Uniformly accurate nonlinear transmission rate models arising from disease spread through pair contacts

Jonathan Wylie[1] and Tom Chou[2]
[1]*Department of Mathematics, City University of Hong Kong, Tat Chee Avenue, Kowloon Tong, Hong Kong*
[2]*Department of Computational Medicine and Department of Mathematics, UCLA, Los Angeles, California 90095, USA*

(Received 29 September 2020; accepted 23 February 2021; published 12 March 2021)

We derive and asymptotically analyze mass-action models for disease spread that include transient pair formation and dissociation. Populations of unpaired susceptible individuals and infected individuals are distinguished from the population of three types of pairs of individuals: both susceptible, one susceptible and one infected, and both infected. Disease transmission can occur only within a pair consisting of one susceptible individual and one infected individual. We use perturbation expansion to formally derive uniformly valid approximations for the dynamics of the total infected and susceptible populations under different conditions including combinations of fast association, fast transmission, and fast dissociation limits. The effective equations are derived from the fundamental mass-action system without implicitly imposing transmission mechanisms, such as those used in frequency-dependent models. Our results represent submodels that show how effective nonlinear transmission can arise from pairing dynamics and are juxtaposed with density-based mass-action and frequency-based models.

DOI: 10.1103/PhysRevE.103.032306

I. INTRODUCTION

Ordinary differential equations (ODEs) have been widely used to model population biology and disease spread in systems where the agents are spatially homogeneous. Canonical mass-action theories include the susceptible-infected-susceptible (SIS), susceptible-infected-recovered (SIR), susceptible-exposed-infected-recovered (SEIR) or susceptible-exposed-infected-susceptible (SEIS), and other models, which have been widely used to provide insight into the dynamics of infected populations [1,2]. Such models are simplified, averaged representations of disease spread within complex, multispecies, and heterogeneous populations. In this paper, we revisit and analyze transmission models and consider the effects of pairing dynamics on infectious disease propagation through a population.

Typically, the transmission rate is assumed to depend on three factors: (a) the rate at which an infected individual contacts other individuals, (b) the proportion of the contacts with susceptible individuals, and (c) the probability that a contact between the infected individual and a susceptible individual leads to the susceptible individual becoming infected. An important factor in determining the contact rate is the relative timescales of the time required for an infected individual to "find" another individual and the time required for behavior that is responsible for transmission.

Two widely used models dominate the literature [2–4]. *Mass-action* transmission models assume that the contact rate between any one infected individual and susceptible individuals is proportional to only the *density* of susceptible individuals. That is, the transmission rate is given by $B_m\rho_S\rho_I$, where ρ_S is the density of susceptible individuals, ρ_I is the density of infected individuals, and B_m is a (constant rate) $\times$ area. Such models are appropriate when the number of contacts per time is not limited by behavior and does not "saturate" at high population densities. In such a limit, the rate of generating new infected individuals is simply proportional to the product of susceptible and infected densities and independent of any behavior that influences the frequency of contacts. Mass-action models are expected to be more accurate for highly contagious diseases such as tuberculosis [5] where frequent, short interactions that occur at rates proportional to population density lead to transmission.

Frequency-based transmission models [6] assume that an infected individual experiences the same number of contacts in a given time period regardless of the density. Given that the populations are homogeneous and that the contact mechanism does not distinguish between susceptible and infected individuals, the proportion of contacts with susceptible individuals will be $\rho_S/(\rho_S+\rho_I)$. The transmission rate will therefore be $B_f\rho_I\rho_S/(\rho_S+\rho_I)$, where B_f is a constant rate specific to an individual's behavior. Frequency-based models are appropriate when behavior (frequency of contacts) that leads to disease transmission is the rate-limiting step. For example, propagation of sexually transmitted diseases is governed through contact frequencies that are behaviorally or socially controlled and not simply proportional to density.

In general, data may not match well with either density-based mass-action or frequency-based models [4]. Although some systems have been shown to match density or frequency models [7], other data exhibit contact rates that are not density or frequency dependent, but fall between these limits as a function of number density [8]. These and many other varied results likely depend on factors such as group size and environment. While difficult to distinguish using data, other authors have devised models in which contact

 ©2021 American Physical Society

or transmission rates scale with density in nontrivial ways [9–11]. Evolutionary game theoretic models can also be cast as density or frequency dependent, leading to different predictions of extinction and coexistence [12]. Other authors have also proposed models with infection rates of the form $B_f B_m \rho_I \rho_S/[B_f + B_m(\rho_S + \rho_I)]$. Such terms are similar to Holling's type II functional response in predator-prey models and asymptotically reproduce the mass-action transmission for low densities and the frequency-based transmission for large densities [13]. While it is unclear what processes give rise to transmission nonlinearity, general forms for density-dependent transmission rates have been proposed and applied to a number of different disease transmission systems [10,14].

It is unclear how these different forms of transmission are mechanistically or mathematically connected. Whether a frequency-like Holling type II model can be theoretically justified or whether an alternative functional response is more natural are also important theoretical questions. Heuristic frequency-based and Holling-type models implicitly incorporate behavior into the dynamics. Qualitatively, one expects that these dynamics might arise from higher-dimensional mass-action ODEs that explicitly include intermediate subpopulations or "reactions" that reflect some behavioral processes.

Here we ask how such behavior-induced frequency-based model can be formally derived from the fundamental mass-action process by considering the simplest mass-action model in which *lone* susceptible and infected individuals associate to form pairs (susceptible-susceptible, susceptible-infected, or infected-infected) [15–19]. These models are similar to the class of household structure models in which groups of individuals form subgroups or communities within which disease transmission spreads faster [20–24]. Pairs can also dissociate into their constituent lone individuals. In these pair-formation-type models, transmission can occur only from an infected individual to a susceptible individual in a susceptible-infected pair. We also include the effects of death and immigration of susceptible individuals, which leads directly to a set of five differential equations: two ODEs for two types of lone individuals and three ODEs for the three different types of pairs. How these five equations can be reduced to *effective* equations under certain conditions will be the topic of our analyses. Previous treatments of the pairing models have been put forth but do not consider certain parameter limits [16], are not systematic [6], implicitly force a frequency-dependent interaction through a "mixing matrix" [6], or only provide approximations at short times [25].

In this paper, we derive, from mass-action models that include pair formation, effective ordinary differential equations for disease spread that are uniformly valid at all times. We first show that if pair dissociation and within-pair transmission is fast, to lowest order, the equations simply reduce to two mass-action-like equations, one for the total infected density and one for the total susceptible density, but with an effective transmission coefficient. If pair association and dissociation are faster than the other processes (death, transmission, and immigration), the resulting effective equations for the total infected and susceptible populations involve terms of rational fractions of polynomials. Such terms represent nonlinearities in the effective transmission rate and can be thought of as interpolations between density- and frequency-based models. These equations further reduce to simpler forms in certain parameter limits.

On the other hand, if association is asymptotically faster than the other process (including dissociation), we show that the leading-order dynamics can only be reduced to three ODEs that bear a number of similarities to models that include an *exposed* subpopulation, such as the susceptible-exposed-infected (SEI) class of models. This type of model, derived from the fundamental mass-action pairing model, reflects a latency period in disease propagation but is still different from the typical SEI-type model.

II. MODELS

We begin by reviewing the basic mass-action, frequency-based, and pairing models for disease propagation.

A. Density-based mass-action model

The simplest mass-action description for the dynamics of the susceptible and infected population densities $\rho_s(t)$ and $\rho_i(t)$ is given by the susceptible-infected model with immigration,

$$\begin{aligned} \frac{d\rho_s(t)}{dt} &= \tilde{\Pi} - \mu_s \rho_s(t) - B_m \rho_s(t)\rho_i(t), \\ \frac{d\rho_i(t)}{dt} &= -\mu_i \rho_i(t) + B_m \rho_s(t)\rho_i(t), \end{aligned} \tag{1}$$

where $\tilde{\Pi}$ represents the rate at which the density of susceptible individuals increases via immigration from outside the region and where μ_s and μ_i are the death rates of susceptible and infected individuals, respectively. If recovery of infected individuals back to the susceptible pool is included, Eq. (1) becomes the standard SIS model when $\tilde{\Pi} = \mu_s = \mu_i = 0$ and the total population is conserved. The steady-state solution to Eqs. (1), $(\rho_s^*, \rho_i^*) = (\tilde{\Pi}/\mu_s, 0)$, exists for all parameters and is linearly stable if the reproduction number

$$R_m := \frac{B_m \tilde{\Pi}}{\mu_s \mu_i} < 1 \tag{2}$$

and linearly unstable if $R_m > 1$. A second steady state $(\rho_s^*, \rho_i^*) = (\mu_i/B_m, \tilde{\Pi}/\mu_i - \mu_s/B_m) > 0$ exists for $R_m > 1$ and is linearly stable. For values of $R_m > 1$, a nonzero infected population can be maintained indefinitely, whereas for $R_m < 1$, the infected population will ultimately die out.

