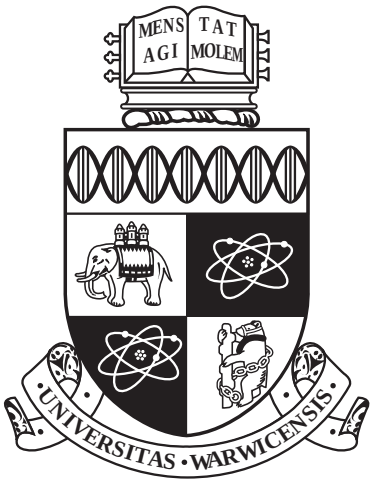

Non-Markovian Network Epidemics

by

Fergus Cooper

Thesis

Submitted to The University of Warwick

Mathematics Institute

April, 2013

THE UNIVERSITY OF
WARWICK

Contents

1 Introduction

Mathematical modelling of epidemics is a large and growing research area. Understanding the spread of infectious disease allows us to develop improved control measures which help to save lives. One particularly well established epidemiological model, and the one on which this report focusses, is the so-called SIR (susceptible–infected–recovered) model.

First proposed by Kermack and McKendrick in 1927[1], the model attempted to capture the phenomenon that, during epidemics, infection rises rapidly then falls. The simplicity of the model allows relative ease in understanding its behaviour, and yet in certain circumstances it has been shown to provide a remarkable fit with actual infection data. In the original 1927 paper, Kermack and McKendrick demonstrated this with plague deaths recorded among the human population of Bombay between December 17, 1905 and July 21, 1906. Although the model is not seen frequently in the literature until a paper by Anderson and May in 1979[2], variations on the original SIR equations have subsequently been used to model a vast range of diseases including cholera[3], bubonic plague[4] and influenza[5].

1.1 Stochastic or continuum model?

When modelling a process such as an epidemic it is necessary to keep in mind the biology underlying the transmission of infection through a population. One problem facing any mathematical model is that there are many different levels upon which the situation may be considered.

Perhaps the most natural would be a model centred around individuals. Each individual, as in the SIR model, may be considered to occupy one, and only one, state from the state space of the model. The state space, $\mathbb{S} = \{S, I, R\}$, for the SIR model contains the states S (susceptible), I (infected) and R (recovered). In this regime, any compartmental model with discrete states carries with it certain assumptions. First, that the transition between states is instantaneous; an individual may be susceptible at one moment and infected the next. Second, that the transition between states is discontinuous; there is no middle ground between, say, being infected and having recovered, other than by means of adding additional states to $\mathbb{S}$. Thus, already, by the choice of model we are removing biological realism even before we make assumptions on how infection and recovery occur.

One could attempt to construct a 'better' model by drilling down to the smaller-scale agents behind infection, perhaps by modelling effects of pathogens on the immune system of each individual. This may confer additional biological realism but would make models significantly more complex, and would mix vastly different length- and timescales, adding additional complication. Whatever model is used, assumptions are made, and so any results or predictions must only be held valid for situations in which the assumptions do not deviate far from reality.

The models presented here will, at their base level, take account only of individuals.

Transmission of, and recovery from, infection will be considered stochastic so that, under the conditions of the model, transition between states will occur with certain probability. Under this regime, given a population and rules determining infection and recovery, an epidemic may be considered as a particular instantiation of the outcome of the stochastic processes, such that each epidemic will probably be different under the exact same rules.

If a population is large, however, we may consider probabilities of changing between states as rates, and derive differential equations. These describe the expected proportion of the (large) population in each state at each time, as the continuum limit of the stochastic processes. Proof that the stochastic processes have a limit in the differential equations will not be given (though it can be done); however, simulations will be presented to lend credibility to such continuum models. The usefulness of a differential equation approach is that it is deterministic; it tells us something about an average epidemic, whereas any one instantiation of the stochastic processes does not give insight into how the next epidemic might unfold.

1.2 Project outline

The goal of this project is to modify the original SIR model in an attempt to make the biological assumptions more realistic. This aim will see four different models discussed:

(a) The original Kermack–McKendrick (K-M) SIR model. This will assume infection is a *mass action* process, and the recovery of infected individuals is *markovian*.

(b) The non-markovian SIR model. Infection will still be assumed a *mass action* process but the *markovian* recovery property will be relaxed.

(c) The pairwise SIR model. This relaxes the original *mass action* assumption but re-introduces *markovian* recovery.

(d) The pairwise non-markovian SIR model. This relaxes both *mass action* and *markovian* recovery.

1.3 Definitions

Mass action

The law of mass action, borrowed from chemistry, states that the rate of a reaction is proportional to the product of the concentrations of the reactants[6, §9].

In our epidemiological context, states in our state space take the place of reactants, proportions of our total population in each state take the place of concentrations, and infection is the reaction which we consider. This means that, under mass action, the rate at which susceptible individuals contract infection is proportional to the product of the proportions of the population infected and susceptible.

Markovian

A stochastic process is said to be markovian if it satisfies the Markov property[7, §1]: the conditional probability distribution of future states of the system depends only on the current state.

To say that recovery is a markovian process would forbid, for instance, the probability of recovery being a function of the time since infection was acquired, because deciding whether a recovery event would occur would require memory of when the infection happened. Translating this to our continuum case, recovery is markovian if it happens at a constant rate.

We begin with a discussion of the properties of the K-M SIR model.

2 The K-M SIR model

2.1 Underlying stochastic model

We proceed with the following assumptions:

- The state space is $\mathbb{S} = \{S, I, R\}$, so that each individual may only be either susceptible, infected or recovered. We assume, therefore, that 'infectious' and 'infected' are synonymous; everyone infected is able to infect others.
- The population is large, of size N, and we neglect the effects of births or deaths, so that N is constant.
- Initially, the population is totally susceptible, bar a small number of infected individuals added at time zero.
- The population can mix uniformly, so that every individual interacts identically with every other; this becomes the mass action assumption when the model is written in its deterministic form.
- Interactions occur at discrete steps in time.
- Infection happens between an infected and a susceptible individual, each timestep, with a constant probability.
- Recovery of an infected individual occurs with constant probability, each timestep; this is the markovian assumption.

2.2 Deterministic ODE model

Converting the above to an ODE (ordinary differential equation) model in continuous time can be done rigorously, as in the original K-M paper[1], but here it will suffice to make a few observations. First, for this deterministic model let us set $N = 1$ so that we now talk about proportions of the total population rather than number of individuals, and create time-dependent variables $S(t)$, $I(t)$ and $R(t)$ to represent the proportions of the population in each state.

Observe that susceptible individuals may only be depleted; there is no mechanism allowing those infected or recovered to re-enter the S state. The law of mass action then states that this depletion happens at a constant rate proportional to $S(t)I(t)$; call the constant of proportionality, β, the rate of infection. Next, the increase in $I(t)$ must balance the decrease in $S(t)$, and depletion from $I(t)$ by recovery happens at a constant rate, call it γ. Finally, the recovered state fills at a rate to balance those coming out of the infected state.

Thus, our system is described by the following ODEs, with overdot representing derivative with respect to time:

$$\begin{aligned} &\text{(i)} \quad \dot{S}(t) = -\beta S(t) I(t) \\ &\text{(ii)} \quad \dot{I}(t) = \beta S(t) I(t) - \gamma I(t) \\ &\text{(iii)} \quad \dot{R}(t) = \gamma I(t) \end{aligned} \tag{1}$$

There are several properties of (1) worth mentioning:

2.3 Initial epidemic growth

With constant infection and recovery rates it is straightforward to calculate the initial growth rate of the epidemic. Taking the equation for $\dot{I}(t)$ in isolation, we can examine the early behaviour when a small amount of infection is invading a totally susceptible population. Notice that, as we neglect births and deaths, $S(t) + I(t) + R(t) = 1$ at any time so that, at an early enough time, $I(t)$ being small means $S(t) \approx 1$. This gives us the following linearised equation for $I(t)$, valid for small t, provided $I(0) \ll 1$:

$$\dot{I}(t) = (\beta - \gamma) I(t) \quad \Rightarrow \quad I(t) = I(0) e^{(\beta - \gamma)t}. \tag{2}$$

This gives the growth rate, r, for the K-M SIR model as $\beta - \gamma$. The infection will invade and initially increase exponentially provided $\beta - \gamma > 0$, and it will fail to invade if not. This invasion criterion can, equivalently, be stated in terms of an important quantity in epidemiology: the basic reproductive ratio, R_0. Defined as 'the average number of secondary cases produced by an average infectious individual in a totally susceptible population'[8], we see that for this simple model $R_0 = \beta/\gamma$. The condition $\beta - \gamma > 0$ is equivalent to $R_0 > 1$.

