A Study of SIR Network models

Thomas Solomon
jjsb59

Department of Mathematics and Statistics
University of Durham

September 2, 2021

Abstract

This paper discusses the use of Networked SIR and SIS Models and their advantages over the traditional form of the SIR and SIS and their extensions. The theoretical basis of these methods is presented and explained and a variety of practical examples are presented to guide the reader's understanding. The performance and applicability of these approaches are also discussed and compared.
All code used to generate our simulations and graphs can be found within:
https://github.com/Tlight1/SIRModelCode.

Acknowledgments
I would like to thank my supervisor Adam for being an exceptional supervisor and providing excellent discussion and support throughout the course of writing this paper.

I would also like to extend a thank you to my father for his support throughout the entirety of the year, giving me an outlet to discuss my mathematics throughout the various lockdowns during the pandemic and helping me present my project in a vaguely coherent manner.

I would also like to give a special thank you to Jessica Bahn for all her encouragement throughout my project, for providing the data for the Love Island experiment but most importantly for saving me from having to watch the show.

Plagiarism Declaration

I declare that this report is a result of my own work. Material from the work of others not involved in the project has been acknowledged and quotations and paraphrases suitably indicated.

Contents

1 Introduction

Throughout history mankind has battled against a huge range of diseases, from the bubonic plague in 1348 England, to the great famine in Ireland beginning in 1845 (caused by a fungal infection ruining the potato crop), through to today where we are in the midst of the Covid-19 pandemic. Scientists and epidemiologists have continually tried to invent solutions and measures to help combat disease and its spread. During the bubonic plague infected households in London were quarantined, cattle infected with bovine TB are immediately slaughtered to avoid passing on the infection and in our present time the UK and other governments have legislated to place the entire country into lockdown in an attempt to minimise the spread of the Covid-19 virus. Today we have access to powerful computers and can draw on a substantial legacy of mathematical and epidemiological knowledge. In this paper I use this knowledge to create mathematical models to understand the factors that affect the spread of an infection amongst a population.

First we need to consider the key aims of modelling a pandemic infection. Primarily we are concerned with 3 things - the predicted duration of the pandemic, the timing and intensity of the peak of the infection and (typically) the maximum number of infections that medical services can manage. One method to model the spread of an infection is the SIR model first described by Bernoulli in 1760.[5] SIR describes the time evolution of a disease spread through a population by first classing each individual as either susceptible, infected or recovered. The model is then generated traditionally by solving three simultaneous ordinary differential equations. This has a complete analytic solution but can also be numerically approximated through finite difference methods such as Euler's method described in section 2. The SIS model however considers the time evolution of a disease throughout a population by classing individuals as either infected or susceptible only. In this paper I will present and compare the advantages and disadvantages of constructing a network model for our SIR and SIS model, making reasonable assumptions and simplifications to our network. [18, see page 2]

2 The SIR & SIS models

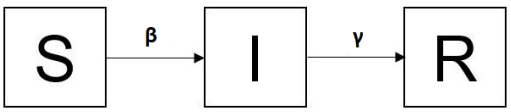

Figure 1: A visual representation of the transition between each of our three states in our SIR model

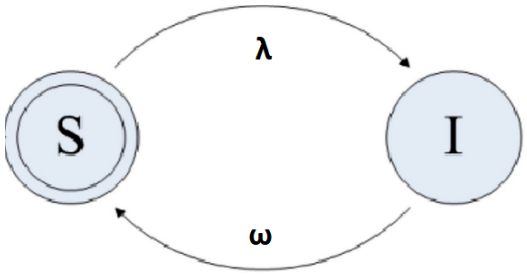

Figure 2: This is a visual representation of the transition between the two states within our SIS model

In this section we define the ODEs which govern the SIR and SIS models. We will split our population into three categories. We can be in either S: susceptible to the virus, I: currently

infected with the virus and critically, infectious if exposed to a susceptible individual or R: the individual has recovered from their infection and is no longer capable of spreading the virus or becoming reinfected. In particular we will be interested in endemic diseases, which are defined as diseases which exist only within a particular group of people or country. This is so we can treat our population as fixed and contained within a particular region where cultural differences should not come into effect.

2.1 The ODE SIR Model

As we are interested in an endemic disease we do not include birth and death rate and treat our population as a constant.
The variables we use are defined as follows. $S(t)$ denotes the population of susceptible individuals at time t, $I(t)$ denotes the population of infected individuals at time t and $R(t)$ is the population of those who have recovered (or died) and by our own model assumptions can no longer become infected or susceptible again. Parameters β and γ represent the infection rate and the recovery rate respectively with $\beta, \gamma \geq 0$. For the rest of this paper $u'(t)$ denotes the 1st derivative of any generic function $u(t)$.
This is the ODE governing the rate of change for our susceptible group S. We have population leaving the group proportionally to the product of remaining susceptible population and infected population normalised by N. This product represents the interactions between infected and susceptible people which is the cause for infection. [4]

$$S'(t) = -\frac{\beta I(t) S(t)}{N}. \tag{1}$$

This is the ODE governing the rate of change for our infected group I. We have the same proportion of people joining our group as we have leaving S. Our second term is the proportion of infected individuals who make a recovery and join the group R [4]

$$I'(t) = \frac{\beta I(t) S(t)}{N} - \gamma I(t). \tag{2}$$

This is the ODE governing the rate of change of our recovered group R. γ represents the proportion of infected individuals who recover in a time step.[4]

$$R'(t) = \gamma I(t), \tag{3}$$

with the initial conditions determined by the initial state of the population at our time $t = 0$. As we are modelling at the beginning of an outbreak we require the recovered population to be zero while the size of the infected population must not be zero otherwise nothing will change in time:

$$S(0) = a, \;\; I(0) = b, \;\; R(0) = 0, \;\; N > a > 0, \;\; N > b > 0. \tag{4}$$

We must also consider the constraint on S, I and R,

$$S(t) + I(t) + R(t) = N, \;\; \forall t, \tag{5}$$

Now we have defined the set of differential equations we need to solve, we can determine the equilibrium points of this system. An equilibrium point is defined as a point in space

where all variables no longer change over time. An equilibrium point can be defined as either stable or unstable. A stable equilibrium means that if we perturb the independent variable(s) some small amount from the equilibrium point, the system will in time return to that same equilibrium state. By contrast, an unstable equilibrium will mean that small perturbations do not result in return to that same equilibrium state as time progresses. In order to find our equilibrium points for the SIR model we need to set all three equations 1, 2 and 3 to zero and solve for appropriate values of S, I and R:

$$0 = -\frac{\beta I(t)S(t)}{N} = \frac{\beta I(t)S(t)}{N} - \gamma I(t) = \gamma I(t). \tag{6}$$

In this particular case, the equation above can only hold true if we set $I = 0$, therefore it is clear that we will reach a state of equilibrium when nobody is infected anymore and only then. This model is idealised from the behaviour of many diseases in reality. Although we will certainly reach an equilibrium point when the infection dies out we will have also have a steady growth of population due to new births who join the susceptible group (e.g. children getting chicken pox). It is interesting to note that a disease will last longer if it is moderately infectious as this allows time for newly susceptible individuals to join the population (see figure 4). By contrast, a highly infectious disease rapidly spreads such that the entire population quickly transitions into either the recovered (or dead) category and thus there will be no more susceptible persons to infect. Similarly, for a disease with a very low infection rate the infected population will join the recovered category before new susceptible individuals can become infected and the virus will die out. This comparison is shown by examining figures 3, 4 and 5.

We can in fact find an implicit solution for S and R by dividing equations (1) by (3) to get

$$\frac{S'(t)}{R'(t)} = \frac{-\beta S}{N\gamma}. \tag{7}$$

Now we can use separation of variables to solve this equation to get:

$$S(t) = S(0)\exp\left(-\frac{\beta}{\gamma}\frac{R(t) - R(0)}{N}\right). \tag{8}$$

We can look at the long term behaviour of our model now as $t \to \infty$. We define for simplicity $s_0 = \frac{S(0)}{N}$ to be the proportion of susceptible individuals at time $t = 0$ while $s_\infty = \frac{S(\infty)}{N}$ and $r_\infty = \frac{R(\infty)}{N}$ are the proportions of susceptible and recovered (or dead) individuals as $t \to \infty$. Now in this limit (as we show in section 2.1.2) we are tending towards our equilibrium state and this can only happen when the proportion of the population in the infectious category reaches zero. We can therefore say

$$s_\infty = 1 - r_\infty = s_0\exp\left(-\frac{\beta}{\gamma}[r(\infty) - r(0)]\right). \tag{9}$$

This equation shows us, subject to the assumption that recovered persons do not rejoin the susceptible group, that at the end of an epidemic not all individuals of the population have been infected: some remain susceptible unless $s_0 = 0$. We can see that the right hand side of our equation only takes values $= 0$ when s_0 is also zero as a property of the exponential

function is it will never reach zero (although the limiting behavior as $e^{-\infty}$ does indeed tend to 0). Therefore the main reason for an epidemic ending is usually a decline in the infectious population not the lack of a susceptible population.

2.1.1 Importance of β and γ

We want to examine what impact the parameter β/γ (β/γ is often called the reproductive ratio) has on the probability of an epidemic outbreak occurring. An outbreak is defined as a sudden spike in the number of endemic cases (which are cases confined to a specific area). If not sufficiently controlled, this can lead to an epidemic. First we rewrite our second Equation (2) for the infectious population as

$$I'(t) = \left(\frac{\beta}{\gamma}\frac{S(t)}{N} - 1\right)\gamma I(t). \tag{10}$$

We are considering the rate of change of the infectious population. It follows that if $\frac{\beta}{\gamma}S(0) < N$ our rate of change at time $t = 0$ will be less than one and, as such, our infectious population will decrease over time and no outbreak will not occur. By contrast, if $\frac{\beta}{\gamma}S(0) > N$, our infectious population will increase and an epidemic outbreak is possible such that infections can increase to a significant proportion of the overall population until $\frac{\beta}{\gamma}S(t) \leq N$. The exact proportion of the population infected is determined by our reproductive ratio and a value close to one will naturally yield a smaller infected population than a larger one. Clearly the relationship between the initial fraction of susceptible individuals and the ratio between γ and β is extremely important to the long term behaviour of our SIR model.

2.1.2 Linear stability analysis

In order to understand something about the long term behaviour and life cycle of a disease, we need to understand more about the equilibrium states of the system. In particular we want to know whether or not the equilibrium points we have found are stable or not (this analysis may be more important when we look at the SIS model in section 2.2). To perform stability analysis we must first find the equilibrium of our system. We require $S'(t) = I'(t) = R'(t) = 0$ giving us a set of 3 simultaneous equations to solve. In the particular case of the SIR model, there is just one equilibrium point occurring when the infected population is zero i.e. $I(t) = 0$. S and R can take any value such that $S + R = N$ and $R, S \leq N$.
Next we make a small perturbation about that equilibrium, i.e. $I(t) = 0 + \epsilon I_1(t)$, $R(t) = R_0 + \epsilon R_1(t)$ and $S(t) = S_0 + \epsilon S_1(t)$ and our equations become -

$$\epsilon S_1'(t) = -\frac{\beta(0 + \epsilon I_1(t))(S_0 + \epsilon S_1(t))}{N} \tag{11}$$

$$\epsilon I_1'(t) = \frac{\beta(0 + \epsilon I_1(t))(S_0 + \epsilon S_1(t))}{N} - \gamma(0 + \epsilon I_1(t)) \tag{12}$$

$$\epsilon R_1'(t) = \gamma(0 + \epsilon I_1(t)) \tag{13}$$

We ignore all terms in ϵ^2 and only consider terms of order ϵ to get the following $O(\epsilon)$ equations.

$$S_1'(t) = -\frac{\beta I_1(t) S_0}{N} \tag{14}$$

$$I_1'(t) = \frac{\beta I_1(t) S_0}{N} - \gamma I_1(t) \tag{15}$$

$$R_1'(t) = \gamma I_1(t) \tag{16}$$

We can express these equations in matrix form as -

$$\frac{d}{dt}\begin{pmatrix} S_1 \\ I_1 \\ R_1 \end{pmatrix} = \begin{pmatrix} 0 & -\frac{\beta S_0}{N} & 0 \\ 0 & \frac{\beta S_0}{N} - \gamma & 0 \\ 0 & \gamma & 0 \end{pmatrix}\begin{pmatrix} S_1 \\ I_1 \\ R_1 \end{pmatrix} \tag{17}$$

where the 3×3 matrix is commonly referred to as the Jacobian matrix $\mathbf{A}$ and this governs our solutions for S_1, I_1 and R_1 which all take the form $C\exp^{\lambda t}$. Substituting this form for the solution back into our equation and rearranging we get -

$$(\mathbf{A} - \lambda \mathbf{I})\begin{pmatrix} S_1 \\ I_1 \\ R_1 \end{pmatrix} = \mathbf{0}$$

This is of course an eigenvalue problem with permissible eigenvalues given by solving $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$. We now substitute our equilibrium point $(s_0, 0, r_0)$ to get our eigenvalue equation as:

$$(-\lambda)\left(\frac{\beta s_0}{N} - \gamma - \lambda\right)(-\lambda) = 0, \tag{18}$$

which is equivalent to

$$\lambda^2\left(\frac{\beta s_0}{N} - \gamma - \lambda\right) = 0. \tag{19}$$

In order for our system to be stable we require that the real part of all our eigenvalues be ≤ 0. The only non-zero eigenvalue is equal to $(\beta s_0)/N - \gamma$. This eigenvalue is positive and hence the equilibrium is unstable if $(\beta s_0)/N > \gamma$. Conversely, the eigenvalue is negative and equilibrium unstable if it is not. An interesting point is that the maximum value of $s_0 = N$ which means our eigenvalue depends purely on the ratio of β/γ and regardless of the value of s_0 $\gamma > \beta$ implies that we always have instability.

2.1.3 SIR example

We have shown that the behaviour of the SIR model is determined by the ratio of the two parameters β and γ which we will now call R. We will now compare the behaviour of three different models setting $N = 20000$, $\gamma = 0.05$ and the initial infected population $I(0) = 10$.

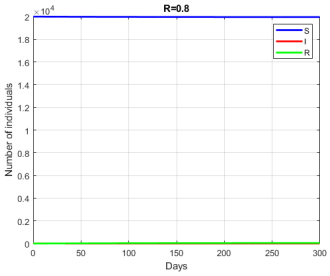

Figure 3: Our first model has been solved with β being chosen such that our Reproductive ratio $R = 0.8$. Note this model is over **300** days

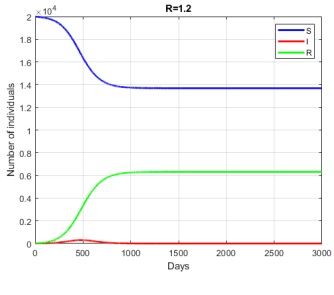

Figure 4: Our second model has β chosen such that the Reproductive Ratio $R = 1.2$ which gives a much longer time span for our disease so we have extended the model to simulate over **3000** days

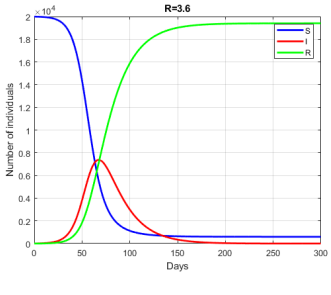

Figure 5: Our third model has β chosen such that the reproductive ratio $R = 3.6$ and the epidemic will end quickly with a very large peak. Note this model is over **300** days

For each model, we will vary β such that R is less than or greater than one and examine the difference in the models' behaviour.

In figure 3 we can see that if R is less than one, no outbreak of the disease will occur. However, as we increase the R ratio even slightly above 1 (see figure 4) we observe a larger proportion of the population - roughly 30 % - becoming infected at some point during the pandemic. This figure is in fact equivalent to the fraction of the population in the final recovered group. While such a small increase in the reproductive ratio R might seem trivial to the public, this clearly shows that some methods of reducing transmission rates even by a small degree can in fact lead a very large difference in the proportion of population infected with a given virus. In the real world, attempts are made to either reduce β by implementing measures such as social distancing, mandatory mask-wearing, restricted travel etc. or increase γ by introducing new treatment methods, antibiotics or even a vaccine.

