

Spatially-implicit modelling of disease-behaviour interactions in the context of non-pharmaceutical interventions

N. Ringa¹, C.T. Bauch²

¹ *Department of Mathematics and Statistical Sciences, Botswana International University of Science and Technology, Private Bag 16, Palapye, Botswana,* ² *Department of Applied Mathematics, University of Waterloo, 200 University Avenue West, Waterloo, ON N2L 3G1, Canada.*

Abstract

Pair approximation models have been used to study the spread of infectious diseases in spatially distributed host populations, and to explore disease control strategies such as vaccination and case isolation. Here we introduce a pair approximation model of individual uptake of non-pharmaceutical interventions (NPIs) for an acute self-limiting infection, where susceptible individuals can learn the NPIs either from other susceptible individuals who are already practicing NPIs (“social learning”), or their uptake of NPIs can be stimulated by being neighbours of an infectious person (“exposure learning”). NPIs include individual measures such as hand-washing and respiratory etiquette. Individuals can also drop the habit of using NPIs at a certain rate. We derive a spatially defined expression of the basic reproduction number R_0 and we also numerically simulate the model equations. We find that exposure learning is generally more efficient than social learning, since exposure learning generates NPI uptake in the individuals at immediate risk of infection. However, if social learning is pre-emptive, beginning a sufficient amount of time before the epidemic, then it can be more effective than exposure learning. Interestingly, varying the initial number of individuals practicing NPIs does not significantly impact the epidemic final size. Also, if initial source infections are surrounded by protective individuals, there are parameter regimes where increasing the initial number of source infections actually decreases the infection peak (instead of increasing it) and makes it occur sooner. The peak prevalence increases with the rate at which individuals drop the habit of using NPIs, but the response of peak prevalence to changes in the forgetting rate are qualitatively different for the two forms of learning. The pair approximation methodology developed here illustrates how analytical approaches for studying interactions between social processes and disease dynamics in a spatially structured population should be further pursued.

Keywords: Pair approximation, network model, social distancing, non-pharmaceutical interventions, transmission model

1. Introduction

Mathematical models in epidemiology often make the assumption that successful control of

epidemics is only determined by the availability and effective deployment of control measures such as vaccination and isolation, whose success largely depends on factors such as quantity

of vaccine and logistical constraints. In recent years, some mathematical models are focusing on endogenously incorporating the impact of human behavioral patterns on the regulation of communicable diseases. Upon gaining awareness about an infectious disease outbreak, susceptible individuals may decrease their infection risk by isolating themselves or reducing interactions with their friends, colleagues, etc, through staying at home and avoiding social contacts. This practice is known as social distancing [16, 22]. Along with social distancing, hand-washing, use of masks, and other respiratory etiquette are further examples of so-called ‘non-pharmaceutical interventions’ (NPIs) that can reduce infection spread [31]. While health-care providers often advise the public on appropriate NPIs, behavioral choices of individual members of the host population partially determine the dynamics and feasibility to control an infectious disease outbreak.

Social distancing and other NPIs have assisted the control of infections such as flu, severe acute respiratory syndrome (SARS) and plague [6, 12, 40, 23, 14, 36, 15, 38, 21, 2]. However despite the availability of a large amount of information about the dangers and risks of sexually transmitted infections (STIs), actions like unsafe sexual behavior and needle sharing during intravenous drug use have been linked to the pandemic-scale dissemination of STIs such as HIV/AIDS [26, 24, 39, 25]. Generally, negligence or relaxation of precautionary measures is brought about by factors such as lack of awareness and engaging in infection-enhancing social practices such as handshakes, hug, kisses, sharing of food and concurrent sexual partner, as well as some cultural practices. The 2014 epidemic outbreak of Ebola in West Africa is an example of how cultural or religious practices, such as engaging in risky rituals and inappropriate handling of the sick or deceased, also influence the dynamics of infectious dis-

eases [28].

Human behavior also plays a role in the regulation of some animal infectious diseases. For instance, the use of dogs in hunting and grazing cattle in countries such as Kenya and Botswana, influences transmission of canine diseases between domestic dogs and the African wild dog [1]. Although culling (slaughtering of infectious or at-risk animals) has been found to effectively control foot and mouth disease, farmers’ resistance towards this intervention measure (because of fear of loss of livestock) often makes it difficult to bring the disease under control.

In [12] the authors explore the impact of social distancing on the spread of an infection by incorporating health status-based contact behavior patterns into a mean-field equations epidemic model. Thus, the transmission dynamics are governed by differing contact levels between individuals of different health types. For example, due to the perceived risk of infection, susceptible individuals are likely to avoid contact with infected individuals, while maintenance contact with recovered individuals may have a less significant impact.

In [15] the authors explore the idea that the adoption of social distancing or other NPIs is driven by the level of information individuals have, such that members of the host population who possess first hand information become more cautious and therefore less susceptible than those who have second hand information. Similarly, individuals who have second hand information are less susceptible than those who possess third hand information, etc. This study was carried out by modeling information transmission and spread of an infection using mean-field equations and individual-based epidemic models. The research also discusses the significance of repeated re-generation of awareness into the population to ensure that most

individuals have access to primary, or close-to-primary, information, which increases the number of individuals who exercise contact precautions and/or NPIs.

In [14] the authors capture the dynamic nature of individuals’ decisions leading to adoption or non-practice of social distancing, by assuming that the network geometry within which the host population resides (particularly the neighborhood size) changes over time, depending on individuals’ perceived risk of infection. Thus, adjustment of individuals’ perceptions about the disease over time results in variation of contact patterns and, therefore, it affects the infection dynamics. Other researchers have explored game-theoretical [33] or rule-based simulation models [35, 37] of social distancing.

The spatial dimension of social distancing has been explored in some of this previous work [15, 14, 35, 37]. Spatial dynamics can be analytically intractable, hence the frequent decision to employ agent-based models. However, one method for implicitly capturing spatial dynamics that often permits analysis is moment closure approximation (MCA). MCAs employ pairs, triples, quadruples, etc., of connected individuals, as model state variables, such that transmission takes place only between connected susceptible and infectious individuals on the network. MCAs are usually comprised by a system of differential equations, where each equation describes time evolution of second order, third order, fourth order, etc., spatial correlations between individual members of the host population. Equations of motion for pairs involve terms in triples, equations of motion for triples involves terms in quadruples, etc. Therefore in order to obtain a closed system of equations, this hierarchy is truncated by techniques referred to as moment closures. Carrying out the closure at the level

of pairs produces a pair approximation model [34, 8, 7, 5, 29, 30, 32, 13, 9, 19, 11, 17].

Here our objective is to demonstrate how pair approximations and analytical expressions for the basic reproduction number can be developed for spatially-structured socio-epidemiological systems. We develop and analyze a pair approximation model and explore the impacts of NPIs on the spread of an infectious disease. We incorporate impacts of NPIs by dividing the susceptible population into susceptible individuals who protect (S_p) (i.e., those who practice NPIs) and individuals who do not (S). State S individuals learn from state S_p and/or infectious (state I) individuals in their network neighborhood, and then decide whether or not to adopt NPIs. The dichotomy between adopting NPIs due to being next to an infectious person, versus learning NPIs from other individuals who adopt NPIs, captures the distinction between practicing NPIs reactively because of an immediate threat due to an infectious neighbour versus changing one’s habits pre-emptively based on observing the actions of other individuals, and forming new habits (such as using hand sanitizers, or using a paper towel to open a bathroom door). We refer to learning NPIs from other state S_p contacts as “social learning”, versus “exposure learning” that occurs from reacting to infection in an immediate network neighbour. Because social learning cannot occur without individuals who have first adopted NPIs through exposure learning in the early stages of the outbreak, exposure learning can be viewed as a first order effect, while social learning can be viewed as a second order effect. We also derive the basic reproduction number and analyze the pair approximation equations to understand how control success depends on epidemiological and behavioral change parameters. The model is described in the following subsection.

2. Model

A state S individual who is neighbouring a state S_p individual transitions to the S_p state at a per capita rate ξ . Similarly, a state S individual neighbouring a state I individual transitions to the S_p state at a per capita rate ρ . These interactions thereby result in susceptible individuals adopting NPIs. Switching from state S_p back to state S occurs at a per capita rate κ , representing forgetting, or complacency.

The rate of infection transmission from an infectious individual to a neighbouring state S_p individual is τ_p , whereas transmission to a neighbouring state S individual occurs at a rate $\tau > \tau_p$. Infected individuals recover at

per capita rate σ . Thus, the state variables of the pair approximation model are numbers of susceptible, protective, infectious and recovered individuals denoted by $[S]$, $[S_p]$, $[I]$ and $[R]$, respectively, and numbers of paired individuals, $[XY]$ where, for instance $[S_p I]$ represents the number of edges comprising of susceptible protective and infectious individuals. We derive equations of motion for our model in Appendix A. We assume that the disease spreads on a regular network in which all nodes have the same degree n , in a population of size N , and we use the binomial ordinary pair approximation (Equation (A.2)), to approximate triples in terms of pairs and singletons, resulting in:

$$\begin{aligned}
\frac{d[S]}{dt} &= -\tau[SI] - \xi[S_p S] + \kappa[S_p] - \rho[SI] \\
\frac{d[S_p]}{dt} &= -\tau_p[S_p I] + \xi[S_p S] - \kappa[S_p] + \rho[SI] \\
\frac{d[I]}{dt} &= \tau[SI] + \tau_p[S_p I] - \sigma[I] \\
\frac{d[R]}{dt} &= \sigma[I] \\
\frac{d[SS]}{dt} &= -2\frac{n-1}{n}\frac{[SS]}{[S]}\left((\tau+\rho)[SI] + \xi[S_p S]\right) + 2\kappa[SS_p] \\
\frac{d[SS_p]}{dt} &= -\frac{n-1}{n}\left((\tau+\rho)\frac{[SI][SS_p]}{[S]} + \tau_p\frac{[SS_p][S_p I]}{[S_p]} - \xi\frac{[SS_p][SS]}{[S]} - \rho\frac{[SI][SS]}{[S]}\right) - \xi[SS_p] \\
&\quad + \kappa([S_p S_p] - [SS_p]) \\
\frac{d[S_p S_p]}{dt} &= -2\frac{n-1}{n}\left(\tau_p\frac{[S_p I][S_p S_p]}{[S_p]} - \rho\frac{[SI][SS_p]}{[S]}\right) + 2\xi[SS_p] - 2\kappa[S_p S_p] \\
\frac{d[SI]}{dt} &= \frac{n-1}{n}\left(\frac{[SI]}{[S]}\left(\tau[SS] - (\tau+\rho)[SI] - \xi[SS_p]\right) + \tau_p\frac{[SS_p][S_p I]}{[S_p]}\right) - (\tau+\sigma+\rho)[SI] + \kappa[S_p I] \\
\frac{d[S_p I]}{dt} &= \frac{n-1}{n}\left(\frac{[SI]}{[S]}\left((\tau+\xi)[SS_p] + \rho[SI]\right) + \frac{[S_p I]}{[S_p]}\tau_p\left([S_p S_p] - [S_p I]\right)\right) - (\tau_p+\sigma+\kappa)[S_p I] \\
&\quad + \rho[SI] \\
\frac{d[SR]}{dt} &= -\frac{n-1}{n}\frac{[SR]}{[S]}\left((\tau+\rho)[SI] + \xi[SS_p]\right) + \sigma[SI] + \kappa[S_p R] \\
\frac{d[S_p R]}{dt} &= -\frac{n-1}{n}\left(\tau_p\frac{[S_p I][S_p R]}{[S_p]} - \frac{[SR]}{[S]}(\rho[SI] + \xi[S_p S])\right) + \sigma[S_p I] - \kappa[S_p R] \\
\frac{d[II]}{dt} &= 2\frac{n-1}{n}\left(\tau\frac{[SI]^2}{[S]} + \tau_p\frac{[S_p I]^2}{[S_p]}\right) + 2\tau[SI] + 2\tau_p[S_p I] - 2\sigma[II] \\
\frac{d[IR]}{dt} &= \frac{n-1}{n}\left(\tau\frac{[SI][SR]}{[S]} + \tau_p\frac{[S_p I][S_p R]}{[S_p]}\right) + \sigma([II] - [IR]) \\
\frac{d[RR]}{dt} &= 2\sigma[IR].
\end{aligned} \tag{1}$$

2.1. The basic reproduction number

The basic reproduction number R_0 is the expected number of secondary infection cases produced by a single infectious individual upon introduction into a wholly susceptible population [10, 5, 3, 20, 27]. An epidemic may occur if $R_0 > 1$, but the infection will die out if $R_0 \leq 1$.