2.4 Final epidemic size

In this setting, we can determine analytically the final size of the epidemic. We define this final size to be $R(\infty)$, the total proportion of the population who ever enter the recovered state. Because every infected individual must recover in finite time, this is equal to the total number of people who ever contract the infection and, additionally, as the system is closed:

$$I(\infty) = 0 \quad \Rightarrow \quad R(\infty) = 1 - S(\infty). \tag{3}$$

To calculate $R(\infty)$, we divide (1)(i) by (1)(iii) which, remembering that $R_0 = \beta/\gamma$, gives:

$$\frac{dS(t)}{dR(t)} = -R_0 S(t) \quad \Rightarrow \quad S(t) = e^{-R_0 R(t)}. \tag{4}$$

Evaluating this at infinity, and by applying (3), we obtain the following transcendental equation for $R(\infty)$:

$$R(\infty) = 1 - e^{-R_0 R(\infty)}. \tag{5}$$

Thus, we see that the final size of the epidemic is a function of only one variable; $R_0 = \beta/\gamma$. If $R_0 \leq 1$, $R(\infty) = 0$ as the infection fails to invade but, as $R_0 \to \infty$, $R(\infty) \to 1$ which indicates that a greater proportion of the population becomes infected if the rate of infection is higher, as one would expect from intuition. $R(\infty)$ increases monotonically with $R(0)$, as can be seen in Figure 1 below.

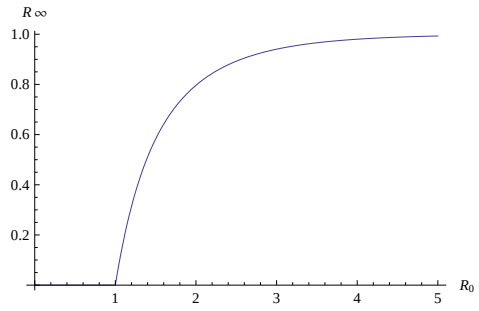

Figure 1: Reproductive ratio R_0 against final epidemic size $R(\infty)$

2.5 Solving the system

There is no closed-form analytical solution for the equations in (1) as functions of time[1], but they can easily be numerically solved with, for instance, `ode45` in Matlab or `NDSolve` in Mathematica. Below are plots of the proportion infected and recovered against time, with results from (2) and (5) included.

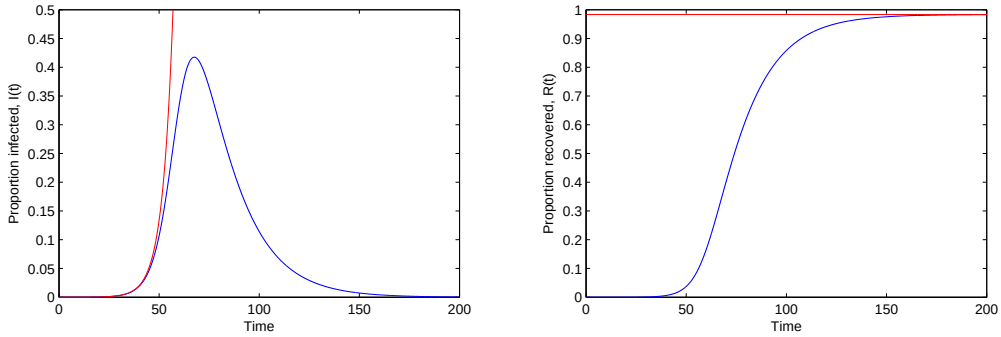

Figure 2: Blue: numerical solution of system (1), with parameters $\beta = 0.25$ and $\gamma = 0.06$. Red: exponential with initial growth rate 0.19 (left) and final epidemic size 0.983 (right).

We see from Figure 2 that the exponential describing the early growth rate does indeed fit the dynamics well for early time. The slowdown in infection is due to the the decrease in the number of susceptible individuals in the population which, due to mass action, reduces the rate at which new cases are created. It is worth noting that the final epidemic size is always < 1; the population is never entirely depleted of susceptible individuals by

an epidemic.

2.6 Simulation of dynamics

In order to demonstrate the usefulness of the deterministic dynamics in predicting the likely outcome of individual epidemics, one can perform stochastic simulations according to the scheme set out at the beginning of this section. There are a few items to note:

- The speed of the simulation reduces significantly with an increase in population size so, for simulations to take a sensible amount of time, only $N = 5000$ is used.
- If the number of initial infected cases is too small, it is likely that infection could die out or not initially increase quickly. It is, therefore, important to pick $I(0)$ sufficiently large. Here, $I(0) = 10$ is used.
- At each timestep in the simulation, each currently susceptible individual must be tested against each currently infected individual to see whether they become infected. The per capita infection probability is now β/N.
- Infection and recovery events are decided by picking uniformly distributed random numbers in the unit interval, with the transition probabilities β/N and γ as thresholds.

Below is a plot showing $I(t)$ as a proportion of N for three such simulations (red), alongside the numerical solution of (1) (blue):

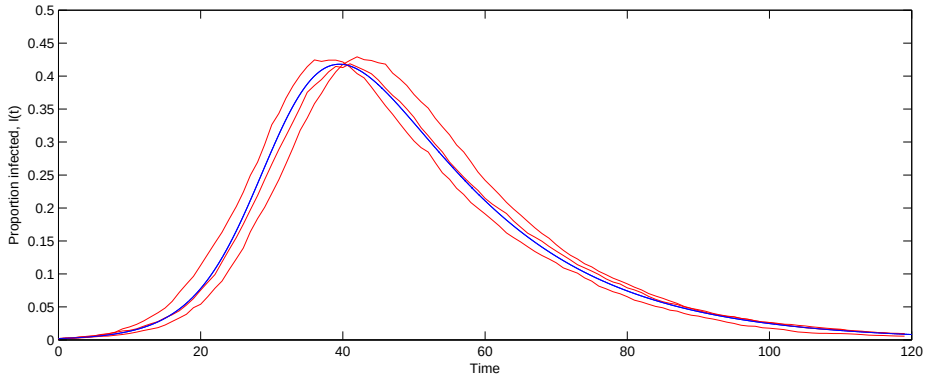

Figure 3: Outcome of simulations with parameters $\beta = 0.25$ and $\gamma = 0.06$

We would expect the simulations to more closely match the deterministic equation as N and $I(0)$ become large, provided $I(0)$ remains small as a proportion of N. Matlab source code for this simulation is presented in the appendix.

We now attempt to increase the biological realism by relaxing the markovian recovery assumption used in this model.

3 Non-markovian SIR model

Of the two assumptions (mass action and markovian recovery) detailed in the introduction, we first relax the latter. To see why, let us examine whether there is any biological underpinning for the markovian recovery assumption.

Turning to the underlying stochastic model, assume that an individual becomes infected at some time $t = t'$. Then, according to the scheme set up in §1, the probability of the individual recovering one timestep δt in the future is precisely γ, which means the probability that they do not recover is $1 - \gamma$. The probability that they still have not recovered after two timesteps must be conditional on them having not recovered after one timestep, and so is $(1 - \gamma)^2$. The leads to the following probability distribution for the chance of an infected individual still being infected n timesteps after infection:

$$\mathbb{P}(\text{still infected at time } t' + n\delta t) = (1 - \gamma)^n. \tag{6}$$

As γ is a positive constant, this starts at 1 (when $n = 0$) and tends to 0 as $n \to \infty$. This, clearly, bears little relation to reality, where one would expect to stay infected for a certain time before having the possibility of recovering. So, instead of a model in which it is possible to recover straight after becoming infected, how can we modify the SIR model to deal with a situation in which, once infected, an individual remains infected for precisely a certain time? Let us denote this constant infectious period by the positive constant ξ.