2.1.4 Solving the ODE via Forward Euler

Our SIR model is governed by three non linear ODEs so we must solve for these numerically. One method for solution is the simplest Runge-Kutta method, the forward Euler method[20].

So first let us consider a Taylor expansion of a function $u(t)$ about the point t,

$$u(t+\delta t) = u(t) + \frac{u'(t)}{1!}\delta t + \frac{u''(t)}{2!}\delta t^2 + ..., \tag{20}$$

which may be written

$$u(t+\delta t) = u(t) + \frac{u'(t)}{1!}\delta t + R(t). \tag{21}$$

where $R(t)$ is a remainder term of order $(\delta t)^2$ which for small enough δt is negligible.
and this provides an explicit formula for estimating the position of our function at some later time step using information we already have. The forward Euler algorithm numerically solves our SIR ODEs as follows -

1. define $u'(t)$.
2. Set our initial conditions u_0.
3. Set the time step dt and how many steps we wish to take n.
4. calculate $u(t)$
   for i=1:n
   $u(t+dt) = u(t) + dt(u'(t))$
   $t = t + dt$
   end
5. output our values for $u(t)$ and end the algorithm.

Note this method is not stable for poorly behaved functions (i.e. those with many turning points in a short interval)[20] and choice of a sufficiently high value for dt will cause the solution to become unstable. A stable solution means that any small changes to our initial conditions or our step size dt will still give the same outcome over time.

2.1.5 The Gillespie algorithm

The Gillespie algorithm is a way to generate a statistically likely trajectory solution for a stochastic differential equation. We transform our SIR model such that transition rates (defined as the rate of change of individuals moving from one state to another) are now replaced by transition probabilities.

This change is reasonable as each infection and recovery is clearly a random event which naturally occurs with some probability. The explicit stochastic form of the SIR model is very cumbersome to express but can be found [25, Section Stochastic Differential equations] and cannot be solved anayltically [6]. The Gillespie algorithm is in fact very similar to our Forward Euler method for an ODE with the caveat that we now introduce a random element and model the change as:

$$u(t+\delta t) = u(t) + P(\delta t(u'(t))), \tag{22}$$

where $P(\delta t(u'(t)))$ represents a Poisson distributed random variable with mean $\delta t(u'(t))$. A Poisson distribution is suitable due to the fact that the change in state for our population is discrete at each time step, with each transmission event occurring a random number of times in each interval with the mean determined by our ODE.

Our transition events E are defined as the event of a variable moving from one state to another (e.g. infected to recovered state), occurring with a given rate $E_r(\mathbf{x}(t))$. Given a set $\mathbf{x}(t) = [S(t), \quad I(t), \quad R(t)]$ of state variables, our algorithm will be as follows [29].

1. Set up the model with our initial conditions $\mathbf{x}(0) = [S(0), \quad I(0), \quad R(0)]$
2. Calculate the event rates $E_r(\mathbf{x}(t))$ these will be governed by the rate of change for each S,I and R in our ODEs.

3. Choose time step δt.

4. For each state generate generate $K = P(E_r \delta t)$ which is the number of times an event has occurred in this time interval - in our case how many have joined or left a particular group.

5. Update the state by

$$\mathbf{x}(t + \delta t) = \mathbf{x}(t) + \sum K \tag{23}$$

   Note we should at this point check that our population has not reached an unrealistic value, e.g. a negative population due to the unbounded nature of the Poisson distribution. In the case that one of our populaltion groups exceeds N we set that group $= N$ and if we have a negative value this is set to $= 0$.

6. Repeat from step 2 until we have reached our stopping criteria, e.g. a population is now zero or until the time period of interest has passed.

Now in our case we should note that our E_r is given by the terms in our ODE for the SIR model. This is because the change in population over time is governed by our ODE and the number of transitions $= E_r$. As such, the rate will change with each iteration as the populations within each category will vary.

2.1.6 Gillespie example

In figures 6 and 7 we show the results of 10 simulations of our SIR model in which we consider a population of 200 individuals with 10 initially infected. We have used $\gamma = 0.2$, $\beta = 3.3 \times 10^{-3}$ which were chosen for illustrative reasons. We have also shown the average of our Gillespie simulations and finally compared alongside the deterministic equation we solved earlier with the same parameters via our forward Euler method see figure 8.

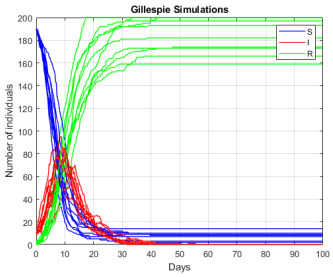

Figure 6: 10 trajectories simulated using the Gillespie algorithm

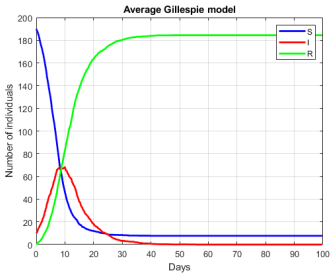

Figure 7: The average of our Gillespie generated trajectories

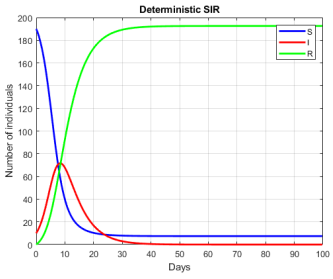

Figure 8: For comparison, the same set of ODEs are solved via our forward Euler method

By inspection, we can clearly see that the averaged Gillespie model in figure 7 is very close to our deterministic model - the average will in fact converge to our deterministic model if

N tends to infinity. Formally this can be shown by considering the ODE for our susceptible group,

$$S'(t) = -\frac{\beta I(t)S(t)}{N}. \tag{24}$$

The right hand side of this equation is our E_r so when performing our Gillespie simulation we do the following

$$S'(t) = -P\left(\frac{\beta I(t)S(t)}{N}\right), \tag{25}$$

where $P(\lambda)$ is a Poisson distributed random variable with mean = λ. The expectation is defined as the sum of N random variables divided by N as $N \to \infty$. Taking the expectation we get:

$$\mathbf{E}[S'(t)] = -\mathbf{E}\left[P\left(\frac{\beta I(t)S(t)}{N}\right)\right] \tag{26}$$

which by the definition of the Poisson distribution gives as expected:

$$\mathbf{E}[S'(t)] = -\frac{\beta I(t)S(t)}{N}. \tag{27}$$

We can use an identical procedure to show convergence for our I and R equations, although similarly the convergence after only ten simulations is also apparent from the figures.

2.1.7 Practical Considerations

It is reasonable to ask why we should use the Gillespie algorithm model when we can already obtain a solution without introducing a stochastic element using the forward Euler method. To answer this question we need to consider the exact purpose of modelling a pandemic infection. Primarily we are concerned with 3 things: the duration of the pandemic, the timing and magnitude of the infection peak and typically the maximum number of infections that the healthcare system can support.

Every estimate obtained via the Gillespie model represents an equally probable trajectory. Some of these trajectories will attain values that exceed the peak of the average curve and conversely some will never reach it. Our deterministic model is essentially the average of all these possible trajectories.

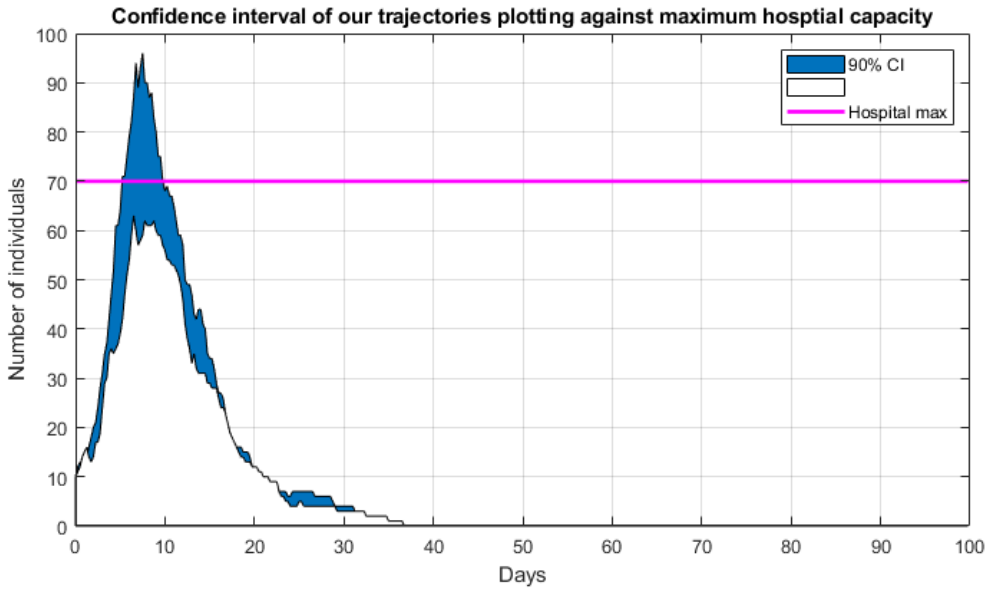

Figure 9: Our shaded region is where 90% of all simulations lie and we have plotted this against the maximum hospital capacity.

This raises the issue of what may happen if the maximum hospital capacity is slightly above the average peak value for infections requiring hospitalisation. If we only consider our deterministic model, we would make the decision that the current level of infection is satisfactory and would not implement any measures to stop the spread of the virus. This however does not take into account that there may be a reasonable proportion of trajectories that in fact have a peak higher then our maximum hospitalisation capacity (see figure 9) in which case we should impose some restrictions to lower the infection rate. The reverse is also true: we may wish to allow the virus to spread more quickly such that the pandemic ends sooner.

2.2 The ODE SIS model

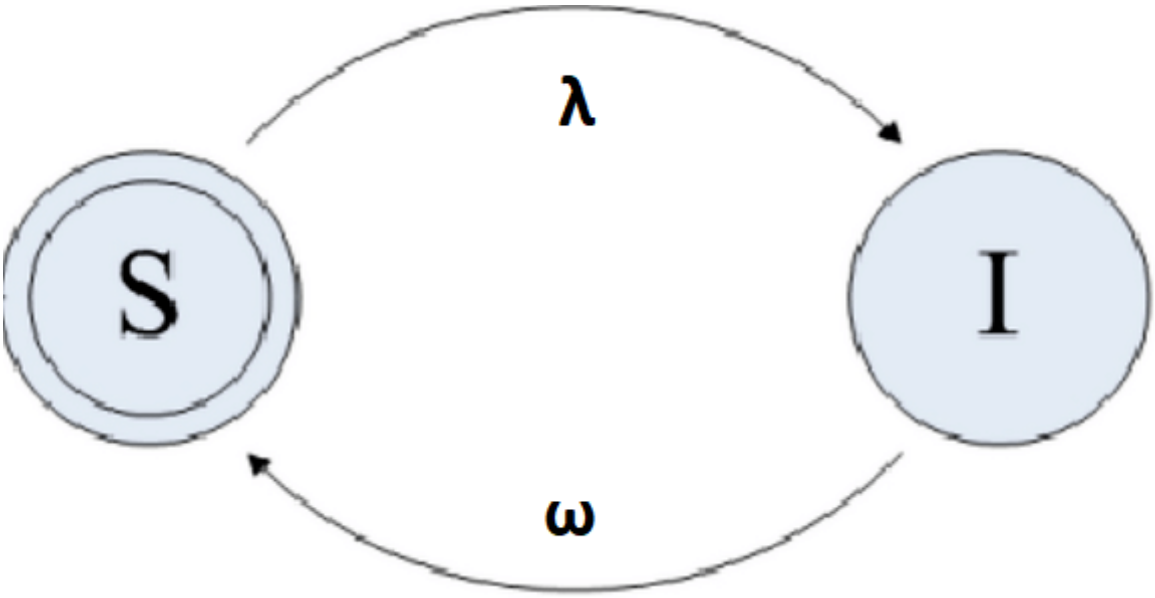

Figure 10: This is a visual representation of the transition between the two states within our SIS model

SIR models are generally used for diseases for which recovery results in lifelong immunity or at least a substantial immunity such that they can be considered immune for the time period we are interested in modelling. For example in the Covid-19 pandemic [15] the immunity period is unknown but is understood to last a minimum of a few months and to provide a resistance to subsequent infection so that for modelling the immediate future we can consider the recovered population as immune. The SIS model is one in which our population can only belong to one of two groups, infected or susceptible, hence the transition path for a given individual is susceptible-infected-susceptible (SIS). Our SIS model is used to model diseases which do not provide any immunity such as most STDs (sexually transmitted diseases), e.g. chlamydia or gonorrhea, where a recovered individual is just as likely to catch it again as somebody who has never been infected. The SIS model is thus governed by only two ODEs for the susceptible population $S(t)$ and the infected population $I(t)$. As we are assuming that individuals do not die from this disease, our population is constant $I(t) + S(t) = N$ and we should expect transition rates to be equal except for a change in sign; this is because one person leaving one group is equivalent to one person joining the other.
We define the ODEs for our SIS model as

$$S'(t) = \omega I - \lambda \frac{SI}{N} \tag{28}$$

$$I'(t) = \lambda \frac{SI}{N} - \omega I. \tag{29}$$

Here ω represents the rate of recovery of the infected population while λ represents the rate of infection. We can find an equilibrium for the system by simply factoring out I in equations 28 and 29 and setting our equations $= 0$:

$$0 = \left(\omega - \lambda \frac{S}{N}\right) I \tag{30}$$

$$0 = \left(\lambda \frac{S}{N} - \omega\right) I \tag{31}$$

If $I = 0$ we have $S = N$ and we have a solution but we also have an equilibrium at $\omega = \lambda \frac{S}{N}$ so this particular model has two clear equilibria. Which of the solutions we tend towards, if any, will depend on the initial conditions and our reproductive ratio $\frac{\lambda}{\omega}$. By way of example, figures 11 – 13 show the effect of the reproductive ratio on our model and the position of the equilibrium for the same initial conditions $S(0)$ and $I(0)$.

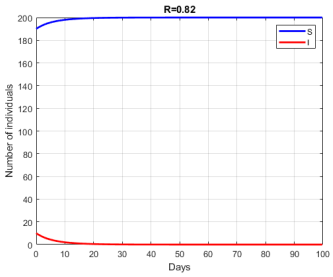

Figure 11: This graph has a reproductive ratio $R = 0.82$. The red line represents the population of infected individuals while the blue represents the susceptible population.

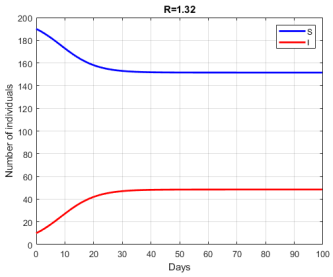

Figure 12: This graph has a reproductive ratio $R = 1.32$. The red line represents the population of infected individuals while the blue represents the susceptible population.

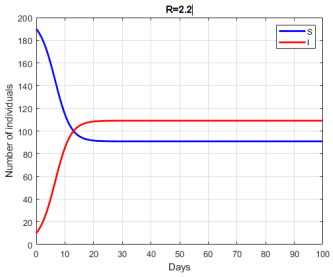

Figure 13: This graph has a reproductive ratio $R = 2.2$. The red line represents the population of infected individuals while the blue represents the susceptible population.

We have solved this numerically using the forward Euler method described in section 2.1.3. We see that the reproductive ratio is playing a very important part in whether or not the virus can sustain itself. For a ratio below 1 infection quickly dies out and we reach an equilibrium for values $R > 1$. Similarly to the SIR model, the SIS model can also be converted into a stochastic differential equation [6] and it would be appropriate to use the Gillespie algorithm to solve for our trajectories.