B. Frequency-dependent model

A typical frequency-based model takes the form

$$\begin{aligned} \frac{d\rho_s}{dt} &= \tilde{\Pi} - \mu_s \rho_s - B_f \frac{\rho_s \rho_i}{\rho_s + \rho_i}, \\ \frac{d\rho_i}{dt} &= -\mu_i \rho_i + B_f \frac{\rho_s \rho_i}{\rho_s + \rho_i}, \end{aligned} \tag{3}$$

which is often used to describe sexually transmitted diseases in which the pair-formation rate is thought to be intrinsic to the individual and largely population density independent. The steady state $(\rho_s^*, \rho_i^*) = (\Pi/\mu_s, 0)$ exists for all parameters

and is linearly stable if

$$R_f := \frac{B_f}{\mu_i} < 1 \tag{4}$$

and linearly unstable if $R_f > 1$. A second steady state

$$(\rho_s^*, \rho_i^*) = \left(\frac{\tilde{\Pi}}{\mu_s + B_f - \mu_i}, \frac{(B_f - \mu_i)\tilde{\Pi}}{\mu_i(\mu_s + B_f - \mu_i)} \right) > 0 \tag{5}$$

exists for $R_f > 1$ and is linearly stable.

An important difference arises between the mass-action and frequency-based models. In the mass-action representation, the reproduction number R_m depends on the influx Π of individuals, so reducing the immigration rate will be an effective strategy in disease control. In the frequency-based model, the reproduction number R_f is independent of the influx.

C. Density-based mass-action pairing model

We now consider the simplest mass-action model that explicitly includes population *densities* of transient pairs

$$\frac{d\rho_s}{dt} = \tilde{\Pi} - \mu_s\rho_s - 2\tilde{a}_{ss}\rho_s^2 - \tilde{a}_{si}\rho_s\rho_i + 2(\mu_{ss} + d_{ss})\rho_{ss} + (\mu_{is} + d_{si})\rho_{si}, \tag{6a}$$

$$\frac{d\rho_i}{dt} = -\mu_i\rho_i - 2\tilde{a}_{ii}\rho_i^2 - \tilde{a}_{si}\rho_s\rho_i + 2(\mu_{ii} + d_{ii})\rho_{ii} + (\mu_{si} + d_{si})\rho_{si}, \tag{6b}$$

$$\frac{d\rho_{ss}}{dt} = -(2\mu_{ss} + d_{ss})\rho_{ss} + \tilde{a}_{ss}\rho_s^2, \tag{6c}$$

$$\frac{d\rho_{si}}{dt} = -(\mu_{is} + \mu_{si} + d_{si} + \beta)\rho_{si} + \tilde{a}_{si}\rho_s\rho_i, \tag{6d}$$

$$\frac{d\rho_{ii}}{dt} = -(2\mu_{ii} + d_{ii})\rho_{ii} + \beta\rho_{si} + \tilde{a}_{ii}\rho_i^2, \tag{6e}$$

where ρ_s and ρ_i are the densities of lone susceptible and infected individuals, respectively. The quantities ρ_{ss}, ρ_{si}, and ρ_{ii} are the densities of susceptible-susceptible, susceptible-infected, and infected-infected pairs, respectively. In this model, transmission can occur only from infected to susceptible individuals who are in a susceptible-infected pair and happens at rate β. The rate of immigration of density of lone susceptible individuals is denoted by $\tilde{\Pi}$. In Eqs. (6a)–(6e), μ_s and μ_i represent the death rates of lone susceptible individuals and lone infected individuals, respectively. The quantities μ_{si} and μ_{is} represent the death rates for a susceptible individual in a susceptible-infected pair and and an infected individual in a susceptible-infected pair, respectively. The quantities μ_{ss} and μ_{ii} represent the death rates for *each individual* in a susceptible-susceptible pair and an infected-infected pair, respectively. Similarly, the quantities d (with the appropriate subscripts) represent the dissociation rates of the indicated pairs. The quantities $\tilde{a}$ (with the appropriate subscripts) represent the association rates *per unit density* of the indicated pairs. The association terms represent interactions between two individuals and involve terms that are quadratic in density. They therefore have units of rate × area.

Both Eqs. (6a) and (6c) contain the term $2\mu_{ss}\rho_{ss}$. The factor of 2 in this term arises since there are two susceptible individuals in a susceptible-susceptible pair and the death rate μ_{ss} denotes the rate of death for each individual in the pair. Similarly, Eqs. (6b) and (6e) contain the term $2\mu_{ii}\rho_{ii}$ representing two individuals who can die.

In order to analyze the full model, we nondimensionalize by multiplying each equation by a reference area A_0,

$$\frac{dN_s}{dt} = \Pi - \mu_s N_s - 2a_{ss}N_s^2 - a_{si}N_sN_i + 2(\mu_{ss} + d_{ss})N_{ss} + (\mu_{is} + d_{si})N_{si}, \tag{7a}$$

$$\frac{dN_i}{dt} = -\mu_i N_i - 2a_{ii}N_i^2 - a_{si}N_sN_i + 2(\mu_{ii} + d_{ii})N_{ii} + (\mu_{si} + d_{si})N_{si}, \tag{7b}$$

$$\frac{dN_{ss}}{dt} = -(2\mu_{ss} + d_{ss})N_{ss} + a_{ss}N_s^2, \tag{7c}$$

$$\frac{dN_{si}}{dt} = -(\mu_{is} + \mu_{si} + d_{si} + \beta)N_{si} + a_{si}N_sN_i, \tag{7d}$$

$$\frac{dN_{ii}}{dt} = -(2\mu_{ii} + d_{ii})N_{ii} + \beta N_{si} + a_{ii}N_i^2, \tag{7e}$$

where $N = \rho A_0$ is the total (dimensionless) population within area A_0, $a \equiv \tilde{a}/A_0$ is the rate of association (with units of time^{-1}), and $\Pi = A_0\tilde{\Pi}$ is the immigration rate (with units of time^{-1}). The reference area A_0 is arbitrary but can be chosen to scale the magnitudes of N and the relative rates a/μ. Under any particular scaling, different limits of the magnitudes of N and a/μ, d/μ can be used to further analyze Eqs. (7a)–(7e). Note that models using bilinear terms representing absolute numbers have been denoted pseudo-mass-action [3], but in the context of this work, the area factor connecting N and ρ is irrelevant. To incorporate frequency-dependent transmission into the density-based mass-action model, as has been often done [6,17], the quadratic pairing terms in Eqs. (6a)–(6e) would be replaced by, e.g., $\tilde{a}_{si}\rho_i\rho_s/(\rho_i + \rho_s) = \tilde{a}_{si}N_iN_s/(N_i + N_s)$, where $\tilde{a}_{si}$ has units of time^{-1}.

III. ASYMPTOTIC ANALYSES AND DISCUSSION

We now analyze the mass-action pairing model in different limits to reduce the model to simpler forms in order to illustrate how pairing and dissociation affect the overall propagation of infection.

A. Fast dissociation and transmission limit

First, consider the simplest case where the dissociation and transmission rates are large by scaling them according to $d_{ss} = \bar{d}_{ss}/\varepsilon$, $d_{si} = \bar{d}_{si}/\varepsilon$, $d_{ii} = \bar{d}_{ii}/\varepsilon$, and $\beta = \bar{\beta}/\varepsilon$, with $\varepsilon \to 0^+$. In this limit, we expect the number or density of pairs to be much smaller than the number of unpaired individuals. Under this transformation, Eqs. (7a)–(7e) give

$$\frac{dN_s}{dt} = \Pi - \mu_s N_s - 2a_{ss}N_s^2 - a_{si}N_sN_i + 2\left(\mu_{ss} + \frac{\bar{d}_{ss}}{\varepsilon}\right)N_{ss} + \left(\mu_{is} + \frac{\bar{d}_{si}}{\varepsilon}\right)N_{si}, \tag{8a}$$

$$\frac{dN_i}{dt} = -\mu_i N_i - 2a_{ii}N_i^2 - a_{si}N_sN_i + 2\left(\mu_{ii} + \frac{\bar{d}_{ii}}{\varepsilon}\right)N_{ii} + \left(\mu_{si} + \frac{\bar{d}_{si}}{\varepsilon}\right)N_{si}, \tag{8b}$$

$$\frac{dN_{ss}}{dt} = -\left(2\mu_{ss} + \frac{\bar{d}_{ss}}{\varepsilon}\right)N_{ss} + a_{ss}N_s^2, \tag{8c}$$

$$\frac{dN_{si}}{dt} = -\left(\mu_{is} + \mu_{si} + \frac{\bar{d}_{si}}{\varepsilon} + \frac{\bar{\beta}}{\varepsilon}\right)N_{si} + a_{si}N_sN_i, \tag{8d}$$

$$\frac{dN_{ii}}{dt} = -\left(2\mu_{ii} + \frac{\bar{d}_{ii}}{\varepsilon}\right)N_{ii} + \frac{\bar{\beta}}{\varepsilon}N_{si} + a_{ii}N_i^2. \tag{8e}$$