3.1 Handling a new infection variable

To move forward, we will now need to keep track of the time that has elapsed since an individual became infected. Let us decide upon the following use of language:

- time: the system time, starting with zero at the start of the epidemic. This will be denoted by the non-negative variable t.
- age: the time which has elapsed, for an infected individual, since infection. This will be denoted by the non-negative variable a.

With this language in place, we construct a two-dimensional variable $i_a(t, a)$ which we use to denote the density of age a infected individuals at time t. We use a lower-case i to distinguish from $I(t)$ which remains the total proportion of the population infected, and the need for subscript a will become apparent later.

In order to construct a deterministic model in this new situation we will need to be able to do calculus on this i_a variable. In particular, we are interested in how it changes over time, so would like its time derivative. Consider a small change in time δt, and let us perform a Taylor expansion to examine how i_a changes if we jump forwards in time (and therefore, at the same rate, in age):

$$i_a(t+\delta t, a+\delta t) = i_a(t,a) + \delta t\left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) i_a(t,a) + O(\delta t^2). \tag{7}$$

Taking the limit as $\delta t \to 0$, we obtain the following form for the derivative:

$$\frac{d}{dt} i_a(t,a) = \lim_{\delta t \to 0}\left(\frac{i_a(t+\delta t, a+\delta t) - i_a(t,a)}{\delta t}\right) = \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right) i_a(t,a). \tag{8}$$

Alternatively, using the standard formula for the total derivative[9], we can write:

$$d\, i_a(t,a) = \frac{\partial i(t,a)}{\partial t} dt + \frac{\partial i(t,a)}{\partial a} da. \tag{9}$$

We note that, as time and age tick at the same rate, i.e. $\frac{da}{dt} = 1$, dividing through by dt gives back the result in (8). This partial differential operator will be used extensively going forward, so let us define the following for ease of notation:

$$\Phi(\cdot) = \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)(\cdot) \tag{10}$$

3.2 The new deterministic model

We are now in a position to write down new differential equations governing the dynamics of the system. As it is only recovery which we seek to change, (1)(i) will remain unchanged but we must modify (1)(ii) and (1)(iii). Let us introduce, first, a general recovery function $\gamma(a)$. A function now of age, it is no longer necessarily constant.

Infection, we assume, is still a mass action process, so still happens at a rate $\beta S(t) I(t)$. Here, $I(t)$ is the total number of infected individuals, which must obey $I(t) = \int_0^\infty i(t,a) da$. Additionally, when an individual is newly infected they are constrained to start with age 0. This information can be supplied in the form of a boundary condition, or we may make use of the Dirac delta function and incorporate it into the main equation. The latter approach is presented here. Finally. the rate of recovery of those infected individuals of age a is simply $\gamma(a)$.

The change in recovered class must be accounted for by all those recovering, at whatever age they are recovering. While we will eventually wish everyone to recover at the same age, ξ, for now we can maintain generality, and the rate of change of recovery will simply occur as the sum over all ages of infected individuals, weighted by the recovery function $\gamma(a)$.

Putting this reasoning together, our system becomes:

$$\begin{aligned}
&\text{(i)} && \dot{S}(t) = -\beta S(t) I(t) \\
&\text{(ii)} && \Phi(i_a(t,a)) = \delta(a)\beta S(t) I(t) - \gamma(a) i_a(t,a) \\
&\text{(iii)} && \dot{R}(t) = \int_0^\infty \gamma(a) i_a(t,a)
\end{aligned} \tag{11}$$

There are, in principle, methods which can be used to numerically solve such systems

which contain PDEs (partial differential equations). Usually, however, it is better (in the sense of providing more precise numerical solutions) to integrate ODEs[10, §1], as derivatives with respect to only one variable are present in the equations. To this end, we head towards deriving a system of DDEs (delay differential equations).

3.3 Analytic solution of the PDE

We can find a solution to (11)(ii) in bulk (ignoring the boundary condition) by means of separation of variables. We first notice that an arbitrary function $f(t-a)$ solves the homogeneous equation $\Phi(i_a) = 0$. Next, we use the following ansatz, for some unknown T and A, to find a particular solution:

$$i_a(t,a) = T(t) \cdot A(a) \tag{12}$$

Using our definition of Φ, this ansatz becomes:

$$\Phi(i_a(t,a)) = T'(t)A(a) + T(t)A'(a) = i_a(t,a)\left(\frac{T'(t)}{T(t)} + \frac{A'(a)}{A(a)}\right). \tag{13}$$

Combining (11) and (13) we have:

$$\frac{T'(t)}{T(t)} + \frac{A'(a)}{A(a)} = -\gamma(a). \tag{14}$$

We can now separate this into two equations, because the right-hand side of (14) is a function only of a which means each term in the left-hand side must also be a function only of a. First, $\frac{T'(t)}{T(t)} = 0$, which gives $T(t) = \text{const}$; however, this constant can be absorbed into the arbitrary function $f(t-a)$. Second, $\frac{A'(a)}{A(a)} = -\gamma(a)$ yields $A(a) = e^{-\int_0^a \gamma(\alpha)d\alpha}$. Putting these together, we have a general solution, in bulk, for $i_a(t,a)$:

$$i_a(t,a) = f(t-a)e^{-\int_0^a \gamma(\alpha)d\alpha}. \tag{15}$$

This is subject to the boundary condition

$$i_a(t,0) = \beta S(t) I(t), \tag{16}$$

so (15) and (16) describe the solution to (11)(ii). In general, we cannot apply the boundary condition to pin down our unknown function $f(t-a)$. We can, however, attempt to describe the early-time behaviour of the epidemic, as we did in the K-M SIR model.

3.4 Initial epidemic growth

If we consider time early in an epidemic, shortly after a small quantity $I(0) \ll 1$ of infection is introduced into a totally susceptible population, we would expect the proportion of infected individuals to initially grow exponentially, at some unknown rate r. This means

that, for small t, $S(t) \approx 1$ and we can approximate to $I(t) = I(0)e^{rt}$. Substituting this approximation into the boundary condition (16), we have:

$$i_a(t,0) = \beta I(0) e^{rt} \overset{(15)}{\widetilde{=}} f(t), \tag{17}$$

and so the unknown $f(t-a)$ is constrained to be $I(0)\beta e^{r(t-a)}$. This updates (15) for small times:

$$i_a(t,a) = I(0)\beta e^{r(t-a)} e^{-\int_0^a \gamma(\alpha)d\alpha}. \tag{18}$$

We now pin down the unknown growth rate r by integrating (18) over all ages:

$$I(0)e^{rt} = I(t) = \int_0^\infty i(t,a)da = \int_0^\infty \left(I(0)\beta e^{r(t-a)} e^{-\int_0^a \gamma(\alpha)d\alpha} \right) da \tag{19}$$

which yields, upon cancellation, the integral equation for r:

$$1 = \beta \int_0^\infty e^{-ra} \overbrace{e^{-\int_0^a \gamma(\alpha)d\alpha}}^{(*)} da. \tag{20}$$

In general, we cannot extract r from (20) as it depends on the arbitrary function γ. We are, however, interested in a particular functional form for the recovery: we want that each infected individual is infected for a constant time ξ before recovering. In this case, our recovery function γ must be such that $\int_0^a \gamma(\alpha)d\alpha = 0$ for all $a < \xi$, and $\int_0^a \gamma(\alpha)d\alpha = \infty$ for all $a \geq \xi$.