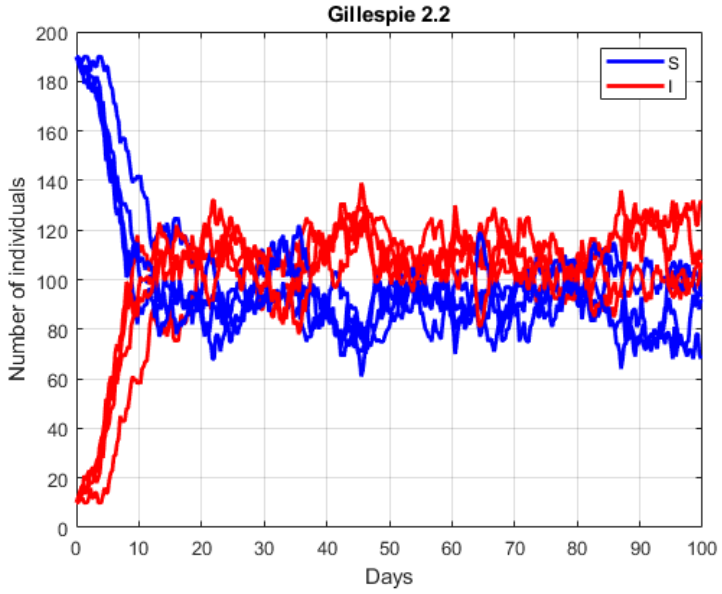

Figure 14: Our reproductive ratio is $R = 2.2$. Our red line represents the infected population while the blue is the susceptible population

In figure 14 we have 4 trajectories simulated via the Gillespie algorithm and plotted on the same graph. Clearly these follow the same basic pattern as our deterministic model but show random behaviour around the equilibrium value as shown in section 2.3. This is what we would expect in reality. Each infection is a random event determined by some probability and we do not expect the smooth curve of our deterministic model, especially with a small population size. If we take the average we should again find that the average tends to the deterministic model as with SIR.

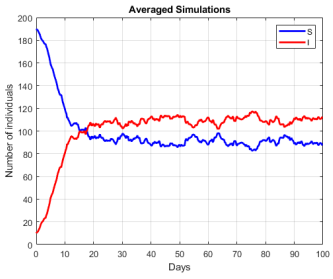

Figure 15: 10 Gillespie simulations

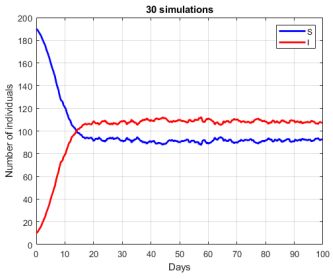

Figure 16: 30 Gillespie simulations

Figure 15 shows the average trajectory after 10 simulations and figure 16 shows the average trajectory after 30. We see clearly that we increasingly tend towards the equilibrium of our deterministic SIS model discussed in section 2.3.

2.3 Linear stability of the SIS model

We perform here a linear stability analysis on the equilibrium points within our SIS model. As before the first step is to find equilibrium values for S and I. We clearly have equilibrium

at $I = 0$ similarly to our SIR model but we also have another equilibrium point when $\lambda\frac{S}{N} = \omega$. This is a more interesting equilibrium point as this corresponds to the infection surviving as opposed to dying out.
We consider a small perturbation around our equilibrium points i.e. $S(t) = S_0 + \epsilon S_1(t)$, $I(t) = I_0 + \epsilon I_1(t)$ to obtain

$$\epsilon S_1'(t) = \omega(I_0 + \epsilon I_1(t)) - \lambda\left(\frac{(S_0 + \epsilon S_1(t))(I_0 + \epsilon I_1(t))}{N}\right) \tag{32}$$

$$\epsilon I_1'(t) = \lambda\left(\frac{S_0 + \epsilon S_1(t))(I_0 + \epsilon I_1(t))}{N}\right) - \omega(I_0 + \epsilon I_1(t)) \tag{33}$$

and now the $O(\epsilon)$ equations are

$$S_1'(t) = \omega\ I_1(t) - \lambda\left(\frac{S_0 I_1(t) + S_1(t) I_0}{N}\right) \tag{34}$$

$$I_1'(t) = \lambda\left(\frac{S_0 I_1(t) + S_1(t) I_0}{N}\right) - \omega I_1(t). \tag{35}$$

We express this equation in matrix form as -

$$\frac{d}{dt}\begin{pmatrix} S_1 \\ I_1 \end{pmatrix} = \begin{pmatrix} -\lambda\left(\frac{I_0}{N}\right) & \omega - \lambda\left(\frac{S_0}{N}\right) \\ \lambda\left(\frac{I_0}{N}\right) & \lambda\left(\frac{S_0}{N}\right) - \omega \end{pmatrix}\begin{pmatrix} S_1 \\ I_1 \end{pmatrix}. \tag{36}$$

As in our stability analysis for the SIR model, we need to solve for the eigenvalues λ^* of the Jacobian matrix. First let us consider the equilibrium $S = N$, $I = 0$. Our eigenvalue equation becomes

$$det\begin{pmatrix} -\lambda^* & \omega - \lambda \\ 0 & \lambda - \omega - \lambda^* \end{pmatrix} = 0 \tag{37}$$

and we must solve for λ^* and check if the equilibrium is stable.

$$\lambda^*(\lambda^* + \omega - \lambda) = 0. \tag{38}$$

We have one eigenvalue of zero and the other is positive when $\lambda > \omega$ indicating unstable equilibrium, but negative if $\omega > \lambda$ in which case the equilibrium is stable.

To check the stability of our other equilibrium point we substitute $S = \omega N/\lambda$ and $I = N(1 - \omega/\lambda)$ into our Jacobian matrix to get the eigenvalue equation:

$$det\begin{pmatrix} \omega - \lambda - \lambda^* & 0 \\ \lambda - \omega & -\lambda^* \end{pmatrix} = 0 \tag{39}$$

yielding the equation

$$\lambda^*(\lambda^* + \lambda - \omega) = 0 \tag{40}$$

We have one eigenvalue of zero and our other will be positive when $\omega > \lambda$ which results in an unstable equilibrium point. If $\lambda > \omega$ we have a negative eigenvalue and we will have a stable equilibrium point.

Clearly, the stability of each equilibrium point depends entirely on ω and λ. Each equilibrium is stable when the other is unstable and vice versa so we should find the system will always tend to whichever equilibrium point is stable over time.

The practical implications of using this Gillespie model over the classic ODE follow an identical argument to that given in section 2.1.6 and so are not repeated here.

3 Network models

3.1 Motivation behind using network models

The construction of any network is grounded within social sciences and within graph theory. A network is constructed by the number of links between individuals and links within groups and these links are represented in graph theory as nodes and edges between these nodes. In social science we are primarily concerned with the reasoning behind certain connections as opposed to the actual structure of our network, i.e. why are some individual groups linked to another. For example, why do we get clusters of certain individuals within a class but not another? The University of Durham would contain a cluster of individuals with few links to the University of Kent: the reason for this would be fairly obvious as they are based far apart, but in many cases the reason is not so obvious and worth investigating. We will be interested in making some reasonable assumptions about the links between individuals (or even groups of individuals) in order to create an accurate simulation of the real world with an emphasis on how we can use this network to improve upon our SIR/SIS models for different infectious diseases.

3.2 Network construction

In any given population we need a method for describing and storing information about the connections between its members. One method is to use an adjacency matrix $\mathbf{A}$ to define our network: if there is a link between nodes i and j we set $A_{ij} = 1$, if not then $A_{ij} = 0$. In the case of our infection models we will always have a symmetric adjacency matrix as the edges are commutative, e.g. if A connects to B, B connects to A. In some graphs you may attach meaning to a node connected to itself but in our specific context that is biologically infeasible as an infected individual cannot infect themselves. Using the adjacency matrix we can calculate the average number of connections per individual by simply summing the lower triangular section of the matrix and dividing by the population,

$$n = \frac{1}{N} \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} A_{ij}. \tag{41}$$

A path of length m means we can move from one node in our network to another by "crossing" m links between nodes to reach it. For example, a path of length 3 from node

A to D would mean we have a link between node A to node B which is linked to node C which is linked to node D. We can gain information regarding the paths of length m by considering A^m. Powers of the adjacency matrix can be used to calculate a measure for the amount of clustering within the network. One such example is given by:

$$\psi = \frac{\text{trace}(A^3)}{||A^2|| - \text{trace}(A^2)}. \tag{42}$$

This is the ratio of the number of triangles within the network over the number of connected triples. A larger value implies a larger amount of clustering. Details are included in section 3.5. [23, see section 2]

We can also determine whether or not our network is fully connected, i.e. is there is a path to any node regardless of the starting node? By considering powers of the adjacency matrix, if $\sum_{m=1}^{\infty} A^m$ has any zero entries this implies our network is divided into two or more distinct components [23, see section 2] and there are no links from one cluster to another.

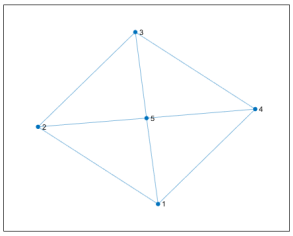

Figure 17: 5 node adjacency graph.

$$\begin{matrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{matrix}$$

Figure 18: 5 node adjacency matrix for figure 17

In figure 17 we have a visual representation of our adjacency matrix for a 5 node network. In developing our own network, we need to consider some social factors to determine the appropriate links between nodes. Of course creating accurate links for every individual within a population is a practical impossibility and would also be subject to significant change over time. Thus, in creating our networks we must make use of generalisations and simplifications over the total population. The interest and focus will of course be primarily on the overall trend of infections as distinct from outcomes for any particular individual. becoming infected.

3.3 Small world network (Watts-Strogatz)

A small world network is a mathematical graph where most nodes are not neighbours of one another but the neighbours of a particular node are likely to be neighbours of each other and all nodes can be reached from any other node by a relatively short path. Formally a small world network is defined to be one in which the average distance between two randomly chosen nodes L grows proportionally to the logarithm of the number of nodes N in the network, $L \propto \log N$, as long as the clustering coefficient is small [30]. Duncan Watts and Steven Strogatz proposed that graphs could be classified according to two independent features, the clustering coefficient and the average shortest path length. Watts and Strogatz found that many real world networks have both a small average shortest path and a higher clustering coefficient then we would expect from a random graph. As such, we introduce the Watts-Strogatz model which maintains the properties of a small world network whilst being applied to a large world network such as the lattice described in section 3.6.[30]

3.4 Algorithm

We are given the population N which will be our desired number of nodes and the mean degree K which represents the number of connections of each node. We also have a parameter $0 < \beta < 1$ which determines the amount of rewiring present within the network. When we rewire an edge we move the node connection from the end to a random node amongst our N nodes. Now we can construct our undirected graph with N nodes and $\frac{NK}{2}$ edges.

1. First we construct a ring lattice by connecting each node to its K nearest neighbours. There are $\frac{K}{2}$ on each side e.g. if our index is 4 and $K = 4$ we connect index 4 to indices 2, 3, 5 and 6. If our nodes are labeled $0 - (N-1)$ we have an edge between node i and j if and only if $0 < |i-j| \mod \left(N - 1 - \frac{K}{2}\right) \leq \frac{K}{2}$.

2. For each node $i = 0, ...N-1$ we take each edge connected to i and then we rewire it to a random node such that we have no self loops with probability β.[1, See Matlab documentation for building the Watts-Strogatz model]

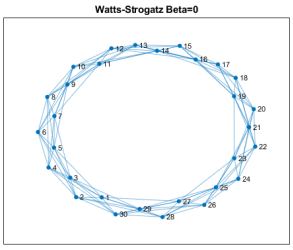

Figure 19: Here the rewiring probability is zero and we get a circle which gives us large world properties i.e. no shortcuts from one node to another.

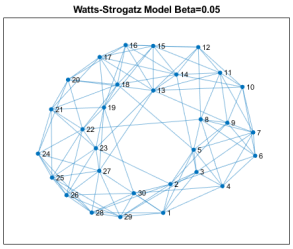

Figure 20: Here the rewiring probability is non-zero and we now have shortcuts from one side of the graph to the other, reducing the average path length from one node to another.

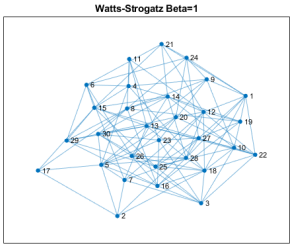

Figure 21: Here we set the rewiring probability $\beta = 1$ to showcase a completely random graph (known as an Erdős–Rényi random graph - see next section) for comparison. In all 3 graphs we have the same mean node degree of 4.

3.5 Clustering and effect on network models

We can define a global clustering coefficent as:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples}} [30]. \tag{43}$$

A triangle consists of three nodes that are fully connected. In a set $\{i, j, k\}$ every node has an edge between the other two. A connected triple however is three nodes $\{i, j, k\}$ such that i is connected to j and j is connected to k. The factor of three is because each triangle is counted three times in a connected triple. The clustering coefficent C indicates how many triples are in fact triangles. In a fully connected graph with $N \geq 3$ we would find $C = 1$ and of course the minimum value is $C = 0$ [30]. The Watts-Strogatz model will construct networks with significant clustering but also small average path length between each node relative to $\log N$. With this in mind, we can create a family of undirected networks with different properties. In figure 19 we have a large world, as we can only take slow "local" steps to travel from one node to another. On the other end of the spectrum we have figure 21 also known as the Erdős–Rényi random graph. This yields a small world graph however, the clustering coefficent tends to zero as $N \to \infty$.[30]
In the context of pandemic spread we can use data gathered about the population and their contacts to ascertain an appropriate value of p and K and thereby create our own small world network, uniquely defined for each disease that we are interested in modelling.

3.6 Lattice construction

One way to construct our network is to consider the graph to be a square lattice. Thus we have a 2 dimensional grid of points in which adjacent individuals are connected such that an individual's network is determined by their position in space. In this way the relevant

connections are clustered; i.e. we expect an outbreak to occur within a local area. This is more akin to what we expect in reality whereas our standard SIR model does not consider geographical location in modelling an outbreak. A well known example of a lattice model construction is the forest fire model [34]. This is a dynamical system which will eventually reach some equilibrium. There are 4 rules for our forest fire model at each time step.

1. A burning cell turns into an empty cell.
2. A tree will burn if at least one of its neighbours is burning.
3. A tree ignites with probability f even if no neighbour is burning.
4. A tree appears in an empty space with probability p.[14, see section 4]

So with this construction, trees and burning are analogous to infection and recovery. We should find that clusters of infection will appear within the population according to the ratio of f and p. We can adapt these rules to form a more accurate representation of our SIR model by making the following changes.

1. A burning cell turns into an empty cell with probability γ.
2. A cell will burn with a probability = $\beta\times$ (number of neighbours which are burning)/total number of neighbours.
3. No cell will burn without a burning neighbour.
4. An empty space can no longer fill with a tree.

Under these rules a burning cell represents an infected node, an empty cell represents a recovered node and all infections can only occur along a chain i.e. you must come into contact with the infection for it to spread. Note that β and γ are the same parameters as in our SIR model.

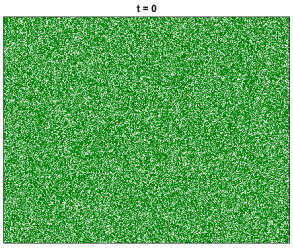

Figure 22: $T = 0$

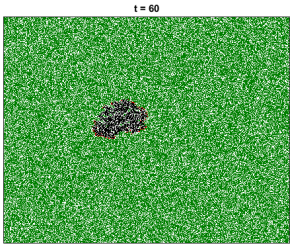

Figure 23: $T = 60$

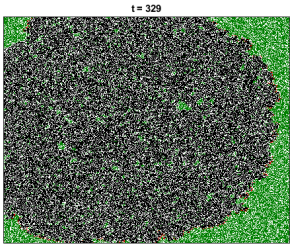

Figure 24: $T = 329$

For the sake of clarity in figs 22–24, I have chosen a very high infection parameter $\beta = 0.5$ and a fairly dense population occupying 70 percent of our total square lattice. Green represents a tree (susceptible individual), black represents a burning tree (infected individual) and white is an empty cell (recovered). Clearly the infection spreads from the origin over time but we nonetheless observe clusters of uninfected individuals surrounded by those who

are infected. This model has practical advantages over our standard SIR model as the latter does not take into account our population network and makes no distinction between densely and sparsely populated areas. For example: Greater Manchester (population 2.9 million) and Wales (population 3.1 million) will be treated almost identically in our SIR model. However in our adapted forest fire model for infection spread, this would result in our density parameter being changed thereby affecting the outcome of our model. This would be a useful and more accurate way to model how infection spreads in different regions.