Therefore, effective disease control reduces R_0 below 1.

Here we use the pair approximation model above to derive an expression for R_0 that incorporates some effects of spatiality and allows us to study the impact of adoption of NPIs on the dynamics of the disease at the initial stage of an outbreak. For simplicity, we derive R_0

for a scenario where individuals start to learn and practice contact precautions during an outbreak ($S_p(0) \ll N$).

The condition under which the infection will spread is

$$\frac{d[I]}{dt} > 0 \Rightarrow \tau[SI] + \tau_p[S_p I] - \sigma[I] > 0, \quad (2)$$

which can be rearranged to yield

$$\frac{\tau[SI] + \tau_p[S_p I]}{\sigma[I]} > 1. \quad (3)$$

Therefore, we write R_0 in terms of (i) susceptibility of individuals who do not protect (captured by a high transmission parameter τ), (ii) susceptibility of individuals who practice NPIs (captured by a low transmission rate τ_p), (iii) the rate of recovery σ , and (iv) the numbers of S - I and S_p - I pairs as well as the overall number of infectious individuals $[I]$:

$$R_0 = \frac{\tau[SI]}{\sigma[I]} + \frac{\tau_p[S_p I]}{\sigma[I]}. \quad (4)$$

Next we express pairs $[SI]$ and $[S_p I]$ in terms of the correlations between state S individuals and their infectious neighbors, and state S_p individuals and their infectious contacts, respectively. The correlation between individuals with status X and Y is given by

$$C_{XY} = \frac{N}{n} \frac{[XY]}{[X][Y]}, \quad (5)$$

where n and N are the number of contacts each individual has and the total population size, respectively. $C_{XY} < 1$ implies avoidance of interaction between state X and state Y individuals, $C_{XY} = 1$ assumes homogeneous mixing, while $C_{XY} > 1$ implies strong correlation between state X and state Y individuals. Note that Equation (5) can be re-written as

$$[XY] = \frac{[n]}{N} [X][Y] C_{XY}, \quad (6)$$

therefore,

$$R_0 = \frac{n}{\sigma N} (\tau[S]C_{SI} + \tau_p[S_p]C_{S_p I}). \quad (7)$$

At the initial stage of an epidemic, we assume that the population is comprised mainly by susceptible individuals, only a few of whom practice NPIs:

$$[S] + [S_p] \approx N, \text{ where } [S_p] \ll [S] \quad (8)$$

We define $s_p \equiv [S_p]/N$, thus the proportion of state S individuals at the beginning of the epidemic is $1 - s_p$, and R_0 becomes:

$$R_0 = \frac{n}{\sigma} \left(\tau(1 - s_p)C_{SI} + \tau_p s_p C_{S_p I} \right). \quad (9)$$

To estimate the values for C_{SI} and $C_{S_p I}$, we use biologically reasonable assumptions about disease spread, as follows. There is a very small number of infectious individuals at the beginning of an outbreak, so $C_{SI} \approx 1$. However as the infection spreads, C_{SI} decreases and the clustering of infected individuals leads to a decrease in the rate of spread, and the disease may die out if there are not enough susceptible individuals in the vicinity of the infected cluster to transmit the disease to. The dynamics of the disease at this point (referred to as the local minimum and denoted by C_{SI}^{min}) determine whether an epidemic will succeed or fail to take off. Thus, we need to evaluate C_{SI}^{min} [18]. Similar reasoning applies to $C_{S_p I}^{min}$. Hence

$$R_0 = \frac{n}{\sigma} \left(\tau(1 - s_p)C_{SI}^{min} + \tau_p s_p C_{S_p I}^{min} \right). \quad (10)$$

The quantities C_{SI}^{min} and $C_{S_p I}^{min}$ are the solutions of $\frac{d}{dt}C_{SI} = 0$ and $\frac{d}{dt}C_{S_p I} = 0$, respectively.

The derivation of these quantities as well as the full expression of R_0 are presented in Appendix B. The full expression of R_0 , Equation

(B.3), is unwieldy but it depends on epidemiological parameters τ, τ_p and σ , NPIs-based parameters ξ, ρ and κ , as well as initial network configuration-dependent correlations $C_{S_p S}$ and $C_{S_p S_p}$. It is clear from Equation (B.3) that a higher rate of recovery σ , reduces R_0 , but the equation is too complicated to directly infer the impacts of other model components. Hence, numerical computations will be used to explore dependence of R_0 on model parameters and initial network configurations (Section 3).

We derive reduced versions of R_0 by considering special cases where individuals adopt NPIs through (a) social learning only (i.e. $\xi > 0 \text{ day}^{-1}$ and $\rho = 0 \text{ day}^{-1}$) and (b) exposure learning only (i.e. $\xi = 0 \text{ day}^{-1}$ and $\rho > 0 \text{ day}^{-1}$). For both of these scenarios we assumed that protective individuals consistently practice NPIs throughout the outbreak so that state S_p individuals do not switch back to state S (i.e. $\kappa = 0 \text{ day}^{-1}$). Furthermore, we assumed that the initial network configuration constitutes one infectious (state I) individual with one protective (state S_p) neighbor who also has one state S_p contact, and the rest of the population is completely susceptible (i.e. state S) such that at the initial stage of the outbreak $[S] \approx N$, where the population size N is very large.

2.1.1. High efficacy NPIs adopted through social learning only

In Appendix B we simplified Equation (B.3) to derive the expression for the basic reproduction number for dynamics in which adoption of NPIs results from social learning only (case (a) above), and that NPIs are highly effective as a control measure ($\tau_p \ll \tau$), and for typical model parameters ($N = 40000$, $s_p = 2/N$, $n = 4$, $\tau = 1 \text{ day}^{-1}$, $\tau_p = 0.0025 \text{ day}^{-1}$,

$\sigma = 0.25 \text{ day}^{-1}$ and $\xi = 0.25 \text{ day}^{-1}$), to obtain

$$R_0 \approx \frac{\tau n \chi - \xi + \sqrt{\tau^2 n^2 \chi^2 + \xi(\xi + 2\tau n \chi)}}{2\sigma}, \quad (11)$$

where

$$\chi \approx \frac{\tau(n-2) + \sqrt{\tau^2(n-2)^2 + 4\tau\tau_p(n-1)}}{2\tau n}.$$

Equation (11) confirms that social learning (ξ) reduces the initial spread of the infection. As expected, a highly transmissible infection (large τ) will increase R_0 . We note that R_0 increases with χ . We discuss in Appendix B that χ approximates the minimum of the correlation function C_{SI}^{min} between susceptible and infectious individuals. Therefore, factors that increase χ should also increase R_0 . The quantity χ increases with the transmission rate to protective susceptible neighbours (τ_p), as well as the number of neighbours per individual (n), confirming the mitigating effects of spatially localized transmission.

2.1.2. High efficacy NPIs adopted through exposure learning only

For comparison to Equation (11), in Appendix B we also simplified Equation (B.3) to derive the expression for the basic reproduction number for dynamics in which adoption of NPIs results from exposure learning only (case (b) above), and that NPIs are highly effective as a control measure ($\tau_p \ll \tau$), and for typical model parameters ($N = 40000$, $s_p = 2/N$, $n = 4$, $\tau = 1 \text{ day}^{-1}$, $\tau_p = 0.0025 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$ and $\xi = 0.25 \text{ day}^{-1}$), to obtain

$$R_0 \approx \frac{\tau n \chi + \sqrt{\left(\tau^2 n^2 \chi + 4\tau_p \rho n\right) \chi}}{2\sigma}, \quad (12)$$

where

$$\chi \approx \frac{\tau(n-2) - \rho + \sqrt{\tau^2(n-2)^2 - 2\tau\rho(n-2) + \rho^2 + 4\tau\tau_p(n-1)}}{2\tau n}.$$

Hence, more rapid adoption of NPIs due exposure learning (ρ) decreases χ , and therefore also decreases R_0 . Spatial structure has a mitigating effect in this case as well.

In Appendix B we show that when adopted NPIs are not strict (such that $\tau_p \ll \tau$ does not hold), then the corresponding expressions of R_0 for cases where individuals practice cautious behavior due to social learning only and exposure learning only are given by Equations (B.4) and (B.7), respectively.

3. Results

3.1. Dependence of R_0 on model parameters and network configuration

The spatial distribution of susceptible individuals (S) and individuals who practice NPIs (S_p) around the infection source at the initial stage of an epidemic have a strong influence on R_0 , as computed from the full expression appearing in Appendix B, Equation (B.3) (Figure 1).

As expected, R_0 decreases with the proportion of protective individuals around the infection source cluster and in the entire population (Figures 1a versus 1c). This finding implies that upon inception of an outbreak it is crucial to identify the infection source promptly, and sensitize members of the host population about the disease and prevention measures, so that they can propagate awareness further through interactions with their spatial neighbours. We note that associating R_0 (and the overall disease dynamics) with specific network configurations in which diseases and social interactions disseminate would not be possible under

a mean-field equations approach, which ignores spatial structure altogether. Spatial processes such as transmission of most infectious diseases and dissemination of information can be understood better by the study of spatially oriented models such as pair approximation models.

Increasing the rate of adopting NPIs, either through exposure learning (ρ) or through social learning (ξ), decreases R_0 . However, increasing the rate of adoption via infectious neighbours reduces R_0 more effectively than increasing the rate of adoption via protective susceptible neighbours (Figure 2a, b). This occurs because being next to infectious neighbors results in prompt adoption of preventative measures and avoidance of infection from that infectious neighbor, and hence it reduces R_0 better than a scenario where adoption of NPIs results from interaction with protective neighbors only, which may leave some infectious individuals in a part of the network with no protective individuals, while other parts of the network may have significant populations of protective individuals, but no infections. However, we note that this only applies when social learning can only begin at the start of an outbreak, and not beforehand as a pre-emptive measure.