This means that (*) in (20) is 1 for $a < \xi$ and 0 for $a \geq \xi$, which simplifies (20) to:

$$1 = \beta \int_0^\xi e^{-ra} da. \tag{21}$$

The initial growth rate r in which we are interested, therefore, is the non-zero solution of the transcendental equation:

$$r = \beta \left(1 - e^{-r\xi} \right). \tag{22}$$

3.5 DDE model

Because of our choice of γ, we have a model in which infected individuals recover with a fixed delay. This allows us to set up a DDE model; the rate of change of I depends on the system one timestep of size ξ in the past as well as on the system at the current time.

The additional information we need is a set of *history functions*; how S, I and R behave in the first time interval $[0, \xi]$. Provided ξ is not too large, the exponential growth of I with rate r derived above will serve as a good approximation to the dynamics in this first time interval.

To construct the model we observe that, at any time t, only those infected individuals of age ξ recover, and are thus the only ones removed from the pool of infection. They were all infected at precisely time $t-\xi$, at which time the rate of infection (from (11)) was $\beta S(t-\xi)I(t-\xi)$. Thus, at each time, infection is gained at a rate $\beta S(t)I(t)$ and lost at a rate $\beta S(t-\xi)I(t-\xi)$. Hence, our DDE model is:

$$\begin{aligned} &\text{(i)} \quad \dot{S}(t) = -\beta S(t)I(t) \\ &\text{(ii)} \quad \dot{I}(t) = \beta S(t)I(t) - \beta S(t-\xi)I(t-\xi) \\ &\text{(iii)} \quad \dot{R}(t) = \beta S(t-\xi)I(t-\xi), \end{aligned} \tag{23}$$

subject to the history functions:

$$\begin{aligned} &\text{(i)} \quad S(t) = 1 - I(0)e^{rt} \quad \forall t \in [0,\xi] \\ &\text{(ii)} \quad I(t) = I(0)e^{rt} \quad \forall t \in [0,\xi] \\ &\text{(iii)} \quad R(t) = 0 \quad \forall t \in [0,\xi]. \end{aligned} \tag{24}$$

In practice, most DDE systems cannot be solved analytically; however, there are several numerical approaches for solving them: `dde23` in Matlab and `NDSolve` in Mathematica are both capable of numerically solving such equations. See appendix for input code from a Mathematica notebook designed to numerically solve (23) subject to (24). Below is a plot of the numerical solution, where values for parameters β, ξ and $I(0)$ have been chosen for illustrative purposes. The blue curve is $S(t)$ and the red is $I(t)$. Notice that explicit calculation of $R(t)$ is not necessary as the equations are still conservative; $R(t) = 1 - S(t) - I(t)$.

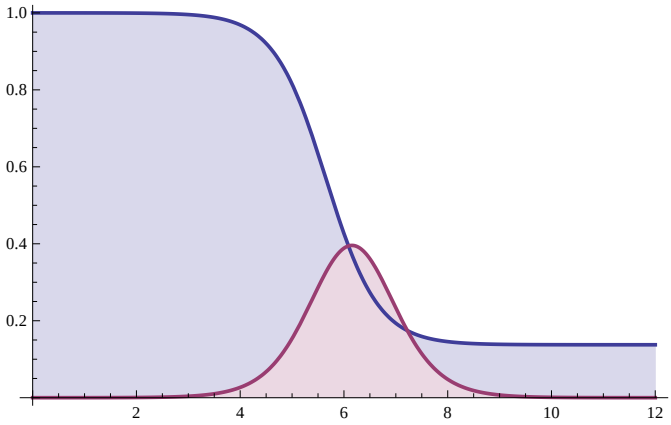

Figure 4: Evolution of (23) subject to (24), with $\beta = 2.3$, $\xi = 1$ and $I(0) = 10^{-5}$

3.6 Simulation of dynamics

As with the K-M SIR model, we can carry out simulations to demonstrate that our deterministic system makes useful predictions about the underlying stochastic processes. This simulation is similar to the previous one, and the code can be found in the appendix.

Below is a plot showing $I(t)$ as a proportion of N for three simulations (red) alongside the numerical solution of (23) (blue):

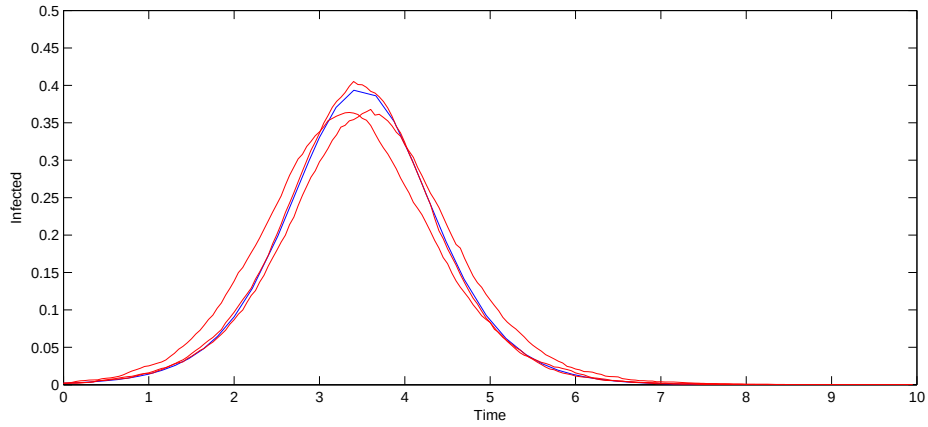

Figure 5: Outcome of simulations with parameters $\beta = 2.3$ and $\xi = 1$, giving $r \approx 1.98$

3.7 Comparison with the K-M SIR model

Before comparing the models like-for-like, let us investigate how the initial growth rate, r, varies with β and ξ.

From (2) and (22), we see that the initial growth rates coincide (with value β) in the absence of recovery. In the K-M SIR model, 'no recovery' is achieved by setting $\gamma = 0$ while in the non-markovian model we set $\xi = \infty$.

Due to the exponential term in (22), an increase in ξ from 1 causes very swift growth of r towards β which means that, for a K-M SIR epidemic to have the same initial growth rate, γ must be almost zero. Thus, comparing cases with equal β and r, with $\xi > 1$, the non-markovian model exhibits significantly reduced total infection, as the K-M SIR recovery mechanism becomes too slow at removing people from the infected state.

In the case where ξ is small, the rate of infection becomes very low. This pushes γ higher in the K-M SIR model if one wishes to maintain the same initial rate of increase, but still the fixed recovery period gives reduced total infection.

In conclusion, if one fixes the transmission rate β and wishes to maintain the same initial rate of infection increase, r, the non-markovian model has significantly reduced total infection.

We now turn our attention to the mass action assumption used in these previous two models.

4 Pairwise SIR model

In certain circumstances the mass action approximation may be warranted. Plague deaths in Bombay, for instance, used to demonstrate Kermack and McKendrick's original model, do make good use of mass action; the underlying vector causing infection is rats. Assuming that rats are fairly uniformly distributed, it is reasonable to assume that infection in humans could be devoid of any structure in human interactions, which would allow mass action to describe the dynamics well.

If, however, we consider other epidemics transmitted human-to-human via contact or proximity, we might be better off taking into account that some members of the population will not interact with others. In order to model SIR dynamics in this way, we turn our attention to networks.

4.1 Graphs and pairwise variables

Let us consider a graph, $\mathbb{G} = \mathbb{G}(V, E)$, which is defined as a finite non-empty set V together with an irreflexive, symmetric relation R on V. We denote by E the set of symmetric pairs in R[11].

Here, V is the *vertex set* and E is the *edge set*. We now consider the population as a graph; each individual is represented by a vertex, so $V = \{v_1 \dots v_N\}$ is the vertex set representing a population with N individuals. The model is, again, compartmental, so that each v_j is assigned a state, either S, I or R. We will use the shorthand $v_j = I$, for instance, to convey that the j^{th} individual is currently infected.

To keep track of which individuals in a population are connected (which, for us, will indicate that infection can pass between them), we introduce the *adjacency matrix* which is an $N \times N$ matrix $G = (G_{jk})$ associated with $\mathbb{G}$, defined as:

$$G_{jk} = \begin{cases} 1 & \text{if vertices } j \text{ and } k \text{ are linked} \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

We assume that infection can travel both ways, so we do not care that the links be ordered. This means G will be symmetric, and so $G = G^T$. In this setting, it does not make sense for an individual to be able to infect themself, so $G_{jj} = 0 \quad \forall j$.