3.7 Data collection methods

Often the largest obstacle when trying to construct a network model for an infectious disease is the act of gathering data. For all but very small population sizes, creating a comprehensive social network is practically impossible due to the volume of data to be gathered. We will also encounter human error as an individual is likely to have many contacts but complete recall is unlikely to be accurate. Further cases may exist in which that information is not always going to be readily volunteered (e.g. sexually transmitted infections). These issues are entirely to do with the data collection but we must exercise care in our methods for gathering data as not all information is relevant when we are constructing our network. We should only be concerned with social contact that has reasonable probability of transferring infection - i.e. how much contact is necessary for infection to be possible? This naturally allows some degree of arbitrariness. Even in cases where the links should be straightforward, such as for STDs, there are different types of relationships which may have greater probability of transmission. For example, heterosexual, homosexual, monogamous, or casual sexual relationships may not carry the same risk and the chance of transmission for each of these links is not necessarily the same. A solution to this problem would be to give each of these particular types of link a predetermined weight. This of course adds more complications to our network model. A consequence of this is that network models must be unique to whichever disease we are interested in. [24, section 2] Three main techniques have been used to gather network information which are infection tracing, complete contact tracing, and diary-based studies. Each have their own benefits and disadvantages and the choice depends on the population and the data we are interested in gathering.

3.7.1 Infection tracing

After an epidemic such as SARS an epidemiologist will place emphasis on determining the origin of infection in each individual case. Each infected individual is linked to the person to whom they transmitted the virus or from whom they were infected. Once the data is collected we have a transmission network consisting of all edges from which the virus has spread. As this in actual fact represents the true transmission events, there is no issue with defining what constitutes sufficient “contact” to form a link and interactions that did not result in an infection are not included in the network. The networks observed are likely to form tree-like networks e.g. one person infects 2 who infects 2 others etc. As such they contain no loops and we are unable to conduct much meaningful analysis on the model. However this is often a key part of disease control policy and can provide useful information

about individuals most involved in the particular transmission of the infection in question. [24, section 3.1]

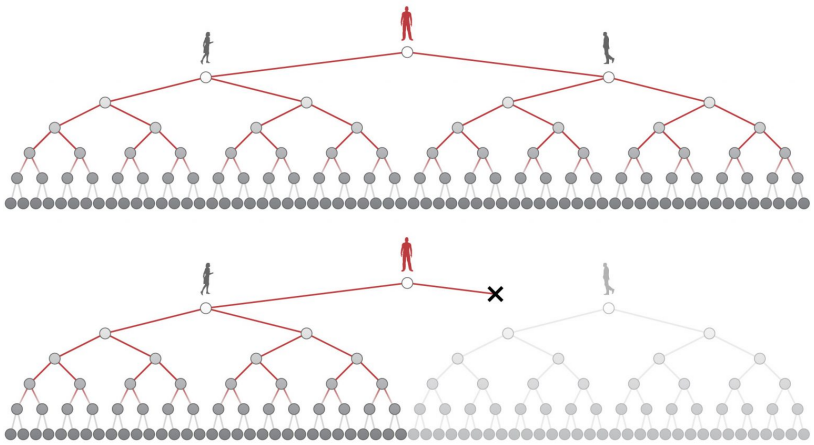

Figure 25: [7]

Figure 25 showcases the fundamental idea of implementing strategies such as social distancing, curfews and quarantining. These tree models can be used to identify where in the transmission chain widespread infection occurs e.g. at school or universities or which age groups.

3.7.2 Contact tracing

Contact tracing aims to find all potential transmission contacts for each individual. This then identifies a new set of individuals who are subject to further tracing. As we aim to identify all transmission routes we again suffer from uncertainties accurately defining network links and some links may be included that are not plausible . e.g. 'entered into the same supermarket'. Contact tracing is also very time consuming and relies entirely on accurate data about personal relationships. It is most commonly and effectively used in the case of STDs where an individual will normally be able to define and recall all contacts which have occurred. This is in fact the same basic method employed by NHS 'test and trace' app but here the contacts are established via Bluetooth devices as distinct from subjects who willingly inform the researcher. The aim of this control technique is to try to get asymptomatic individuals either treated or quarantined to "cut" the tree earlier and so be left with a smaller subset of the fully connected model. A good example of this is the study of sexual networks in Colorado Springs [21, see discussion] which highlighted core groups of individuals as epicentres of infection and "longer distance" contacts who link different subsets of the network together thereby allowing transmission to spread. [24, section 3.2].

3.7.3 Diary-based studies

Networks based on tracing are very labour intensive and rely on the individual's ability to recall their contacts. In diary based studies, subjects record contacts systematically (like a personal diary). Thus the workload is shifted onto the individual and this allows a larger number of individuals to be sampled in detail. There are however issues with this as each

individual will be responsible for determining their own connections and the definition of what constitutes a contact for infection purposes is subjective and increasingly arbitrary. In addition, it can be very difficult for the coordinator to link the information together as the names/identifiers for each connection may not be uniquely recorded. Unless the whole group are part of a similar circle e.g. small community workplace, local school we will likely find several unconnected sub networks which constitute an individual's personal network. [24, section 3.3].

3.8 Comparison with the SIR and method

In our ODE model we have two parameters β and γ to consider and their ratio R determines the proportions of individuals which join the recovered and infected groups. In our network model each node in the susceptible group will have a probability P of becoming infected which changes at each time step depending on the number of infected nodes connected to it. P corresponds exactly to $\beta I(t)$ from our ODE model. In effect, we should find that for any group of nodes we will be solving the ODE SIR equations on a smaller scale. An infected node will recover with probability $= \gamma$ as the recovery of a node is independent of any of the adjacent nodes. Thus recovery takes place in essentially the same way as in our ODE model. Before we start using our network model a natural question arises - given the same assumptions as our SIR model, and treating the entire population as connected, do we get similar results?

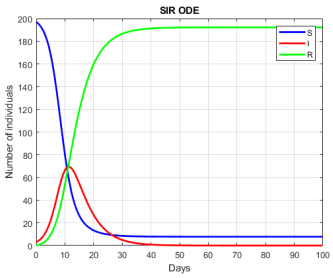

Figure 26: Original SIR

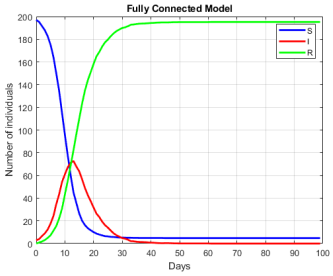

Figure 27: Fully connected model averaged ten times, in this simulation we have made every node in our model connected to every other node in the population.

From figures 26 and 27 it appears that making the same assumptions as those for our SIR ODE model we observe the same outcome. Let us now consider how we will choose the probabilities for a particular node to transition from the susceptible to infected and infected to recovered. We already know the rate of change of these populations from equations 1, 2 and 3 so they correspond to the following probabilities:

$$p_i = \frac{\beta I(t)}{N}, p_r = \gamma. \tag{44}$$

As these probabilities correspond exactly to the rate of change in our ODE for the SIR

model we should expect the average to be identical to our ODE model.

The simulated result shown in figure 27 strongly suggests that the network model is in fact suitable. However the network model is far more flexible. We are free to choose our connections and implement restriction strategies within the population to examine the effect this might have on infections. It is however more computationally intense. Needless to say this is entirely the purpose of modelling pandemics so government strategists can gain a reasonable understanding of the likely events if they impose restrictions (e.g. curfew, lockdown). Note that our fully connected network model is still a stochastic process as the event of an infection is random and we still maintain the same benefits provided by our Gillespie ODE SIR model. The idea of using a social network to combat pandemics is not new and several previous attempts have been made to construct such a network (e.g. NHS test and trace) in order to isolate a virus as effectively as possible. In terms of our network this means that a node which has a connecting edge to an infected node will then change to have no edges (i.e. cannot infect another): the practical implications and socio-economic issues of such a strategy are certainly a contentious issue and the advantages and disadvantages of this approach are subject to dispute.

3.8.1 STD spread through a population

A clear example in which networks can provide better modelling than our original ODE type is in the spread of STDs (sexually transmitted diseases). Here the only way to become infected is through sexual contact with another person so clearly the structure of the model is the key to understanding the spread of infection. Traditionally the data required for this model can be very difficult to acquire as individual sexual history is often not readily volunteered. There will be many individuals who do not form links often while others will change partner frequently or maintain several links at once. Naturally within this model the more links you have the higher your chance of infection will become.

3.8.2 Method

In order to model the infection throughout this population we need to know initial conditions, that is the social network at the start i.e. who are partners and the average time for somebody to find a new partner. As time progresses, new links will be made and old links will disappear.
We can relate our parameters within this model to our ODE SIS model as the probabilities of becoming infected are determined by βI. The probability of a node becoming infected will be given by:

$$\frac{\beta \times \text{number of adjacent infected}}{\text{total number of nodes} + 1}. \tag{45}$$

This is identical to the SIS rate of susceptible population becoming infected and similarly the rate of change from the infected category is given by ωI so the probability of recovery is ω. However unlike our fully connected model, the number of adjacent infected nodes will vary depending on the particular node so that our probability will be unique for each

node. Note that our recovery probability is simply ω as return to the susceptible group is independent of the number of individuals in it. As before if the model is fully connected we expect no change from our SIS model.

For our first model we will make the rather bold assumption that everybody within the population is a possible sexual partner of everybody else (e.g. everybody is bisexual with no preference either way) so we treat the entire population as equal. [2]

We ran a simulation of a small network of ten individuals in a fairly promiscuous group. As time changes the network links will also change according to some relinking parameter δ which is simply the probability of a link breaking and being reformed. Each new link is completely random, so the model has no preference as to with whom the next link is formed once a link is broken. Initially we assume 3 infected people and each link has a probability of being broken and of a new link being formed at each time step. In the model we have decided not to include a "waiting" period between forming new links. Further, for simplicity no more links can be formed than the number we started with. We have allowed the same link to be formed several times to indicate a heavy preference or more sexual contact between any particular pair. A non-infected individual has a chance of becoming infected β which depends upon the frequency of sexual contact with an infected person in a given time step.

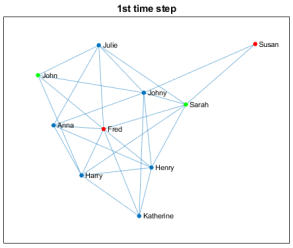

Figure 28: Time step=1. Red represents an infected node, blue represents a susceptible node and green represents a recovered node.

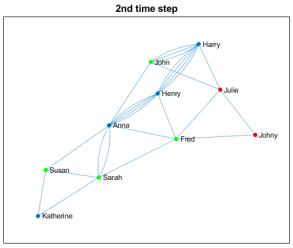

Figure 29: Time step=2. Matlab will rearrange the position of each node to avoid visual confusion on the node links.

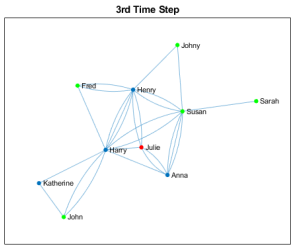

Figure 30: Time step=3.

In figures 28–30 we can clearly see that as time is changing the links between the individuals are also changing, allowing possibility for infections to emerge from different sources. The model is simulating disease spread between these ten people such that once recovered, the chance of infection becomes zero. This however is not a very realistic scenario for an STI virus such as chlamydia where recovery offers no lasting immunity from the virus. We conduct the experiment again as an SIS model and expect to find some equilibrium point determined by our infection parameters.

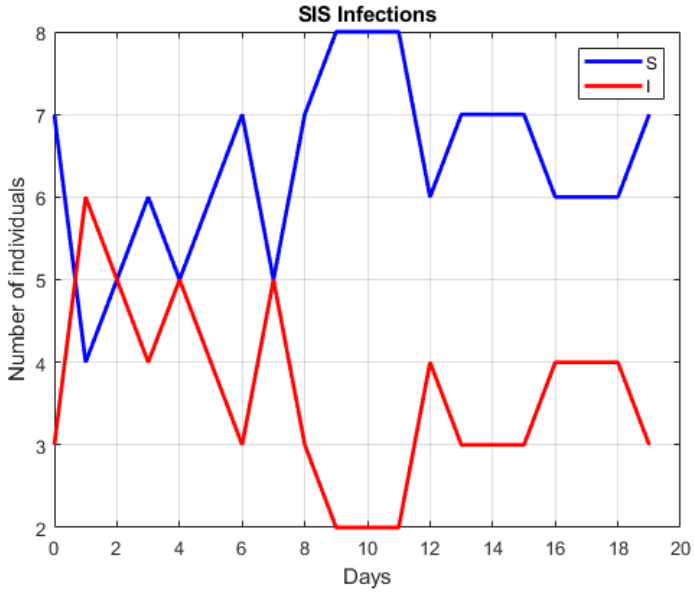

Figure 31: Our reproductive ratio is 2.2

In figure 31 we see that the infected population over time is varying between 6 and 2 which implies the equilibrium point of the population depends upon the particular network between the individuals. Even in this small example we can see that the infected population will not die out although it will vary over time, similarly to what we expect from our SIS model.

For purposes of illustration, let us conduct the same experiment but not allow the network structure to vary over time. The results and corresponding network layout are shown in figures 32 and 33.

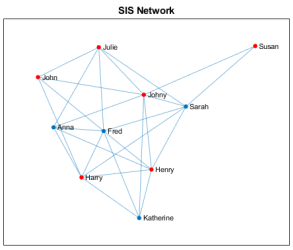

Figure 32: Model layout for our SIS infection model.

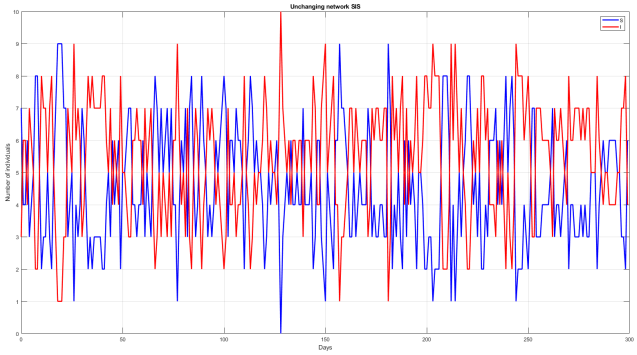

Figure 33: Resulting SIS graph for our unchanging network.

With the network structure remaining the same we appear to have random behaviour of both S and I about an approximate mean of 5. Our experiment uses just ten individuals so the variation in figure 33 seems very large. As we have a small population, extreme values for our population become far more likely. However, when we expand our network to include a larger population size the law of large numbers comes into play. This states that as a sample size grows, its mean gets closer to the average of the whole population [33].

Thus if we conduct this experiment several times our infection mean should converge to the real life equilibrium of the population: either the infection dies out or we have a steady amount of infected population. We have conducted the experiment again using another network shown in figures 34 and 35.

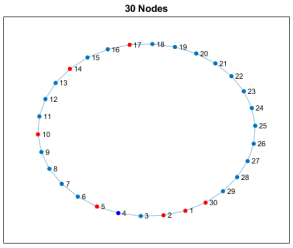

Figure 34: In this model layout we have allowed each node to be connected to its two neighbours so the whole population form a ring, red represents an infected node while blue is susceptible.