3.2. Numerical analysis of pair approximation differential equations

3.2.1. Social and exposure learning during an outbreak

Numerical analysis of our model was carried out in *MATLAB* using the ode45 solver. Similar to the results from the R_0 derivation, numerical simulation of Equation (1) shows that if both forms of learning can begin only during

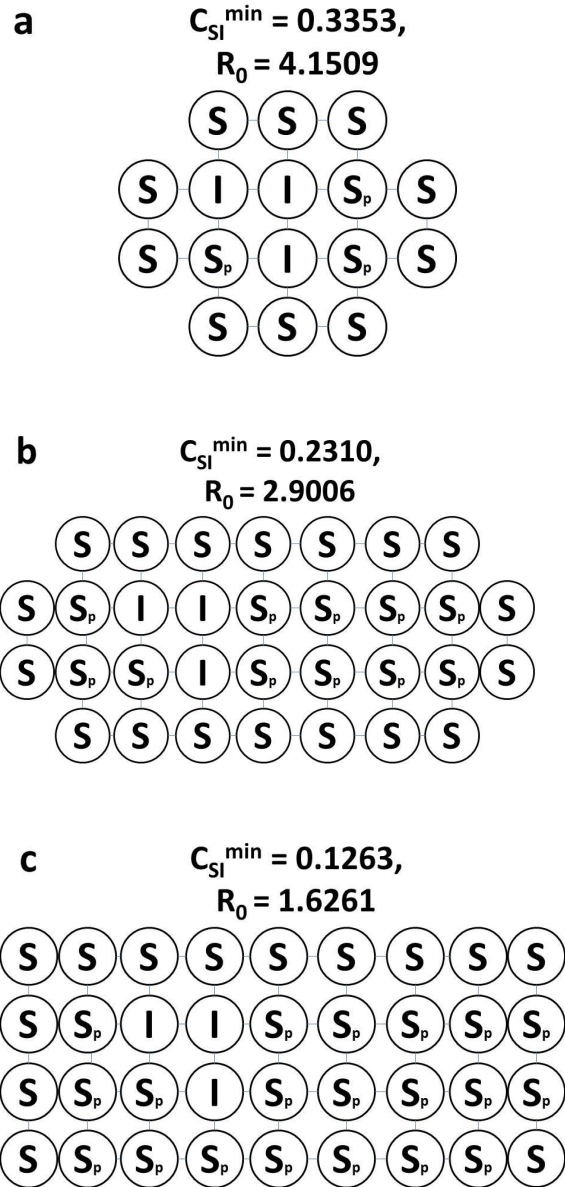


Figure 1: Typical network distributions of susceptible contacts, S , neighbors who practice social distancing techniques, S_p (as well as the respective calculations of the basic reproduction number) around the initial infection source, where all other members of the host population are fully susceptible (i.e. state S). The population size is $N = 40000$, each individual has $n = 4$ neighbors and model parameters are $\tau = 0.75 \text{ day}^{-1}$, $\tau_p = 0.1 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$, $\xi = \rho = 0.5 \text{ day}^{-1}$ and $\kappa = 0.01 \text{ day}^{-1}$.

an outbreak, then NPIs adopted due to exposure learning (ρ) have a much larger impact

on the size of the epidemic peak, than NPIs adopted due to social learning (ξ) (Figure 3i,l

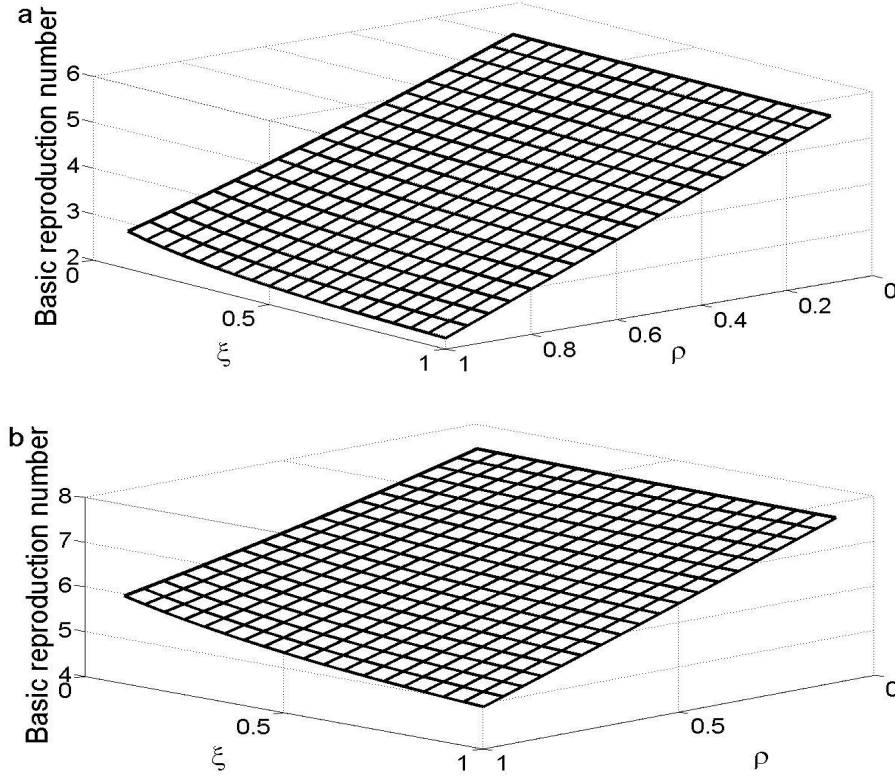


Figure 2: The basic reproduction number as a function of social learning from protective contacts at a rate ξ , and from infectious contacts at a rate ρ , where the transmission rate to protective individuals is $\tau_p = 0.1 \text{ day}^{-1}$ (a) and $\tau_p = 0.5 \text{ day}^{-1}$ (b). In all these plots $N = 40000$, $n = 4$, $\tau = 0.75 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$, $C_{S_p S_p} = 0$, $C_{S_p S} = 3/4$, $\kappa = 0 \text{ day}^{-1}$ and $s_p = 1/N$.

versus 3c,f). We also find that, adoption of NPIs stimulated by neighbouring infectious individuals leads to a higher number of protective individuals throughout and at the end of the epidemic, than practice of NPIs due to social learning (Figures 3h,k versus Figures 3b,e). This occurs because in this situation, learning from a neighbouring protective susceptible contact may not reach the parts of the network that need to be protected, resulting in those parts of the network being infected before they can adopt NPIs.

Increasing the initial number of infection source points with n protective neighbours (i.e. completely surrounded by individuals who

practice NPIs) generally increases the infection peak (Figures 3a,g,j). This confirms intuition. However, contrary to phenomena normally observed in homogeneous-mixing models, increasing the number of infection source points who are surrounded by highly protective individuals can actually decrease the infection peak in other parameter regimes ($\tau_p = 0.1 \text{ day}^{-1}$, see Figures 3d and 3f). On the other hand, when NPIs are not strictly practiced (leading to a relatively high value of τ_p), increasing the number of infection source points will increase the infection peak, as usual, even when each of the initial infection sources are surrounded by a large proportion of protective contacts (see Supple-

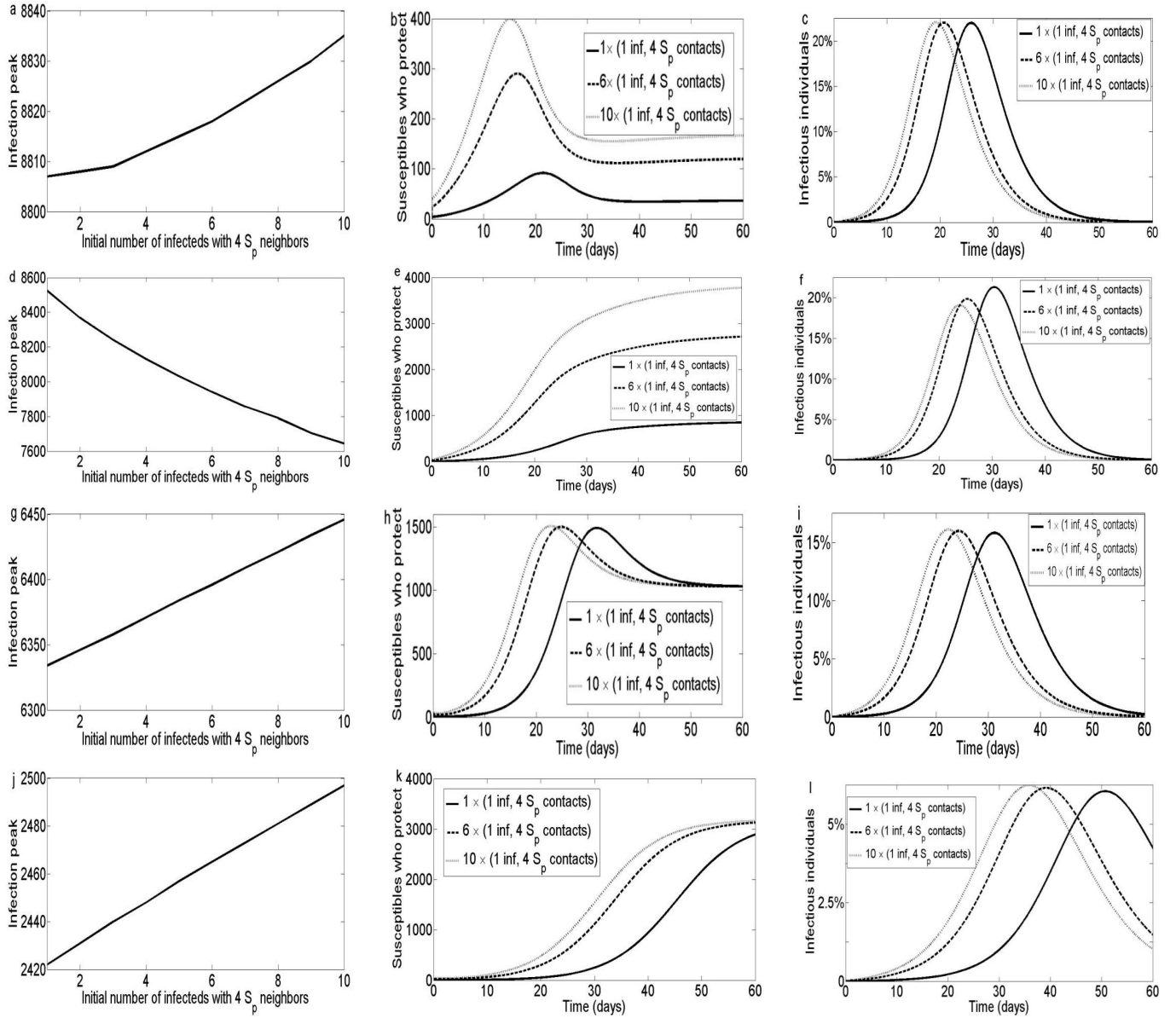


Figure 3: Infection peak versus initial distribution of *single infected individuals with 4 state S_p neighbors* (a, d, g, j), time series for susceptible individuals who protect (b, e, h, k) and time series for infectious individuals (c, f, i, l), varying the number of *1 infected node plus 4 S_p neighbors* at the beginning of the outbreak (the rest of the population is fully susceptible). In (a to f) $\xi = 0.25 \text{ day}^{-1}$, $\rho = 0 \text{ day}^{-1}$; in (g to l) $\xi = 0 \text{ day}^{-1}$, $\rho = 0.25 \text{ day}^{-1}$; in (a, b, c and g, h, i) $\tau_p = 0.6 \text{ day}^{-1}$; in (d, e, f and j, k, l) $\tau_p = 0.1 \text{ day}^{-1}$. Model parameters common to all graphs are $\tau = 0.8 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$ and $\kappa = 0 \text{ day}^{-1}$.

mentary Material, Figure C.1a).