We are now in a position to define the variables we will use in our model. The key concept behind this model will be that infection can only happen between a susceptible and an infected individual if they are linked. Thus, we will need language to describe not just the number of individuals in each state but the number of pairs and triples, which are strings of individuals linked on $\mathbb{G}$.

In the underlying stochastic model, at each time we will care about the number of $S-S$ pairs (the number of connections which have a susceptible vertex at each end), the number of $S-I$ pairs (the number of connections which have a susceptible vertex at one end and an infected vertex at the other), and so on. To write down a deterministic model,

however, we will be interested in the expected values of such quantities. To that end, we define the following time-dependent variables, where each variable is still time-dependent; however, we omit the explicit reminder of this ($[X]$ rather than $[X](t)$) in order to keep the equations reasonably succinct:

$$\begin{aligned}
[X] &= \mathbb{E}\left(\sum_{j=1}^{N} \mathbb{I}\{v_j \in V \,:\, v_j = X\}\right) \\
[XY] &= \mathbb{E}\left(\sum_{j,k=1}^{N} \mathbb{I}\{v_j, v_k \in V \,:\, v_j = X,\, v_k = Y,\, G_{jk} = 1\}\right) \\
[XYZ] &= \mathbb{E}\left(\sum_{j,k,l=1}^{N} \mathbb{I}\{v_j, v_k, v_l \in V \,:\, v_j = X,\, v_k = Y,\, v_l = Z,\, G_{jk} = 1,\, G_{kl} = 1\}\right)
\end{aligned} \tag{26}$$

where each of X, Y and Z can be any of S, I or R, $\mathbb{E}$ represents the expectation, and $\mathbb{I}$ is the indicator function.

4.2 Constructing the deterministic pairwise SIR model

In this section, we will revert to markovian recovery in order to familiarize ourselves with the workings of this model before relaxing the assumption again later. For the infection dynamics, this means that recovery happens again at constant rate γ, but now infection transform $S-I$ pairs into $I-I$ pairs at a constant rate, τ.

Thus, the number of susceptible individuals $[S]$ is depleted at rate $\tau[SI]$, which is also the rate of increase of infected individuals $[I]$. Infection is, as in Chapter 2, depleted at rate $\gamma[I]$, which is the rate of increase in $[R]$. So far, this looks very similar to the K-M SIR model, but we must also form expressions for the dynamics of the pairs.

For clarity of notation, let us adopt the notation that $[X \hookrightarrow YZ]$ indicates the action of a vertex in state X onto a $Y-Z$ pair. Then, $[SS]$ is depleted by $[I \hookrightarrow SS]$ and $[SS \hookleftarrow I]$. $[SI]$ is replenished by $[SS \hookleftarrow I]$ but depleted by $[I \hookrightarrow SI]$, $[S \hookleftarrow I]$ and recovery of the I vertex in the $S-I$ pair. The other changes in doubles can be derived similarly and, putting them all together, this leads to the following system describing the dynamics of an average epidemic:

$$\begin{aligned}
[\dot{S}] &= -\tau[SI], & [\dot{SS}] &= -2\tau[SSI], \\
[\dot{I}] &= \tau[SI] - \gamma[I], & [\dot{SI}] &= \tau([SSI] - [ISI] - [SI]) - \gamma[SI], \\
[\dot{R}] &= \gamma[I], & [\dot{SR}] &= -\tau[ISR] + \gamma[SI], \\
& & [\dot{II}] &= 2\tau([ISI] + [SI]) - 2\gamma[II], \\
& & [\dot{IR}] &= \tau[ISR] + \gamma([II] - [IR]), \\
& & [\dot{RR}] &= \gamma[IR].
\end{aligned} \tag{27}$$

4.3 Closure assumption

This system completely describes the dynamics of such epidemics, but it is unclosed. There are several ways in which we could proceed from here. We could continue to write down equations for the triples in terms of quadruples; however, this would become very untidy and we would still be left with an unclosed system. Instead, we will make a closure assumption by writing triples in terms of pairs. Several possibilities exist for how to do this, but here we will assume that the graph is n-regular; that is, each vertex is connected to precisely n other vertices.

Under this assumption, let us derive an approximate form for an arbitrary triple.

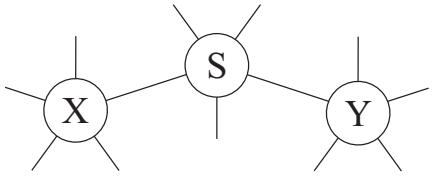

Figure 6: Vertex configuration

We are considering the vertex configuration pictured in Figure 6; $X - S - Y$ triples in an n-regular graph, $\mathbb{G}$. We seek an expression for $[ASB]$ in terms of $[SA]$, $[SB]$ and n.

First, we need to calculate $\mathbb{P}(X = A)$. As $\mathbb{G}$ is n-regular, there are $n[S]$ links emanating from all vertices in state S. Of these, precisely $[SX]$ link to a vertex in state X. Therefore, if $\mathbb{G}$ has N vertices, provided nN is large we have:

$$\mathbb{P}(X = A) = \frac{[SA]}{n[S]}, \text{ and similarly: } \mathbb{P}(Y = B) = \frac{[SB]}{n[S]} \tag{28}$$

Now, we consider $\mathbb{P}((X = A) \wedge (Y = B))$. We will assume that the two probabilities are independent of each other(**), so that we may simply multiply the results in (28). Under this assumption, we have:

$$\mathbb{P}((X = A) \wedge (Y = B)) = \mathbb{P}(X = A)\mathbb{P}(Y = B) = \frac{[SA][SB]}{n^2[S]^2} \tag{29}$$

Finally, around each vertex in state S there are $n(n-1)$ triples: n choices for the first vertex in the triple and $n-1$ choices remaining for the third vertex. Thus, in total, there are $n(n-1)[S]$ triples in $\mathbb{G}$ which have centre vertex in state S. Equation (29) tells us the probability that, in each possible triple, one vertex is in state A and one is in state B so we have, writing $\eta = \frac{(n-1)}{n}$:

$$[ASB] = \eta\frac{[SA][SB]}{[S]} \tag{30}$$

We have assumed that the probabilities (28) are independent. This need not necessarily be the case, and others have dealt with more sophisticated methods of providing a closure assumption for dependent probabilities[12]; however, it is reported that this assumption is

numerically extremely accurate for SIR dynamics[13], and so we will make use of it here.

4.4 Initial growth rate and final epidemic size

In this situation, we can easily reason that the initial growth rate for the epidemic will be $r = (n-2)\tau - \gamma$. If we only had one infected individual the rate would be just $n\tau - \gamma$ but, because we assume an epidemic starts with multiple infected individuals, each one must be connected to one vertex which infected them and, on average, one which they have infected. Thus, the average number of susceptible vertices adjacent to each initially infected individual is $n-2$; hence the factor which appears in the initial growth rate.

The final epidemic size can be calculated analytically[12][13], but it requires much more algebraically messy manipulations of the equations in (27) and so we shall not present it here. Of relevance, though, is that the final epidemic size calculated in (5) is obtained again from this model if we hold $\beta = n\tau$ constant while taking $n \to \infty$.

4.5 Solving the pairwise SIR model

We note that the system is still closed, in the sense that $[S] + [I] + [R] = 1$. Notice also that in order to fully describe the dynamics of $[S]$, $[I]$ and $[R]$, we need only concern ourselves with the pairs $[SS]$ and $[SI]$. Combining the original system (27) with the closure assumption (30), our dynamics are fully described by the following reduced system:

$$\begin{aligned} [\dot{S}] &= -\tau[SI], & [\dot{SS}] &= -2\tau\eta\frac{[SS][SI]}{[S]}, \\ [\dot{I}] &= \tau[SI] - \gamma[I], & [\dot{SI}] &= \tau\left(\eta\frac{[SS][SI] - [SI]^2}{[S]} - [SI]\right) - \gamma[SI]. \end{aligned} \tag{31}$$

This system can be solved in the same way as the K-M SIR system, for instance using `ode45` in Matlab or `NDSolve` in Mathematica. The dynamics are very similar to those of the K-M SIR model; however, infection spreads much more slowly in the pairwise case, with the final size of the epidemic being smaller. This is precisely what we would expect from intuition, as the population is now no longer able to mix uniformly; only certain individuals are connected, so infection is less freely able to propagate through the population.