We then expand the populations in the form

$$\begin{aligned} N_s &= N_s^{(0)} + \varepsilon N_s^{(1)} + \cdots, \\ N_i &= N_i^{(0)} + \varepsilon N_i^{(1)} + \cdots, \\ N_{ss} &= N_{ss}^{(0)} + \varepsilon N_{ss}^{(1)} + \cdots, \\ N_{si} &= N_{si}^{(0)} + \varepsilon N_{si}^{(1)} + \cdots, \\ N_{ii} &= N_{ii}^{(0)} + \varepsilon N_{ii}^{(1)} + \cdots \end{aligned} \tag{9}$$

and substitute them into Eqs. (8a)–(8e) to find, at $O(1/\varepsilon)$,

$$0 = 2\bar{d}_{ss}N_{ss}^{(0)} + \bar{d}_{si}N_{si}^{(0)}, \tag{10a}$$

$$0 = 2\bar{d}_{ii}N_{ii}^{(0)} + \bar{d}_{si}N_{si}^{(0)}, \tag{10b}$$

$$0 = -\bar{d}_{ss}N_{ss}^{(0)}, \tag{10c}$$

$$0 = -(\bar{d}_{si} + \bar{\beta})N_{si}^{(0)}, \tag{10d}$$

$$0 = -\bar{d}_{ii}N_{ii}^{(0)} + \bar{\beta}N_{si}^{(0)}. \tag{10e}$$

Equations (10a)–(10e) can be solved to obtain the leading-order solution $N_{ss}^{(0)} = N_{si}^{(0)} = N_{ii}^{(0)} = 0$. Note that at $O(\varepsilon^{-1})$, the equations do not determine $N_s^{(0)}$ and $N_i^{(0)}$. To do so, we must consider Eqs. (8a)–(8e) at $O(1)$:

$$\begin{aligned} \frac{dN_s^{(0)}}{dt} &= \Pi - \mu_sN_s^{(0)} - 2a_{ss}N_s^{(0)2} - a_{si}N_s^{(0)}N_i^{(0)} \\ &\quad + 2\bar{d}_{ss}N_{ss}^{(1)} + \bar{d}_{si}N_{si}^{(1)}, \end{aligned} \tag{11a}$$

$$\begin{aligned} \frac{dN_i^{(0)}}{dt} &= -\mu_iN_i^{(0)} - 2a_{ii}N_i^{(0)2} - a_{si}N_s^{(0)}N_i^{(0)} \\ &\quad + 2\bar{d}_{ii}N_{ii}^{(1)} + \bar{d}_{si}N_{si}^{(1)}, \end{aligned} \tag{11b}$$

$$0 = -\bar{d}_{ss}N_{ss}^{(1)} + a_{ss}N_s^{(0)2}, \tag{11c}$$

$$0 = -(\bar{d}_{si} + \bar{\beta})N_{si}^{(1)} + a_{si}N_s^{(0)}N_i^{(0)}, \tag{11d}$$

$$0 = -\bar{d}_{ii}N_{ii}^{(1)} + \bar{\beta}N_{si}^{(1)} + a_{ii}N_i^{(0)2}. \tag{11e}$$

Equations (11c)–(11e) can be solved to yield

$$\begin{aligned} N_{ss}^{(1)} &= \frac{a_{ss}}{\bar{d}_{ss}}N_s^{(0)2}, \\ N_{si}^{(1)} &= \frac{a_{si}}{\bar{\beta} + \bar{d}_{si}}N_s^{(0)}N_i^{(0)}, \\ N_{ii}^{(1)} &= \frac{\bar{\beta}}{\bar{d}_{ii}}N_{si}^{(1)} + \frac{a_{ii}}{\bar{d}_{ii}}N_i^{(0)2}. \end{aligned} \tag{12}$$

Upon substitution of the expressions in Eq. (12) into Eqs. (11a) and (11b), we find the ODEs for the leading-order approximations of the number of isolated susceptible individuals and infected individuals

$$\begin{aligned} \frac{dN_s^{(0)}}{dt} &= \Pi - \mu_sN_s^{(0)} - \left(\frac{a_{si}\beta}{\beta + d_{si}}\right)N_s^{(0)}N_i^{(0)}, \\ \frac{dN_i^{(0)}}{dt} &= -\mu_iN_i^{(0)} + \left(\frac{a_{si}\beta}{\beta + d_{si}}\right)N_s^{(0)}N_i^{(0)}, \end{aligned} \tag{13}$$

wherein an *effective* transmission rate can be defined as

$$B_{\text{eff}} := \frac{a_{si}\beta}{\beta + d_{si}} = \frac{a_{si}\bar{\beta}}{\bar{\beta} + \bar{d}_{si}}. \tag{14}$$

In this limit, the effective equations for infected individuals and susceptible individuals retain the mass-action form, but with a modified transmission parameter. The pair-formation process mediates the disease transmission through the association rate a_{si}. For $\bar{\beta} \ll \bar{d}_{si}$, the rate limiting step is transmission within a susceptible-infected pair. When intra-pair transmission is fast, $\bar{\beta} \gg \bar{d}_{si}$, the overall transmission rate $B_{\text{eff}} \approx a_{si}$ approaches the association rate itself. In this limit, the five-dimensional mass-action pairing equations reduce to a two-dimensional mass-action model with a modified transmission rate. Note that if we were to use the frequency-dependent variant of the pairing model, the form would also be preserved to lowest order with the corresponding transmission term $B_{\text{eff}}N_s^{(0)}N_i^{(0)}/(N_i^{(0)} + N_s^{(0)})$.

B. Fast dissociation and association limit

Now consider the limit where *both* the association and dissociation coefficients are significantly larger than the death and infection rates and define $a_{ss} = \bar{a}_{ss}/\varepsilon$, $a_{si} = \bar{a}_{si}/\varepsilon$, $a_{ii} = \bar{a}_{ii}/\varepsilon$, $d_{ss} = \bar{d}_{ss}/\varepsilon$, $d_{si} = \bar{d}_{si}/\varepsilon$, and $d_{ii} = \bar{d}_{ii}/\varepsilon$, with $\varepsilon \to 0^+$. We also perform a linear transformation on Eqs. (7a) and (7b) so that they describe *total* susceptible and infected populations and are independent of ε:

$$\frac{d}{dt}(N_s + 2N_{ss} + N_{si}) = \Pi - \mu_sN_s - 2\mu_{ss}N_{ss} - (\mu_{si} + \beta)N_{si},$$

$$\frac{d}{dt}(N_i + 2N_{ii} + N_{si}) = -\mu_iN_i - 2\mu_{ii}N_{ii} - (\mu_{is} - \beta)N_{si},$$

$$\begin{aligned} \frac{dN_{ss}}{dt} &= -\left(2\mu_{ss} + \frac{\bar{d}_{ss}}{\varepsilon}\right)N_{ss} + \frac{\bar{a}_{ss}}{\varepsilon}N_s^2, \\ \frac{dN_{si}}{dt} &= -\left(\mu_{is} + \mu_{si} + \frac{\bar{d}_{si}}{\varepsilon} + \beta\right)N_{si} + \frac{\bar{a}_{si}}{\varepsilon}N_sN_i, \\ \frac{dN_{ii}}{dt} &= -\left(2\mu_{ii} + \frac{\bar{d}_{ii}}{\varepsilon}\right)N_{ii} + \beta N_{si} + \frac{\bar{a}_{ii}}{\varepsilon}N_i^2. \end{aligned} \tag{15}$$

We now substitute the expansion in Eqs. (9) into Eqs. (15) and keep only the $O(1)$ terms to find

$$\begin{aligned} &\frac{d}{dt}\left(N_s^{(0)} + 2N_{ss}^{(0)} + N_{si}^{(0)}\right) \\ &\quad = \Pi - \mu_sN_s^{(0)} - 2\mu_{ss}N_{ss}^{(0)} - (\mu_{si} + \beta)N_{si}^{(0)}, \end{aligned} \tag{16a}$$

$$\begin{aligned} &\frac{d}{dt}\left(N_i^{(0)} + 2N_{ii}^{(0)} + N_{si}^{(0)}\right) \\ &\quad = -\mu_iN_i^{(0)} - 2\mu_{ii}N_{ii}^{(0)} - (\mu_{is} - \beta)N_{si}^{(0)}, \end{aligned} \tag{16b}$$

$$0 = -\bar{d}_{ss} N_{ss}^{(0)} + \bar{a}_{ss} N_s^{(0)2}, \tag{16c}$$

$$0 = -\bar{d}_{si} N_{si}^{(0)} + \bar{a}_{si} N_s^{(0)} N_i^{(0)}, \tag{16d}$$

$$0 = -\bar{d}_{ii} N_{ii}^{(0)} + \bar{a}_{ii} N_i^{(0)2}. \tag{16e}$$

By using Eqs. (16c)–(16e) to eliminate $N_{ss}^{(0)}$, $N_{si}^{(0)}$, and $N_{ii}^{(0)}$ from Eqs. (16a) and (16b), we obtain

$$\frac{d}{dt}\left(N_s^{(0)} + 2\kappa_{ss} N_s^{(0)2} + \kappa_{si} N_s^{(0)} N_i^{(0)}\right) = \Pi - \mu_s N_s^{(0)} - 2\mu_{ss}\kappa_{ss} N_s^{(0)2} - (\mu_{si} + \beta)\kappa_{si} N_s^{(0)} N_i^{(0)},$$

$$\frac{d}{dt}\left(N_i^{(0)} + 2\kappa_{ii} N_i^{(0)2} + \kappa_{si} N_s^{(0)} N_i^{(0)}\right) = -\mu_i N_i^{(0)} - 2\mu_{ii}\kappa_{ii} N_i^{(0)2} - (\mu_{is} - \beta)\kappa_{si} N_s^{(0)} N_i^{(0)}, \tag{17}$$

where $\kappa_{ss} = \bar{a}_{ss}/\bar{d}_{ss}$, $\kappa_{si} \equiv \kappa_{is} = \bar{a}_{si}/\bar{d}_{si}$, and $\kappa_{ii} = \bar{a}_{ii}/\bar{d}_{ii}$.