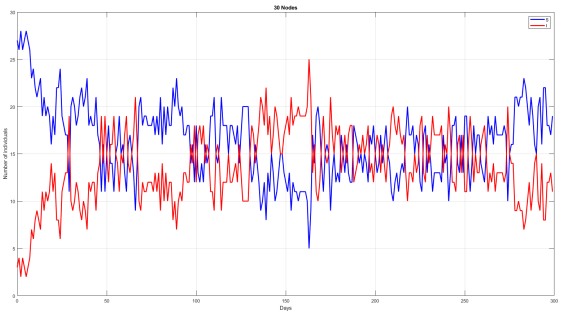

Figure 35: this is our SIS graph for the unchanging circular network after simulating the infection spread through our network for 300 days

In figure 35 we can clearly see that we have reached stable long term behaviour around our equilibrium state so that we will observe random fluctuations around that state as the infection spreads. We have reached a stable point where our rate of change for each state is zero i.e. $\frac{dS}{dt} = \frac{dI}{dt} = 0$ with all fluctuations being entirely due to the stochastic element of our model. The only other naturally occurring steady state is when the number of infections reach zero.

3.9 Network model for Love Island

Love Island is a reality TV show where single people spend several weeks together in an effort to find a partner and eventually be voted as the public's favourite couple and thereby win prize money. In this section, I will simulate this year's show and introduce an infectious disease into the population and examine what happens over time. The show works as follows: each week the contestants choose who they would like to partner with and couples form accordingly. The host will give the contestants reasons to switch partners either by involving them in specific tasks or introducing new people for the contestants to choose from. Contestants entering a new week that do not have a partner are removed from the show. At the end of the series, the remaining contestants are put to a public vote to decide the winner.

We will make the following assumptions in our model. Each pair will have sufficient contact to be at risk of infection from an infected partner. They will not however be aware of their infection or be able to seek treatment during the course of the show so their chance of recovery is zero. We will also assume that the only chance of infection is from their current partner i.e. there will only be 1 link or no links for each individual in our network.

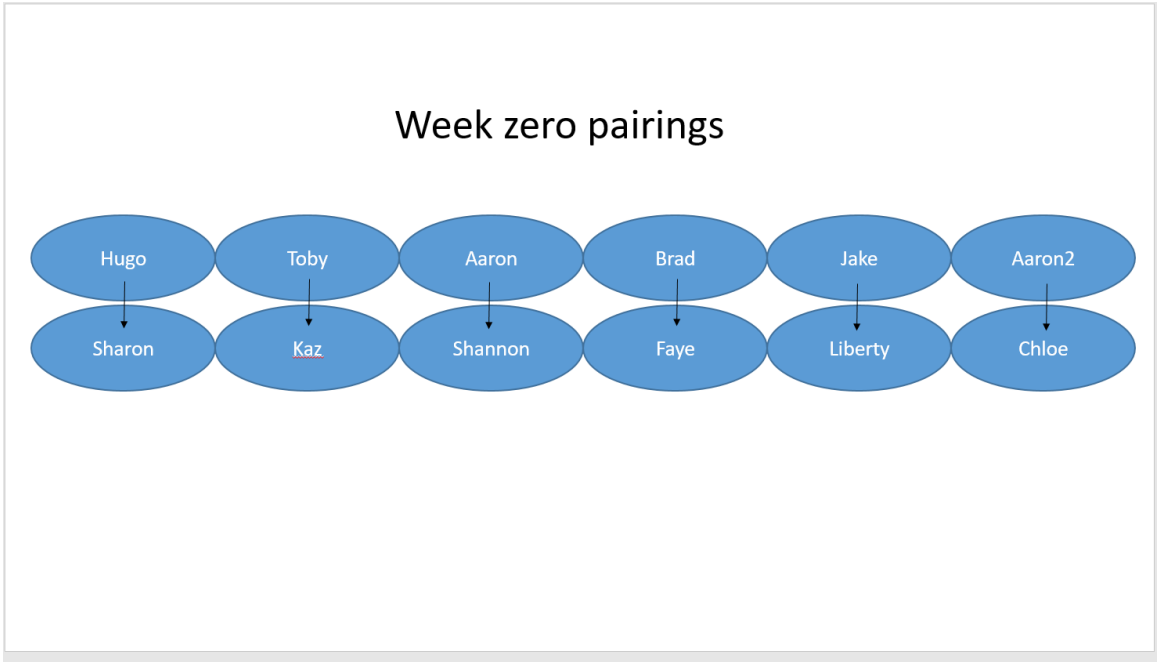

Figure 36: Love Island network week zero

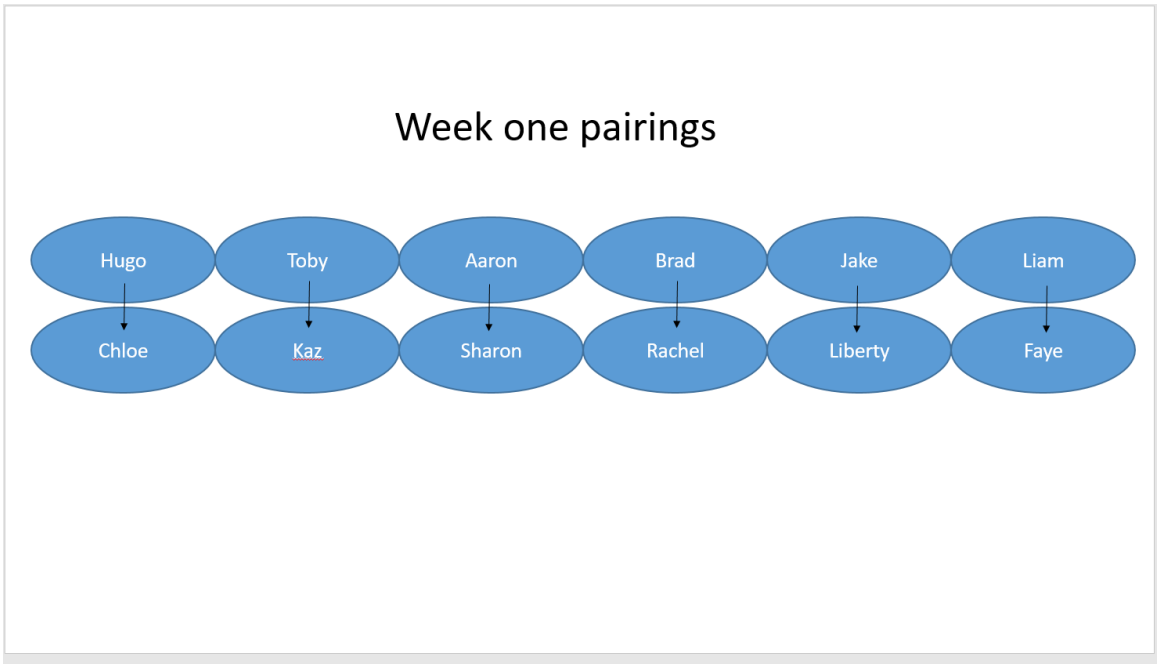

Figure 37: Love Island network week one

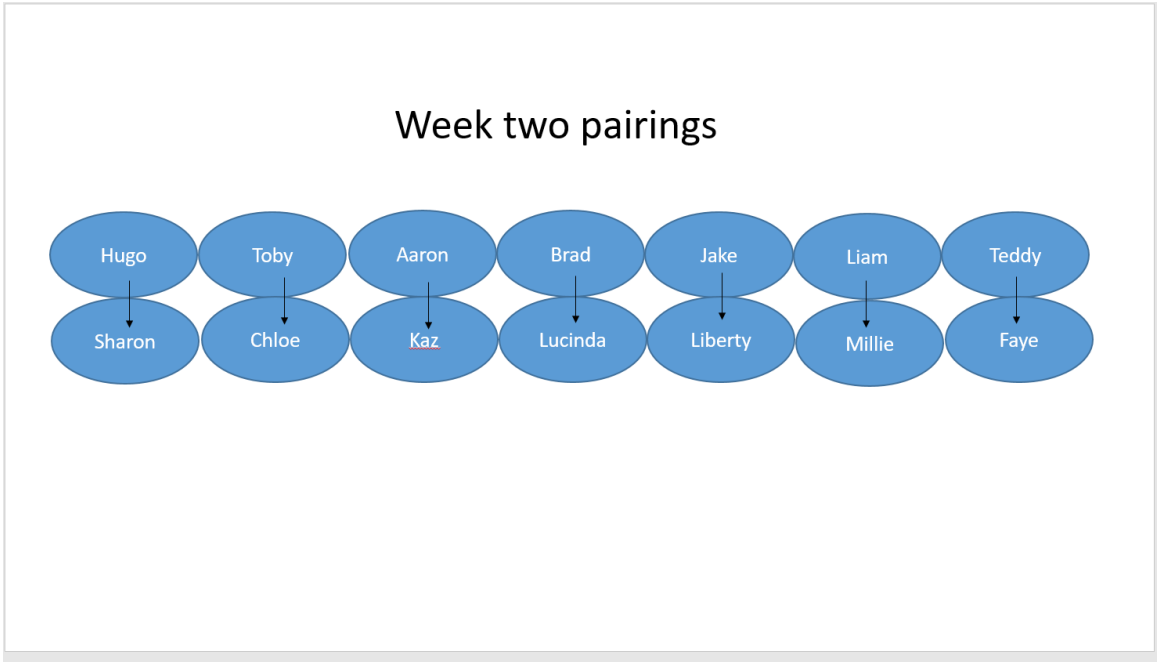

Figure 38: Love Island network week two

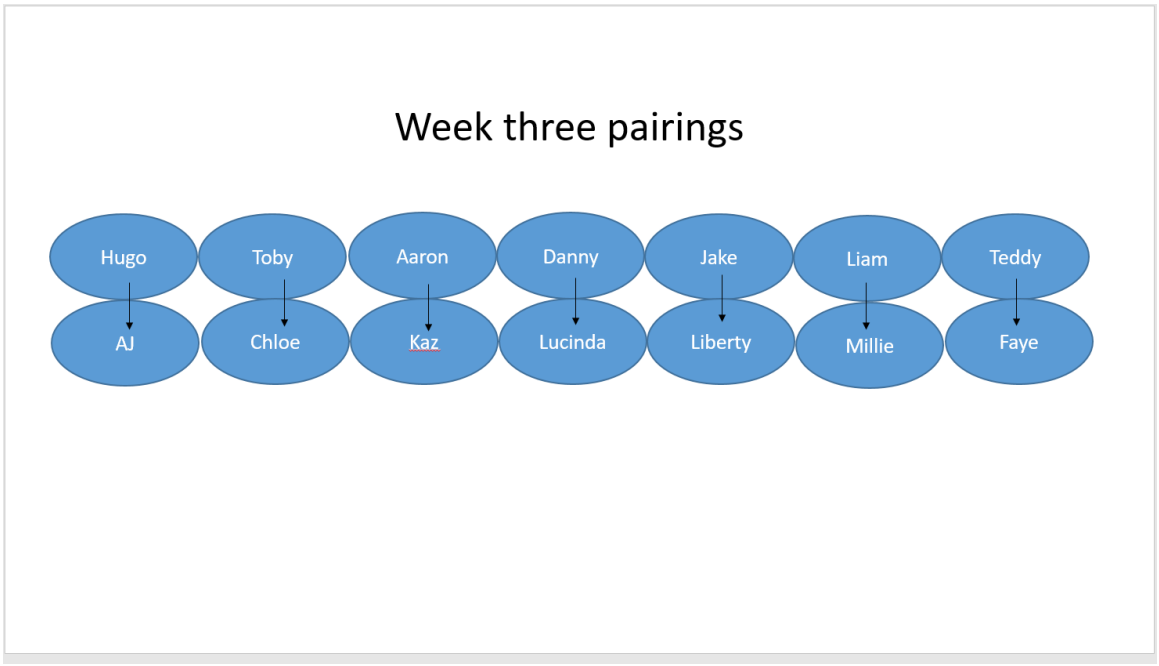

Figure 39: Love Island network week three

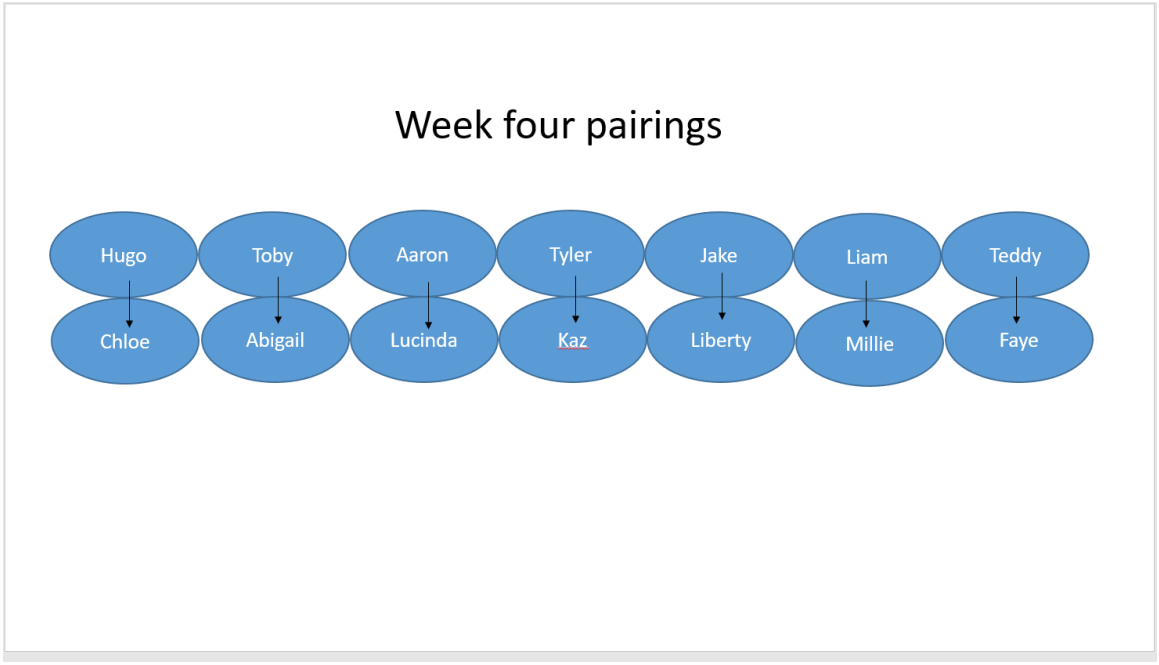

Figure 40: Love Island network week four

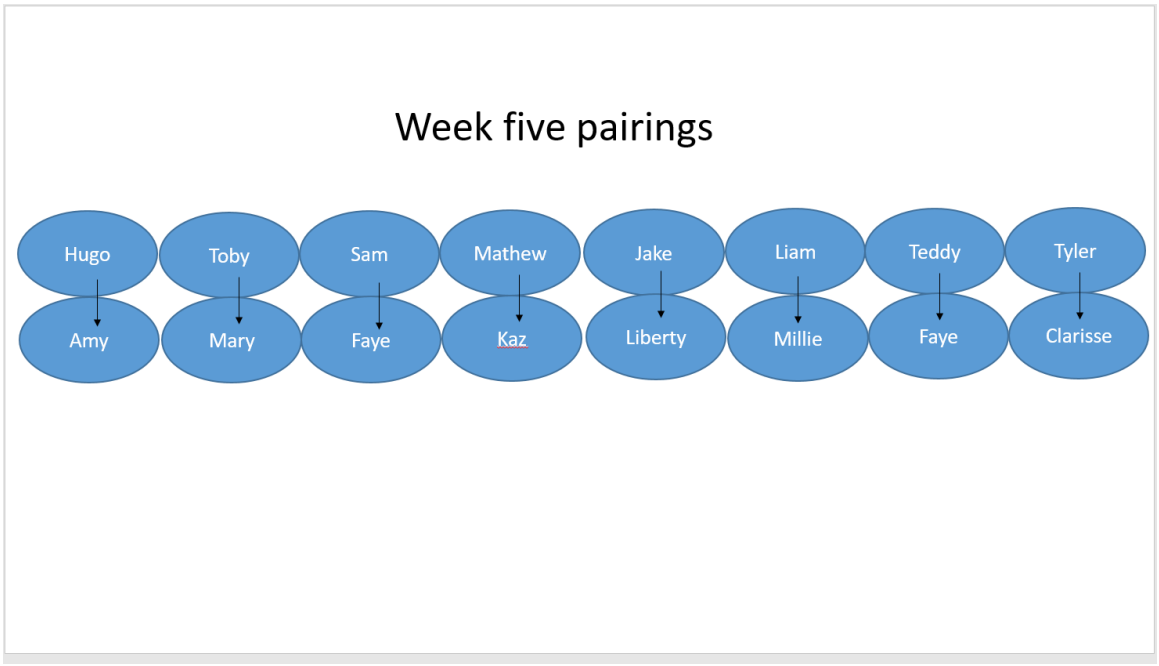

Figure 41: Love Island network week five

We are simulating a "what if" scenario and as such already know the exact pairings and therefore know how the network develops over time. We have shown the shape of the network in figures 36– 41 with each unpaired node excluded for visual simplicity. We will introduce the herpes virus into 33% of our population. We make the assumption that each couple will have multiple instances of sexual contact while coupled in which case herpes has a 50-70% transmission probability [3]. We now examine how the virus spreads and what

factors affect transmission within the network.

