}}dt,\end{aligned}$$

where $\mathbf{1}_{\{T<\infty\}}$ is the indicator function that takes the value 1 for $T < \infty$. Letting $c = v_1 - 1, d = 1 - v_0$, and $k = \beta_{aa}(v_1 - v_0)$, we can rewrite this as follows:

$$\begin{aligned}\mathbb{E}[T^n \,|\, T < \infty] &= \frac{nk}{\beta_{aa}} \int_0^\infty \frac{t^{n-1} e^{-kt}}{c + de^{-kt}}\, dt \\ &= \frac{-n!}{\beta_{aa} k^{n-1} d}\, \mathrm{Li}_n\left(\frac{-d}{c}\right) \\ &= \frac{n!}{(\beta_{aa})^n (v_0 - 1)(v_1 - v_0)^{n-1}} \mathrm{Li}_n\left(\frac{v_0 - 1}{v_1 - 1}\right),\end{aligned} \tag{A1}$$

where $\mathrm{Li}_n(z)$ is the polylogarithm function of order n. The conditional expected value of the first passage time is obtained by setting $n = 1$ in this expression,

$$\mathbb{E}[T \,|\, T < \infty] = \frac{1}{\beta_{aa}(1 - v_0)} \log\left(\frac{v_1 - v_0}{v_1 - 1}\right). \tag{A2}$$

Our result is for all moments of the distribution; to our knowledge, only the first moment has been reported earlier in Ref. [22] (in the context of a problem in population genetics).

APPENDIX B: FINITE-SIZE CORRECTIONS TO PROBABILITY OF SPILLOVER

The probability of spillover, as calculated in Eq. (15), is valid only in the limit of $N_a, N_h \to \infty$. Deviations from this result are expected for finite system sizes, which we report here. We make the assumption that the linear birth-death process provides accurate statistics of small outbreaks in finite-size systems. This is a valid assumption since small outbreaks are $o(N)$ and, thus, their distribution is independent of the total system size provided $N \gg 1$. Using, Eq. (17), the assumption is stated mathematically as follows:

$$P^N[\text{spill}|\text{small outbreak}] = \begin{cases} 1 - \mathcal{R}_0^{aa} v_0 & \text{if } \mathcal{R}_0^{aa} \leqslant 1, \\ 1 - v_0 & \text{if } \mathcal{R}_0^{aa} > 1. \end{cases} \tag{B1a}$$

$$P^N[\text{small outbreak}] = \frac{1}{\mathcal{R}_0^{aa}}. \tag{B1b}$$

Now we calculate the finite-size equivalent of $P[\text{spill} \,|\, \text{large outbreak}]$ using the hazard function. For this calculation, we ignore the fluctuations around the mean and assume that the animal epidemic obeys the deterministic SIR. Before the first primary infection, the entire human population is susceptible and thus $S_{h,1}(t) = \nu N_h$.

$$\begin{aligned}P^N[\text{spill} \,|\, \text{large outbreak}] &= 1 - \exp\left\{-\int_0^\infty \frac{\beta_{ah} S_{h,1} I_a}{N_a} dt\right\} \\ &= 1 - \exp\left\{-N_a \mathcal{R}_0^{ah} f_a\right\},\end{aligned} \tag{B2a}$$

where

$$f_a = \lim_{N_a \to \infty} \frac{\mathbb{E}[R_a(\infty)]}{N_a} \tag{B2b}$$

is obtained by solving the final size equation for a simple SIR,

$$1 - f_a = e^{-\mathcal{R}_0^{aa} f_a}. \tag{B2c}$$

From Eq. (B2a), $P^N[\text{spill} \,|\, \text{large outbreak}] \to 1$ as $N_a \to \infty$ in agreement with the large system size limit. Using the law of total probability, we now arrive at the probability of spillover with finite-size corrections.

$$\begin{aligned}&P^N\left[\text{spill}; \mathcal{R}_0^{aa} \leqslant 1\right] = 1 - v_0 \\ &P^N\left[\text{spill}; \mathcal{R}_0^{aa} > 1\right] = 1 - v_0 - \left(1 - \frac{1}{\mathcal{R}_0^{aa}}\right) \exp\left\{-N_a \mathcal{R}_0^{ah} f_a\right\}.\end{aligned} \tag{B3}$$

Figure 5 shows the comparison of finite-size corrections as calculated using Eq. (B3) with stochastic simulations.