1. Steady states and stability

The most convenient way to determine the steady states and/or further analyze Eqs. (17) is to unpack them in terms of $N_i^{(0)}$ and $N_s^{(0)}$ and write them in the form

$$\frac{d}{dt}\begin{bmatrix} N_s^{(0)} \\ N_i^{(0)} \end{bmatrix} = \mathbf{M}^{-1}\begin{bmatrix} \Pi - \mu_s N_s^{(0)} - 2\mu_{ss}\kappa_{ss} N_s^{(0)2} - (\mu_{si} + \beta)\kappa_{si} N_s^{(0)} N_i^{(0)} \\ -\mu_i N_i^{(0)} - 2\mu_{ii}\kappa_{ii} N_i^{(0)2} - (\mu_{is} - \beta)\kappa_{si} N_s^{(0)} N_i^{(0)} \end{bmatrix}, \tag{18}$$

where

$$\mathbf{M} = \begin{bmatrix} 1 + 4\kappa_{ss} N_s^{(0)} + \kappa_{si} N_i^{(0)} & \kappa_{si} N_s^{(0)} \\ \kappa_{si} N_i^{(0)} & 1 + 4\kappa_{ii} N_i^{(0)} + \kappa_{si} N_s^{(0)} \end{bmatrix}.$$

Note that since $N_s^{(0)}, N_i^{(0)}, \kappa_{ss}, \kappa_{si}, \kappa_{ii} \geqslant 0$, the eigenvalues of $\mathbf{M}$ can never be zero and $\mathbf{M}$ is invertible.

We can readily show that the system of equations always supports an infection-free steady-state solution

$$\left(N_s^{(0)}, N_i^{(0)}\right) = \left(\frac{-\mu_s + \sqrt{\mu_s^2 + 8\mu_{ss}\kappa_{ss}\Pi}}{4\mu_{ss}\kappa_{ss}}, 0\right) \tag{19}$$

and that this solution is linearly stable if

$$R := \frac{\kappa_{si}(\beta - \mu_{is})\left(-\mu_s + \sqrt{\mu_s^2 + 8\mu_{ss}\kappa_{ss}\Pi}\right)}{4\mu_i\mu_{ss}\kappa_{ss}} < 1 \tag{20}$$

and linearly unstable if $R > 1$. Another stable solution with positive $N_s^{(0)}$ and $N_i^{(0)}$ will arise if $R > 1$. This solution structure closely mirrors that of the mass-action and frequency-dependent models.

2. Comparison to mass-action and frequency-based models

In order to compare Eqs. (17) or (18) to the simpler classic models, it is preferable to rewrite the equations in terms of the leading-order expressions for the total susceptible and infected populations

$$N_S^{(0)} = N_s^{(0)} + 2N_{ss}^{(0)} + N_{si}^{(0)}, \qquad N_I^{(0)} = N_i^{(0)} + 2N_{ii}^{(0)} + N_{si}^{(0)}. \tag{21}$$

Again using Eqs. (16c)–(16e) to eliminate $N_{ss}^{(0)}$, $N_{si}^{(0)}$, and $N_{ii}^{(0)}$, we find

$$N_S^{(0)} = N_s^{(0)} + 2\kappa_{ss} N_s^{(0)2} + \kappa_{si} N_s^{(0)} N_i^{(0)}, \tag{22a}$$

$$N_I^{(0)} = N_i^{(0)} + 2\kappa_{ii} N_i^{(0)2} + \kappa_{si} N_s^{(0)} N_i^{(0)}. \tag{22b}$$

Next we need to express the quantities $N_s^{(0)}$ and $N_i^{(0)}$ in terms of $N_S^{(0)}$ and $N_I^{(0)}$. Solving Eq. (22b) for $N_i^{(0)}$ and substituting the result into Eq. (22a), we find a quartic equation for $N_s^{(0)}$,

$$2\kappa_{ss}\left(4\kappa_{ii}\kappa_{ss} - \kappa_{si}^2\right) N_s^{(0)4} + \left(8\kappa_{ii}\kappa_{ss} - 2\kappa_{si}\kappa_{ss} - \kappa_{si}^2\right) N_s^{(0)3} + \left(\kappa_{si}^2 (N_S^{(0)} - N_I^{(0)}) - 8N_S^{(0)}\kappa_{ii}\kappa_{ss} + 2\kappa_{ii} - \kappa_{si}\right) N_s^{(0)2} + N_S^{(0)}(\kappa_{si} - 4\kappa_{ii}) N_s^{(0)} + 2N_S^{(0)2}\kappa_{ii} = 0. \tag{23}$$

We can readily show that only one of the four roots gives values of $N_s^{(0)}$ and $N_i^{(0)}$ that are both positive when $N_S^{(0)}$ and $N_I^{(0)}$ are positive. Upon using this physical root for $N_s^{(0)}$ as functions of $N_S^{(0)}$ and $N_I^{(0)}$ in Eq. (22a), we find the unique physical root for $N_i^{(0)}$, expressed in terms of $N_S^{(0)}$ and $N_I^{(0)}$. Explicit formulas for the solution of a quartic are known, so we can express $N_s^{(0)} \equiv F_S(N_S^{(0)}, N_I^{(0)})$ and $N_i^{(0)} \equiv F_I(N_S^{(0)}, N_I^{(0)})$ as functions F_S and F_I that are obtained by the procedure described above. We can then rewrite

$$\frac{dN_S^{(0)}}{dt} = \Pi - \mu_s N_S^{(0)} + 2(\mu_s - \mu_{ss})\kappa_{ss} F_S^2 + (\mu_s - \mu_{si} - \beta)\kappa_{si} F_S F_I,$$

$$\frac{dN_I^{(0)}}{dt} = -\mu_i N_I^{(0)} + 2(\mu_i - \mu_{ii})\kappa_{ii} F_I^2 + (\mu_i - \mu_{is} + \beta)\kappa_{si} F_S F_I. \tag{24}$$

Although $F_S(N_S^{(0)}, N_I^{(0)})$ and $F_I(N_S^{(0)}, N_I^{(0)})$ are unwieldy functions of $N_S^{(0)}$ and $N_I^{(0)}$, Eqs. (24) represent a systematic projection of the original five-dimensional problem to two closed equations describing the total susceptible and infected populations $N_S^{(0)}$ and $N_I^{(0)}$. Note that we retained immigration and death in our general model, but in the limit where $\Pi = \mu = 0$, the total population is conserved at all times and $N_S^{(0)} + N_I^{(0)} = \text{const}$ as in the standard SIS model. Equations (24) can be further simplified in different limits of rate parameters as described below.

3. Low-density asymptotics

Consider the solutions to $N_s^{(0)}$ and $N_i^{(0)}$ in the limit where the populations in the reference area A_0 are small, $N_S^{(0)}, N_I^{(0)} \ll 1$. Upon Taylor expansion of the solutions to Eqs. (22a) and (22b), we find $F_S(N_S^{(0)}, N_I^{(0)}) \approx N_S^{(0)} - (\kappa_{si} N_I^{(0)} + 2\kappa_{ss} N_S^{(0)}) N_S^{(0)} + O(N_{S,I}^{(0)3})$ and $F_I(N_S^{(0)}, N_I^{(0)}) \approx$

$N_{\rm I}^{(0)} - (\kappa_{\rm si}N_{\rm S}^{(0)} + 2\kappa_{\rm ii}N_{\rm I}^{(0)})N_{\rm I}^{(0)} + O(N_{\rm S,I}^{(0)3})$, and Eqs. (24) to lowest order become

$$\begin{aligned}\frac{dN_{\rm S}^{(0)}}{dt} &\approx \Pi - \mu_{\rm s}N_{\rm S}^{(0)} + 2(\mu_{\rm s}-\mu_{\rm ss})\kappa_{\rm ss}N_{\rm S}^{(0)2} \\ &\quad + (\mu_{\rm s}-\mu_{\rm si}-\beta)\kappa_{\rm si}N_{\rm S}^{(0)}N_{\rm I}^{(0)}, \\ \frac{dN_{\rm I}^{(0)}}{dt} &\approx -\mu_{\rm i}N_{\rm I}^{(0)} + 2(\mu_{\rm i}-\mu_{\rm ii})\kappa_{\rm ii}N_{\rm I}^{(0)2} \\ &\quad + (\mu_{\rm i}-\mu_{\rm is}+\beta)\kappa_{\rm si}N_{\rm S}^{(0)}N_{\rm I}^{(0)}. \end{aligned} \tag{25}$$

The dynamics in this low-density limit are dominated by immigration and death but are also qualitatively different from those of the standard mass-action model in that Eqs. (25) contain $N_{\rm S}^{(0)2}$ and $N_{\rm I}^{(0)2}$ terms. These quadratic terms arise from the difference in death rates between paired and unpaired susceptible individuals $\mu_{\rm s}-\mu_{\rm ss}$ and paired and unpaired infected individuals $\mu_{\rm i}-\mu_{\rm ii}$. However, if we assume that the death rate is independent of the pairing status, i.e., $\mu_{\rm ss}=\mu_{\rm s}$, $\mu_{\rm ii}=\mu_{\rm i}$, $\mu_{\rm si}=\mu_{\rm s}$, and $\mu_{\rm is}=\mu_{\rm i}$, we obtain the standard mass-action model with $B_{\rm m}\propto\kappa_{\rm si}\beta$.