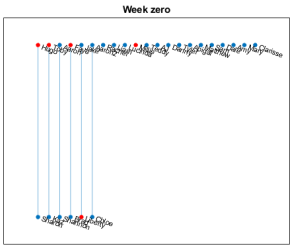

Figure 42: Love Island network week zero infections

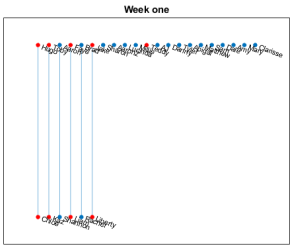

Figure 43: Love Island network week one infections

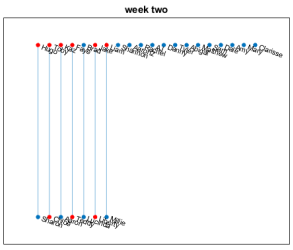

Figure 44: Love Island network week two infections

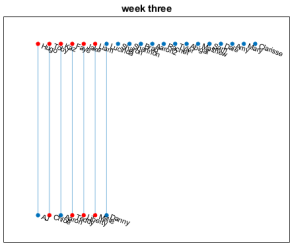

Figure 45: Love Island network week three infections

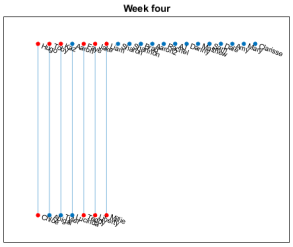

Figure 46: Love Island network week four infections

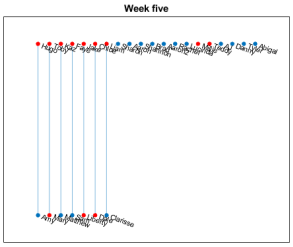

Figure 47: Love Island network week five infections

In figures 42–47 we show an example of how the infection spreads through the network with a random initial set of infected individuals (where a red node indicates an infected individual). We conducted this simulation 20 times with just 10% of our population initially infected to find our average SI graph (see figure 48). However an inspection of the network itself shows that some of the individuals create far more connections then other members: they are considered a "core group" for infection spread. Let us compare figure 48 with a similar graph where we ensure that Hugo, who changes partner each week is always amongst the initially infected 10% (see figure 49). We also repeat the same test but with 33% of our population initially infected and also where both Toby and Hugo are amongst this group (figures 50 and 51).

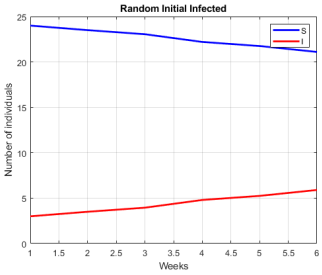

Figure 48: Love Island SIS Random 10 %

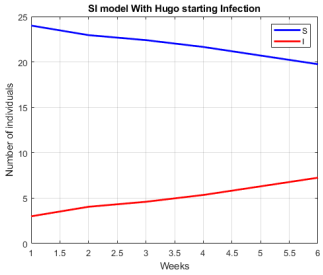

Figure 49: Love Island SIS 10% + Hugo

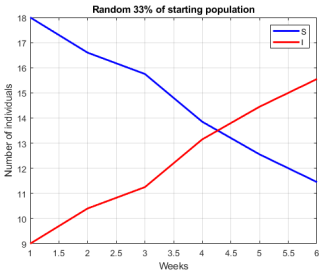

Figure 50: Love Island SIS Random 33%

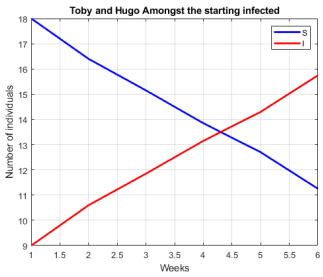

Figure 51: Love Island SIS Random 33% + Hugo and Toby

In both cases we find that the average number of infections has been increased by inclusion of our "core" members but perhaps not by so much as we might expect. This is likely due to the relatively small network we have and the fact that many of the contestants remain single for the majority of the show or only partner once before being removed from the show. This makes infection spread less likely for those outside the initial 12 contestants who have the most opportunities to encounter an infected individual. If we now assume only people outside the initial 12 contestants can be infected we can assume we will have a much lower amount of infections.

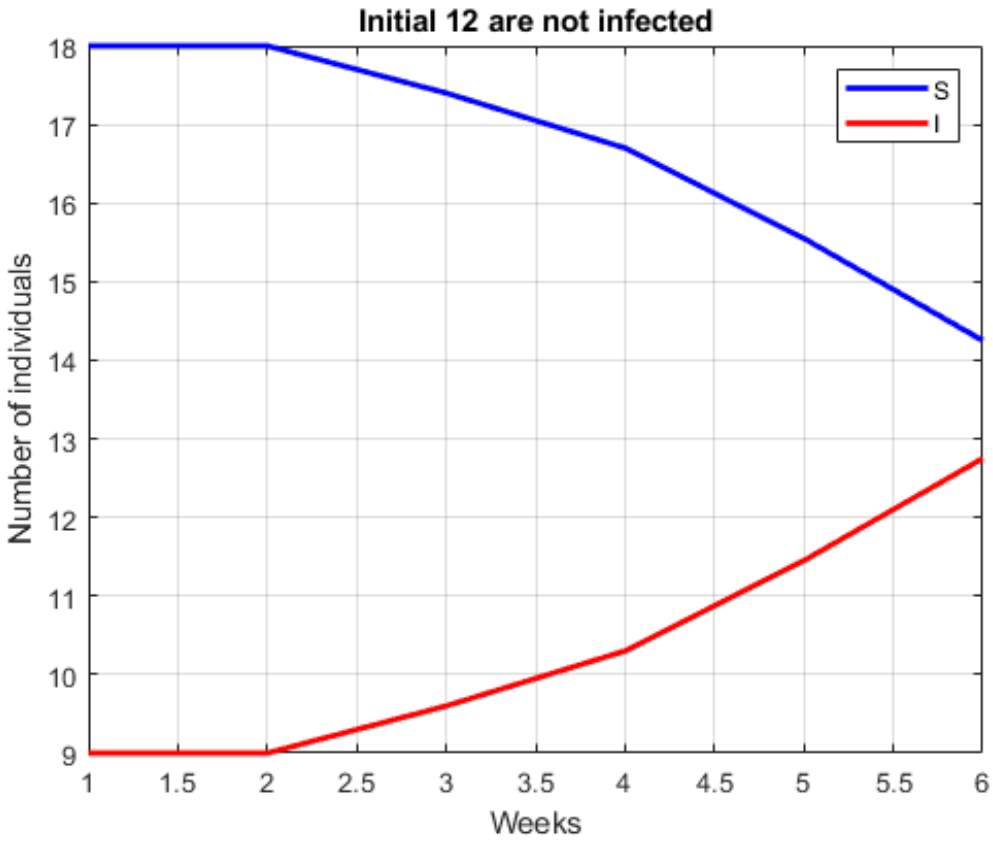

Figure 52: Love Island spread without initial infection amongst starting contestants

We can clearly see the effect this has by figure 52. If our initial infection does not start amongst the "core" group, the infection does not get the chance to spread at all until they begin pairing up and forming links. Many of these individuals only form one or two links before they stop making connections so we certainly do see what we expect.
In conclusion we can say your chances to form many pairs and the likelihood to encounter an infection is much higher if you belong to the "core" group and these "core" groups are of course essential for infection spread [22].

We have now shown that the core group of individuals are predominantly responsible for infection spread. To end our analysis on our prospective romantics, we would like to pose the question "If infected, which individual in our model will thereby cause maximum infection within the population? Who is our super spreader?" At first glance, we might be inclined to say it is somebody like Hugo who regularly switches partner. However the answer may not be so clear cut as all of his partners may never form another romantic relationship and as such the number of infections will have a maximum = the number of partners Hugo has. To answer this question we have simulated our infection spread model ten times and averaged the total number of infections caused by each individual who starts with the infection.

Hugo	Sharon	Toby	Kaz	Aaron	Shannon	Faye	Brad	Jake	Liberty	Aaron2	Chloe
3.8	3.1	4	4.1	3.4	3	3.8	4.2	2	2	2.1	3.1

From our starting 12 contestants it seems that Brad is actually the person most likely to initiate maximum infection spread. This inherently does not seem obvious as Brad himself only lasts in the show for 2 weeks and therefore does not have many opportunities to spread infection, however his pairings such as Faye and Lucinda do in fact make new pairings after they pair with Brad, subsequently causing more infected links to the other members who switch partners often. Interestingly Hugo who has a new pairing each week does not score as highly even though he would be the obvious choice for our super spreader. Using this network model we can gather information about the individuals long term connections that otherwise would not be clear at all.

4 Pair formation models for sexually transmitted infections

When we consider an airborne virus we make the assumption that contact is instantaneous and that contact is sufficient to transmit the virus to everyone in the population. However in the transmission of STIs the situation is different and we cannot make the same assumption. If members of our population form long lasting, monogamous relationships the chance of transmission is naturally affected. If a couple are both in the susceptible category they are protected from infection until they form a new relationship. Similarly, if someone is in partnership with an infected partner they are liable to become infected, recover and then become reinfected. Therefore the dynamics of these relationships is key in establishing a faithful model of transmission.

4.1 Partnership dynamics

First we need to construct a model for the partnership dynamics of the population itself [26]. We subdivide our population into two groups: single people $X(t)$ and pairs of individuals $P(t)$. The population size is $N = X + 2P$ as a pair consists of two individuals. Singles form new pairs at a rate ρ and pairs separate at a rate ω. In a changing population we have a recruitment rate of B and individuals leave the population at a rate η. We make the straightforward assumption that $X(t) \geq 0$ and $P(t) \geq 0 \; \forall t \geq 0$. Now we have the information needed to describe the pair formation process by the following pair of differential equations:

$$\frac{dX(t)}{dt} = B - \rho X - \eta X + 2\omega P + 2\eta P. \tag{46}$$

[26, see section 2] where B represents the new population joining, $-\rho X$ is the proportion of population forming a couple and joining P, $-\eta X$ is the proportion of X leaving the overall population and $2\omega P$ is the proportion of individuals joining X from P.
Our equation for the rate of change of our pairs is given as:

$$\frac{dP(t)}{dt} = \frac{1}{2}\rho X - \omega P - 2\eta P. \tag{47}$$

[26, see section 2]

where $(1/2)\rho X$ represents the number of individuals joining P from the X group, the $1/2$

is due to a pair requiring two individuals and $-\omega P$ is the proportion of pairs breaking up and joining the group X.

A brief explanation of the final two terms in each equation is relevant as their origin is not intuitive. Consider η is the probability of a person leaving the population i.e. dying or leaving the dating market. Now each person in a partnership still retains that probability, however if one leaves their partner automatically joins the single population. For each partnership however there are two ways for that relationship to end i.e. the man leaves and the girl becomes single or vice versa (in heterosexual relationships), because of this we actually have $2\eta P$ leaving our population. In our first equation for each $2\eta P$ that leaves the P population we get one person joining X not twice the amount as one person has left the model altogether, hence we have $2\eta P$ joining X.

Let us solve for the equilibrium of this system by setting our left hand side equal to zero.

$$0 = B + X(-\rho - \eta) + P(\rho + \eta), \tag{48}$$

$$0 = \frac{X\rho}{2} - \omega P - 2\eta P. \tag{49}$$

Rearranging top and bottom to make P and X the subject we have:

$$X = \frac{B + P(\rho + \eta)}{\rho + \eta}, \tag{50}$$

$$P = \frac{X\rho}{2(\omega + 2\eta)}. \tag{51}$$

Substitute equations (51) into (50) and rearrange to make X the subject and we get:

$$X^* = \frac{B(\omega + 2\eta)}{\eta(\rho + \omega + 2\eta)} \tag{52}$$

Substituting (52) back into (51) we get our value for P as:

$$P^* = \frac{B\rho}{2\eta(\rho + \omega + 2\eta)} \tag{53}$$

As $N = 2P + X$ in equilibrium we may express N as:

$$N = \frac{B}{\eta}\left(\frac{\omega + 2\eta}{\rho + \omega + 2\eta} + \frac{2\rho}{2(\rho + \omega + 2\eta)}\right). \tag{54}$$

and the term inside our brackets = 1 so that in equilibrium we have $N = \frac{B}{\eta}$. This implies a proportion $x^* = \frac{\omega+2\eta}{\rho+\omega+2\eta}$ of our population are single. The proportion of people in a partnership is given by $p^* = \frac{\rho}{\rho+\omega+2\eta}$. We will always assume that the pair formation and separation process is in equilibrium and therefore can use the proportions of single and paired individuals within a population sample to estimate our η and ω from the data gathered.

We can use these equations to derive other relevant quantities regarding the life course of an individual, for example the mean duration of partnerships is given by

$$D = \frac{\omega}{\omega + 2\eta}. \tag{55}$$

[26, see section 2]

As the expected lifetime of an individual is given by $\frac{1}{\eta}$, we can compute the expected number M of lifetime partners per individual as:

$$M = \frac{\rho(\omega + 2\eta)}{\eta(\rho + \omega + 2\eta)}. \tag{56}$$

[26, see section 2]

4.2 Linear Stability Analysis

We should ensure that the equilibrium points found for our partnership dynamic ODEs are stable. We perform a linear stability analysis by first expanding $X(t) = X^* + \epsilon X_1(t)$ and $P(t) = P^* + \epsilon\ P_1(t)$ and substituting into our differential equations we get:

$$\epsilon\frac{dX_1(t)}{dt} = B - \rho(X^* + \epsilon X_1(t) - \eta(X^* + \epsilon X_1) + 2\omega(P^* + \epsilon P_1) + 2\eta(P^* + \epsilon P_1). \tag{57}$$

giving our order ϵ equation as:

$$\frac{dX_1(t)}{dt} = -(\rho + \eta)X_1(t) + 2P_1(\omega + \eta). \tag{58}$$

For our equation for P our order ϵ equation becomes

$$\frac{dP_1(t)}{dt} = \frac{\rho X_1}{2} - P_1(\omega + 2\eta). \tag{59}$$

Now we express this is in matrix form to find our Jacobian:

$$\frac{d}{dt}\begin{pmatrix} X_1 \\ P_1 \end{pmatrix} = \begin{pmatrix} -(\rho + \eta) & 2(\omega + \eta) \\ \frac{\rho}{2} & -(\omega + 2\eta) \end{pmatrix}\begin{pmatrix} X_1 \\ P_1 \end{pmatrix}. \tag{60}$$

Next we need to solve the eigenvalue problem for our jacobian matrix:

$$\lambda^2 + \lambda(\rho + 3\eta + \omega) + (\eta\omega + 2\eta^2 + \rho\eta) = 0. \tag{61}$$

Finally we will use the quadratic formula to solve for λ.

$$\lambda = \frac{-(\rho + 3\eta + \omega) \pm \sqrt{(\rho + 3\eta + \omega)^2 - 4(\eta\omega + 2\eta^2 + \rho\eta)}}{2}. \tag{62}$$

By factoring out $\rho + 3\eta + \omega$ we get

$$\lambda = \frac{-(\rho + 3\eta + \omega) \pm (\rho + 3\eta + \omega)\sqrt{(1 - \frac{4(\eta\omega + 2\eta^2 + \rho\eta)}{(\rho + 3\eta + \omega)}}}{2}. \tag{63}$$

Now ρ, ω and η are all greater than zero so both eigenvalues λ are always negative. Therefore this equilibrium is stable.