Because of having to keep track of which individuals are connected in the system, it becomes much more computationally time-intensive to carry out simulations of these dynamics. For that reason, none are presented here and we move on to relaxing the non-markovian assumption used in this model.

5 Pairwise non-markovian SIR model

Using the tools and language built up in the non-markovian and pairwise SIR models, we aim to relax both the mass action and markovian recovery assumptions in the original K-M SIR model.

Relaxing markovian recovery, we can derive differential equations where, in complete generality, both $\tau(a)$ (infection across an $S-I$ pair) and $\gamma(a)$ (recovery of a vertex in state I) are arbitrary functions of age. This means that, to relax mass action as well, we must consider $[X]$, $[XY]$ and $[XYZ]$ where X, Y and Z can be any of S, I, i_a, or R. Thus, variables of the form $[Si_a]$, which represents the expected density of $S-I$ pairs where the infected individual has age a at time t, are of importance. Terms of this type (containing our two-dimensional variable $i_a(t,a)$) will, as in the non-markovian SIR model, involve our differential operator, Φ.

5.1 Deriving the non-markovian pairwise SIR PDE model

As with the standard pairwise SIR model, we can see that only equations for $[\dot{S}]$, $\Phi([S])$, $[\dot{SS}]$ and $\Phi([Si_a])$ will be necessary to fully describe the temporal dynamics of $[S]$, $[I]$ and $[R]$. We will, therefore, describe in some detail the derivation of these four equations, but will state later the full system including differential equations for all pairs.

$[S]$:

- Depleted only by infection: $[S \leftarrow i_a]$, but this is at rate $\tau(a)$ and can happen at any age a.

$$[\dot{S}] = -\int_0^\infty \tau(\alpha)[Si_\alpha]d\alpha \tag{32}$$

$[i_a]$:

- Replenished by infection: $[S \leftarrow i_a]$ but when an individual becomes newly infected they must start with age $a = 0$.
- Depleted through recovery: $[i_a] \rightarrow [R]$, which happens at rate $\gamma(a)$.

$$\Phi([i_a]) = \delta(a)\int_0^\infty \tau(\alpha)[Si_\alpha]d\alpha - \gamma(a)[i_a] \tag{33}$$

$[SS]$:

- Depleted by infection: $[i_a \hookrightarrow SS]$ and $[SS \leftarrow i_a]$, at rate $\tau(a)$, at any age a.

$$[\dot{SS}] = -2\int_0^\infty \tau(\alpha)[SSi_\alpha]d\alpha \tag{34}$$

$[Si_a]$**:**

- Depleted by: (i) $[S \leftarrow i_a]$, at rate $\tau(a)$.
- Depleted by: (ii) $[i_b \hookrightarrow Si_a]$, at rate $\tau(b)$, but for any age b.
- Depleted by: (iii) $[Si_a] \to [SR]$, at rate $\gamma(a)$.
- Replenished by: (iv) $[SS \leftarrow i_a]$, at rate $\tau(a)$, for any age a, but new infection must start with age 0.

$$\begin{aligned} &\text{(i) and (ii)} && -\tau(a)[Si_a] - \int_0^\infty \tau(\alpha)[i_\alpha Si_a]d\alpha \\ &\text{(iii)} && -\gamma(a)[Si_a] \\ &\text{(iv)} && \delta(a)\int_0^\infty \tau(\alpha)[SSi_\alpha]d\alpha \end{aligned} \tag{35}$$

Collecting these together gives:

$$\Phi([Si_a]) = \delta(a)\int_0^\infty \tau(\alpha)[SSi_\alpha]d\alpha - \int_0^\infty \tau(\alpha)[i_\alpha Si_a]d\alpha - (\tau(a)+\gamma(a))[Si_a] \tag{36}$$

5.2 Full set of equations

Some of these equations contain triples, so are not closed. We can apply our closure assumption (30) to close the system, and we will proceed with $\tau(a) = \tau = \text{ const.}$

Core equations

$$\begin{aligned} &\text{(i)} && [\dot{S}] = -\tau\int_0^\infty [Si_\alpha]d\alpha \\ &\text{(ii)} && \Phi([i_a]) = \delta(a)\tau\int_0^\infty [Si_\alpha]d\alpha \quad -\gamma(a)[i_a] \\ &\text{(iii)} && [\dot{SS}] = -2\tau\eta\frac{[SS]}{[S]}\int_0^\infty [Si_\alpha]d\alpha \\ &\text{(iv)} && \Phi([Si_a]) = \eta\tau\left(\frac{\delta(a)[SS]-[Si_a]}{[S]}\right)\int_0^\infty [Si_\alpha]d\alpha \quad -(\tau+\gamma(a))[Si_a] \end{aligned} \tag{37}$$

Secondary equations

$$
\begin{aligned}
&\text{(v)} && [\dot{R}] = \int_0^\infty \gamma(\alpha)[i_\alpha]d\alpha \\
&\text{(vi)} && [\dot{SR}] = \int_0^\infty \left(\gamma(\alpha) - \eta\tau\frac{[SR]}{[S]}\right)[Si_\alpha]d\alpha \\
&\text{(vii)} && \Phi([Ii_a]) = \frac{\tau\eta}{[S]}\left([Si_a] + \delta(a)[SI]\right)\int_0^\infty [Si_\alpha]d\alpha \\
& && \qquad - \int_0^\infty (\gamma(a) + \gamma(\alpha))[i_\alpha i_a]d\alpha \quad + \tau[Si_a] \\
&\text{(viii)} && \Phi([Ri_a]) = \eta\tau\frac{\delta(a)[SR]}{[S]}\int_0^\infty [Si_\alpha]d\alpha + \int_0^\infty \gamma(\alpha)[i_\alpha i_a]d\alpha \quad - \gamma(a)[Ri_a] \\
&\text{(ix)} && [\dot{RR}] = \int_0^\infty \gamma(\alpha)[Ri_\alpha]d\alpha
\end{aligned}
\tag{38}
$$

Now that we have our closed-form differential equations (37) describing the behaviour of our system, we next aim to solve them. We will proceed in the same manner as with the non-markovian SIR model; take the particular functional form for γ in which we are interested (a fixed recovery time ξ), find history functions for the interval $[0, \xi]$ and write down a DDE system which can be numerically solved for the temporal dynamics of S, I and R in this regime. The main step will be to calculate the initial epidemic growth rate, r, in this setting.

5.3 Initial epidemic growth

In order to calculate r, we will need to consider equations (37)(ii) and (37)(iv) at early time. We will solve them in bulk (ignoring boundary conditions), then make an ansatz which will allow us to integrate the equations and find the unknown initial growth rate.

As we start our epidemic with a very small proportion of the population infected, we know that $[i_a]$ and $[Si_a]$ will be small. Equation (37)(ii) is already linear in $[i_a]$, in bulk, but we must linearise equation (37)(iv). This is straightforward as we can simply discard the $[Si_a]\int_0^\infty [Si_\alpha]d\alpha$ term, which is $O([Si_a]^2)$.

We can simplify the boundary conditions as well; at early time when there is nearly no infection, $[SS] \approx nN$ and $[S] \approx N$. Switching notation, we can also write $[SI] = \int_0^\infty [Si_\alpha]d\alpha$, which will make the equations somewhat less cumbersome.