4. High-density asymptotics

If $N_{\rm S}^{(0)}, N_{\rm I}^{(0)} \gg 1$, and hence $N_{\rm s}^{(0)}, N_{\rm i}^{(0)} \gg 1$, the physical solutions to Eqs. (22a) and (22b) are approximately

$$\begin{aligned} F_{\rm S}(N_{\rm S}^{(0)},N_{\rm I}^{(0)}) &\approx \sqrt{\frac{N_{\rm I}^{(0)} + (2K-1)N_{\rm S}^{(0)} - \sqrt{\left(N_{\rm I}^{(0)}-N_{\rm S}^{(0)}\right)^2 + 4KN_{\rm S}^{(0)}N_{\rm I}^{(0)}}}{4\kappa_{\rm ss}(K-1)}}, \\ F_{\rm I}(N_{\rm S}^{(0)},N_{\rm I}^{(0)}) &\approx \sqrt{\frac{N_{\rm S}^{(0)} + (2K-1)N_{\rm I}^{(0)} - \sqrt{\left(N_{\rm I}^{(0)}-N_{\rm S}^{(0)}\right)^2 + 4KN_{\rm S}^{(0)}N_{\rm I}^{(0)}}}{4\kappa_{\rm ii}(K-1)}}, \end{aligned} \tag{26}$$

where $K \equiv 4\kappa_{\rm ss}\kappa_{\rm ii}/\kappa_{\rm si}^2$. Upon substituting Eqs. (26) into Eqs. (24), we find the effective equations for $N_{\rm S}^{(0)}, N_{\rm I}^{(0)} \gg 1$. In this case, even if $\mu_{\rm ss}=\mu_{\rm s}$, $\mu_{\rm ii}=\mu_{\rm i}$, $\mu_{\rm si}=\mu_{\rm s}$, and $\mu_{\rm is}=\mu_{\rm i}$, the effective model differs significantly in form from both the mass-action and frequency-dependent models. In Fig. 1 we compare the exact solutions of $N_{\rm S}(t)$ and $N_{\rm I}(t)$ from Eqs. (7a)–(7e) to $N_{\rm S}^{(0)}(t)$ and $N_{\rm I}^{(0)}(t)$ derived from solving Eqs. (24) using Eqs. (26). There is excellent agreement at all times.

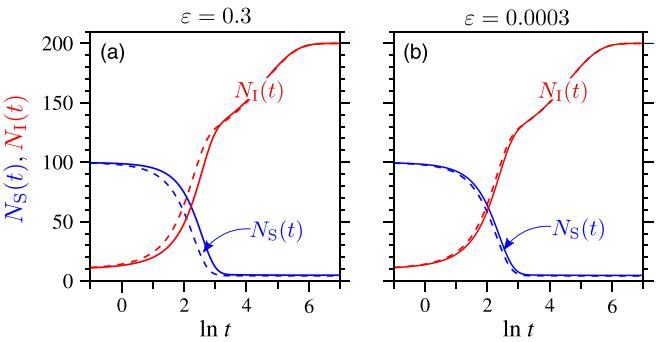

FIG. 1. Fast association and dissociation in the high-density limit. Comparison of the numerical solution of Eqs. (7a)–(7e) with the numerical solution of the high-density asymptotic approximation derived from using Eqs. (26) in Eqs. (24). We plot the total susceptible and infected populations $N_{\rm s}+N_{\rm si}+2N_{\rm ss}$ and $N_{\rm i}+N_{\rm si}+2N_{\rm ii}$, derived from Eqs. (7a)–(7e) (solid blue and red curves) versus $N_{\rm S}^{(0)}$ and $N_{\rm I}^{(0)}$ from Eqs. (26) and (24) (dashed blue and dashed red curves) as functions of $\ln t$. (a) The parameters used are $\bar{a}=\bar{d}=1$, $\varepsilon=0.3$, $\mu_{\rm s}=\mu_{\rm ss}=\mu_{\rm si}=0$, $\mu_{\rm i}=\mu_{\rm ii}=\mu_{\rm is}=0.01$, $\Pi=2$, and $\beta=0.5$, with initial conditions $N_{\rm s}^{(0)}(0)=100$ and $N_{\rm i}^{(0)}(0)=10$. (b) Same parameters and initial conditions as in (a) but with $\varepsilon=0.0003$. In both plots, the decreasing and increasing curves indicate $N_{\rm S}^{(0)}(t)$ and $N_{\rm I}^{(0)}(t)$, respectively. The asymptotic approximations are quite accurate even for $\varepsilon=0.3$.

5. Equal association rates and equal dissociation rates

A further simplification can be made in the special case in which both the pairing and dissociation rates are equal to each other for all types of pairs. This implies that the dissociation coefficients for each of the pairings are the same $d_{\rm ss}=d_{\rm ii}=d_{\rm si}$. For association, there are three possible pairings: susceptible-susceptible, infected-infected, and susceptible-infected. A pair with one infected individual and one susceptible individual can combinatorially arise in two ways so $a_{\rm si}=2a_{\rm ss}=2a_{\rm ii}$. Thus, $\kappa_{\rm si}=2\kappa_{\rm ss}=2\kappa_{\rm ii}\equiv\kappa$ and $K=1$. The physical solution to Eqs. (22a) and (22b) then reduces to

$$\begin{aligned} &F_{\rm S}\left(N_{\rm S}^{(0)},N_{\rm I}^{(0)}\right) \\ &\quad = \frac{N_{\rm S}^{(0)}}{4\kappa\left(N_{\rm S}^{(0)}+N_{\rm I}^{(0)}\right)}\left(\sqrt{8\kappa\left(N_{\rm I}^{(0)}+N_{\rm S}^{(0)}\right)+1}-1\right), \\ &F_{\rm I}\left(N_{\rm S}^{(0)},N_{\rm I}^{(0)}\right) \\ &\quad = \frac{N_{\rm I}^{(0)}}{4\kappa\left(N_{\rm S}^{(0)}+N_{\rm I}^{(0)}\right)}\left(\sqrt{8\kappa\left(N_{\rm I}^{(0)}+N_{\rm S}^{(0)}\right)+1}-1\right). \end{aligned} \tag{27}$$

Using these expressions, Eqs. (24) in the $N_{\rm S}^{(0)}+N_{\rm I}^{(0)} \gg 1$ limit simplify to

$$\begin{aligned} \frac{dN_{\rm S}^{(0)}}{dt} &= \Pi - \mu_{\rm s}N_{\rm S}^{(0)} + (\mu_{\rm s}-\mu_{\rm ss})\frac{N_{\rm S}^{(0)2}}{N_{\rm S}^{(0)}+N_{\rm I}^{(0)}} \\ &\quad + \frac{(\mu_{\rm s}-\mu_{\rm si}-\beta)}{2}\frac{N_{\rm S}^{(0)}N_{\rm I}^{(0)}}{N_{\rm S}^{(0)}+N_{\rm I}^{(0)}}, \\ \frac{dN_{\rm I}^{(0)}}{dt} &= -\mu_{\rm i}N_{\rm I}^{(0)} + (\mu_{\rm i}-\mu_{\rm ii})\frac{N_{\rm I}^{(0)2}}{N_{\rm S}^{(0)}+N_{\rm I}^{(0)}} \\ &\quad + \frac{(\mu_{\rm i}-\mu_{\rm is}+\beta)}{2}\frac{N_{\rm S}^{(0)}N_{\rm I}^{(0)}}{N_{\rm S}^{(0)}+N_{\rm I}^{(0)}}, \end{aligned} \tag{28}$$

which is similar to a frequency-dependent model with effective transmission rate $B_f = \beta/2$. Thus, we have found a specific limit where pairing and unpairing dynamics within a mass-action model reduces it to an effective frequency-dependent model.