4.3 Sexually transmitted infections within a pair model

We now consider the SIS model in which a recovered individual immediately rejoins the susceptible group. We can have infected singles or susceptible singles; we will denote both with X_i where $i = 0$ is susceptible and $i = 1$ is infected. Now with our pairs we can have 3 scenarios: both are susceptible, one is infected, or both. We will indicate this distinction by P_{00} for both susceptible, P_{01} for one infected and P_{11} for both infected. Each parameter in our system of ODEs corresponds to the same parameter in our separate SIS and pair formation model. We express our system of equations as:

$$\frac{dX_0}{dt} = B - \rho X_0 - \eta X_0 + \omega(2P_{00} + P_{01}) + \eta(2P_{00} + P_{01}) + \gamma X_1. \tag{64}$$

Here B is the number of susceptible people joining the population, ρX_0 is the number of single people who are susceptible joining from either P_{00} or P_{01}. While ηX_0 is the proportion who leave the population altogether. We have $\omega(2P_{00} + P_{01})$ as the proportion of people separating from the P_{00}, P_{01} groups and joining X_0 respectively. We have $\eta(2P_{00} + P_{01})$ as the proportion of people joining the X_0 group because their partner left the population altogether. Finally γX_1 is the number of single infected people who have recovered and rejoined the susceptible group.
Our equation for the infected single population X_1 is

$$\frac{dX_1}{dt} = -\rho X_1 - \eta X_1 + \omega(2P_{11} + P_{01}) + \eta(2P_{11} + P_{01}) - \gamma X_1. \tag{65}$$

Here $-\rho X_1$ is the infected population joining either P_{01} or P_{11}, $-\eta X_1$ is the proportion of infected single people leaving the population altogether. We have $\omega(2P + 11 + P_{01})$ as the proportion of people leaving P_{11} and P_{01} to join the X_1 group. While $\eta(2P_{11} + P_{01})$ is the proportion of infected individuals joining X_1 because their partner left the population altogether. Finally $-\gamma X_1$ is the proportion of the infected single population who recover and rejoin the susceptible group X_0.
Our equation for the uninfected pairs P_{00} is as follows

$$\frac{dP_{00}}{dt} = \frac{1}{2}\frac{\rho X_0^2}{X_0 + X_1} - \omega P_{00} - 2\eta P_{00} + \gamma P_{01}. \tag{66}$$

We have $\frac{\rho X_0^2}{X_0+X_1}$ as the amount of pairs formed with both individuals coming from the X_0 group. We have $-\omega P_{00}$ as the number of pairs who separate and rejoin the X_0 group. We have $-2\eta P_{00}$ as the number of pairs leaving due to an individual leaving the population altogether. Finally γP_{01} is the amount of pairs with a single infected partner P_{01} who has recovered and joined the susceptible group.
Our equation for pairs who have only one infected individual P_{01} is as follows

$$\frac{dP_{01}}{dt} = \frac{\rho X_0 X_1}{X_0 + X_1} - \omega P_{01} - 2\eta P_{01} - \gamma P_{01} - \beta P_{01} + 2\gamma P_{11}, \tag{67}$$

Here $\frac{\rho X_0 X_1}{2(X_0+X_1)}$ is the proportion of pairs formed by a coupling with one individual from X_0 and the other from X_1. We have ωP_{01} as the number of pairs who separate and join the X_0

and X_1 group respectively. We have $-2\eta P_{01}$ as the proportion of couples who break up in this group due to the partner leaving the population. We have $-\gamma P_{01}$ as the proportion of population which join P_{00} due to the infected individual recovering and becoming susceptible again. We have $-\beta P_{01}$ as the proportion of couples who join P_{11} due to becoming infected from their partner. Finally $2\gamma P_{11}$ is the proportion of couples joining from P_{11} due to one of the individuals recovering and joining the susceptible group.
Our equation for pairs with both individuals being infected P_{11} is as follows

$$\frac{dP_{11}}{dt} = \frac{1}{2}\frac{\rho X_1^2}{X_0 + X_1} - \omega P_{11} - 2\eta P_{11} - 2\gamma P_{11} + \beta P_{01}. \tag{68}$$

Here $\frac{\rho X_1^2}{X_0+X_1}$ is the proportion of couples made with both individuals belonging to the X_1 group. We have $-\omega P_{11}$ as the proportion of couples who break up and join the X_1 category. We have $-2\eta P_{11}$ as the proportion of couples who separate due to one of the partners leaving the population altogether. Here $-\gamma P_{11}$ is the proportion of population leaving the P_{11} category due to recovery and at least one individual joining the susceptible group. Finally βP_{01} is the proportion of couples in the P_{01} category where the susceptible individual has become infected from their partner.

In our model β represents the transmission rate in discordant pairs P_{01} and γ is the natural clearance rate. We note that β is composed of a rate of sexual intercourse during a partnership and a transmission probability per act.[26, see section 3].
If we use our assumption that the pair formation process is in equilibrium we get $X_0 = X^* - X_1$ and $P_{00} = P^* - P_{01} - P_{11}$ so we can reduce our system to 3 equations -

$$\frac{dX_1}{dt} = -\left(\rho + \gamma + \eta\right) X_1 + (\omega + \eta)(2P_{11} + P_{01}) \tag{69}$$

$$\frac{dP_{01}}{dt} = \rho X_1 (1 - \frac{X_1}{X^*}) - (\omega + 2\eta + \beta + \gamma) P_{01} + 2\gamma P_{11} \tag{70}$$

$$\frac{dP_{11}}{dt} = \frac{\rho X_1^2}{2X^*} - (\omega + 2\eta + 2\gamma) P_{11} + \beta P_{01} \tag{71}$$

The following segment closely follows [26, Section 4] regarding the reproduction number R. For the case of $\gamma = 0$ (i.e. no return to the susceptible group) we require three pieces of information to calculate our basic reproduction number$R$. The first is the probability of moving into the single state after being infected by a partner. The second is the number M of partners in the remaining lifetime of an individual and the third is the probability b of infecting a susceptible partner. We can calculate $b = \frac{\beta}{\omega+2\eta+\beta+\gamma}$. We have already shown $M = \frac{\rho(\omega+2\eta)}{\eta(\rho+\omega+2\eta)}$. The rate of individuals moving from P_{11} to the single state is $\frac{\omega+\eta}{\omega+2\eta+2\gamma}$ while the probability of first moving to P_{01} and then into the single state is given by $\frac{\gamma(\omega+\eta)}{(\omega+2\eta+2\gamma)(\omega+2\eta+\beta+\gamma)}$. As $\gamma = 0$ we get our product of bM as [26][see section 4]

$$R = \frac{\beta\rho(\omega + \eta)}{\eta(\rho + \omega + 2\eta)(\omega + 2\eta + \beta)}. \tag{72}$$

As with our SIS model our reproductive number R value being ≤ 1 will determine whether the disease will die out or cause an epidemic.

If we consider recovery $\gamma > 0$ then an infected individual can return to the susceptible group while in a partnership, they can therefore become reinfected and the infection can bounce back and forth prolonging the epidemic. The question now becomes, how do these infections count toward calculating R ? The issue lies in that, if in one couple, we had one person infected, the partner becomes infected, then recovers and returns to the susceptible group and then becomes infected again, our case reproduction number would be one since the initial case only infected one person. However the basic reproduction number would be 2 as the partner actually became infected twice.

A proposed solution is to introduce the reproduction number R_C and the partnership reproduction number R_p [19]. R_C is defined as the average number of secondary cases i.e. other individuals a primary case will infect during their infectious period. For this number an infection of the same partner who has recovered from an earlier infection does not count. Therefore $R_C \leq R$. The average number of secondary partnerships consisting of two infected individuals that a typical infected individual will produce during their lifetime is the partnership reproduction number R_p. If there were no re-infections within partnerships then $R = R_C$.

4.4 Partnership extension

The partnership dynamics of a population can be expanded to include different types of relationships. We can introduce long term or short term partnerships into our model [26, see section 5]. Shorter term relationships in disease models becomes much more important when we deal with a disease having a shorter infectious period as this keeps transmissions high. However in infections having a much longer duration of infection the distinction becomes less important. Many studies [28] have compared our pair formation models with instantaneous contact models i.e. spontaneous infection [12]. Generally we find instantaneous models overestimate endemic prevalence [12]. If pair formation dynamics are fast compared to the transmission dynamics does the instantaneous transmission model become a good approximation for the pair formation model? A weakness of this model is that in reality people can have more than one partner, or can have instantaneous contact as well (affairs, open relationship etc). Our model does not take this into account. A possible method of inclusion is to add a spontaneous infection term for our individuals with concurrent relationships. There is another model type known as the triple model [12, see figure 1] and this includes the possibility that an individual can have 2 partners at any given time.

4.5 A chlamydia spread model

We are in a position where we can use our partnership model with an actual disease so we need appropriate values for our parameters. First let us consider the transmission rate per act for the STI infection, chlamydia. According to a study in 2011 in Los Angeles [11] the transmission rate per act was estimated to be 4.5% and the average number of acts per week being 1.6 according to another study in the UK [31]. In our model we set our time step to

be 6 hours and therefore our β is $\frac{0.045\times1.6}{7\times4}$. Chlamydia does not have any natural clearance rate, in fact the chance of spontaneously clearing an infection is entirely negligible. We can however remove the infection easily with the use of antibiotics so what we need to consider is how often people become aware of their infection and subsequently seek out treatment. We can say that an infected person becomes recovered when they receive a test and they are no longer in their infectious period (at least if they are responsible) as they cease sexual activity. We need to know how often people will get tested. According to government advice sexually active individuals should be tested every 3 to 6 months [10], so we will make an extremely bold assumption and assume that everybody in our model is responsible and follows government advice receiving a test on average every four and a half months. We can calculate the clearance rate per 6 hours as $\frac{1}{135\times4}$. We will only be interested in modelling a fixed population so we will let $B = \eta = 0$. Our breakup rate ω is harder to estimate as there generally is a very high variation in breakup length, some relationships lasting just a week or a few months whilst some will last 20 years. In order to estimate this we have taken some information provided by dating.com [17] which says that 67% of their users have experienced a breakup in the year 2020: we can therefore calculate our break up rate as $\frac{0.67}{365\times4}$. Finally we need to consider the average time elapsed for a new single to form a new partnership. Again this data is difficult to source, so we have found a study [27] which states people form new partnerships within 7 months 25% of the time, within one year 8 months 50% of the time and within 3 years 6 months 75% of the time. We will assume that anybody else will form a relationship within 5 years. The average number of months is $0.25\times(7+20+42+60)$ as an upper bound and our lower bound is $0.25\times(0+7+20+42)$ so our average will be 24.75 months corresponding to a rate of $\rho = \frac{1}{742.5\times4}$.

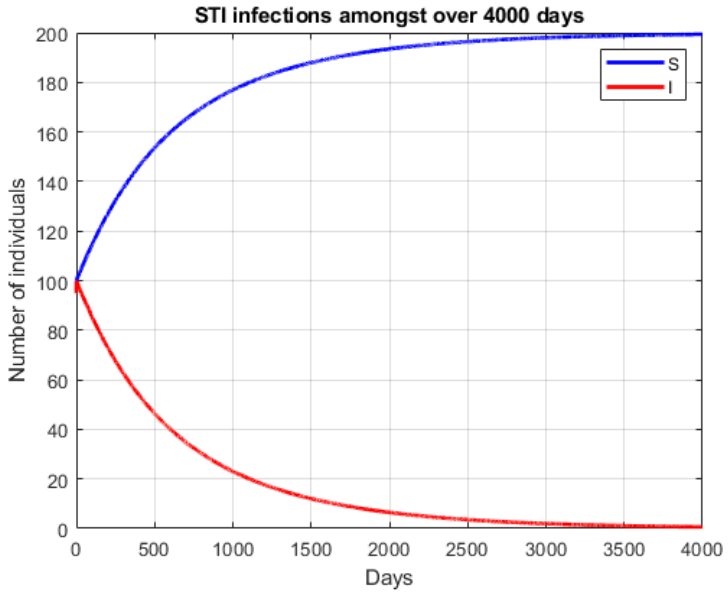

Figure 53: Decay of infections over time

In figure 53 we can clearly see that over time our infection slowly dies out from our population. Let us see if this matches with what we expect if our model was fully connected as in our original SIS model using the same infection and recovery rates.

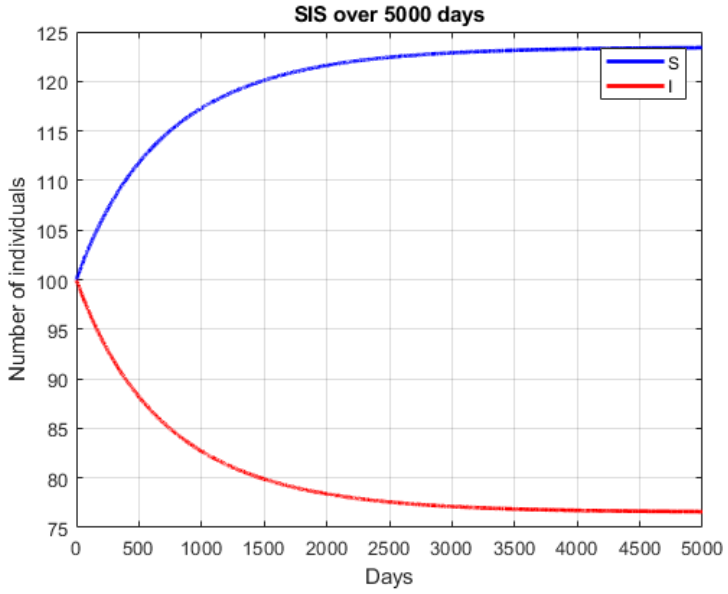

Figure 54: Infection increase over time

Now it is quite clear from figure 54 that under the assumption that there is free mixing of the population we actually do have an epidemic occuring and the infection will reach a non zero equilibrium point, in this case roughly 40% of our total population. Data gathered implies that in real life, of our total population, only 0.3828% have the infection [13]. Now there are many reasons for this rather massive difference: a clear one is of course the assumption in our model that everybody has an equal chance of being infected. Also the data used references [13] all age groups and it is widely known that chlamydia is a much more common infection amongst the younger population. I think we can say with confidence that there is an issue with the data gathered and the information used when constructing our parameters. The studies used all involved different participants and the location age and sex of each group has a huge influence on sexual activity. For example a study regarding average sexual partners in the US found in Louisiana an average of 15.7 while Utah was only 2.6. This is likely due to the prominence of Catholicism which encourages abstinence until marriage [9]. Clearly each source of data is heavily biased to the specific region and culture of that region. Some time spent examining other sources shows there are a large number of separate studies which give different results for length of partnerships and frequency of partnerships [8] which makes creating such a dynamic all the more challenging.

Currently our pair dynamic model seems completely unrealistic as chlamydia as an infection most certainly has not died out as it has in our model. This is due to incorrect estimates for pair formation and relationship separations as discussed above. In particular for our model the data we have counts “long term” relationships only and does not consider “short” term relationships or simple spontaneous encounters (affairs, one night stands etc). Let us see what happens if we try to account for these types of relationships by increasing the pair formation rate and the pair separation rate.