Cumulative infections over a period of two

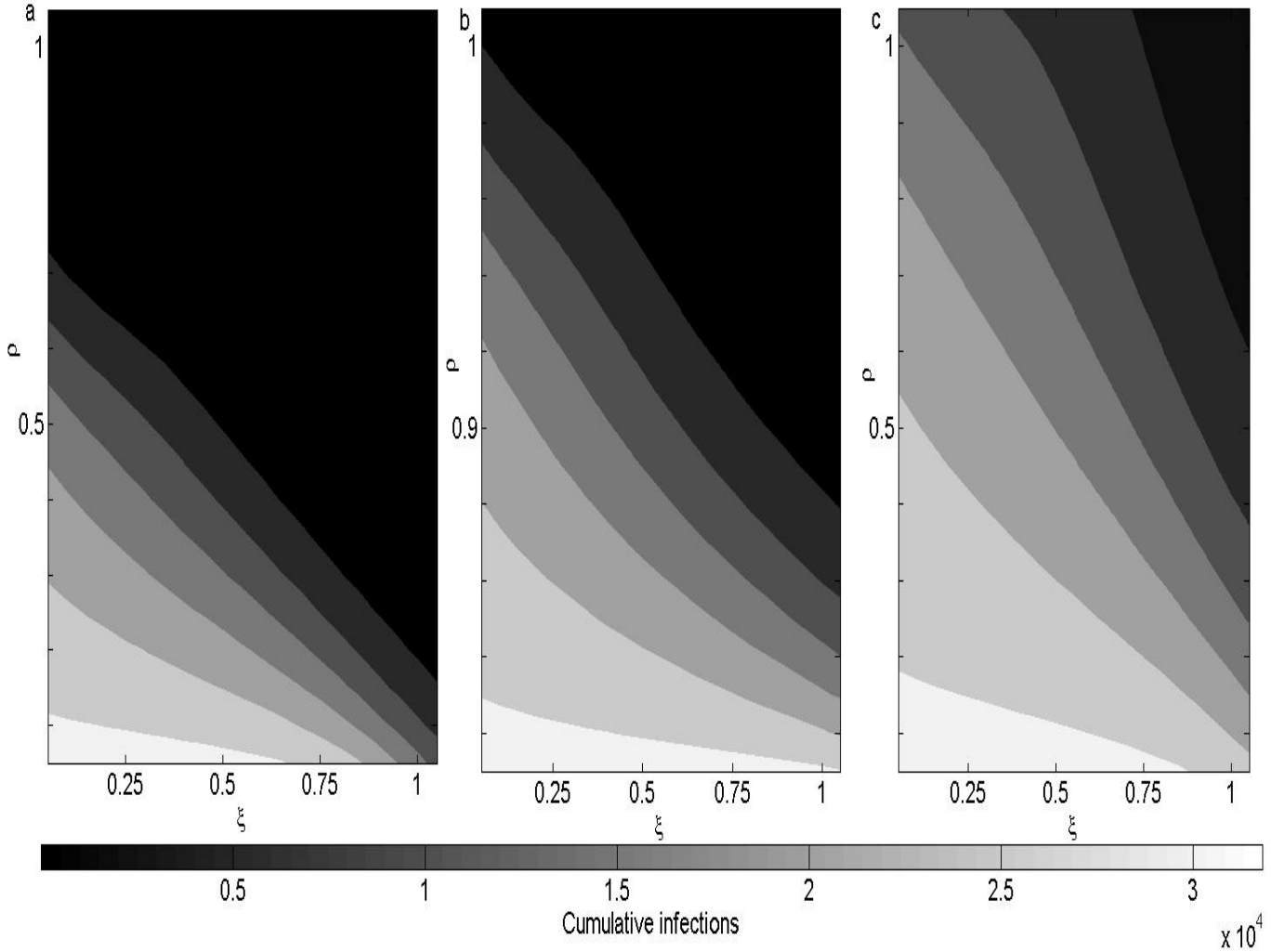


Figure 4: Cumulative infections as a function of social learning from both infectious and state S_p neighbors at rates ρ and ξ , respectively, where the initial conditions are 1 infected node and 1 state S_p neighbor while the rest of the population is fully susceptible (i.e. state S), and $\tau_p = 0.1 \text{ day}^{-1}$ (a), $\tau_p = 0.2 \text{ day}^{-1}$ (b), $\tau_p = 0.3 \text{ day}^{-1}$ (c). Other model parameters are $\tau = 0.8 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$ and $\kappa = 0 \text{ day}^{-1}$.

months decrease with adoption of precautionary behavior due to social learning at a rate ξ , and exposure learning at a rate ρ , but, also as observed in Figures 2 and 3, the decrease in cumulative infections is more profound when individuals use exposure learning than social learning (Figure 4). Furthermore, adoption of NPIs only moderately decreases cumulative in-

fections if upon learning about the disease and becoming protective, the new state S_p individuals practice less effective precautionary measures, resulting in an increased rate of transmission to state S_p individuals, τ_p (Figure 4a versus Figure 4b versus Figure 4c). The asymmetry between effects of ξ and ρ , and the impact of complacency, on cumulative infections,

shown in Figure 4, are even more obvious in the corresponding surface plot of the same model outcomes (see Supplementary Material, Figure C.2).

If individuals who practice NPIs lose this habit (captured by conversion from state S_p to state S at a ‘forgetting’ rate κ) then the population susceptibility increases, leading to a large number of infection cases and occurrence of large epidemic outbreaks, which are characterized by high infection peaks (Figure 5). However, with respect to the governing model equations used in this paper, the response of infection peaks to changes in the rates of forgetting is qualitatively different for the two types of learning: κ scales linearly or logistically with infection peak, depending on whether only ρ or only ξ , respectively, is active. This stems from the fact that per capita success of exposure learning—which operates only in immediate neighbours of an infected node—depends less on population prevalence than social learning. In particular, when the rate of forgetting is low, then the infection peak is roughly the same for both types of learning. However, when the rate of forgetting is high, then infection peaks are very high under social learning, but only moderately high under exposure learning. This occurs because effective social learning requires large parts of the network to be ‘ready’ for any infections which may enter the area by having high and stable populations of protective susceptible individuals, and large rates of forgetting prevent this from happening.

3.2.2. *Social learning before and during an outbreak, and exposure learning during an outbreak*

Many of the behaviours that fall under the rubric of NPIs, such as hand-washing and respiratory etiquette, are learned preventively and are practiced in a population even before an

epidemic. This builds up the proportion of protective individuals before introduction of the disease. Thus, the effectiveness of social learning may thus be considerably improved, although it is not clear how far in advance social learning must begin for it to be useful. In this subsection we consider scenarios where social learning can occur both prior to and after the introduction of an infection. In particular, we contrast a scenario where only social learning is practiced (but social learning begins to spread before the epidemic starts), to a scenario where only exposure learning takes place, and we compare their performance.

In the absence of exposure learning ($\rho = 0$), model simulations confirm how social learning before an epidemic creates large pools of protective individuals before the epidemic begins, leading to a decreased epidemic final size (Figure 6). At baseline parameter values, introducing social learning as early as possible is a highly effective way of decreasing the epidemic final size; increasing the initial number of protective individuals, S_p , also works but is less effective than stimulating social learning as early as possible (Figure 6a-c). Either of these measures is made more effective when social learning is more rapid (large ξ , Figure 6c versus b and a).

In contrast to observations made in most of the simulations in the previous subsections, social learning reduces the epidemic final size more effectively than exposure learning, except when social learning is not introduced soon enough before the epidemic, or when there are not enough initial protective individuals (Figure 6a-c). In either of these two exceptional cases, there is an insufficient pool of protective susceptible individuals in the population at the beginning of the epidemic, for social learning to be effective.

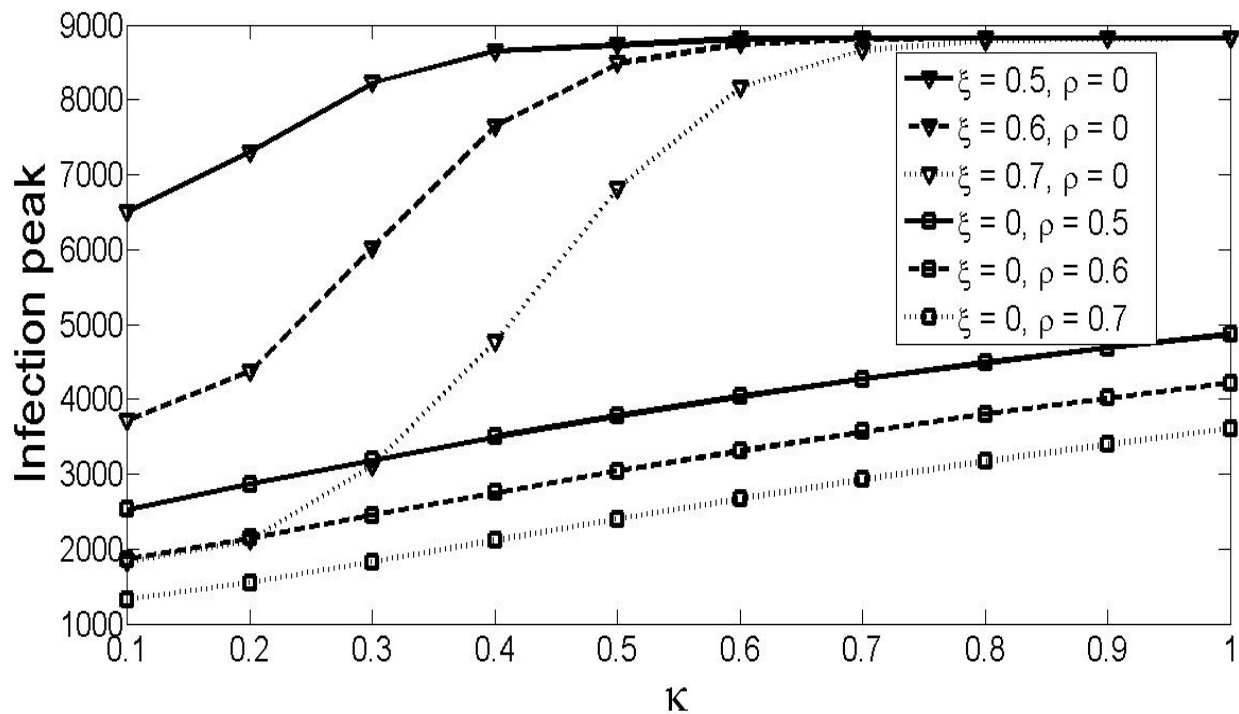


Figure 5: Infection peak versus the rate at which protective susceptible individuals forget, κ , varying regimes for social contagion parameters ξ and ρ . Initial conditions are 1 infected node and 2 state S_p neighbors while the rest of the population is fully susceptible (i.e state S). Other model parameters are $\tau = 0.8 \text{ day}^{-1}$, $\tau_p = 0.3 \text{ day}^{-1}$ and $\sigma = 0.25 \text{ day}^{-1}$.