C. Fast association limit

Now consider a different limit in which the fast dissociation constraint is relaxed and assume only the association rates are significantly larger than all other (dissociation, death, and infection) rates. Upon defining $a_{ss} = \bar{a}_{ss}/\varepsilon$, $a_{si} = \bar{a}_{si}/\varepsilon$, and $a_{ii} = \bar{a}_{ii}/\varepsilon$, with $\varepsilon \to 0^+$, Eqs. (7a)–(7e) become

$$\begin{aligned}
\frac{dN_s}{dt} &= \Pi - \mu_s N_s - 2\frac{\bar{a}_{ss}}{\varepsilon}N_s^2 - \frac{\bar{a}_{si}}{\varepsilon}N_s N_i \\
&\quad + 2(\mu_{ss} + d_{ss})N_{ss} + (\mu_{is} + d_{si})N_{si}, \\
\frac{dN_i}{dt} &= -\mu_i N_i - 2\frac{\bar{a}_{ii}}{\varepsilon}N_i^2 - \frac{\bar{a}_{si}}{\varepsilon}N_s N_i \\
&\quad + 2(\mu_{ii} + d_{ii})N_{ii} + (\mu_{si} + d_{si})N_{si}, \\
\frac{dN_{ss}}{dt} &= -(2\mu_{ss} + d_{ss})N_{ss} + \frac{\bar{a}_{ss}}{\varepsilon}N_s^2, \\
\frac{dN_{si}}{dt} &= -(\mu_{is} + \mu_{si} + d_{si} + \beta)N_{si} + \frac{\bar{a}_{si}}{\varepsilon}N_s N_i, \\
\frac{dN_{ii}}{dt} &= -(2\mu_{ii} + d_{ii})N_{ii} + \beta N_{si} + \frac{\bar{a}_{ii}}{\varepsilon}N_i^2.
\end{aligned} \tag{29}$$

Since the fast association $\bar{a}/\varepsilon$ terms are products of single populations N_i and N_s, we expect an expansion in powers of $\sqrt{\varepsilon}$. Substituting the expansion

$$\begin{aligned}
N_s &= N_s^{(0)} + \varepsilon^{1/2}N_s^{(1)} + \varepsilon N_s^{(2)} + \cdots, \\
N_i &= N_i^{(0)} + \varepsilon^{1/2}N_i^{(1)} + \varepsilon N_i^{(2)} + \cdots, \\
N_{ss} &= N_{ss}^{(0)} + \varepsilon^{1/2}N_{ss}^{(1)} + \varepsilon N_{ss}^{(2)} + \cdots, \\
N_{si} &= N_{si}^{(0)} + \varepsilon^{1/2}N_{si}^{(1)} + \varepsilon N_{si}^{(2)} + \cdots, \\
N_{ii} &= N_{ii}^{(0)} + \varepsilon^{1/2}N_{ii}^{(1)} + \varepsilon N_{ii}^{(2)} + \cdots
\end{aligned} \tag{30}$$

into Eqs. (29) and retaining only terms of size $O(\varepsilon^{-1})$, we find $N_s^{(0)} = N_i^{(0)} = 0$. Next, collecting terms of size $O(1)$, we obtain

$$\begin{aligned}
0 &= \Pi - 2\bar{a}_{ss}N_s^{(1)2} - \bar{a}_{si}N_s^{(1)}N_i^{(1)} \\
&\quad + 2(\mu_{ss} + d_{ss})N_{ss}^{(0)} + (\mu_{is} + d_{si})N_{si}^{(0)}, && \text{(31a)} \\
0 &= -2\bar{a}_{ii}N_i^{(1)2} - \bar{a}_{si}N_s^{(1)}N_i^{(1)} \\
&\quad + 2(\mu_{ii} + d_{ii})N_{ii}^{(0)} + (\mu_{si} + d_{si})N_{si}^{(0)}, && \text{(31b)} \\
\frac{dN_{ss}^{(0)}}{dt} &= -(2\mu_{ss} + d_{ss})N_{ss}^{(0)} + \bar{a}_{ss}N_s^{(1)2}, && \text{(31c)} \\
\frac{dN_{si}^{(0)}}{dt} &= -(\mu_{is} + \mu_{si} + d_{si} + \beta)N_{si}^{(0)} + \bar{a}_{si}N_s^{(1)}N_i^{(1)}, && \text{(31d)} \\
\frac{dN_{ii}^{(0)}}{dt} &= -(2\mu_{ii} + d_{ii})N_{ii}^{(0)} + \beta N_{si}^{(0)} + \bar{a}_{ii}N_i^{(1)2}. && \text{(31e)}
\end{aligned}$$

Upon solving Eqs. (31a) and (31b), we find

$$\begin{aligned}
N_s^{(1)} &= \sqrt{\frac{P_i + (2f-1)P_s - \sqrt{(P_i - P_s)^2 + 4fP_iP_s}}{4\bar{a}_{ss}(f-1)}}, \\
N_i^{(1)} &= \sqrt{\frac{P_s + (2f-1)P_i - \sqrt{(P_i - P_s)^2 + 4fP_iP_s}}{4\bar{a}_{ii}(f-1)}},
\end{aligned} \tag{32}$$

where $f \equiv 4\bar{a}_{ss}\bar{a}_{ii}/\bar{a}_{si}^2$ and

$$\begin{aligned}
P_s &= 2(\mu_{ss} + d_{ss})N_{ss}^{(0)} + (\mu_{is} + d_{si})N_{si}^{(0)} + \Pi \\
P_i &= 2(\mu_{ii} + d_{ii})N_{ii}^{(0)} + (\mu_{si} + d_{si})N_{si}^{(0)}.
\end{aligned} \tag{33}$$

Thus, to lowest order in the fast association limit, the infected population is $N_I^{(0)} \approx 2N_{ii}^{(0)} + N_{si}^{(0)}$. In what follows, it will be useful to define the susceptible individuals who are in susceptible-infected pairs, $N_E^{(0)} \equiv N_{si}^{(0)}$, as an exposed population. Analogously, the unexposed susceptible population *not* in mixed pairs is dominated by susceptible-susceptible pairs and is $N_S^{(0)} \approx 2N_{ss}^{(0)}$.

Rewriting Eqs. (31c)–(31e) using Eqs. (32), we find

$$\begin{aligned}
\frac{dN_S^{(0)}}{dt} &= -(2\mu_{ss} + d_{ss})N_S^{(0)} \\
&\quad + \frac{P_i + (2f-1)P_s - \sqrt{(P_i - P_s)^2 + 4fP_iP_s}}{2(f-1)}, \\
\frac{dN_E^{(0)}}{dt} &= -(\mu_{is} + \mu_{si} + d_{si} + \beta)N_E^{(0)} \\
&\quad + \frac{\sqrt{(P_i - P_s)^2 + 4fP_sP_i} - (P_i + P_s)}{2(f-1)}, \\
\frac{dN_I^{(0)}}{dt} &= -\mu_{ii}N_I^{(0)} + (\mu_{ii} - \mu_{is} + \beta)N_E^{(0)},
\end{aligned} \tag{34}$$

where P_s and P_i can also be expressed as

$$\begin{aligned}
P_s &= (\mu_{ss} + d_{ss})N_S^{(0)} + (\mu_{is} + d_{si})N_E^{(0)} + \Pi, \\
P_i &= (\mu_{ii} + d_{ii})N_I^{(0)} + (\mu_{si} - \mu_{ii} + d_{si} - d_{ii})N_E^{(0)}.
\end{aligned} \tag{35}$$

Equations (34) and (35) constitute a self-contained system of equations for the three subpopulations $N_S^{(0)}(t)$, $N_E^{(0)}(t)$, and $N_I^{(0)}(t)$.

An alternative formulation is to group all susceptible individuals and write

$$\begin{aligned}
\frac{d}{dt}\left(N_S^{(0)} + N_E^{(0)}\right) &= \Pi - \mu_{ss}\left(N_S^{(0)} + N_E^{(0)}\right) \\
&\quad + (\mu_{ss} - \mu_{si} - \beta)N_E^{(0)}, \\
\frac{dN_E^{(0)}}{dt} &= -(\mu_{is} + \mu_{si} + d_{si} + \beta)N_E^{(0)} \\
&\quad + \frac{\sqrt{(P_i - P_s)^2 + 4fP_sP_i} - (P_i + P_s)}{2(f-1)}, \\
\frac{dN_I^{(0)}}{dt} &= -\mu_{ii}N_I^{(0)} + (\mu_{ii} - \mu_{is} + \beta)N_E^{(0)}.
\end{aligned} \tag{36}$$

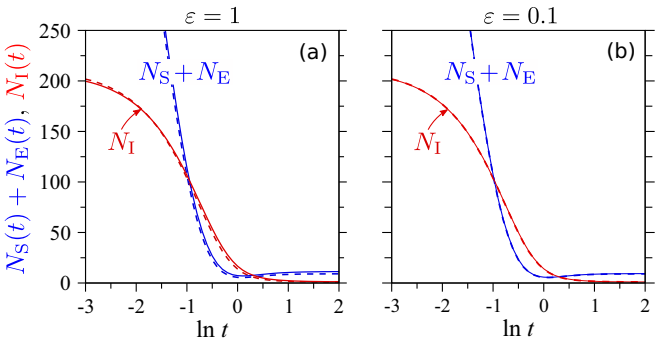

FIG. 2. Fast association limit. Comparison of numerical solution of Eqs. (7a)–(7e) to the numerical solution of the high association rate approximation [Eqs. (36)]. The exact solutions of $N_s + N_{si} + 2N_{ss}$ and $N_i + N_{si} + 2N_{ii}$ [from Eqs. (7a)–(7e), solid blue and red, respectively] are compared with the corresponding quantities $N_S + N_E$ and N_I found from numerically integrating Eqs. (36). (a) The parameters used are $\bar{a} = 1$, $d = 1$, $\mu_s = 0.01$, $\mu_i = \mu_{is} = \mu_{si} = 0.05$, $\mu_{ii} = \mu_{ss} = 5$, $\Pi = 50$, and $\beta = 100$, with initial conditions $N_s(0) = N_i(0) = 0$, $N_{ss}(0) = 500$, $N_{si}(0) = 100$, and $N_{ii} = 10$. These parameters correspond to $R_0 < 1$ and an infection that dies out. The corresponding initial conditions for Eqs. (36) are $N_S(0) = 1000$, $N_E(0) = 100$, and $N_I(0) = 120$. (a) The approximation is highly accurate even for $\varepsilon = 1$. (b) Same parameters as in (a) but with $\varepsilon = 0.1$, for which the approximation is indistinguishable, in this plot, from the full numerical solution.