Hence the PDEs we are left with to solve, at early time, are:

$$
\begin{aligned}
&\text{(i)} && \Phi([i_a]) = -\gamma(a)[i_a] \\
&\text{(ii)} && \Phi([Si_a]) = -(\tau + \gamma(a))[Si_a],
\end{aligned}
\tag{39}
$$

subject to the boundary conditions:

$$\begin{aligned} &\text{(i)} \quad [i_a](t,0) = \tau[SI](t) \\ &\text{(ii)} \quad [Si_a](t,0) = (n-1)\tau[SI](t), \end{aligned} \tag{40}$$

Both of the bulk PDEs in (39) are in the same form as the bulk PDE (11), and so the solutions, obtained by separation of variables, will be the same:

$$\begin{aligned} &\text{(i)} \quad [i_a](t,a) = f(t-a)e^{-\int_0^a \gamma(\alpha)d\alpha} \\ &\text{(ii)} \quad [Si_a](t,a) = g(t-a)e^{-\tau a}e^{-\int_0^a \gamma(\alpha)d\alpha} \end{aligned} \tag{41}$$

for arbitrary functions $f(t-a)$ and $g(t-a)$.

We now make the ansatz that the infection initially grows exponentially; that is, for small time, $[SI] = k_2 e^{rt}$ for some constant k_2 and some unknown growth rate r. Using the initial conditions (40), this gives us that $f(t-a) = \tau k_2 e^{r(t-a)}$ and $g(t-a) = (n-1)\tau k_2 e^{r(t-a)}$, in the same manner as for the non-markovian SIR model. This pins down both of our arbitrary functions $f(t-a)$ and $g(t-a)$, and updates our solutions as follows:

$$\begin{aligned} &\text{(i)} \quad [i_a](t,a) = \tau k_2 e^{r(t-a)}e^{-\int_0^a \gamma(\alpha)d\alpha} \\ &\text{(ii)} \quad [Si_a](t,a) = (n-1)\tau k_2 e^{r(t-a)}e^{-\tau a}e^{-\int_0^a \gamma(\alpha)d\alpha}. \end{aligned} \tag{42}$$

Integrating (42)(ii) over age gives us the unknown growth rate, r:

$$k_2 e^{rt} = [SI](t) = \int_0^\infty [Si_a](t,a)da = \int_0^\infty \left((n-1)\tau k_2 e^{r(t-a)}e^{-\tau a}e^{-\int_0^a \gamma(\alpha)d\alpha}\right)da. \tag{43}$$

We can simplify $e^{-\int_0^a \gamma(\alpha)d\alpha}$ in the same way as in the non-markovian SIR model and, cancelling $k_2 e^{rt}$ from both sides, we are left with the following integral equation for r:

$$1 = \int_0^\xi (n-1)\tau e^{-(r+\tau)a}da, \tag{44}$$

which simplifies to give the following transcendental equation for r:

$$r + \tau = (n-1)\tau\left(1 - e^{-(r+\tau)\xi}\right). \tag{45}$$

It is worth noting that this equation coincides with the initial growth rate in the markovian pairwise SIR model under the condition of no recovery. If there is no recovery ($\gamma = 0$ or, equivalently, $\xi = \infty$), the initial growth rate in both models is $r = (n-2)\tau$. This gives credibility to the answer and also reassures us that our ansatz was reasonable.

5.4 History functions for DDE model

We require a history function for each of the $[S]$, $[I]$, $[SS]$ and $[SI]$ equations as these are the four necessary to fully solve for the temporal behaviour of this non-markovian pairwise

SIR model. We have, however, already got two: $[SI](t) = k_2 e^{rt}$, where $k_2 = [SI](0)$, and so $[SS](t) = n - k_2 e^{rt}$. It remains to find the history function for $[I]$.

Clearly, $[I]$ will have the same initial growth as $[SI]$, so $[I] \propto e^{rt}$, but what is the initial condition $[I](0)$? Call the constant of proportionality k_1. It turns out that, with the assumptions we have made in this regime, k_1 is determined by integrating (42)(i) over age:

$$k_1 e^{rt} = [I](t) = \int_0^\infty [i_a](t,a)da = \int_0^\infty \left(\tau k_2 e^{r(t-a)} e^{-\int_0^a \gamma(\alpha)d\alpha}\right) da \tag{46}$$

Simplifying this in the same manner as (43), we obtain the following expression for the ratio $k_1 : k_2$, i.e. $[I](0) : [SI](0)$:

$$\frac{k_1}{k_2} = \frac{\tau}{r}\left(1 - e^{-\xi r}\right). \tag{47}$$

We now have all the information necessary for the four history functions:

$$\begin{aligned}
&\text{(i)} && [S](t) = 1 - k_1 e^{rt} && \forall t \in [0,\xi] \\
&\text{(ii)} && [I](t) = k_1 e^{rt} && \forall t \in [0,\xi] \\
&\text{(iii)} && [SS](t) = n - k_2 e^{rt} && \forall t \in [0,\xi] \\
&\text{(iv)} && [SI](t) = k_2 e^{rt} && \forall t \in [0,\xi].
\end{aligned} \tag{48}$$

5.5 DDE model

In principle we should now be able to write down a DDE system as we did in the case of the mass-action non-markovian SIR model.

The equations corresponding to the first three history functions in (48) are easy to write down:

$$\begin{aligned}
&\text{(i)} && [\dot{S}](t) = -\tau[SI](t) \\
&\text{(ii)} && [\dot{I}](t) = \tau[SI](t) - \tau[SI](t-\xi) \\
&\text{(iii)} && [\dot{SS}](t) = -2\tau[SSI](t),
\end{aligned} \tag{49}$$

but the fourth is not. Only $\tau[SSI](t-\xi)$ will involve any delay in the fourth equation as it is the only term which creates new $S-I$ pairs at time $t-\xi$. The problem is that in the interval $[t-\xi, t]$ it is possible for those $S-I$ pairs to become $I-I$ pairs due to infection across their link.

It is, therefore, more difficult to pin down the form for this final equation, so the resolution of this is not presented here.

This concludes our treatment of the four different SIR models.

6 Conclusions and extensions

6.1 What do these different models tell us?

We have now explored a basic epidemiological model and have adapted it in three different ways, each designed to relax different assumptions.

What is clear is that, despite the changes made to the original K-M SIR model, each new model has roughly the same qualitative behaviour. All four models capture the increase and subsequent decrease in the proportion of the population infected during an epidemic.

One might argue that, as there is no biological justification for assuming markovian recovery, one should always attempt to use a more sophisticated model with biological underpinning. As we have seen, though, this complicates the problem significantly, while leaving the result qualitatively very similar. Therefore, in the case of the markovian recovery assumption it remains unclear (and, indeed, subjective) whether or not either model is 'better' than the other.

Regarding the mass-action assumption, we can say something slightly stronger. There are cases (such as bubonic plague) in which mass action might describe the epidemic dynamics well, whereas epidemics which spread through contact are likely to be modelled much more accurately if a pairwise model is used. In this regard, then, the choice of model depends very much on the specific epidemic being studied.

In conclusion, one must always hope to understand the biology behind the epidemic being modelled in order to best set up an epidemiological model to make accurate predictions. It should be noted, however, that non-markovian pairwise models do allow mathematical investigation, and so are a potentially useful avenue in modelling epidemics.

6.2 Where can we go from here?

There are several ways in which this investigation could be taken forward.

With the two markovian models we discussed it is straightforward to include additional information about the population. For instance, if a particular epidemic is likely to occur over quite a large timescale, one might wish to add birth and death rates to each variable. Additionally, the markovian models can include a variety of extra components such as vaccination or immunity[13]. Finally, if a disease is potentially fatal, and a significant proportion of the population might become infected, death of individuals will reduce the population size. All of these amendments could, in theory, be incorporated into non-markovian pairwise models.

Another avenue worth exploration is altering the rate of infection, τ. When deriving the PDE system (37) we initially allowed $\tau(a)$ to be an arbitrary function of age. This was relaxed in order that τ be taken outside the integrals, which enabled our ansatz to work. In reality, it is unlikely that an infected individual is as infectious at all times between catching a disease and recovering from it; each disease will have its own infectiousness

profile $\tau(a)$ which could be incorporated to make the dynamics of certain infections more realistic.