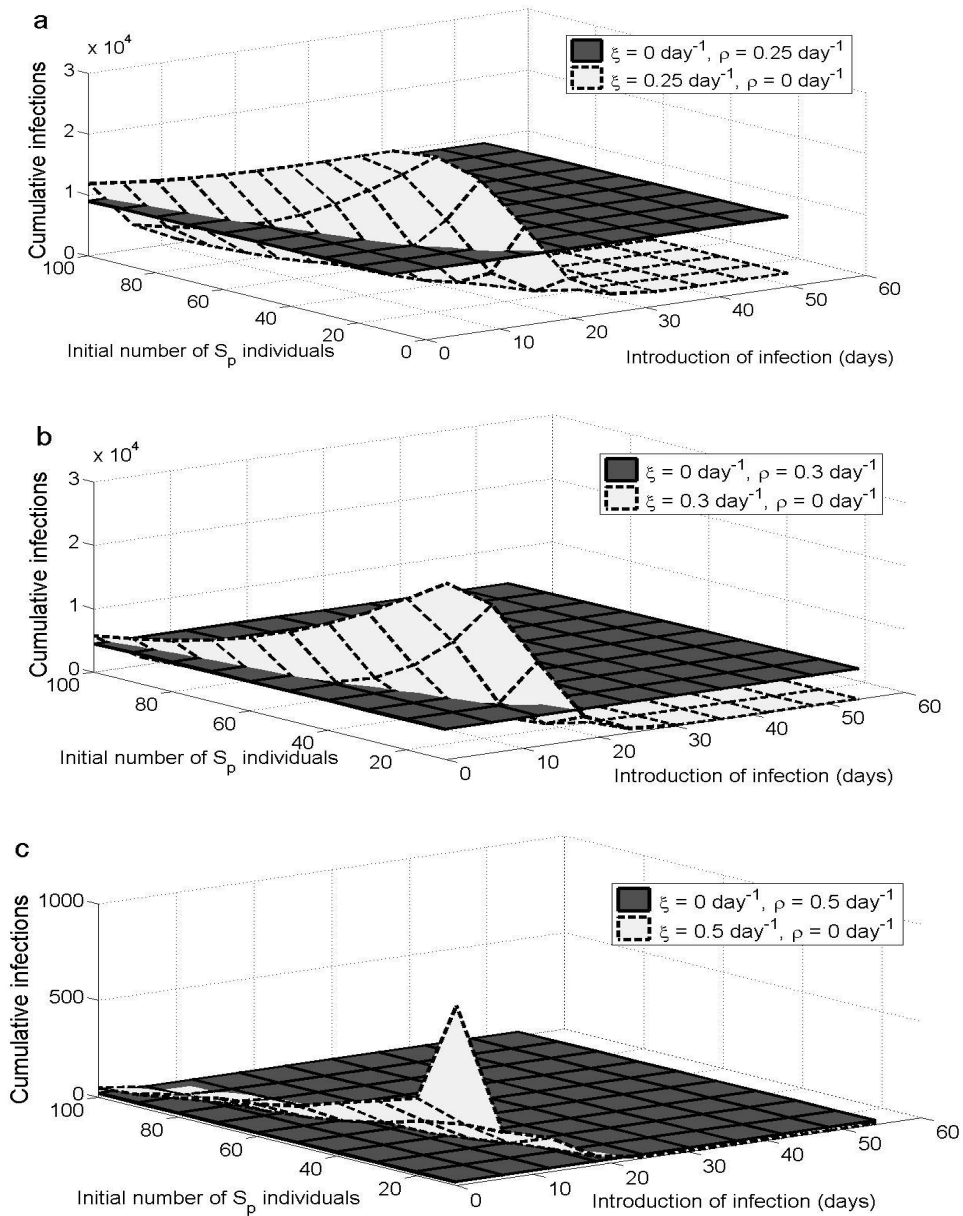


Figure 6: Cumulative infections as a function of the initial number of state S_p individuals and the time at which the infection is introduced, varying ξ and ρ , for the scenario of exposure learning only (dark grey surface) and social learning only (light grey surface). Other model parameters are $\tau = 0.8 \text{ day}^{-1}$, $\tau_p = 0.001 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$ and $\kappa = 0 \text{ day}^{-1}$.

4. Discussion

NPIs partly determine the feasibility of infection control for many infectious diseases, es-

pecially ones where pharmaceutical interventions are not yet available. Here, we constructed a pair approximation model of a self-

limiting infectious disease where individuals can choose to adopt NPIs either in response to learning it from other susceptible individuals, or having been stimulated to learn it from neighbouring an infectious person. Our objective was to demonstrate how pair approximation methods might be useful for studying socio-epidemiological processes in spatially structured populations.

We found that the impact of NPIs depends on the structure of the initial network configuration, particularly, the number and the neighborhood distribution of infectious, susceptible and individuals who practice NPIs, at the beginning of an outbreak. Both social learning and exposure learning lead to a decrease in the final size. At baseline parameter values, exposure learning is much more effective than social learning if social learning can only begin during the outbreak. However, social learning can outperform exposure learning if social learning begins early enough before the epidemic (although the initial number of protective individuals is not as important). While peak disease prevalence increases with the rate at which protective susceptible individuals stop the habit of practicing NPIs, the response of the infection peak to the rate of forgetting is qualitatively different for the two types of learning. We also found that, under certain parameter regimes, if infection source points are initially surrounded by protective individuals, increasing the number of infection source points at the beginning of an outbreak actually decreases the infection peak. This phenomena would not be revealed by the non-spatial, mean-field equations models.

Our model makes several simplifying assumptions. The model is based on the assumption that disease propagation and spatially localized learning take place only between connected neighbors on a regular network. In real

life, networks within which infections spread are more complex, and mean-field effects (such as mass media) may be important. Future work could extend the pair approximation model to account for these effects. On the other hand, the importance of higher-order spatial correlations in many spatial systems is known. In spatially-structured epidemic systems in particular it may be necessary to use triple approximations instead of just pair approximations in order to capture dynamics of the full spatially explicit model with a high degree of accuracy [4]. Our paper did not evaluate the importance of higher-order correlations in spatial socio-epidemiological dynamics, and this aspect is left for future research as well.

In Section 3.2.2 we found that social learning is more effective in reducing cumulative infections than exposure learning, provided there is a sufficient pool of protective susceptible individuals in the population at the beginning of the epidemic. Future work could include derivation of the threshold for the initial population size of protective susceptible individuals, above which social learning will be more effective than exposure learning. Also, in general, social networks are structurally different from networks in which infections spread. Thus, future work could also develop pair approximations for dual-level networks consisting of both a social network and a disease spread network.

In conclusion, we have shown how pair approximation models that incorporate both spatial transmission of diseases and impacts of NPI decision-making can be developed and analyzed. Future research using this methodology might yield insights regarding infection control in spatially-structured socio-epidemiological systems.

References

- [1] K.A. Alexander and J.W. McNutt. Human behavior influences infectious disease emergence at the human-animal interface. *Frontiers in Ecology and the Environment*, 8(10):522–526, 2010.
- [2] M.C. Auld. Estimating behavioral response to the aids epidemic. *Contributions to Economic Analysis and Policy*, 5(12), 2006.
- [3] N. Bacaer. Approximation of the basic reproduction number for vector-borne disease with periodic vector population. *Bulleting of Mathematical Biology*, 69:1067–1091, 2007.
- [4] C. T. Bauch. A versatile ode approximation to a network model for the spread of sexually transmitted diseases. *Journal of Mathematical Biology*, 45(5):375–395, 2002.
- [5] C.T. Bauch. The spread of infectious diseases in spatially structured populations: an invasyory pair approximation. *Mathematical Biosciences*, 198:217–237, 2005.
- [6] C.T. Bauch, A. d’Onofrio, and P. Manfredi. Behavioral epidemiology of infectious diseases: an overview: Modeling the interplay between human behavior and the spread of infectious diseases. *Springer-Verlag*, pages 1–19, 2013.
- [7] C.T. Bauch and A.P. Galvani. Using network models to approximate spatial point-process models. *Mathematical Biosciences*, 184:101–114, 2003.
- [8] C.T. Bauch and D.A. Rand. A moment closure model for sexually transmitted disease transmission through a concurrent partnership network. *The Royal Society*, 267:2019–2027, 2000.
- [9] J. Benoit, A. Nunes, and M. Telo da Gama. Pair approximation models for disease spread. *The European Physical Journal B*, 1:177–181, 2006.
- [10] K. Dietz. The estimation of the basic reproduction number for infectious diseases. *Statistical Methods in Medical Research*, 2(1):23–41, 1993.
- [11] S.P. Ellner. Pair approximation for lattice models with multiple interaction scales. *Journal of Theoretical Biology*, 210:435–447, 2001.
- [12] E.P. Fenichel, C. Castillo-Chavez, M.G. Geddia, G. Chowell, P.A. Gonzalez Parra, G.J. Hickling, G. Holloway, R. Horan, B. Morin, C. Perrings, M. Springborn, L. Velazquez, and C. Villalobos. Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences*, 108(15):6306–6311, 2011.
- [13] N.M. Ferguson, C.A. Donnelly, and R.M. Anderson. The foot and mouth epidemic in great britain: pattern of spread and impact of interventions. *Science*, 292:1155–1160, 2001.
- [14] M.J. Ferrari, S. Bansal, L.A. Meyers, and O.N. Bjørnstad. Network frailty and the geometry of head immunity. *Proceedings of the Royal Society B*, 273:2743–2748, 2006.
- [15] S. Funk, E. Gilad, C. Watkins, and V.A.A. Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 106(16):6872–6877, 2009.

- [16] R.J. Glass, L. M. Glass, W.E. Beyeler, and H.J. Min. Targeted social distancing designs for pandemic influenza. *Emerging Infectious Diseases*, 12(11):1671–1681, 2006.
- [17] D. Hiebeler. Moment equations and dynamics of a household sis epidemiological model. *Bulletin of Mathematical Biology*, 68:1315–1333, 2006.
- [18] M.J. Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of The Royal Society of London B*, 22(266):859–867, 1999.
- [19] M.J. Keeling, D.A. Rand, and A.J. Morris. Correlation models for childhood epidemics. *Proceedings of The Royal Society of London B*, 264:1149–1156, 1997.
- [20] J. Li, D. Blakeley, and R.J. Smith? The failure of \mathbb{R}_0 . *Computational and Mathematical Methods in Medicine*, 12:1–17, 2011.
- [21] C.N.L. Macpherson. Human behavior and the epidemiology of parasitic zoonoses. *International Journal for Parasitology*, 35(11-12):1319–1331, 2005.
- [22] S. Maharaj and A. Kleczkowski. Controlling epidemic spread by social distancing: Do it well or not at all. *BMC Public Health*, 12(679):1–16, 2012.
- [23] L. Mao and Y. Yang. Coupling infectious diseases, human preventive behavior, and networks—a conceptual framework for epidemic modeling. *Social Science Medicine*, 74(2):167–175, 2012.
- [24] J.P. McGowan, S.S. Shah, C.E. Ganea, S. Blum, J.A. Ernst, K.L. Irwin, N. Olivo, and P.J. Weidle. Risk behavior for transmission of human immunodeficiency virus (hiv) among hiv- seropositive individuals in an urban setting. *Clinical Infectious Diseases*, 38:122–127, 2004.
- [25] T. Modie-Moroka. Intimate partner violence and sexually risky behavior in botswana: implications for hiv prevention. *Health Care for Women International*, 30:230–231, 2009.
- [26] S.S. Morse. Factors in the emergence of infectious diseases. *Emerging Infectious Diseases*, 1(1):7–15, 1995.
- [27] S. Mushayabasa, C.P. Bhunu, and M. Dhlamini. Impact of vaccination and culling on controlling foot and mouth disease: a mathematical modeling approach. *World of Journal Vaccines*, 1:156–161, 2011.
- [28] S. O. Oyeyemi, E. Gabarron, and R. Wynn. Ebola, twitter, and misinformation: a dangerous combination? *British Medical Journal*, 349(g6178), 2014.
- [29] P.E. Parham and N.M. Ferguson. Space and contact networks: capturing of the locality of disease transmission. *Journal of Royal Society*, 3:483–493, 2005.
- [30] P.E. Parham, B.K. Singh, and N.M. Ferguson. Analytical approximation of spatial epidemic models of foot and mouth disease. *Theoretical Population Biology*, 72:349–368, 2008.
- [31] F.M. Pillemer, R.J. Blendon, A.M. Zaslavsky, and B.Y. Lee. Predicting support for non-pharmaceutical interventions during infectious outbreaks: a four region analysis. *Disasters*, 39(1):125–145, 2014.
- [32] D.A. Rand. Correlation equations and pair approximations for spatial ecologies.

CWI Quarterly, 12(3 and 4):329–368, 1999.

- [33] C.T. Reluga. Game theory of social distancing in response to an epidemic. *Plos Computational Biology*, 6(5):e1000793, 2010.
- [34] N. Ringa and C.T. Bauch. Dynamics and control of foot and mouth disease in endemic countries: a pair approximation model. *Journal of Theoretical Biology*, 357:150–159, 2014.
- [35] A. Rizzo, M. Frasca, and M. Porfiri. Effect of individual behavior on epidemic spreading in activity-driven networks. *Physical Review E*, 90(042801), 2014.
- [36] M. Salathe and S. Bonhoeffer. The effect of opinion clustering on disease outbreaks. *Journal of the Royal Society Interface*, 5:1505–1508, 2008.
- [37] L.B. Shaw and I.B. Schwartz. Fluctuating epidemics on adaptive networks. *Physical Review E*, 77(066101), 2008.
- [38] R.L. Stoneburner and D. Low-Beer. Population-level hiv decline and behavioral risk avoidance in uganda. *Science*, 304:714–718, 2004.
- [39] R.R. Swenson, W.S. Hadley, C.D. Houck, S.K. Dance, and L.K. Brown. Who accepts a rapid hiv antibody test? the role of race/ethnicity and hiv risk behavior among community adolescents. *Journal of Adolescent Health*, 48:527–529, 2011.
- [40] J.M. Tchuente and C.T. Bauch. Dynamics of an infectious disease where media coverage influences transmission. *ISRN Biomathematics*, 2012.