In the limit of vanishingly small $\mathcal{R}_0^{ah}$, it is important to consider the limit of $\mathcal{R}_0^{ah} N_a$ as $N_a \to \infty$. Let $\xi = \mathcal{R}_0^{ah} N_a$. The probability of spillover presented in Fig. 4 assumes the limit of $\xi \to \infty$. For finite sizes, the probability of spillover simplifies to

$$\begin{aligned}&\lim_{\substack{\mathcal{R}_0^{ah} \to 0 \\ N_a \to \infty}} P^N\left[\text{spill}; \mathcal{R}_0^{aa} \leqslant 1\right] = 0 \\ &\lim_{\substack{\mathcal{R}_0^{ah} \to 0 \\ N_a \to \infty}} P^N\left[\text{spill}; \mathcal{R}_0^{aa} > 1\right] \\ &\qquad = \left(1 - \frac{1}{\mathcal{R}_0^{aa}}\right)[1 - \exp\{-\xi f_a\}].\end{aligned} \tag{B4}$$

Thus, depending on the value of ξ, the limiting value for the probability of spillover when $\mathcal{R}_0^{aa} > 1$ can assume any value in the range $[0, 1 - 1/\mathcal{R}_0^{aa}]$. Thus, if $\mathcal{R}_0^{aa} \gg 1$ and $\mathcal{R}_0^{ah} \ll 1$, then the probability of spillover is indeterminate if there is no information about the scale of $\mathcal{R}_0^{ah} N_a$.

APPENDIX C: ASYMPTOTIC SCALING NEAR THE CRITICAL THRESHOLD

The scaling of the outbreak sizes near the critical threshold can be investigated through the singularity analysis of the associated generated function $H(z)$ [30]. The dominant singularity ζ of the PGF determines the asymptotic form for $P(n)$ which is the probability of having an outbreak of size n. If a given PGF can be expanded around the singularity such that

$$H(z) \sim \left(1 - \frac{z}{\zeta}\right)^\alpha, \tag{C1}$$

then

$$P(n) \sim \frac{\zeta^{-n} n^{-\alpha - 1}}{\Gamma(-\alpha)}, \quad n \to \infty, \tag{C2}$$

where $\alpha \notin \mathbb{Z}_{>0}$. The asymptotic form for $P(n)$ can be derived by substituting Eq. (C1) in the Cauchy integral formula [Eq. (28)] and making the following substitution:

$$z \mapsto \zeta\left(1 + \frac{t}{n}\right). \tag{C3}$$

Thus, the singularity determines the exponential factor and the asymptotic form of the generating function determines the power-law exponent. By rescaling the function $H(z) \to H(z\zeta)$, the calculation of the power-law exponent is simplified since the singularity is now located at $z = 1$. We now apply this analysis to the generating function $H_h(z)$.

Let $\Delta_a = 1 - \mathcal{R}_0^{aa}$ and $\Delta_h = 1 - \mathcal{R}_0^{hh}$ be the distances from the critical thresholds. We first calculate the scaling near

the threshold $\mathcal{R}_0^{aa} = 1$, i.e., $|\Delta_a| < |\Delta_h|$ and $|\Delta_a| \ll 1$. We assume that the parameters are such that the singularities of the generating function $H_h(z)$ are far apart. The dominant singularity near the chosen threshold is given by

$$\zeta_1 = \left(1 + \frac{\left(\sqrt{\mathcal{R}_0^{aa}} - 1\right)^2}{\mathcal{R}_0^{ah}}\right)\left(1 - \mathcal{R}_0^{hh}\frac{\left(\sqrt{\mathcal{R}_0^{aa}} - 1\right)^2}{\mathcal{R}_0^{ah}}\right) = 1 + \Delta_h\left(\frac{\Delta_a^2}{4\mathcal{R}_0^{ah}} + \mathcal{O}\left(\Delta_a^3\right)\right). \tag{C4}$$

The singularity ζ_1 determines the exponential prefactor. To obtain the power-law scaling, the generating function can be analyzed at the critical point ($\mathcal{R}_0^{aa} = 1$ in this case) without loss of generality. At the critical point $\zeta_1 = 1$ and the PGF $H_{h,p}(x)$ simplifies as follows:

$$H_{h,p}(z) = \frac{2 + \mathcal{R}_0^{ah}(1-z) - \sqrt{\mathcal{R}_0^{ah}(1-z)\left(4 + \mathcal{R}_0^{ah}(1-z)\right)}}{2}. \tag{C5}$$