In the case $\bar{a}_{si}^2 \to 4\bar{a}_{ss}\bar{a}_{ii}$ ($f \to 1$), we apply l'Hôpital's rule to Eqs. (34) to further simplify them to

$$\begin{aligned}
\frac{dN_S^{(0)}}{dt} &= -(2\mu_{ss} + d_{ss})N_S^{(0)} + \frac{P_s^2}{P_s + P_i},\\
\frac{dN_E^{(0)}}{dt} &= -(\mu_{is} + \mu_{si} + d_{si} + \beta)N_E^{(0)} + \frac{P_s P_i}{P_s + P_i},\\
\frac{dN_I^{(0)}}{dt} &= -\mu_{ii}N_I^{(0)} + (\mu_{ii} - \mu_{is} + \beta)N_E^{(0)},
\end{aligned} \tag{37}$$

which are reminiscent of simple SEI-type models [26–28]. A comparison between N_S and N_I derived from the exact equations (7a)–(7e) and those derived from solving Eqs. (37) is given in Fig. 2. The approximations are accurate for all valid parameter regimes across all times.

Linearization of Eqs. (34) and (37) about the disease-free equilibrium point $(N_S^*, N_E^*, N_I^*) = (\Pi/\mu_{ss}, 0, 0)$ yields [omitting the (0) notation]

$$\begin{aligned}
\frac{d\delta N_S}{dt} &= -\mu_{ss}\delta N_S + (\mu_{ii} + d_{ii} + \mu_{is} - \mu_{si})\delta N_E\\
&\quad - (\mu_{ii} + d_{ii})\delta N_I,\\
\frac{d\delta N_E}{dt} &= -(\mu_{is} + \mu_{ii} + d_{ii} + \beta)\delta N_E + (\mu_{ii} + d_{ii})\delta N_I,\\
\frac{d\delta N_I}{dt} &= (\mu_{ii} - \mu_{is} + \beta)\delta N_E - \mu_{ii}\delta N_I,
\end{aligned} \tag{38}$$

where $(\delta N_S, \delta N_E, \delta N_I)$ are deviations about (N_S^*, N_E^*, N_I^*). Eigenvalues of Eqs. (38) indicate instability whenever

$$\beta d_{ii} > \mu_{is}(d_{ii} + 2\mu_{ii}). \tag{39}$$

Fast association and transmission limit

Finally, within this general fast association limit, we can also explore the fast transmission limit. Setting both $a \to \bar{a}/\varepsilon and\ \beta \to \bar{\beta}/\varepsilon$ in Eqs. (7a)–(7e) and using the expansion given in Eqs. (30), we find, to $O(\varepsilon^{-1})$, $N_s^{(0)} = N_i^{(0)} = N_{si}^{(0)} = 0$. To $O(\varepsilon^{1/2})$, we find $N_{si}^{(1)} = 0$, while to order $O(1)$, Eqs. (7a)–(7e) become

$$0 = \Pi - 2\bar{a}_{ss}N_s^{(1)2} - \bar{a}_{si}N_s^{(1)}N_i^{(1)} + 2(\mu_{ss} + d_{ss})N_{ss}^{(0)}, \tag{40a}$$

$$0 = -2\bar{a}_{ii}N_i^{(1)2} - \bar{a}_{si}N_s^{(1)}N_i^{(1)} + 2(\mu_{ii} + d_{ii})N_{ii}^{(0)}, \tag{40b}$$

$$\frac{dN_{ss}^{(0)}}{dt} = -(2\mu_{ss} + d_{ss})N_{ss}^{(0)} + \bar{a}_{ss}N_s^{(1)2}, \tag{40c}$$

$$0 = -\bar{\beta}N_{si}^{(2)} + \bar{a}_{si}N_s^{(1)}N_i^{(1)}, \tag{40d}$$

$$\frac{dN_{ii}^{(0)}}{dt} = -(2\mu_{ii} + d_{ii})N_{ii}^{(0)} + \bar{\beta}N_{si}^{(2)} + \bar{a}_{ii}N_i^{(1)2}. \tag{40e}$$

Upon solving Eqs. (40a), (40b), and (40d), we find the same solution for $N_i^{(1)}$ and $N_s^{(1)}$ as given in Eqs. (32) as well as

$$N_{si}^{(2)} = \frac{\bar{a}_{si}}{\bar{\beta}}N_i^{(1)}N_s^{(1)}. \tag{41}$$

The terms P_s and P_i are given by Eqs. (35) except now $N_{si}^{(0)} = 0$, leading to

$$\begin{aligned}
P_s &= 2(\mu_{ss} + d_{ss})N_{ss}^{(0)} + \Pi = (\mu_{ss} + d_{ss})N_S^{(0)} + \Pi,\\
P_i &= 2(\mu_{ii} + d_{ii})N_{ii}^{(0)} = (\mu_{ii} + d_{ii})N_I^{(0)}.
\end{aligned} \tag{42}$$

Since $N_E^{(0)} \equiv N_{si}^{(0)} = 0$, substitution of Eqs. (42) into Eqs. (32) and (41) allows us to write Eqs. (40c) and (40e) as a closed system of equations for $N_I^{(0)} \approx 2N_{ii}^{(0)}$ and $N_S^{(0)} \approx 2N_{ss}^{(0)}$, respectively. The stability of the disease-free equilibrium at $(N_S^*, N_I^*) = (\frac{\Pi}{2\mu_{ss}}, 0)$ can be analyzed by using Eqs. (42) in Eqs. (32). Upon Taylor expanding Eqs. (40c) and (40e) about (N_S^*, N_I^*), we find

$$\begin{aligned}
\frac{d\delta N_S}{dt} &= -\mu_{ss}\delta N_S - (\mu_{ii} + d_{ii})\delta N_I,\\
\frac{d\delta N_I}{dt} &= d_{ii}\delta N_I.
\end{aligned} \tag{43}$$

In the fast association and transmission limit, the disease-free fixed point is always linearly unstable with growth rate d_{ii} representing the rate limiting dissociation step that allows further disease spread outside of doubly infected pairs.

IV. SUMMARY AND CONCLUSIONS

We have revisited the canonical mass-action susceptible-infected disease transmission models and systematically incorporated pairing dynamics. The purpose was to rigorously find uniformly valid effective equations from mass-action models with pairing and unpairing steps. After nondimensionalization of the five fundamental mass-action equations, we found parameter regimes that allow us to develop uniformly valid approximations to the total infected and susceptible populations. Our results were compared with lower-dimensional mass-action and frequency-dependent models *without* pairing.

First, in the fast transmission and pair dissociation limit, we found that the mass-action pairing model reduces to a standard mass-action susceptible-infected (SI) model [Eqs. (13)] without pairing, but with an effective disease transmission rate given by Eq. (14).

Next, when pair formation and breakup were assumed to be fast, we found effective equations for the total susceptible and infected populations. Although the two resulting ODEs can be unwieldy, this system differs fundamentally from the basic SI model. However, if death rates do not depend on the pairing status, we show that, in the low-density limit, the simple mass-action response is recovered [Eq. (25)]. In this low-density limit, the pairing dynamics do not affect the leading-order form of the functional response. However, in the high-density limit, a frequency-dependent response is recovered [Eq. (28)] if the association and dissociation rates are the same for each of the three different types of pairs. Under these assumptions, we showed that, for finite densities, a Holling type II response does not arise. Nevertheless, we derived a simple functional response that contains the same number of parameters as a model using Holling's type II response but with a clear mathematical justification.

Finally, we relaxed the fast dissociation constraint and assumed that only the association rates are large. In this case, we could reduce the five-dimensional system of mass-action equations only to a three-dimensional system that includes susceptible individuals, infected individuals, and an exposed population describing susceptible individuals in susceptible-infected pairs [Eqs. (34) or (37)]. These equations share features with the canonical SEI-type models [26,27,29].

Although the two- or three-dimensional system of equations we derived are typically more complicated in form, our formulas allow for straightforward incorporation of the effects of pair formation and dissociation in a self-consistent uniformly valid way in a number of limits. We have also numerically compared our solutions with those from the full five-dimensional mass-action system and found excellent agreement in the limits analyzed (Figs. 1 and 2).