Appendix A. Derivation of the equation of motion for $[SI]$

Here we demonstrate the construction of equations of motion for pair approximation models by deriving $\frac{d[SI]}{dt}$.

In moment closure approximations the equation of motion for a state variable $g(t)$, is determined by expressing $\frac{dg(t)}{dt}$ as a function of the sum of all events that affect the state variable. That is

$$\frac{dg(t)}{dt} = \sum_{\epsilon \in \text{events}} r(\epsilon) \Delta g(\epsilon), \quad (\text{A.1})$$

where $r(\epsilon)$ is the rate of event ϵ and $\Delta g(\epsilon)$ is the change this event causes in $g(t)$. Equation (A.1) is referred to as the master equation. As illustrated below, at each node on the network the rates $r(\epsilon)$ and change $\Delta g(\epsilon)$ are expressed in terms of their population-averaged values and the deviations of those values from the expected means at a given node. The summation over each node is carried out in such a way that any significant stochasticity is incorporated in the evaluation of a state variable while the remaining stochasticity can be treated as random noise and may be discarded. We illustrate this concept below.

The time evolution of the number of $S-I$ pairs is determined by the following events.

Infection at a rate τ of a susceptible, S individual by their infectious, I neighbour (in a $S-I$ pair) converts S into I , i.e. $SI \mapsto II$, where \mapsto means 'transformed to'. This process *destroys* a $S-I$ pair. Similarly, infection at a rate τ of a susceptible individual 'from the left' in a triple $I-S-I$, i.e. $I \leftrightarrow SI$ also *destroys* a $S-I$ pair. However a $S-I$ pair is *created* as a susceptible individual is infected at the same rate, τ , in a triple $I-S-S$.

Transmission of the disease from an infectious, I individual to their protective contact, S_p at a rate τ_p in a triple S - S_p - I transforms S_p to I , and therefore *create* a S - I pair.

Recovery of the infectious individual at a rate, σ in a pair S - I implies $SI \mapsto SR$. Therefore the process *destroys* S - I .

Adoption of NPIs at a rate ξ , by a state S individual as a result of social learning from their state S_p neighbor in a triple S_p - S - I converts S to S_p , and therefore it *destroys* a S - I link.

Adoption of NPIs at a rate ρ , by a state S individual due to being in the neighborhood of a state I individual in a pair S - I or a triple I - S - I converts S to S_p , and therefore it *destroys* a S - I link.

$$\begin{aligned} \frac{d[SI]}{dt} = & \sum_{\zeta_{xy}=SS} \tau(n_{xy}(I))(+1) + \sum_{\zeta_{xy}=SI} \tau(n_{xy}(I))(-1) + \sum_{\zeta_{xy}=SS_p} \tau_p(n_{xy}(I))(+1) + \sum_{\zeta_x=S} \sigma(n_x(I))(-1) \\ & + \sum_{\zeta_{xy}=SI} \psi(n_{xy}(I))(-1) + \sum_{\zeta_{xy}=SI} \xi(n_{xy}(S_p))(-1) + \sum_{\zeta_{xy}=SI} \rho(n_{xy}(I))(-1) + \sum_{\zeta_x=I} \kappa(n_x(S_p))(+1). \end{aligned}$$

The *positive* (+) and *negative* (-) signs in the master equation above indicate that the corresponding events *create* or *destroy* S - I pairs, respectively.

Next we replace quantities such as $n_x(I)$ and $n_{xy}(I)$ by their population-averaged values (means) plus the stochastic deviations of those quantities from the means at nodes x and pairs xy . Let $n(i|j)$ be the population-averaged value of $n_x(i)$ when $\zeta_x = j$ and let $n(i|jk)$ be the population-averaged value of $n_{xy}(i)$ when $\zeta_{xy} = jk$. Then at each node we replace $n_{xy}(I)$ by $n(I|SI) + \theta_{xy}(I|SI)$ where $\theta_{xy}(I|SI)$ represents the stochastic fluctuation from the mean. The resulting expression is then simplified by taking out constants such as $n(I|SI)$ and the

Stopping the use of NPIs by a protective individual at a rate κ in a S_p - I pair increases their susceptibility, i.e. S_p converts to S , and the process *creates* a S - I pair.

To demonstrate the next steps of the derivation of the equation of motion for $[SI]$, the following notations will be useful:

$n_x(i)$: number of state i neighbours of a node x ;

$n_{xy}(i)$: number of state i neighbours of a node x , which has node y as a neighbour;

ζ_x : disease state of node x ;

ζ_{xy} : disease state of an edge involving x and y .

Using this notation, the master equation for the dynamics of $[SI]$ can now be represented as the sum of all the events (listed above) as

model parameters out of the sums and further noting that terms such as $\sum_{\zeta_x=S} \theta_x(I|S)$ that represent fluctuations are zero by definition. Furthermore the following identities (which apply to all network types):

$$\begin{aligned} n(i|jk) &= \frac{[ijk]}{[jk]}; \quad n(i|ji) = 1 + \frac{[iji]}{[ji]}; \quad n(i|j) = \frac{[ij]}{[j]} \\ \text{and } n(i|i) &= 1 + \frac{[ii]}{[i]}, \end{aligned}$$

enable us to write the equation of motion for $[SI]$ as

$$\frac{d[SI]}{dt} = \tau([ISS] - [ISI] - [SI]) + \tau_p[SS_pI] - \sigma[SI] - \xi[S_pSI] + \kappa[S_pI] - \rho([ISI] + [SI]).$$

We assume the disease spreads a regular net-

work where neighbors of an individual are themselves conditionally independent, therefore, third order correlations take the form

$$n(i|jk) = \frac{(n-1)}{n}n(i|j) \text{ and } n(i|ji) = 1 + \frac{(n-1)}{n}n(i|j).$$

That is, to close the system (i.e. approximate higher-order moments by lower-order moments) of equations we use the binomial ordinary pair approximation (OPA):

$$[ijk] = \frac{(n-1)}{n} \frac{[ij][jk]}{[j]}. \quad (\text{A.2})$$

The binomial OPA is based on the idea that the disease state of a node j is directly influenced by the states of two of its indirectly connected neighbors i and k . That is, there are *no triangles* in the network.

$$\begin{aligned} \frac{d}{dt}C_{SI} = & \tau(n-1)\frac{[S]}{[N]}C_{SS}C_{SI} - \tau(n-1)\frac{[I]}{N}C_{SI}^2 - \tau C_{SI} + \tau_p(n-1)\frac{[S_p]}{N}C_{S_pS}C_{S_pI} - \sigma C_{SI} - \xi(n-1)\frac{[S_p]}{N}C_{S_pS}C_{SI} + \\ & \kappa\frac{[S_p]}{[S]}C_{S_pI} - \rho(n-1)\frac{[I]}{N}C_{SI}^2 - \rho C_{SI} - \tau n\frac{[S]}{N}C_{SI}^2 - \tau_p n\frac{[S_p]}{N}C_{S_pI}C_{SI} + \sigma C_{SI} + \tau n\frac{[I]}{N}C_{SI}^2 + \xi n\frac{[S_p]}{N}C_{S_pS}C_{SI} - \\ & \kappa\frac{[S_p]}{[S]}C_{SI} + \rho n\frac{[I]}{N}C_{SI}^2. \end{aligned}$$

Similarly, the equation of motion for the correlation between susceptible individuals

$$\begin{aligned} \frac{d}{dt}C_{S_pI} = & \tau(n-1)\frac{[S]}{[N]}C_{SS_p}C_{SI} + \tau_p(n-1)\frac{[S_p]}{N}C_{S_pS_p}C_{S_pI} - \tau_p(n-1)\frac{[I]}{N}C_{S_pI}^2 - \tau_p C_{S_pI} - \sigma C_{S_pI} + \xi(n-1) \\ & \frac{[S]}{N}C_{S_pS}C_{SI} - \kappa C_{S_pI} + \rho\frac{[S]}{[S_p]}C_{SI} - \tau_p n\frac{[S_p]}{N}C_{S_pI}^2 - \tau n\frac{[S]}{N}C_{S_pI}C_{SI} + \sigma C_{S_pI} + \tau_p n\frac{[I]}{N}C_{S_pI}^2 - \xi n\frac{[S]}{N}C_{S_pS}C_{S_pI} + \kappa C_{S_pI}. \end{aligned}$$

We make biologically reasonable assumptions about the disease to simplify the equations above as follows. At the beginning of the epidemic there are very few infectious individuals (initial inoculation: $[I] = 1$) such that the entire population, N comprises almost only of susceptible individuals who protect, $[S_p]$ and those who do not protect, $[S]$ i.e. $[I] \ll N$ (where total population N is very large) and $[S_p] + [S] \approx N$. We remind the reader that our

Appendix B. Derivation of the basic reproduction number

Here we derive the expressions for C_{SI}^{min} and $C_{S_pI}^{min}$, and present expressions of the basic reproduction number resulting from a number of disease scenarios.

$C_{SI} = \frac{N}{n} \frac{[SI]}{[S][I]}$, therefore the equation of motion for C_{SI} is given by

$$\begin{aligned} \frac{d}{dt}C_{SI} = & \frac{N}{n} \left(\frac{1}{[S][I]} \frac{d}{dt}[SI] + \right. \\ & \left. \frac{[SI]}{[S][I]} \left(-\frac{1}{[I]} \frac{d}{dt}[I] - \frac{1}{[S]} \frac{d}{dt}[S] \right) \right). \end{aligned}$$

We substitute the equations of motion for the number of susceptible-infectious pairs, $[SI]$, the number of infectious individuals, $[I]$, and the number of susceptible individuals, $[S]$ from Equation (1) into the equation above, and approximate triples by the OPA to show that

who protect and infectious individuals can be written as

derivation of R_0 is based on the idea that both social and exposure learning take place only after the disease has been introduced. Therefore, we assume that at the initial stage of an outbreak the public has little information about the disease and only a small proportion of the population is aware and can decide to practice NPIs. Thus, at the beginning of an outbreak $[S_p]/N = s_p$, $[S]/N = 1 - s_p$, where $0 < s_p \ll 1$. Therefore, we can simplify fur-

ther the equations of motion for C_{SI} and $C_{S_p I}$ above by equating to zero all terms in which the denominator is large and $[I]$ is the numerator.