In terms of mass action, our models assumed the underlying graph was n-regular. Of course, this is not always going to be the case; normally in networks we would expect clustering. Thinking in terms of human interaction, some people are much more gregarious than others. These people are likely to have a large number of connections which makes them both more likely to become infected and more effective at disseminating further infection. The assumption of n-regularity was convenient for providing a simple closure assumption, but closure assumptions better tuned to certain networks should be incorporated into models for use in those circumstances.

Finally, we have taken no account of closed loops in networks. The number of closed triangles is a quantity of importance in epidemiological modelling[12], and account of this could be taken by means of altering the closure assumption we made.

I hope that this exploration of SIR dynamics demonstrates that new models can be investigated with relatively simple mathematical tools, and I would urge interested readers to investigate further the models presented here.

7 Appendix A: Matlab and Mathematica code

7.1 K-M SIR simulation

```
% Model parameters
N  = 5000;
I0 = 10;
b  = 0.25/N;
g  = 0.06;
T  = 120;

% Compartment vectors
Sus = zeros(1,T);
Inf = zeros(1,T);
Time= 0:1:T-1;

% Initial conditions
Sus(1) = N-I0;
Inf(1) = I0;

for i=1:T-1
    % Localise variables to reduce vector accesses
    tempInf = 0;
    currInf = Inf(i);
    currSus = Sus(i);

    if currSus > 0
        Sus(i+1) = currSus;

        % Break if no remaining infection
        if currInf == 0
            break;
        end

        for j=1:currSus
            % Vector of random numbers
            randMat = rand(1,currInf);
            % Check if one is below threshold
            if (sum(randMat < b) > 0)
                %Sus(i+1) = Sus(i+1) - 1;
                tempInf = tempInf + 1;
            end
        end
        Sus(i+1) = Sus(i+1) - tempInf;
    end

    % Calculate # inf who recover
    randMat = rand(1,currInf);
    newRec = sum(randMat < g);
    Inf(i+1) = Inf(i) + tempInf - newRec;
end
```

7.2 Non-markovian SIR simulation

```
% Model parameters
steps = 20;
N     = 5000;
I0    = 10;
b     = 2.3/(N*steps);
g     = 0.06;
T     = 10;
inver = 1/steps;

% Compartment vectors
Sus = zeros(1,steps*T);
Inf = zeros(1,steps*T);
Time= 0:inver:T-inver;
Prev= zeros(1,steps*T);

% Initial conditions
Sus(1) = N-I0;
Inf(1) = I0;
Prev(1+steps) = I0;

for i=1:steps*T-1
    % Local variables to reduce index calls
    tempInf = 0;
    currInf = Inf(i);
    currSus = Sus(i);

    if currSus > 0
        Sus(i+1) = currSus;

        % Break if no remaining infection
        if currInf == 0
            break;
        end

        for j=1:currSus
            % Vector of random numbers
            randMat = rand(1,currInf);
            % Check if one is below threshold
            if (sum(randMat < b) > 0)
                %Sus(i+1) = Sus(i+1) - 1;
                tempInf = tempInf + 1;
            end
        end
        Sus(i+1) = Sus(i+1) - tempInf;
    end

    % Update inf vector
    Prev(i+1+steps) = tempInf;
    Inf(i+1) = currInf + tempInf - Prev(i+1);
end
```

7.3 Sample use of NDSolve for DDE system

```
ξ = 1;
I0 = 10^-5;
tend = 12;
r = (x /. NSolve[x == β * (1 - Exp[-x ξ])])[[2]];
eqns = {
   Sus'[t] == -β*Sus[t]*Inf[t],
   Inf'[t] == β*Sus[t]*Inf[t] - β*Sus[t - ξ]*Inf[t - ξ],
   Sus[t /; t ≤ ξ] == 1 - I0 Exp[r t],
   Inf[t /; t ≤ ξ] == I0 Exp[r t]
  };
sol = NDSolve[eqns, {Sus[t], Inf[t]}, {t, 0, tend}];
Plot[Evaluate[{Sus[t], Inf[t]} /. sol], {t, 0, tend}, PlotStyle → Thick, Filling → Axis]
```

7.4 Sample use of ode45 for ODE system

```
function sir_solver = nr_sir_m
% Parameter declarations
N    = 10e6;        % Population size
n    = 5;           % Contacts per node (n-regular graph assumed)
tau  = 0.05;        % Transmission rate of infection accross link
gam  = 0.06;        % Recovery rate of I node (const --> Markovian)
I0   = 10;          % Initial number of infected individuals
MaxT = 200;         % Simulation length

% Set initial conditions
inCon = zeros(7,1);
inCon(1) = N-I0; inCon(2) = I0;
inCon(3) = n*(0.5*N - I0); inCon(4) = n;

% Solve the equaton
options = odeset('RelTol', 1e-4);           % Custom tolerance for solver

[t, pop]=ode45(@de_nr_sir_m,[0 MaxT], inCon, options, N, n, tau, gam);

S=pop(:,1); I=pop(:,2);          % Pure vertex vectors
SS=pop(:,3); SI=pop(:,4);        % Link vectors

% Function storring RHS of differential equations for ode45 to use
function diff=de_nr_sir_m(t, pop, N, n, tau, gam)

S=pop(1); I=pop(2);              % Pure vertex types
SS=pop(3); SI=pop(4);            % Link types

frac=(n-1)/n;       % Fraction used in closure assumption
diff=zeros(4,1);    % Vector for RHS of differential equations

% Differential equations
diff(1)=-tau*SI;
diff(2)=tau*SI-gam*I;
diff(3)=-2*tau*frac*SS*SI/S;
diff(4)=tau*frac*(SS*SI/S - SI*SI/S) - (tau+gam)*SI;
```

Based on code written by M. Keeling and P. Rohani[14, §7]

Bibliography

References

[1] W. Kermack and A. McKendrick, "A contribution to the mathematical theory of epidemics," *Proc. R. Soc. Lond. A*, vol. 115, pp. 700–721, August 1927.

[2] R. Anderson and R. May, "Population biology of infectious diseases: Part I," *Nature*, vol. 280, pp. 361–367, August 1979.

[3] A. Alex, H. Gaff, S. Lenhart, and E. Schaefer, "An age-structured model for the spread of epidemic cholera: Analysis and simulation," *Nonlinear Analysis: Real World Applications*, vol. 16, July 2011.

[4] M. Keeling and C. Gilligan, "Bubonic plague: a metapopulation model of a zoonosis," *Proc. R. Soc. Lond. B*, vol. 267, pp. 2219–2230, November 2000.

[5] R. Casagrandia, L. Bolzonib, S. Levinc, and V. Andreasen, "The SIRC model and influenza A," *Mathematical Biosciences*, vol. 200, pp. 152–169, April 2006.

[6] A. Negi and S. Anand, *Textbook of physical chemistry.* New Delhi: New Age International, 1985.

[7] W. Feller, *Introduction to Probability Theory and Its Applications*, vol. 2. Wiley, 2nd ed., 1971.

[8] M. Keeling, *Population Dynamics (Lecture Notes).* University of Warwick, 2011.

[9] Wolfram Research, *Total derivative.* `http://mathworld.wolfram.com/TotalDerivative.html`, 2013.

[10] G. Smith, *Numerical Solution of Partial Differential Equations: Finite Difference Methods.* Great Clarendon Street, Oxford: Oxford University Press, 3rd ed., 1985.

[11] G. Chartrand, *Introductory Graph Theory.* Dover Publications, 1977.

[12] M. Keeling, "The effects of local spatial structure on epidemiological invasions," *Proc. R. Soc. Lond. B*, vol. 266, p. 859867, April 1999.

[13] T. House, "Modelling epidemics on networks," *Contemporary Physics*, vol. 53, no. 3, pp. 213–225, 2012.

[14] M. Keeling and P. Rohani, *Modelling Infectious Diseases in Humans and Animals.* Princeton University Press, 2007.