Our asymptotic analysis can be straightforwardly extended to more complex disease models such as SIS- and SIR-type models that incorporate structured populations, incubation periods [30,31], and other processes such as birth and aging [32]. For structured models describing, e.g., population densities in age or time since infection τ, we can similarly assume fast association or dissociation rates proportional to $1/\varepsilon$ (which are now functions of τ) and expand the partial differential equations describing $N_{\rm ss}(t)$, $N_{\rm i}(\tau,t)$, $N_{\rm s}(\tau,t)$, $N_{\rm is}(\tau,t)$, and $N_{\rm ii}(\tau_1,\tau_2,t)$ in powers of ε. It would also be interesting to combine our asymptotic approaches with models of particle coagulation and fragmentation [33] to analyze disease dynamics occurring under group interactions or household structures [20,21,23,34,35]. For example, under epidemic conditions, one may consider contact subgroups [36] or distributions of pair contact durations [37], with household structure representing slow dissociation and public or casual contacts representing short-lived pairing [37]. Our reduced effective equations may also admit accurate closed-form analytic solutions previously derived for the standard SIR model [38]. Pair interactions could also be used to model certain properties of infections across networks [39,40] and spatial nodes [41], and our results may also provide insight into how to connect network models to effective ODE representations.

Our work may extend to other applications such as mass-action chemical reaction models in which an enzyme and substrate must first associate before a reaction can occur. A classic example in which related asymptotic analyses have been applied is Michaelis-Menten kinetics, in which an inner and outer solution are pieced together to describe substrate and product concentrations at short and long times [42]. In our problem, we have considered only the "outer" solutions, yet for all cases studied, our lowest-order approximations are valid at all times. Our analysis also provides possible avenues for a more detailed reexamination of bimolecular interactions in mass-action chemical kinetics in certain reaction rate limits.

ACKNOWLEDGMENTS

The authors thank X. Cheng for useful feedback on the manuscript. J.W. was supported by the Research Grants Council of Hong Kong Special Administrative Region through Project No. CityU11306115. T.C. was supported by the NSF through Grant No. DMS-1814364 and the Army Research Office through Grant No. W911NF-18-1-0345.

[1] J. D. Murray, *Mathematical Biology: I. An Introduction*, 3rd ed. (Springer, New York, 2002).

[2] R. M. Anderson, *Population Dynamics of Infectious Diseases: Theory and Applications* (Chapman and Hall, London, 1982).

[3] M. C. M. de Jong, O. Diekmann, and J. A. P. Heesterbeek, in *Epidemic Models: Their Structure and Relation to Data*, edited by D. Mollison (Cambridge University Press, Cambridge, 1995), pp. 84–94.

[4] H. McCallum, N. Barlow, and J. Hone, How should pathogen transmission be modelled? Trends Ecol. Evol. **16**, 295 (2001).

[5] S. M. Blower and T. Chou, Modeling the emergence of the "hot zones": Tuberculosis and the amplification dynamics of drug resistance, Nat. Med. **10**, 1111 (2004).

[6] J. O. Lloyd-Smith, W. M. Getz, and H. V. Westerhoff, Frequency-dependent incidence in sexually-transmitted disease models: Portrayal of pair-based transmission and effects of illness on contact behaviour, Proc. R. Soc. B **271**, 625 (2004).

[7] E. S. Devenish-Nelson, S. A. Richards, S. Harris, C. Soulsbury, and P. A. Stephens, Demonstrating frequency-dependent transmission of sarcoptic mange in red foxes, Biol. Lett. **10**, 20140524 (2014).

[8] P. C. Cross, T. G. Creech, M. R. Ebinger, K. Manlove, K. Irvine, J. Henningsen, J. Rogerson, B. M. Scurlock, and S. Creel, Female elk contacts are neither frequency nor density dependent, Ecology **94**, 2076 (2013).

[9] C. J. Rhodes and R. M. Anderson, Contact rate calculation for a basic epidemic model, Math. Biosci. **216**, 56 (2008).

[10] A. Fenton, J. P. Fairburn, R. Norman, and P. J. Hudson, Parasite transmission: Reconciling theory and reality, J. Anim. Ecol. **71**, 893 (2002).
[11] H. Hu, K. Nigmatulina, and P. Eckhoff, The scaling of contact rates with population density for the infectious disease models, Math. Biosci. **244**, 125 (2013).
[12] V. Křivan, T. E. Galanthay, and R. Cressman, Beyond replicator dynamics: From frequency to density dependent models of evolutionary games, J. Theor. Biol. **455**, 232 (2018).
[13] J. H. P. Dawes and M. O. Souza, A derivation of Holling's type I, II and III functional responses in predator-prey systems, J. Theor. Biol. **327**, 11 (2013).
[14] M. E. Hochberg, Non-linear transmission rates and the dynamics of infectious disease, J. Theor. Biol. **153**, 301 (1991).
[15] K. P. Hadeler, R. Waldstätter, and A. Wörz-Busekros, Models for pair formation in bisexual populations, J. Math. Biol. **26**, 635 (1988).
[16] J. A. P. Heesterbeek and J. A. J. Metz, The saturating contact rate in marriage and epidemic models, J. Math. Biol. **31**, 529 (1993).
[17] M. Kretzschmar and K. Dietz, The effect of pair formation and variable infectivity on the spread of an infection without recovery, Math. Biosci. **148**, 83 (1998).
[18] M. I. Chen and A. C. Ghani, Populations and partnerships: Insights from metapopulation and pair models into the epidemiology of gonorrhoea and other sexually transmitted infections, Sex. Transm. Infect. **86**, 433 (2010).
[19] M. Kretzschmar and J. C. M. Heijne, Pair formation models for sexually transmitted infections: A primer, Infect. Dis. Model. **2**, 368 (2017).
[20] G. Ghoshal, L. M. Sander, and I. M. Sokolov, SIS epidemics with household structure: The self-consistent field method, Math. Biosci. **190**, 71 (2004).
[21] P. J. Dodd and N. M. Ferguson, Approximate disease dynamics in household-structured populations, J. R. Soc. Interface **4**, 1103 (2007).
[22] C. Fraser, Estimating individual and household reproduction numbers in an emerging epidemic, PLoS One **2**, e758 (2007).
[23] T. House and M. J. Keeling, Deterministic epidemic models with explicit household structure, Math. Biosci. **213**, 29 (2008).
[24] G. Yan, Z.-Q. Fu, J. Ren, and W.-X. Wang, Collective synchronization induced by epidemic dynamics on complex networks with communities, Phys. Rev. E **75**, 016108 (2007).
[25] C. Llensa, D. Juher, and J. Saldaña, On the early epidemic dynamics for pairwise models, J. Theor. Biol. **352**, 71 (2014).
[26] M. Y. Li, J. R. Graef, L. Wang, and J. Karsai, Global dynamics of a SEIR model with varying total population size, Math. Biosci. **160**, 191 (1999).
[27] H. W. Hethcote, The mathematics of infectious diseases, SIAM Rev. **42**, 599 (2000).
[28] P. Yan and S. Liu, SEIR epidemic model with delay, ANZIAM J. **48**, 119 (2006).
[29] R. M. Anderson and R. M. May, *Infectious Diseases of Humans, Dynamics and Control* (Oxford University Press, Oxford, 1991).
[30] J. Qin, C. You, Q. Lin, T. Hu, S. Yu, and X.-H. Zhou, Estimation of incubation period distribution of COVID-19 using disease onset forward time: A novel cross-sectional and forward follow-up study, Sci. Adv. **6**, eabc1202 (2020).
[31] L. Böttcher, M. Xia, and T. Chou, Why case fatality ratios can be misleading: Individual- and population-based mortality estimates and factors influencing them, Phys. Biol. **17**, 065003 (2020).
[32] T. Chou and C. D. Greenman, A hierarchical kinetic theory of birth, death and fission in age-structured interacting populations, J. Stat. Phys. **164**, 49 (2016).
[33] M. R. D'Orsogna, Q. Lei, and T. Chou, First assembly times and equilibration in stochastic coagulation-fragmentation, J. Chem. Phys. **143**, 014112 (2015).
[34] D. E. Hiebeler and A. K. Criner, Partially mixed household epidemiological model with clustered resistant individuals, Phys. Rev. E **75**, 022901 (2007).
[35] M. J. Keeling, T. House, A. J. Cooper, and L. Pellis, Systematic Approximations to susceptible-infectious-susceptible dynamics on networks, PLoS Comput. Biol. **12**, e1005296 (2016).
[36] C. Bauch and D. A. Rand, A moment closure model for sexually transmitted disease transmission through a concurrent partnership network, Proc. R. Soc. B **267**, 2019 (2000).
[37] M. Xiridou, R. Geskus, J. De Wit, R. Coutinho, and M. Kretzschmar, The contribution of steady and casual partnerships to the incidence of HIV infection among homosexual men in Amsterdam, AIDS **17**, 1029 (2003).
[38] N. S. Barlow and S. J. Weinstein, Accurate closed-form solution of the SIR epidemic model, Physica D **408**, 132540 (2020).
[39] C.-R. Cai, Z.-X. Wu, and J.-Y. Guan, Effective degree Markov-chain approach for discrete-time epidemic processes on uncorrelated networks, Phys. Rev. E **90**, 052803 (2014).
[40] I. A. Perez, M. A. Di Muro, C. E. La Rocca, and L. A. Braunstein, Disease spreading with social distancing: A prevention strategy in disordered multiplex networks, Phys. Rev. E **102**, 022310 (2020).
[41] J. Joo and J. L. Lebowitz, Pair approximation of the stochastic susceptible-infected-recovered-susceptible epidemic model on the hypercubic lattice, Phys. Rev. E **70**, 036114 (2004).
[42] J. Keener and J. Sneyd, *Mathematical Physiology* (Springer, Berlin, 1998).