Although it may be necessary to also derive the equations of motion for three other correlation functions (for C_{SS} , $C_{S_p S}$ and $C_{S_p S_p}$) that appear in the equations of motion for C_{SI} and $C_{S_p I}$ above, we estimate them from the network configuration of the disease at the beginning of an outbreak. This step will help explore the relationship between the initial network configuration of the population with respect to the disease, and the evolution of the epidemic outbreak. We assume that at the beginning of the outbreak susceptible neighboring individuals mix homogeneously. That is, $C_{SS} \approx 1$.

$$\frac{d}{dt}C_{SI} = \tau(n-1)(1-s_p)C_{SI} - \tau C_{SI} + \tau_p(n-1)s_p C_{S_p S} C_{S_p I} + \xi s_p C_{S_p S} C_{SI} + \kappa s_p C_{S_p I} - \rho C_{SI} - \tau n(1-s_p)C_{SI}^2 - \tau_p n s_p C_{S_p I} C_{SI} - \kappa s_p C_{SI}$$

and

$$\frac{d}{dt}C_{S_p I} = \tau(n-1)(1-s_p)C_{S_p S} C_{SI} + \tau_p(n-1)s_p C_{S_p S_p} C_{S_p I} - \tau_p C_{S_p I} + \xi(n-1)(1-s_p)C_{S_p S} C_{SI} + \rho \frac{1-s_p}{s_p} C_{SI} - \tau_p n s_p C_{S_p I}^2 - \tau n(1-s_p)C_{S_p I} C_{SI} - \xi n(1-s_p)C_{S_p S} C_{S_p I} .$$

Solving $\frac{d}{dt}C_{SI} = 0$ and $\frac{d}{dt}C_{S_p I} = 0$ yields:

$$C_{SI}^{min} = \frac{R - \tau_p n s_p C_{S_p I}^{min} + \sqrt{(R - \tau_p n s_p C_{S_p I}^{min})^2 + 4\tau n s_p (1-s_p) \left(\tau_p (n-1) C_{S_p S} + \kappa \right) C_{S_p I}^{min}}}{2\tau n (1-s_p)}, \quad (\text{B.1})$$

$$C_{S_p I}^{min} = \frac{T - \tau n (1-s_p) C_{SI}^{min} + \sqrt{\left(T - \tau n (1-s_p) C_{SI}^{min} \right)^2 + 4\tau_p n s_p (1-s_p) \left((n-1)(\tau + \xi) C_{S_p S} + \rho / s_p \right) C_{SI}^{min}}}{2\tau_p n s_p}, \quad (\text{B.2})$$

where

$$R = \tau(n-1)(1-s_p) - \tau - \rho - \kappa s_p + \xi s_p C_{S_p S}$$

and $T = \tau_p(n-1)s_p C_{S_p S_p} - \tau_p - \xi n(1-s_p)C_{S_p S}$.

Simplifying assumptions

This value remains constant throughout calculations of the basic reproduction number for the initial network configurations considered in this paper, but the same does not hold for $C_{S_p S}$ and $C_{S_p S_p}$, since the latter take different values depending on various initial distributions of state S and state S_p individuals around the infection source. Since at the beginning of the outbreak s_p is very small, it is reasonable to assume that $[S] \approx N$. Therefore, at the initial stage of an infection $C_{S_p S} = \frac{N}{n} \frac{[S_p S]}{[S_p][S]} \approx \frac{1}{n} \frac{[S_p S]}{[S_p]}$ while $C_{S_p S_p} = \frac{N}{n} \frac{[S_p S_p]}{[S_p][S_p]}$. These values are calculated from actual initial network configurations as shown in Figure 1. We proceed with the derivation of C_{SI}^{min} and $C_{S_p I}^{min}$.

Note that now

Note that at the beginning of an outbreak the initial network configuration constitutes very few susceptible individuals who practice NPIs, so that while $s_p \approx O(1/N)$ (for a network where the infection source has at least 1 protec-

tive contact), $C_{S_p I}$ is large (i.e. $C_{S_p I}^{min} = O(N)$). Therefore we can simplify Equation (B.1) by assuming that $s_p C_{S_p I}^{min} \approx 1$, so that

$$C_{S_p I}^{min} \approx \frac{R - \tau_p n + \sqrt{(R - \tau_p n)^2 + 4\tau n(1 - s_p) \left(\tau_p(n-1)C_{S_p S} + \kappa \right)}}{2\tau n(1 - s_p)} = \chi$$

is a constant determined by the initial propor-

tion of susceptible individuals who protect, s_p , the model parameters $\tau, \tau_p, \sigma, \rho, \xi$ and κ , as well as the initial network configuration-specific values of correlation functions $C_{S_p S}$ and $C_{S_p S_p}$. Substituting χ and the resulting expression of $C_{S_p I}^{min}$ into Equation (10) yields

$$R_0 \approx \frac{T + n\tau(1 - s_p)\chi + \sqrt{\left(T - \tau n(1 - s_p)\chi \right)^2 + 4\tau_p n s_p(1 - s_p) \left((n-1)(\tau + \xi)C_{S_p S} + \rho/s_p \right) \chi}}{2\sigma}, \quad (\text{B.3})$$

where

$$T = \tau_p(n-1)s_p C_{S_p S_p} - \tau_p - \xi n(1 - s_p)C_{S_p S},$$

$$\chi \approx \frac{R - \tau_p n + \sqrt{(R - \tau_p n)^2 + 4\tau n(1 - s_p) \left(\tau_p(n-1)C_{S_p S} + \kappa \right)}}{2\tau n(1 - s_p)}$$

and

$$R = \tau(n-1)(1 - s_p) - \tau - \rho - \kappa s_p + \xi s_p C_{S_p S}.$$

Simulation results involving the basic reproduction number (in the Results section) are based on Equation (B.3).

Below we explore other scenarios of the disease to present simpler expressions of R_0 . We consider cases where protective susceptible individuals are assumed to maintain the habit of practicing NPIs throughout the epidemic outbreak (i.e. $\kappa = 0$). Also let the initial network configuration constitute 1 infectious individual with 1 state S_p neighbor (who also has 1 state S_p contact) and let the rest of the population be state S individuals so that for a large population size, initially $[S] \approx N$. Thus, the properties of the network at the initial stage of an

outbreak are $s_p = 2/N, C_{S_p S} = 5/2n, C_{S_p S_p} = N/4n$.

(a) *Adoption of NPIs through social learning only*

Here individuals are assumed to learn about the disease, and therefore adopt NPIs, from their contacts who already practice preventative behavior, and not from their infectious neighbors. That is we let $\xi > 0$, and $\rho = 0$.

If the population size, N is very large, and there are very few state S_p individuals at the beginning of the outbreak, so that $s_p = \frac{[S_p]}{N} \approx 0$, then we can simplify R by noting that $1 - s_p \approx 1$ and $\xi s_p C_{S_p S} \approx 0$ (since $0 \leq \xi \leq 1$ and $C_{S_p S} = O(1)$). Therefore

$$R \approx \tau(n-2).$$

Also, we simplify T by further noting that $s_p C_{S_p S_p} \approx O(1)$ and $n C_{S_p S} \approx O(1)$, so that

$$T \approx \tau_p(n-2) - \xi.$$

Thus, the expression of the basic reproduction number can be written as

$$R_0 \approx \frac{\tau_p(n-2) - \xi + \tau n \chi + \sqrt{\left(\tau_p(n-2) - \xi - \tau n \chi \right)^2 + 4\tau_p s_p(n-1)(\tau + \xi)\chi}}{2\sigma}, \quad (\text{B.4})$$

$$\text{where } \chi \approx \frac{\tau(n-2) - \tau_p n + \sqrt{\left(\tau(n-2) - \tau_p n\right)^2 + 4\tau\tau_p(n-1)}}{2\tau n}.$$

High efficacy case

Here we estimate the expression of R_0 for a case where individuals acquire preventive behavior through interaction with contacts who practice NPIs only, and the adopted NPIs are highly effective (i.e. τ_p is small so that $\tau_p^2 \approx 0$). Therefore,

$$\left(\tau_p(n-2) - \xi - \tau n\chi\right)^2 = \tau^2 n^2 \chi^2 - 2\tau_p(n-$$

$$R_0 \approx \frac{\tau_p(n-2) - \xi + \tau n\chi + \sqrt{\tau^2 n^2 \chi^2 - 2\tau_p(n-2)(\xi + \tau n\chi) + \xi(\xi + 2\tau n\chi) + 4\tau_p s_p(n-1)(\tau + \xi)\chi}}{2\sigma}, \quad (\text{B.5})$$

where

$$\chi \approx \frac{\tau(n-2) - \tau_p n + \sqrt{\tau(n-2) \left(\tau(n-2) - 2\tau_p n\right) + 4\tau\tau_p(n-1)}}{2\tau n}.$$

Model parameter-based R_0 for high efficacy case

We simplify Equation (B.5) further by prescribing a reasonable model parameter regime. Let $N = 40000$, $s_p = 2/N$, $n = 4$, $\tau = 1 \text{ day}^{-1}$, $\tau_p = 0.0025 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$ and $\xi = 0.25 \text{ day}^{-1}$. Then the following features of Equation (B.5) become apparent:

$$\begin{aligned} \tau(n-2) &\gg \tau_p n; \\ \tau(n-2) &\gg 2\tau_p n; \\ \xi &\gg \tau_p(n-2); \\ \tau n\chi &\gg \tau_p(n-2); \\ \tau^2 n^2 \chi^2 &\gg 4\tau_p s_p(n-1)(\tau + \xi)\chi; \\ \tau^2 n^2 \chi^2 &\gg 2\tau_p(n-2)(\xi + \tau n\chi); \\ \xi(\xi + 2\tau n\chi) &\gg 4\tau_p s_p(n-1)(\tau + \xi)\chi; \end{aligned}$$

$$2)(\xi + \tau n\chi) + \xi(\xi + 2\tau n\chi) + \tau_p^2(n-2)^2 \approx \tau^2 n^2 \chi^2 - 2\tau_p(n-2)(\xi + \tau n\chi) + \xi(\xi + 2\tau n\chi)$$

and

$$\begin{aligned} \left(\tau(n-2) - \tau_p n\right)^2 &= \tau^2(n-2)^2 - 2\tau\tau_p n(n-2) + \tau_p^2 n^2 \approx \tau^2(n-2)^2 - 2\tau\tau_p n(n-2) = \\ &= \tau(n-2) \left(\tau(n-2) - 2\tau_p n\right). \end{aligned}$$

The resulting expression of the basic reproduction number is

$$\xi(\xi + 2\tau n\chi) \gg 2\tau_p(n-2)(\xi + \tau n\chi).$$

We use these observations to cancel terms of Equation (B.5) that are insignificant (as per the prescribed parameter regime) to write the expression of the basic reproduction number as

$$R_0 \approx \frac{\tau n\chi - \xi + \sqrt{\tau^2 n^2 \chi^2 + \xi(\xi + 2\tau n\chi)}}{2\sigma}, \quad (\text{B.6})$$

$$\text{where } \chi \approx \frac{\tau(n-2) + \sqrt{\tau^2(n-2)^2 + 4\tau\tau_p(n-1)}}{2\tau n}.$$

(b) *Adoption of NPIs due to exposure learning only*

Here we consider a scenario where individual members of the population gain awareness about the disease, and in turn adopt NPIs, due to being next to infectious contacts only. Thus, we assume $\rho > 0$ and $\xi = 0$.