For further simplification, let $z\hat{H}_{h,s}(z)$ be denoted by $\tilde{H}_{h,s}(z)$. Making the substitution (C3) and performing a series expansion in fractional powers of $(-t/n)$ gives

$$\tilde{H}_{h,s}(1 + t/n) \sim 1 + \frac{t}{\Delta_h n}. \tag{C6}$$

Using (26), we obtain

$$H_h(1 + t/n) \sim 1 + \frac{\mathcal{R}_0^{ah}}{2\Delta_h}\left(\frac{-t}{n}\right) - \sqrt{\frac{\mathcal{R}_0^{ah}}{\Delta_h}}\left(\frac{-t}{n}\right)^{1/2} - \frac{1}{8}\left(\frac{\mathcal{R}_0^{ah}}{\Delta_h}\right)^{3/2}\left(\frac{-t}{n}\right)^{3/2}. \tag{C7}$$

By using the Cauchy integral formula on the asymptotic expansion of $H_h(z)$, we obtain the asymptotic probability $P_1(n)$

$$P_1(n) \sim n^{-3/2}, \quad \zeta_1 = 1 \tag{C8a}$$

at the threshold boundary $\mathcal{R}_0^{aa} = 1, \mathcal{R}_0^{hh} \neq 1$. Using the exponential prefactor obtained in Eq. (C4) we arrive at the asymptotic scaling for large n near $\mathcal{R}_0^{aa} = 1$,

$$P_1(n) \sim \zeta_1^{-n} n^{-3/2}. \tag{C8b}$$

This scaling of outbreak sizes near $\mathcal{R}_0^{aa} = 1$ is shown in Fig. 8(b). Note that the scaling can be guessed by looking at the leading term in the expansion, which in Eq. (C7) is $(-t/n)^{1/2}$. Similarly, performing the same steps of analysis near the critical point of $\mathcal{R}_0^{hh} = 1$, we obtain

$$P_2(n) \sim \zeta_2^{-n} n^{-3/2}, \tag{C9}$$

where

$$\zeta_2 = 1 + \frac{\Delta_h^2}{4}. \tag{C10}$$

This scaling of outbreak sizes near $\mathcal{R}_0^{hh} = 1$ is shown in Fig. 8(d). Near the multicritical point $\mathcal{R}_0^{aa} = \mathcal{R}_0^{hh} = 1$, the function has a unique singularity if the value of the function $\tilde{H}_{h,s}(z)$ at its singularity ζ_2 coincides with the singularity of the function $H_{h,p}(z)$, i.e.,

$$1 + \frac{\left(\sqrt{\mathcal{R}_0^{aa}} - 1\right)^2}{\mathcal{R}_0^{ah}} = \frac{\mathcal{R}_0^{hh} + 1}{2\mathcal{R}_0^{hh}}, \tag{C11}$$

which simplifies to

$$\Delta_h = \frac{\Delta_a^2}{2\mathcal{R}_0^{ah}} + \mathcal{O}\left(\Delta_a^3\right) \tag{C12}$$

for $\Delta_a, \Delta_h \ll 1$. The unique singularity is given by ζ_2. Thus, the correction to the pure power law would be ζ_2^{-n} but only on the curve given by Eq. (C12). Next, we extract the power-law scaling at the threshold. For $\mathcal{R}_0^{aa} = \mathcal{R}_0^{hh} = 1$,

$$\tilde{H}_{h,s}(z) = 1 - \sqrt{1-z}$$
$$H_{h,p}(z) = \frac{2 + \mathcal{R}_0^{ah}(1-z) - \sqrt{\mathcal{R}_0^{ah}(1-z)\left(4 + \mathcal{R}_0^{ah}(1-z)\right)}}{2}, \tag{C13}$$

whose functional composition yields

$$H_h(z) = H_{h,p}(\tilde{H}_{h,s}(z))$$
$$H_h(z) = \frac{2 + \mathcal{R}_0^{ah}\sqrt{1-z} - \sqrt{\mathcal{R}_0^{ah}\sqrt{1-z}\left(4 + \mathcal{R}_0^{ah}\sqrt{1-z}\right)}}{2}.$$