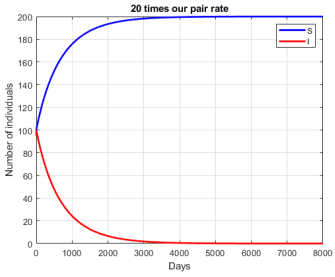

Figure 55: Our pair rate is multiplied by 20

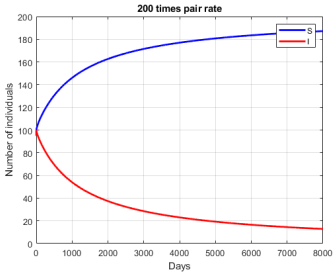

Figure 56: Our pair rate is multiplied by 200

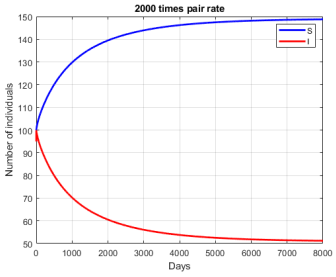

Figure 57: Our pair rate is multiplied by 2000

For our graphs 55, 56, 57 we have set our pair separation rate to be 50 times our original separation rate $\omega = 50 \times \frac{0.67}{365\times4}$. Clearly we can see that as we increase our pairing rate we will reach a critical point somewhere between 20 and 200 times our original where the infection will not die out and the infection will be constant throughout the population. This is what we would expect as when we increase the rate of pair formation we expect a higher equilibrium. Now let us examine the effect of the separation rate ω on our model.

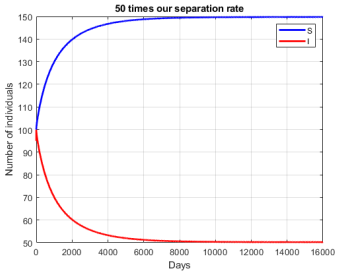

Figure 58: Our separation rate is multiplied by 50

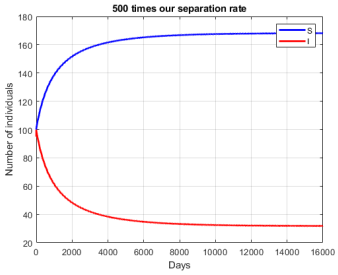

Figure 59: Our separation rate is multiplied by 500

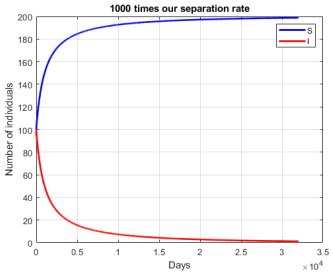

Figure 60: Our separation rate is multiplied by 1000

In figures 58, 59 and 60 we have kept our pair rate at 2000 times for all 3 graphs. We can see clearly that as our separation rate increases we actually get a lower equilibrium point for our infections as each infected partner has less "time" in the paired state to infect their partner. We again however have an optimum separation rate as a separation rate too low will also result in the virus dying out: see figures 61–63.

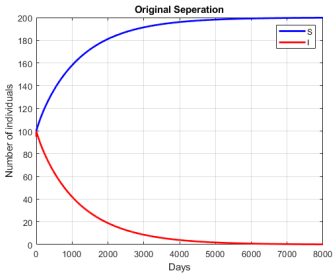

Figure 61: Our separation rate is our original

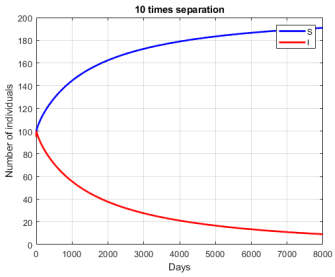

Figure 62: Our separation rate is multiplied by 10

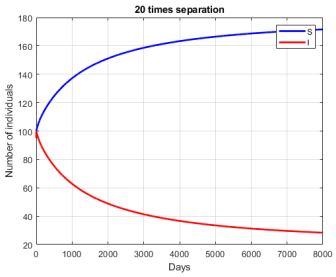

Figure 63: Our separation rate is multiplied by 20

Due to the complexity of the differential equations we will struggle to find an analytical

solution for our optimum ω. We can however find an optimal solution numerically rather easily. We can employ a crude version of the midpoint method to find our optimum as we know we only have one maximum. This method does however have the issue that we may accidentally "skip" past our maximum. We can however certainly estimate our maximum via numerical trial and error and plot ω as a function of the maximum.

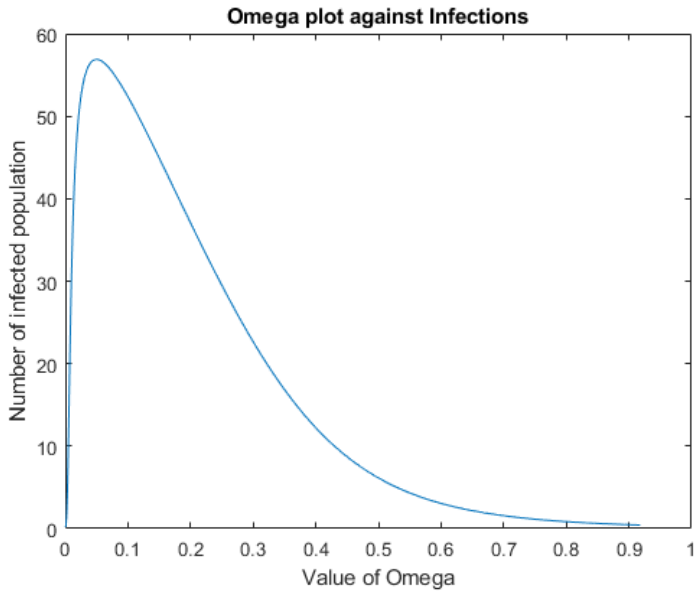

Figure 64: this is a plot of ω against infections

We can get a visual representation of our optimum ω from our graph in figure 64. However we can find the maximum from our plot which is 56.886 when $\omega = 0.05$. We can perform similar tests for each parameter within our system, however ideally we would like to find data which gives accurate starting parameters so that we can estimate parameters that match more closely with given data.

In an effort to try to improve our model we will consider the break up rate ω to be variable i.e. a function of time. According to Facebook data gathered on relationship status changes, the largest amount of breakups occur in the period leading up to spring while the least occur over summer [32]. With this in mind, we will model the break up rate as a sinusoidal function with our optimized value of ω being our average. So $\omega(t) = 0.05 + 0.05a\sin(t)$ where a is the weight we give to the sin function.

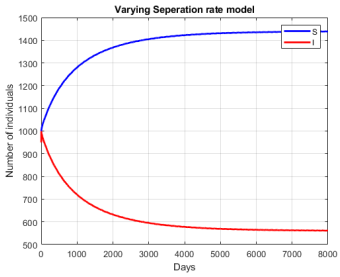

Figure 65: the yearly separation rate is now a function of time

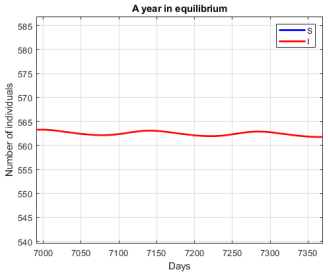

Figure 66: here we have zoomed in on the equilibrium over the course of a year to examine the oscillatory behavior

On inspection of figure 65 it is difficult to see any difference from varying ω. However in our second figure 66 we can see clearly that we have oscillatory behaviour over the course of the year. This is more in keeping with how we expect infection equilibrium to change over the course of the year [16]. To take this into account seems a sensible approach to improving our model. We could perform a similar adjustment with our partnership pairing rate by also making that vary over time.

4.6 Advantages and disadvantages

This model can provide an excellent representation of real world STI infection spread as it takes into consideration the actual pair formation dynamics required to understand it. However the model does not take into account other methods of transmission apart from couples (for example HIV may be spread via needles) and we make the assumption that the single population has zero sexual activity within that group. The main disadvantage of using this is the difficulty in acquiring accurate data that is representative of larger population groups and their specific partnership dynamics. For this reason, this model would be much more appropriate for modelling STI spread amongst a smaller subset of the entire population. Examples might be students at University or people confined within a particular city. This is because culture surrounding sexual activity varies greatly as you travel around the UK let alone the world and as such yield vastly different partnership dynamics.

5 Conclusion

In this project we have explored several of the current techniques available to epidemiologists to construct models for infection spread. We have implemented and tested practical improvements to the basic SIR and SIS models through use of the Gillespie algorithm. The resulting model is more flexible and allows decision makers to be better informed about possible and probable infection trajectories. We have further examined in detail the modelling of infections through the use of networks and demonstrated different methods of network construction.

The primary motivation for the use of network models is the ability to create precise links between individuals and groups, enabling identification of where transmission occurs and the possibility of preventing infection spread. In practice however, severe difficulties are often encountered in the actual task of gathering the required information.

These network simulations are often only feasible when employed on smaller scales, for example in modelling the spread of STIs. To extend our study of STIs, we examined pair formation models and defined ODEs which model the partnership dynamics within a population. We conclude that there is no single ideal model of infection spread. The choice made needs to be based upon the specific disease we are interested in, the availability of data regarding social networks, the disease-specific infection and recovery rates and the specific population or region of interest. Societal and geographical differences have a huge impact on infection spread and must be considered when formulating our model.
Possible means to extend our SIR model could be through the introduction of new categories which attempt to simulate the characteristics of the individuals of our population, such as their age. We could further improve our network simulation by introducing a waiting time between infection and recovery as we cannot realistically expect somebody to leave their infectious stage near instantaneously. The pair formation model we have presented is already an extension of the SIS model, however we could extend this further and include more categories such as occupation or age which may have an influence on infection spread.

6 References

[1] Build Watts-Strogatz Small World Graph Model - MATLAB & Simulink Example - MathWorks United Kingdom.

[2] Epidemic Models (SIR), (SIS), and Sexually Transmitted Diseases : Networks Course blog for INFO 2040/CS 2850/Econ 2040/SOC 2090. https://blogs.cornell.edu/info2040/2019/12/04/epidemic-models-sir-sis-and-sexually-transmitted-diseases/.

[3] Protected Herpes Transmission - STDCenterNY. https://stdcenterny.com/articles/protected-herpes-transmission.html.

[4] SIR and SIRS models — Generic Model documentation. https://docs.idmod.org/projects/emod-generic/en/latest/model-sir.html.

[5] SIR models of epidemics. https://tb.ethz.ch/education/learningmaterials/modelingcourse/level-1-modules/SIR.html.

[6] Stochastic simulation of a simple epidemic. http://www.biosym.uzh.ch/modules/models/ETHZ/StochasticSimulation/sir_stoch.xhtml.

[7] You Can Help Break the Chain of Transmission. https://siobhanroberts.com/you-can-help-break-the-chain-of-transmission/.

[8] The verdict is in: This is the ideal number of sexual partners to have in your lifetime, Sept. 2017. https://www.mindbodygreen.com/articles/the-ideal-number-of-sexual-partners-for-men-women.

[9] Average Number of Sexual Partners: By Gender, Country, State, and, Aug. 2018. https://www.healthline.com/health/healthy-sex/average-number-of-sexual-partners.

[10] Which STD Tests Should I Get? | Prevention | STDs | CDC, July 2021. https://www.cdc.gov/std/prevention/screeningreccs.htm.

[11] C. Althaus and N. Low. P1-S5.02 Towards more robust estimates of the per sex act transmission probability of Chlamydia trachomatis. *Sexually Transmitted Infections*, 87(Suppl 1):A175–A175, July 2011. Publisher: BMJ Publishing Group Ltd Section: Poster Sessions.

[12] C. L. Althaus, J. C. M. Heijne, S. A. Herzog, A. Roellin, and N. Low. Individual and Population Level Effects of Partner Notification for Chlamydia trachomatis. *PLoS ONE*, 7(12):e51438, Dec. 2012.

[13] L. G. Association. Chlamydia diagnosis rate per 100,000 population (all ages) in England. Publisher: Local Government Association, Local Government House, Smith Square, London, SW1P 3HZ, 020 7664 3000, http://www.local.gov.uk, info@local.gov.uk.

[14] S. Bukhori. *Forest Fire Model*. IntechOpen, Dec. 2017. Publication Title: Forest Fire.

[15] I. Cooper, A. Mondal, and C. G. Antonopoulos. A SIR model assumption for the spread of COVID-19 in different communities. *Chaos, Solitons, and Fractals*, 139:110057, Oct. 2020.

[16] V. J. Cornelisse, E. P. F. Chow, M. Y. Chen, C. S. Bradshaw, and C. K. Fairley. Summer heat: a cross-sectional analysis of seasonal differences in sexual behaviour and sexually transmissible diseases in Melbourne, Australia. *Sexually Transmitted Infections*, 92(4):286–291, June 2016. Publisher: BMJ Publishing Group Ltd Section: Epidemiology.

[17] Dating.com. Dating.com Reveals That 2020 Has Been The Year Of Break Ups. https://www.prnewswire.com/news-releases/datingcom-reveals-that-2020-has-been-the-year-of-break-ups-301135831.html.

[18] G. P. Garnett and R. M. Anderson. Sexually Transmitted Diseases And Sexual Behavior: Insights From Mathematical Models. *The Journal of Infectious Diseases*, 174(Supplement_2):S150–S161, Oct. 1996.

[19] J. C. M. Heijne, S. A. Herzog, C. L. Althaus, N. Low, and M. Kretzschmar. Case and partnership reproduction numbers for a curable sexually transmitted infection. *Journal of Theoretical Biology*, 331:38–47, Aug. 2013.

[20] D. Houcque. Ordinary Differential Equations (ODE). page 12. https://www.mccormick.northwestern.edu/docs/efirst/ode.pdf.

[21] A. Jolly, S. Muth, J. Wylie, and J. Potterat. Sexual Networks and Sexually Transmitted Infections: A Tale of Two Cities. *Journal of urban health : bulletin of the New York Academy of Medicine*, 78:433–45, Oct. 2001.

[22] A. M. Jolly and J. L. Wylie. Gonorrhoea and chlamydia core groups and sexual networks in Manitoba. *Sexually Transmitted Infections*, 78(suppl 1):i145–i151, Apr. 2002. Publisher: The Medical Society for the Study of Venereal Disease Section: Symposium.

[23] M. J. Keeling and K. T. Eames. Networks and epidemic models. *Journal of The Royal Society Interface*, 2(4):295–307, Sept. 2005. Publisher: Royal Society.

[24] M. J. Keeling and K. T. Eames. Networks and epidemic models. *Journal of the Royal Society Interface*, 2(4):295–307, Sept. 2005.

[25] A. A. King. Exact simulation of birth-death processes via the Gillespie algorithm, May 2017. https://kingaa.github.io/clim-dis/stochsim/gillespie.html#simple-stochastic-si-epidemic.

[26] M. Kretzschmar and J. C. Heijne. Pair formation models for sexually transmitted infections: A primer. *Infectious Disease Modelling*, 2(3):368–378, July 2017.

[27] Nanaya. Groundbreaking Study Predicts How Long People Stay Single. https://www.prnewswire.com/news-releases/groundbreaking-study-predicts-how-long-people-stay-single-300795880.html.

[28] J. B. S. Ong, X. Fu, G. K. K. Lee, and M. I.-C. Chen. Comparability of Results from Pair and Classical Model Formulations for Different Sexually Transmitted Infections. *PLOS ONE*, 7(6):e39575, June 2012. Publisher: Public Library of Science.

[29] J. M. A. Padgett and S. Ilie. An adaptive tau-leaping method for stochastic simulations of reaction-diffusion systems. *AIP Advances*, 6(3):035217, Mar. 2016. Publisher: American Institute of Physics.

[30] M. A. Porter. Small-world network. *Scholarpedia*, 7(2):1739, Feb. 2012. http://www.scholarpedia.org/article/Small-world_network.

[31] A. Rabbett and Z. Mulroy. Scientists reveal how often 'normal' couples have sex and for how long, May 2019. Section: UK & World News https://www.cambridge-news.co.uk/news/uk-world-news/scientists-reveal-how-often-normal-16222670.

[32] R. Santos. Most Popular Times of The Year For Breaking Up, Apr. 2011. Section: Love & Sex https://www.marieclaire.com/breakup-calendar.

[33] T. I. Team. Law Of Large Numbers Definition. https://www.investopedia.com/terms/l/lawoflargenumbers.asp.

[34] N. Zhang and J. Tang. Forest Fires Model and SIR Model Used in Spread of Ebola Virus in Prediction and Prevention. *International Journal of Scientific Research in Science and Technology*, 2:222, Aug. 2016.