Applying similar arguments as in part (a) above, we simplify the original values of R , T and χ so that $R \approx \tau(n-2) - \rho$, $T \approx \tau_p(n-2)$ and

$$R_0 \approx \frac{\tau_p(n-2) + \tau n\chi + \sqrt{\left(\tau_p(n-2) - \tau n\chi\right)^2 + 4\tau_p n s_p \left(\tau(n-1)C_{S_p S} + \rho/s_p\right)\chi}}{2\sigma}, \quad (\text{B.7})$$

where

$$\chi \approx \frac{\tau(n-2) - \rho - \tau_p n + \sqrt{\left(\tau(n-2) - \rho - \tau_p n\right)^2 + 4\tau\tau_p(n-1)}}{2\tau n}.$$

High efficacy case

Applying the condition for a high efficacy case (i.e. $\tau_p \ll \tau$) and simplifying assumptions also applied in part (a) to a scenario where susceptible individuals adopt positive behavior through exposure learning only, transforms Equation (B.7) to

$$R_0 \approx \frac{\tau_p(n-2) + \tau n\chi + \sqrt{\left[\tau^2 n^2 \chi - 2\tau\tau_p n(n-2) + 4\tau\tau_p s_p(n-1) + 4\tau_p \rho n\right]\chi}}{2\sigma}, \quad (\text{B.8})$$

where

$$\chi \approx \frac{\tau(n-2) - \rho - \tau_p n + \sqrt{\tau^2(n-2)^2 - 2\tau(n-2)(\rho + \tau_p n) + \rho(\rho + 2\tau_p n) + 4\tau\tau_p(n-1)}}{2\tau n}.$$

Model parameter-based R_0 for high efficacy case

We prescribe the same parameter regime used in part (a) above, but note that here $\xi = 0$. Thus, $N = 40000$, $s_p = 2/N$, $n = 4$, $\tau = 1 \text{ day}^{-1}$, $\tau_p = 0.0025 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$ and $\rho = 0.25 \text{ day}^{-1}$. We cancel out parts of Equation (B.8) that are numerically less significant so that the expression of the basic re-

production number becomes

$$R_0 \approx \frac{\tau n\chi + \sqrt{\left(\tau^2 n^2 \chi + 4\tau_p \rho n\right)\chi}}{2\sigma}, \quad (\text{B.9})$$

where

$$\chi \approx \frac{\tau(n-2) - \rho + \sqrt{\tau^2(n-2)^2 - 2\tau\rho(n-2) + \rho^2 + 4\tau\tau_p(n-1)}}{2\tau n}.$$

Scenarios considered for the development of the above expressions of the basic reproduction number, are summarized in Table B.1.

Table B.1: Summary of expressions of the basic reproduction number R_0 developed in this paper

(a) General expression of R_0	Equation (10)
(b) Expression of R_0 used in simulation results in this manuscript: obtained by assuming that initially the proportion of susceptible individuals who practice NPIs is very small $s_p \approx O(1/N)$	Equation (B.3)
(c) Simplification of R_0 in Part (b) above by further assumptions: adoption of NPIs is through <i>social learning only</i> (i.e. $\xi > 0, \rho = 0$); once adopted NPIs are practised consistently (i.e. $\kappa = 0$); at initial stage there is 1 state I with 1 state S_p contact who has 1 state S_p , and the rest of the population is of state S	Equation (B.4)
(d) Simplification of R_0 in Part (c) above by a further assumption: <i>high efficacy NPIs</i> (i.e. $\tau_p \approx 0$)	Equation (B.5)
(e) Simplification of R_0 in Part (d) above by cancelling out insignificant terms dependent on the parameter regime: $N = 40000$; initially $s_p = 2/N$; $n = 4$; $\tau = 1$; $\tau_p = 0.0025$; $\sigma = 0.25$; $\xi = 0.25$	Equation (B.6)
(f) Simplification of R_0 in Part (b) above by further assumptions: adoption of NPIs is through <i>exposure learning only</i> (i.e. $\xi = 0, \rho > 0$); other conditions are as in Part (c) above	Equation (B.7)
(g) Simplification of R_0 in Part (f) above by a further assumption: <i>high efficacy NPIs</i> (i.e. $\tau_p \approx 0$) acquired through exposure learning only	Equation (B.8)
(h) Simplification of R_0 in Part (g) above by cancelling out insignificant terms dependent on the parameter regime: $N = 40000$; initially $s_p = 2/N$; $n = 4$; $\tau = 1$; $\tau_p = 0.0025$; $\sigma = 0.25$; $\xi = 0.25$	Equation (B.9)

Appendix C. Supplementary Material

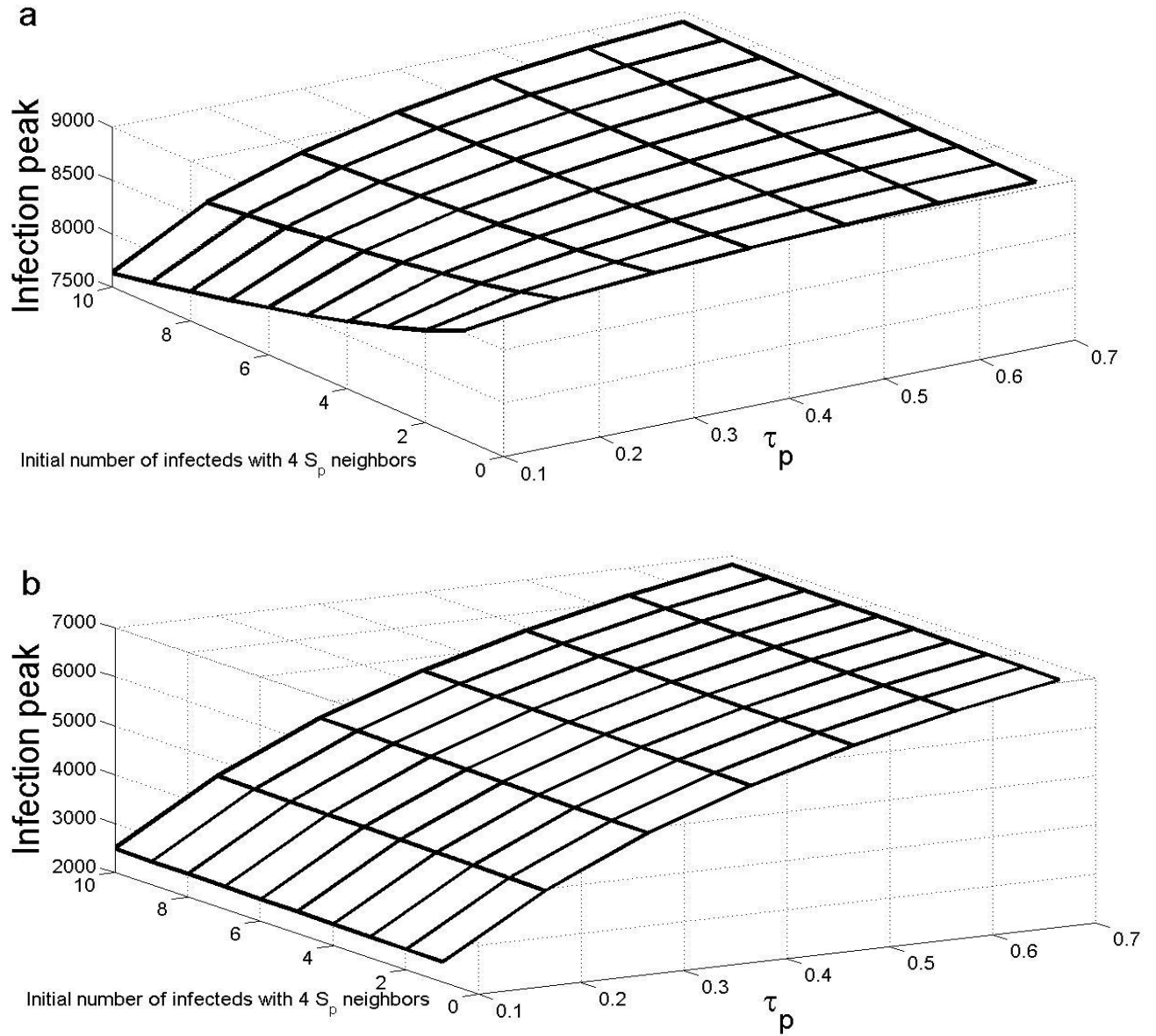


Figure C.1: Infection peak versus rate of disease transmission to protective individuals, τ_p , and the initial distribution of *single infected individuals with 4 state S_p neighbors* (and all other members of the host population are fully susceptible, S), where $\xi = 0.25 \text{ day}^{-1}$, $\rho = 0 \text{ day}^{-1}$ (a) and $\xi = 0 \text{ day}^{-1}$, $\rho = 0.25 \text{ day}^{-1}$ (b). Other model parameters are $\tau = 0.8 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$ and $\kappa = 0 \text{ day}^{-1}$.

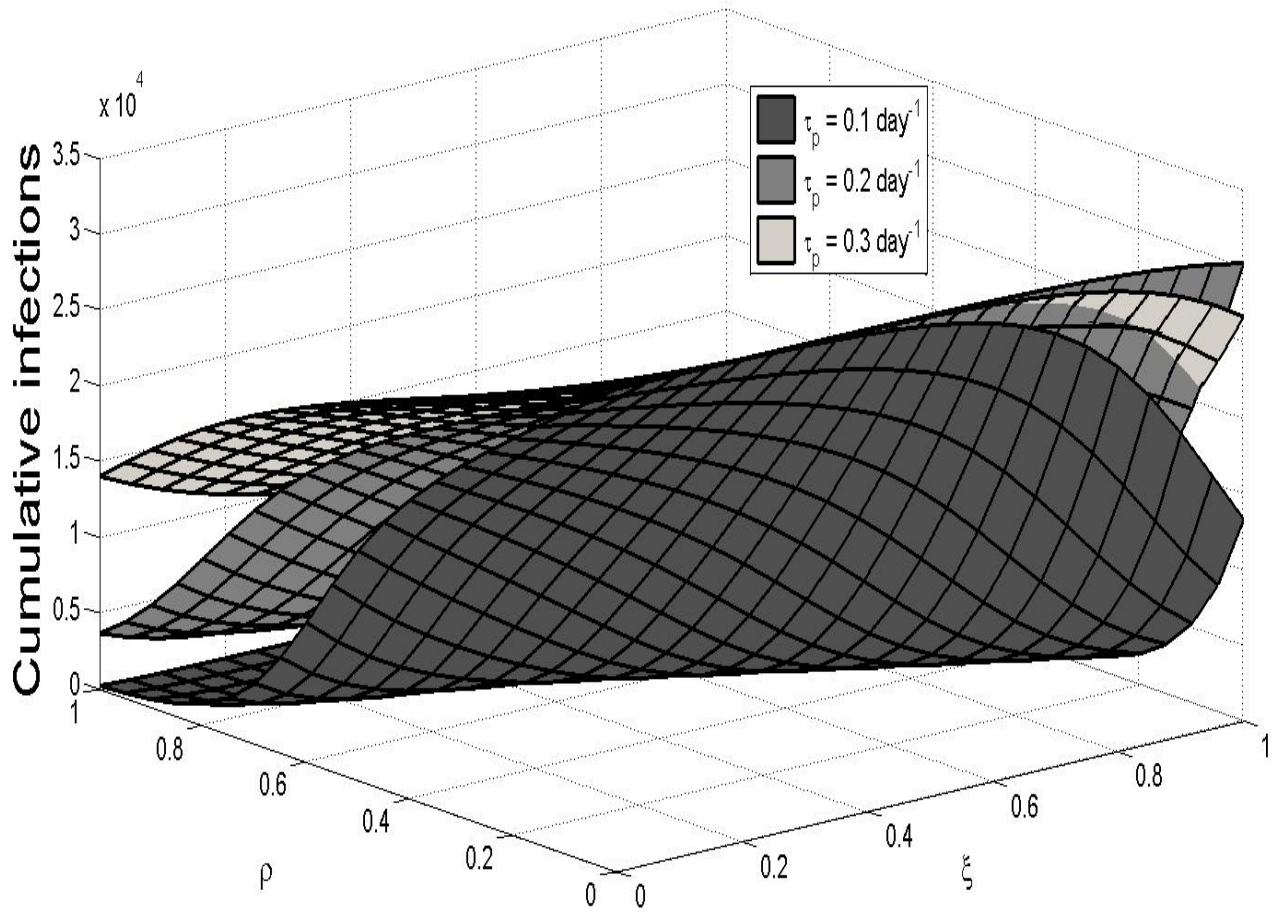


Figure C.2: Cumulative infections as a function of social learning from both infectious and state S_p neighbors at rates ρ and ξ , respectively, where the initial conditions are *1 infected node and 1 state S_p neighbor* while the rest of the population is fully susceptible (i.e. state S). Model parameters are $\tau = 0.8 \text{ day}^{-1}$, $\sigma = 0.25 \text{ day}^{-1}$ and $\kappa = 0 \text{ day}^{-1}$.