Substituting (C3) and performing a series expansion in fractional powers $(-t/n)$, we obtain the $(-t/n)^{1/4}$ as the leading term. Using the Cauchy integral formula, the asymptotic scaling is given by

$$P_3(n) \sim n^{-5/4}, \quad \zeta_3 = 1. \tag{C14a}$$

Away from the multicritical threshold but staying on the curve (C12), the asymptotic form is

$$P_3(n) \sim \zeta_3^{-n} n^{-5/4}, \tag{C14b}$$

where $\zeta_3 = \zeta_2$ as defined in Eq. (C10). This scaling of outbreak sizes at the multicritical point $\mathcal{R}_0^{aa} = 1, \mathcal{R}_0^{hh} = 1$ is shown in Fig. 8(c).

[1] James Dickson Murray, *Mathematical Biology*, Vol. 2 (Springer, Berlin, 2002).

[2] Håkan Andersson and Tom Britton, *Stochastic Epidemic Models and Their Statistical Analysis*, Vol. 4 (Springer, New York, 2000).

[3] F. Brauerand, P. Van den Driessche, J. Wu, and L. J. S. Allen, *Mathematical Epidemiology* (Springer, Berlin, 2008).

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

[5] Eben Kenah and James M. Robins, Phys. Rev. E **76**, 036113 (2007).

[6] N. T. J. Bailey, *The Elements of Stochastic Processes with Applications to the Natural Sciences* (Wiley-Interscience, New York, 1990).

[7] K. B. Athreya and P. E. Ney, *Branching Processes* (Springer-Verlag, Berlin, 1972).

[8] E. Ben-Naim and P. L. Krapivsky, Eur. Phys. J. B **85**, 1 (2012).

[9] M. E. J. Woolhouse and S. Gowtage-Sequeria, Emerg. Infect. Dis. **11**, 1842 (2005).
[10] T. Kuiken *et al*., Science **309**, 1680 (2005).
[11] J. O. Lloyd-Smith *et al*., Science **326**, 1362 (2009).
[12] N. D. Wolfe, C. P. Dunavan, and J. Diamond, Nature **447**, 279 (2007).
[13] S. S. Morse *et al*., Lancet **380**, 1956 (2012).
[14] L. J. S. Allen *et al*., Nat. Resour. Model. **25**, 5 (2012).
[15] S. K. Collinge and C. Ray, *Disease Ecology: Community Structure and Pathogen Dynamics* (Oxford University Press, Oxford, 2006).
[16] J. E. Childs, J. S. Mackenzie, and J. A. Richt, *Wildlife and Emerging Zoonotic Diseases: The Biology, Circumstances and Consequences of Cross-Species Transmission: The Biology, Circumstances, and Consequences of Cross-species Transmission*, Vol. 315 (Springer-Verlag, Berlin, Heidelberg, 2007).
[17] Paul R. Epstein, Am. J. Public Health **85**, 168 (1995).
[18] William B. Karesh *et al*., Lancet **380**, 1936 (2012).
[19] Frank Ball, J. Appl. Probab. **20**, 227 (1983).
[20] Tibor Antal and P. L. Krapivsky, J. Stat. Mech. (2011) P08018.
[21] D. A. Griffiths, J. Appl. Probab. **9**, 65 (1972).
[22] S. Karlin and S. Tavaré, J. Appl. Probab. **19**, 477 (1982).
[23] M. J. Keeling and P. Rohani, *Modeling Infectious Diseases in Humans and Animals* (Princeton University Press, Princeton, NJ, 2008).
[24] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).
[25] Tibor Antal and P. L. Krapivsky, J. Stat. Mech. (2012) P07018.
[26] Rustom Antia, Roland R. Regoes, Jacob C. Koella, and Carl T. Bergstrom, Nature **426**, 658 (2003).
[27] Olivier Restif *et al*., Ecol. Lett. **15**, 1083 (2012).
[28] James L. N. Wood *et al*., Philos. Trans. R. Soc. Lond. B Biol. Sci. **367**, 2881 (2012).
[29] Jason Hindes, Sarabjeet Singh, Christopher R. Myers, and David J. Schneider, Phys. Rev. E **88**, 012809 (2013).
[30] Philippe Flajolet and Robert Sedgewick, *Analytic Combinatorics* (Cambridge University Press, Cambridge, 2009).